IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# EARLY AUTISM DETECTION USING MACHINE LEARNING

*Dr.S. Raja Mohamed,*

*Assistant Professor,*
*Department of Computer Science and Engineering,*
*KIT-Kalaignar karunanidhi Institute of Technology,(Autonomous)*
*Coimbatore, India.*
*rajadelip@gmail.com*

*Priyanka.R*
*Master of Engineering,*
*Department of Computer Science and Engineering,*
*KIT-Kalaignar karunanidhi Institute of Technology,(Autonomous)*
*Coimbatore, India*
*Sripriyanka.r8@gmail.com*

Abstract— *Autism Disease research project focuses on a comprehensive study involving Exploratory Data Analysis (EDA) and the application of Artificial Neural Networks (ANN) for the classification of Autism Spectrum Disorder patients. The primary objectives of this study are to gain insights into the characteristics of autistic patients through EDA and to develop an accurate classification model using ANN for the identification of individuals with ASD. The initial phase of the project involves Exploratory Data Analysis, wherein we explore and analyze the dataset containing various features related to ASD patients. The goal is to uncover patterns, trends, and key insights that may contribute to a better understanding of the characteristics associated with ASD. Subsequently, the study transitions into the application of Artificial Neural Networks, a machine learning technique, for the classification of individuals into two categories: "ASD" and "non-ASD." The neural network is trained on a labelled dataset, utilizing the information gathered from the EDA phase. The objective is to harness the predictive capabilities of ANN to accurately identify and classify new instances of ASD based on their characteristic features. Additionally, the research incorporates the fine-tuning of the ANN model, optimizing its architecture and parameters to achieve optimal performance in ASD classification. The outcomes of this research may contribute valuable insights to the field of autism diagnosis, offering a more understanding of the characteristics associated with ASD and paving the way for more effective and accurate diagnostic tools.*

**Keywords: Autism Detection, Disease Diagnosis, Artificial Neural Networks, ,Exploratory Data Analysis , Accurate diagnostic tools, Neural Disease.**

## I INTRODUCTION

One major advancement in neural science is the early detection of Autism disease. Autism Spectrum Disorder (ASD) is a complex neuron developmental condition characterized by a spectrum of challenges in social interaction, communication, and repetitive behaviours. Diagnosing ASD requires a thorough examination of behavioural patterns and characteristics, often relying on traditional methods such as the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS). However, the evolving landscape of technology and data science presents an opportunity to augment and enhance diagnostic processes.

This research project embarks on a comprehensive study leveraging two fundamental pillars in data science: Exploratory Data Analysis (EDA) and Artificial Neural Networks (ANN). The overarching aim is twofold: to gain profound insights into the distinctive characteristics exhibited by individuals with ASD through rigorous EDA, and subsequently, to construct a robust classification model using ANN for the precise identification of individuals within the ASD spectrum. The research extends its focus to the fine-tuning of the ANN model, a meticulous process aimed at optimizing the architecture and parameters. This phase seeks to attain an optimal balance that ensures both model efficiency and accuracy in ASD classification. Emphasizing advanced neural network techniques, the project confronts the complexity inherent in ASD datasets, striving to improve predictive capabilities and refine the understanding of the diagnostic features

## II RELATED WORK

In order to determine how and under what circumstances of autism diease throught out the world , we performed a survey using the Python programming language and the sets of reports of the normal and autism detected patients. This has to do with the prediction and prevention of neural disease that was discussed earlier.

The current system for ASD diagnosis relies primarily on behavioral assessments and clinical observations. However, this system has inherent challenges, including subjectivity, delayed diagnosis, and limited scalability. Timely ASD detection is vital for providing necessary intervention and support.

The strength of ANNs lies in their ability to learn from data. During the training phase, the network adjusts its weights and biases using optimization algorithms like gradient descent. This iterative process aims to minimize the difference between the predicted output and the actual target, refining the network's ability to generalize to new, unseen data.

The following are some current talks and developments about the use of machine learning and Artificial neural networks in Medical Applications[12][13].

To identify and classify the different or distinct number of diseases affected by Alzheimer's disease, heart failure, breast cancer and pneumonia diseases ,multiple authors in the cited publications created multiple ANN network models. It is common to come across different pairs of chromosomes and inherit from parents genes involved in causing ASD. To ascertain if the patient is healthy, the ANN model was developed, processed, and analyzed. Every one of their numerous illnesses is analysed by the report and medication will be provided for future prediction.

