Detecting Autism Spectrum Disorder Using Machine Learning Techniques

S. Srividhya[1], D.Vidhya[2]

[1]PG Student, Nandha College of Technology, Erode, Tamilnadu, India

[2]Assistant Professor, Nandha College of Technology, Erode, Tamilnadu, India

ABSTRACT

Autism Spectrum Disorder (ASD) is a neuro-jumble in which an individual lifelongly affects connection and correspondence with others. It encompasses a wide range of symptoms and severity levels, making it a heterogeneous disorder. While the exact causes of ASD remain elusive, research suggests a combination of genetic and environmental factors play a role in its development. Early diagnosis and intervention are crucial for improving outcomes and enhancing the quality of life for individuals with ASD. This abstract highlights the multifaceted nature of ASD and underscores the importance of ongoing research and support services to address the needs of individuals affected by this disorder. Chemical imbalance can be analyzed at any stage in once life and is supposed to be a "conduct sickness" in light of the fact that in the initial two years of life side effects normally show up. As indicated by the ASD issue begins with youth and keeps on continuing onward on into pre-adulthood and adulthood. Moved with the ascent being used of AI strategies in the examination aspects of clinical conclusion, in this paper there is an endeavor to investigate the likelihood to utilize Credulous Bayes, Backing Vector Machine, Calculated Relapse, KNN, Brain Organization and Convolutional Brain Organization for anticipating and investigation of ASD issues in a youngster, teenagers, and grown-ups. The proposed methods are assessed on openly accessible three different non-clinically ASD datasets. In the wake of applying different AI procedures and taking care of missing qualities, results unequivocally recommend that MLP put together expectation models work better with respect to every one of these datasets with higher precision and accuracy for Mentally unbalanced Range Issue Evaluating in Information for Grown-up, Youngsters, and Teenagers separately.

Key words : autism spectrum disorder , MLP , machine learning , classification

INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that has garnered significant attention in both scientific research and public awareness in recent decades. It is characterized by a wide range of symptoms, including difficulties in social interaction, communication challenges, and repetitive behaviors. The term "spectrum" reflects the diversity of experiences and severity levels among individuals with ASD, ranging from mild to severe. Understanding the underlying mechanisms of ASD, including its etiology, neurological basis, and socio-cultural factors, is essential for developing effective interventions and support strategies. This introduction provides a glimpse into the multifaceted nature of ASD and sets the stage for exploring its complexities in subsequent discussions. Autism Spectrum Disorder (ASD) , is a neurological formative problem. It influences how individuals impart and cooperate with others, as well as how they act and learn. The side effects and signs show up at the point when a kid is exceptionally youthful. It is a long lasting condition also, can't be totally restored. That's what an investigation discovered 33% of youngsters with challenges other than ASD have some ASD side effects while not gathering the full characterization measures. ASD has a critical monetary effect both due to the expansion in the quantity of ASD cases around the world, furthermore, the time and costs engaged with diagnosing a patient. Early identification of ASD can help both patient what's more, medical care specialist organizations by recommending appropriate treatment and additionally prescription required and in this manner diminishing the drawn out costs related with postponed analysis. Then again the conventional clinical techniques, for example, Chemical imbalance Symptomatic Meeting Changed (ADIR) and Mental imbalance Demonstrative Perception Timetable Overhauled (ADOS-R), are tedious and awkward. The kid who are excessively youthful and has deferred discourse issue generally score 25% of the complete ADI-R things since the verbal areas

can't be addressed precisely for the patient. Furthermore, directing interviewwith a parental figure by a prepared inspector takes 90 to 150 minutes which is lumbering and frequently misses information Presently a days, AI has been applied to recognize different illnesses including discouragement and ASD. The essential goals of applying AI methods are to further develop determination exactness and diminish finding opportunity of a case all together to give speedier admittance to medical care administrations. Since the finding system of a case includes concocting the right class (ASD, No-ASD) in light of the information case highlights, this cycle can be credited as a grouping task in AI. In this paper, we apply different order methods to acquire further developed exactness on the consequences of identifying ASD cases for every one of the four datasets.