Software is offered by Siriwan Sunsirikul, TiraneeAchalakul and associates [14] as a means of early detection and categorizing the stages of Autism disease. This approach will decrease the autism in future. Picture capture, image preprocessing, segmentation, feature extraction, and classification are some of its phases.

Niket Amoda and associates [15] provide highly automated, reasonably priced, and precise image processing-based solutions. The six main components of the solution are as follows: the data will be collected and preprocessed comes first. using EDA the reports will be evaluated . The collected and evaluated dateset then divided into segments in the subsequent step to extract the most important details. The calculation is completed in the third stage determined by the chromosome number and genetically analysis .At last, classification is done in the fourth step. Applying a neural network that has been trained using the feature outcomes of extraction. In general, the deep training approach learning models on ever-larger and publicly accessible accessible trained datasets provide an obvious path for extensive global smartphone-aided detection of autism diseases.

## METHODOLOGY

### *3.1. PROPOSED METHODOLOGY*

Advancing Autism Spectrum Disorder Prediction through Integrative Data Science. In response to the evolving landscape of autism spectrum disorder (ASD) diagnosis, this research proposes an innovative system that amalgamates Exploratory Data Analysis (EDA) with cutting-edge Artificial Neural Networks (ANN). The primary goal is to enhance the accuracy and efficiency of ASD classification, providing a nuanced understanding of the intricate characteristics associated with this neuron developmental condition.The dataset is divided into two subdirectories, test and train, each containing data used for training and testing. There are two additional folders: the sick folder, which contains data of ASD detected patients in poor health, and the healthy folder, which contains data of the normal patient's excellent health. Since we are spared the difficulties posed by unbalanced data, there is a lower likelihood of bias towards any one class. Through the integration of Exploratory Data Analysis and Artificial Neural Networks, this study aspires to present a holistic and nuanced approach to understanding and classifying ASD patients. The anticipated outcomes hold the potential to contribute significant insights to the realm of autism diagnosis, fostering a deeper understanding of the spectrum's characteristics. Moreover, this research envisions paving the way for more effective and accurate diagnostic tools that align with the evolving landscape of data-driven methodologies.

## MODULES

In the proposed methodology, early detection of autism disease is implemented by using ensemble learning with the EDA and Artificial neural network algorithms. Artificial neutral network Algorithm widespread popularity stems from its user-friendly nature and adaptability, enabling it to tackle both classification and regression problems effectively. The algorithm's strength lies in its ability to handle complex datasets and mitigate overfitting, making it a valuable tool for various predictive tasks in machine learning. One of the most important features of the Artificial neural network is that it can handle the data set containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. It performs better for classification and regression tasks. In this tutorial, we will understand the working of ANN and implement ANN on a classification task.

## DATA PREPROCESSING

Acquire a comprehensive dataset containing diverse features related to ASD patients. Preprocess the data by handling missing values, encoding categorical variables, and normalizing numerical features.

## EXPLORATORY DATA ANALYSIS (EDA)

Perform in-depth statistical and visual analysis of the dataset to uncover patterns, trends, and correlations. Utilize tools like pandas, NumPy, and Matplotlib/Seaborn in Python for data manipulation and visualization.

## FEATURE ENGINEERING

Enhance the dataset through feature engineering, creating new features that may capture relevant information for ASD classification. This could involve transforming existing features or generating new ones based on domain knowledge.

## FEATURE SELECTION:

Use statistical methods or machine learning techniques to select the most relevant features for ASD classification. This step optimizes the model's performance by focusing on the most discriminative attributes.

## ARTIFICIAL NEURAL NETWORK (ANN) ARCHITECTURE

Design the architecture of the neural network, considering the input layer, hidden layers, and output layer. Determine the number of neurons, activation functions, and the overall topology of the network.

## MODEL TRAINING

Train the ANN on the labelled dataset. Utilize Optimization algorithms such as stochastic gradient descent to minimize the error between predicted and actual outputs. Implement regularization techniques to prevent over fitting.

## MODEL EVALUATION

Assess the performance of the trained model using metrics like accuracy, precision, recall, and F1 score. Utilize cross-validation to ensure robustness, and validate the model on external datasets to gauge generalization capability.

## FINE-TUNING

Fine-tune the hyperparameters of the ANN, adjusting learning rates, batch sizes, and other parameters to optimize the model's accuracy and efficiency.

The term ANN Artificial Neural Networks (ANNs) represent a cornerstone in the realm of artificial intelligence, drawing inspiration from the intricate networks of neurons in the human brain. Developed to emulate cognitive processes, ANNs have proven highly effective in solving complex problems, learning patterns, and making predictions across diverse domains. This section delves into the fundamental aspects of Artificial Neural Networks, elucidating their structure, functioning, and applications. Each artificial neuron within a neural network processes information through a set of connections, each assigned a weight that determines its significance. The weighted sum of inputs, along with a bias, passes through an activation function, introducing non-linearity and enabling the network to learn complex relationships within the data. The strength of ANNs lies in their ability to learn from data. During the training phase, the network adjusts its weights and biases using optimization algorithms like gradient descent. This iterative process aims to minimize the difference between the predicted output and the actual target, refining the network's ability to generalize to new, unseen data.

Information flows in one direction, from input to output, without loops or cycles. Commonly used for pattern recognition and classification. Incorporate feedback loops, allowing them to process sequential data, making them suitable for tasks like natural language processing and time-series analysis. Specialized for processing grid-like data, such as images. CNNs use convolution layers to extract hierarchical features.

## CONCLUSION

As in "Feature Selection" exploration, we have seen the attribute named 'result' has such a powerful presence in the ASD dataset; all other attributes have little to no contribution in deciding the final class. In the Figure 18 below, we have drawn a swarm plot with 'result' asx-axis and ASD-class (say 'yes' is 1 and 'no' is 0) as y-axis, and reconfirmed the underlying association between the given variables where the target class is easily separable.


Figure Training : Accuracy vs Training Loss

During pre-processing of the data, we dropped about 95 of rows of data due to its 'NaN' entries. Ideally one should try to retain as much data as possible, as there could always be some valuable insight that could be lost. We usually try to fill the missing entries by 'median filling' or with a fancier approach, one can also run a supervised learning model to 'predict' the 'NaN' value. But that approach is not applicable for our dataset as many inputs that was missing are of categorical type and it is not feasible to make a median of a category).While there are some advanced procedures like 'imputation process' to take care of these issues of missing values of categorical types without having them dropped, we didn't explore those techniques in this current work. Even without implementing more sophisticated techniques like 'imputation' in order to retain valuable information, we managed to find very efficient models that can classify new instances with accuracy score1.The reason scores close to1. 0 is because a single feature is almost entirely responsible for deciding the output. This means that the output can be determined by a high confidence. The present work is certainly quite interesting and is a noble area where technology can be used to actually change lives.

## REFERENCES

[1]Brian Godsey, Think Like a Data Scientist Manning, ISBN: 9781633430273

[2]H. Brink, J. Richards, M. Fetherolf, Real World Machine Learning, Manning, 9781617291920 ISBN:

[3]D. Cielen, A.Meysman, M. Ali, Introducing Data Science, Manning 9781633430037. ISBN:

[4]J. Grus, Data Science From Scratch First Principles With Python, O'Reilly 9781491901427 ISBN:

[5]A. G´eron, Hands-On Machine Learning with Scikit-Learn & Tensor Flow, O'Reilly ISBN: 9781491962299

[6]T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Second Edition, Springer

[7] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, ISBN 9781461471370

[8] Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6.Taichung City, Taiwan, ACM.

[9] Thabtah, F. (2017). ASDTests.A mobile app for ASD screening. www.asdtests.com [accessed December 20th, 2017].

[10] Thabtah, F. (2017). Machine Learning in Autistic Spectrum Disorder Behavioural Re- search: A Review. Informatics for Health and Social Care Journal. December, 2017 (in press)

[11] M. Eni, I. Dinstein, M. Ilan, I. Menashe, G. Meiri and Y. Zigel, "Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network," in IEEE Access, vol. 8, pp. 139489 139500, 2020, doi: 10.1109/ACCESS.2020.3012532.

[12] A. Gorodetski, I. Dinstein and Y. Zigel, "Speaker diarization during noisy clinical diagnoses of autism," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 2593-2596, doi: 10.1109/EMBC.2019.8857247.