MULTI LAYER PERCEPTRON (MLP)

A Multi-Layer Perceptron (MLP) is a type of artificial neural network that consists of multiple layers of interconnected nodes or neurons. It is one of the most common and versatile architectures used in machine learning and deep learning. Each neuron in an MLP is connected to every neuron in the subsequent layer, forming a feedforward network where information flows from the input layer through one or more hidden layers to the output layer. MLPs are capable of learning complex non-linear relationships in data and are widely employed in various tasks such as classification, regression, and pattern recognition. Training an MLP involves adjusting the weights and biases of its connections through iterative optimization algorithms like backpropagation, allowing the network to gradually improve its performance on a given task. Despite their simplicity compared to more advanced architectures like convolutional neural networks (CNNs) or recurrent neural networks (RNNs), MLPs remain a fundamental building block in the field of deep learning, offering flexibility and scalability in modeling diverse datasets and problem domains.

MLPs are highly customizable, with parameters such as the number of layers, the number of neurons in each layer, and the activation functions offering degrees of freedom for model design. However, their performance can be sensitive to the choice of hyperparameters and the size and quality of the training data. Overfitting, where the model learns to memorize the training data rather than generalize to new, unseen data, is a common challenge in training MLPs, often requiring regularization techniques

such as dropout or weight decay to mitigate. Despite these challenges, MLPs remain a powerful tool in machine learning, particularly in domains where interpretability and flexibility are paramount, such as in finance, healthcare, and natural language processing. Ongoing research continues to explore novel architectures and training methodologies to enhance the capabilities and efficiency of MLPs in addressing increasingly complex real-world problems.

MACHINE LEARNINIG

Machine learning is a branch of artificial intelligence that focuses on developing algorithms and models capable of learning from data and making predictions or decisions without explicit programming. It encompasses a variety of techniques, including supervised learning, unsupervised learning, and reinforcement learning, each suited to different types of tasks and data. In supervised learning, algorithms learn from labeled examples to make predictions or classifications on unseen data. Unsupervised learning involves finding patterns and structures in unlabeled data, while reinforcement learning uses feedback from the environment to learn optimal behaviors or strategies. Machine learning has revolutionized numerous industries, from healthcare to finance to autonomous vehicles, by enabling computers to extract insights and patterns from vast amounts of data, leading to more accurate predictions, improved decision-making, and automation of complex tasks. Ongoing advancements in machine learning algorithms, coupled with the availability of big data and computational resources, continue to drive innovation and reshape the way we approach problems in the modern world.

Recent advancements in machine learning, particularly in deep learning, have propelled the field to new heights, allowing for the creation of highly complex models capable of processing and understanding data in ways previously thought impossible. Deep learning, a subfield of machine learning, employs neural networks with multiple layers to automatically extract hierarchical representations of data, leading to state-of-the-art performance in tasks such as image recognition, natural language processing, and speech recognition. The availability of open-source libraries and frameworks, such as TensorFlow and PyTorch, has democratized access to advanced machine learning tools, fostering innovation and collaboration among researchers and practitioners worldwide. However, challenges such as model interpretability, bias, and ethical

considerations remain areas of active research and debate within the machine learning community. As machine learning continues to evolve and integrate with other fields such as robotics, healthcare, and cybersecurity, its impact on society is expected to grow exponentially, shaping the way we live, work, and interact with technology in the years to come.

CLASSIFICATION

Classification is a fundamental task in machine learning that involves categorizing input data into predefined classes or categories based on their features or attributes. The goal of classification algorithms is to learn a mapping from input data to output labels, enabling the automatic labeling or prediction of new, unseen instances. Supervised learning techniques, such as logistic regression, decision trees, support vector machines (SVM), and neural networks, are commonly used for classification tasks. These algorithms learn from labeled training data, where each example is associated with a known class label, and then generalize to accurately classify new data points. Classification finds applications in various domains, including image recognition, spam detection, sentiment analysis, medical diagnosis, and fraud detection, where the ability to automatically classify data into distinct categories is essential for decision-making and problem-solving.

Ongoing research in classification algorithms focuses on improving accuracy, scalability, and interpretability to address the evolving needs of diverse application scenarios.Furthermore, classification algorithms can handle binary or multiclass classification tasks, depending on the number of classes involved. In binary classification, the data is divided into two categories, while in multiclass classification, there are more than two possible classes. Performance evaluation of classification models typically involves metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve, among others, to assess the model's ability to correctly classify instances across different classes. Challenges in classification include dealing with imbalanced datasets, where one class is significantly underrepresented, as well as handling noisy or ambiguous data. Advanced techniques such as ensemble learning, feature engineering, and model selection help improve the robustness and generalization ability of classification models in real-world scenarios. As classification remains a cornerstone of machine learning, ongoing research continues to push the boundaries of algorithmic innovation and application domains, ensuring its relevance and effectiveness in addressing complex data analysis tasks.

RELATED WORKS

Bonnie Auyeung et.al says Bleeding edge wellbeing experts need a "red flag" device to help their dynamic about whether to make a reference for a full demonstrative evaluation for a mental imbalance range condition (ASC) in kids and grown-ups. The point was to distinguish 10 things on the Chemical imbalance Range Remainder (AQ) (Grown-up, Juvenile, and Kid adaptations) and on the Quantitative Agenda for Chemical imbalance in Babies (Q-Talk) with great test exactness. [1] Astha Baranwal et.al says The creators' essential inspiration in suggested mental imbalance metaphysics, which is the significant desire for creating calculations in view of subfield and is basically utilized for ascertaining accuracy, speed, and adaptability. [2] S Aristocrat Cohen et.al expresses Right now there are no concise, self-regulated instruments for estimating how much a grown-up with ordinary insight has the attributes related with the mentally unbalanced range. In this paper, we report on another instrument to survey this: the Chemical imbalance Range Remainder (AQ).

[3] M Duda et.al says In spite of the fact that chemical imbalance range jumble (ASD) and consideration shortfall hyperactivity jumble (ADHD) keep on ascending in predominance, together influencing >10% of the present pediatric populace, the strategies for conclusion stay abstract, lumbering and time serious. With holes up of a year between starting doubt and conclusion, important time where medicines and social intercessions could be applied is lost as these problems stay undetected. [4] Uğur Erkan et.al says Mentally unbalanced Range Problem (ASD) is a problem related with hereditary and neurological parts prompting challenges in friendly cooperation and correspondence. As indicated by measurements of WHO, the quantity of patients determined to have ASD is steadily expanding. The majority of the momentum concentrates on center around clinical conclusion, information assortment and mind pictures examination, however don't zero in on the finding of ASD in light of machine learning.[5] Gerald D Fischbach et.al with an end goal to recognize all over again hereditary

variations that add to the general gamble of mental imbalance, the Simons Establishment Chemical imbalance Exploration Drive (SFARI) has accumulated an exceptional example called the Simons Simplex Assortment (SSC). In excess of 2000 families have been assessed to date. By and large, probands in the ongoing example display moderate to extreme mentally unbalanced side effects with somewhat minimal scholarly handicap. An intelligent data set has been made to work with connections between's clinical, hereditary, and neurobiological data.[6]

METHODOLOGY

Proposed Method for Autism Spectrum Disorder (ASD) using MLP:

1. Data Collection: Gather a comprehensive dataset comprising a range of features relevant to ASD diagnosis, including demographic information, behavioral assessments, medical history, and genetic factors. Ensure the dataset is diverse and representative of different ASD profiles and severity levels.

2. Data Preprocessing: Clean the dataset by handling missing values, normalizing numerical features, and encoding categorical variables. Perform feature selection or extraction to reduce dimensionality and focus on the most informative features for ASD diagnosis.

3. Model Architecture: Design a Multi-Layer Perceptron (MLP) neural network architecture tailored for ASD classification. Configure the input layer to accommodate the processed features, followed by one or more hidden layers with appropriate activation functions (e.g., ReLU) and neuron units. Customize the output layer to represent binary (ASD/non-ASD) or multiclass (ASD severity levels) classification.

4. Model Training: Split the dataset into training, validation, and test sets to train and evaluate the MLP model. Utilize training techniques such as mini-batch gradient descent and regularization (e.g., dropout) to prevent overfitting and improve generalization. Optimize hyperparameters, including learning rate, batch size, and number of epochs, through cross-validation or grid search.

5. Model Evaluation: Assess the performance of the trained MLP model using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve. Analyze the confusion matrix to examine the model's classification performance across

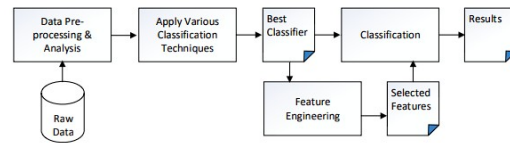different ASD categories and identify any misclassifications.

Fig 1 : autism detection

6. Interpretability and Validation: Interpret the learned representations within the MLP model to understand the key features contributing to ASD classification decisions. Validate the model's predictions through clinical validation studies involving independent datasets and expert evaluations to ensure its reliability and generalizability in real-world settings.

7. Deployment and Integration: Deploy the trained MLP model as a diagnostic tool for ASD screening or support system in clinical settings, educational institutions, or community centers. Integrate the model into existing healthcare infrastructure or mobile applications for accessible and scalable use by healthcare professionals, caregivers, and individuals with ASD.

8. Continuous Improvement: Continuously monitor and update the MLP model based on feedback from users, advancements in research, and changes in diagnostic criteria or guidelines for ASD. Incorporate new data and insights to enhance the model's accuracy, robustness, and usability over time, thereby improving its effectiveness in aiding ASD diagnosis and intervention.

The dataset consist of following attributes and its types with description

Attribute	Type	Description
Age	Number	Age in months/years
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with PDD	Boolean (yes or no)	Whether any immediate family member

		has a PDD
Why taken the screening	Meta	The person can write short reason for completing the task
Country of residence	String	List of countries in text format

Table 1 dataset description

Evaluation matrix

For a given dataset and a predictive model, every data point will lie on one of the below four categories.
– True Positive (TP): The individual having ASD and is correctly predicted as having ASD.
– True Negative (TN): The individual not having ASD and was correctly predicted as not having ASD.
– False Positive (FP): The individual not having ASD, is incorrectly predicted as having ASD.
– False Negative (FN): The individual having ASD, is incorrectly predicted as not having ASD.

Accuracy:

It is the measure of correct predictions made by the classifier. Accuracy is the number of correctly identified predictions by total number of predictions:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Precision:

It measures the accuracy of positive predictions. It is the ratio of true positive out of the total observed positive.

$$Precision = \frac{TP}{TP + FP}$$

Recall/Sensitivity: This is also called true positive rate. It is the proportion of samples that are genuinely positive by all positive results obtained during the test.

$$Recall = \frac{TP}{TP + FN}$$

F-**Measure:** The F-score (or F-measure) considers both the precision and the recall of the test to compute the score. The traditional or balanced F-score (F1 score) is the harmonic mean of the precision and recall:

$$F - Measure\ (F1) = \frac{2 \times Precision}{Precision + Recall} \quad (4)$$

RESULTS AND DISCUSSION

In this part, we think about the expectation exactness of this paper with best in class research. The greater part of the past exploration works depend on adaptation 1 ASD dataset . Among them just Baranwal et al. have thought about include decrease while keeping the exactness most extreme. Supposedly, Thabtah et al. dealt with variant 2 juvenile and grown-up datasets as it were, what's more, applied strategic relapse classifier. The creators applied chi-squared and data gain include positioning methods to distinguish the main ascribes furthermore, accomplished almost 100% exactness for juvenile dataset and 97.58% exactness for grown-up dataset. Our proposed model consist of the accuracy , precision , recall , f measure in the maximum efficient output we could derive form the MLP algorithm.

ALGORITHM	RECALL	ACCURACY	F1
SVM	91	90	91
MLP	99	99	98
RML	92	91	90
CNN	93	92	93
RF	89	89	89
LR	90	91	90
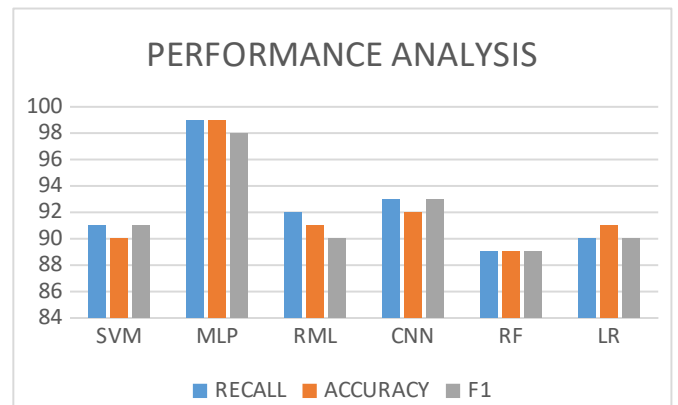
CONCLUSION

In this paper, we have dissected the ASD datasets. We apply most famous five component choice techniques to infer less highlights from ASD datasets yet keeping up with cutthroat execution. We find that Alleviation F highlight choice strategy beats among others. In our trial arrangement, we increment the property numbers bit by bit and afterward apply different characterization strategies. We find that MLP beats among any remaining classifiers utilizing our procedure and approach. The proposed technique utilizing Multi-layer Perceptron (MLP) for Mental imbalance Range Problem (ASD) characterization holds guarantee as an important device in supporting early finding and mediation for people with ASD. By bridling the force of AI and brain organizations, this approach offers an information driven and objective method for recognizing ASD designs in light of different arrangements of elements. The heartiness and adaptability of MLP models make them reasonable for taking care of complicated and heterogeneous ASD datasets, while their interpretability works with bits of knowledge into the basic variables adding to ASD grouping choices. Nonetheless, it is basic to recognize the restrictions and difficulties related with this methodology, including the requirement for enormous and different datasets, model interpretability, and moral contemplations encompassing information security and predisposition. By the by, with progressing headways in AI calculations and medical care innovation, combined with interdisciplinary coordinated effort between specialists, clinicians, and partners, MLP-based strategies for ASD arrangement can possibly fundamentally upgrade demonstrative precision, customized therapy arranging, and backing administrations for people with ASD and their families. Proceeded with examination and approval endeavors are fundamental to understand the maximum capacity of this methodology in further developing results and personal satisfaction for people impacted by ASD.

References

1. Allison, C., Auyeung, B., Baron-Cohen, S.: Toward Brief "Red Flags" for Autism Screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 Cases and 3,000 Controls. Tech. rep. (2012)

2. Baranwal, A., Vanitha, M.: Autistic spectrum disorder screening: Prediction with machine learning models. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) pp. 1 – 7 (2020)

3. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E.: The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. Journal of Autism and Developmental Disorders 31(1), 5–17 (2001)

4. Basu, K.: Autism Detection in Adults Dataset. URL shorturl.at/ekyTW

5. Chawla, N.V.: Data Mining for Imbalanced Datasets: An Overview. In: Data Mining and Knowledge Discovery Handbook, pp. 875–886. Springer US, Boston, MA (2009)

6. Duda, M., Ma, R., Haber, N., Wall, D.P.: Use of machine learning for behavioral distinction of autism and ADHD. Translational Psychiatry 6(2), e732–e732 (2016)

7. Erkan, U., Thanh, D.: Autism spectrum disorder detection with machine learning methods. Current Psychiatry Reviews 15, 297 – 308 (2019)

8. Fischbach, G., Neuron, C.L.: The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Elsevier (2010)

9. Geschwind, D., Sowinski, J., Lord, C.: Letters to the Editor 463. The American Journal, 2001

10. Hossain, M.D., Kabir, M.A.: Detecting child autism using classification techniques. In: Proceedings of the 17th World Congress on Medical and Health Informatics (MedInfo 2019), vol. 264, pp. 1447–1448. IOS (2019)

11. Islam, M.R., Kabir, M.A., Ahmed, A., Kamal, A.R.M., Wang, H., Ulhaq, A.: Depression detection from social network data using machine learning techniques. Health information science and systems 6(1), 8 (2018)

12. Le Couteur, A.L., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., Rutter, M., Bailey, A.: Autism as a strongly genetic disorder evidence from a british twin Study. Psychological Medicine 25(1), 63–77 (1995)

13. Lord, C., Rutter, M., Le Couteur, A.: Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders 24(5), 659–685 (1994)

14. McNamara, B., Lora, C., Yang, D., Flores, F., Daly, P.: Machine Learning Classification of Adults with Autism Spectrum Disorder (accessed June 25, 2020). URL shorturl.at/hEFJY

15. Parikh, M.N., Li, H., He, L.: Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data. Frontiers in Computational Neuroscience 13 (2019)

16. Pratap, A., Kanimozhiselvi, C.S., Vijayakumar, R., Pramod, K.V.: Soft Computing Models for the Predictive Grading of Childhood Autism-A Comparative Study. Tech. Rep. 4 (2014) 12 Md Delowar Hossain et al.

17. Raj, S., Masood, S.: Analysis and detection of autism spectrum disorder using machine learning techniques. Procedia Computer Science 167, 994 – 1004 (2020)

18. Thabtah, F.: Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. dl.acm.org Part F129311, 1– 6 (2017)

19. Thabtah, F.: ASD Tests. A mobile app for ASD screening. (accessed January 10, 2020). URL https://www. asdtests.com/

20. Thabtah, F.: ASD Dataset (accessed January 13, 2020). URL https://fadifayez.com/autism-datasets/

21. Thabtah, F.: ASD Dataset- UCI machine learning repository, 2017 (accessed January 13, 2020). URL https: //archive.ics.