[13] N. Rusli, S. N. Sidek, H. M. Yusof, N. I. Ishak, M. Khalid and A. A. A. Dzulkarnain, "Implementation of Wavelet Analysis on Thermal Images for Affective States Recognition of Children With Autism Spectrum Disorder," in IEEE Access, vol. 8, pp. 120818-120834, 2020, doi: 10.1109/ACCESS.2020.3006004.

[14] L. Mertz, "New Quantitative Approach to Autism

Diagnosis," in IEEE Pulse, vol. 10, no. 2, pp. 34-36, March-April 2019, doi: 10.1109/MPULS.2019.2899766.

[15] G. Shushma and I. J. Jacob, "Autism Spectrum Disorder detection using AI Algorithm," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2022, pp. 1-5, doi: 10.1109/ICAIS53314.2022.9743011.

[16] G. Nie et al., "An Immersive Computer-Mediated Caregiver-Child Interaction System for Young Children With Autism Spectrum Disorder," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 884-893, 2021, doi: 10.1109/TNSRE.2021.3077480.

[17] R. Haweel et al., "A Novel Grading System for Autism Severity Level Using Task-Based Functional MRI: A Response to Speech Study," in IEEE Access, vol. 9, pp. 100570-100582, 2021, doi: 10.1109/ACCESS.2021.3097606.

[18] M. Robles et al., "A Virtual Reality Based System for the Screening and Classification of Autism," in IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 5, pp. 2168-2178, May 2022, doi: 10.1109/TVCG.2022.3150489.

[19] Z.-A. Huang, Z. Zhu, C. H. Yau and K. C. Tan, "Identifying Autism Spectrum Disorder From Resting-State fMRI in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 7, pp. 2847-2861, July 2021, doi: 10.1109/TNNLS.2020.3007943.

[20] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment Analysis is a Big Suitcase," IEEE Intelligent Systems, vol. 32, no. 6, pp. 74-80, 2017.

[21] E. Kim and R. Klinger, "A Survey on Sentiment and Emotion Analysis for Computational Literary Studies," arXiv preprint arXiv: 1808.03137, p. 45, 2018.

[22] Y. Ma and J.C. Principe, "A Taxonomy for Neural Memory Networks," IEEE Trans. Neural Networks and Learning Systems, 2019.

[23] M. Peters et al., "Deep Contextualized Word Representations," Proc. 2018 Conf. North American Chapter of the ACL: Human Language Technologies, pp. 2227-2237, 2018.

[24] A. Giannakopoulos, C. Musat, A. Hossmann, and M. Baeriswyl, "Unsupervised Aspect Term Extraction with B-LSTM & CRF using Automatically Labelled Datasets," Proc. 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 180–188, 2017.

[25] J. Hashemi et al., "Computer Vision Analysis for Quantification of Autism Risk Behaviors," in IEEE Transactions on Affective Computing, vol. 12, no. 1, pp. 215-226, 1 Jan.-March 2021, doi: 10.1109/TAFFC.2018.2868196.

[26] S. Tummala, "Deep Learning Framework using Siamese Diagnosis of Autism from Brain Magnetic Resonance Imaging," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9418143.

[27] C. Li, X. Guo, and Q. Mei, "Deep Memory Networks for Attitude Identification," Proc. 10th ACM Int'l Conf. Web Search and Data Mining, pp. 671-680, 2017.

[28] F. J. Alves, E. A. De Carvalho, J. Aguilar, L. L. De Brito and G. S. Bastos, "Applied Behavior Analysis for the Treatment of Autism: A Systematic Review of Assistive Technologies," in IEEE Access, vol. 8, pp. 118664-118672, 2020, doi: 10.1109/ACCESS.2020.3005296.

[29] J. C. Epifânio and L. F. Da Silva, "Scrutinizing Reviews on Computer Science Technologies for Autism: Issues and Challenges," in IEEE Access, vol. 8, pp. 32802-32815, 2020, doi: 10.1109/ACCESS.2020.2973097.

[30] F. J. Alves, E. A. De Carvalho, J. Aguilar, L. L. De Brito and G. S. Bastos, "Applied Behavior Analysis for the Treatment of Autism: A Systematic Review of Assistive Technologies," in IEEE Access, vol. 8, pp. 118664-118672, 2020, doi: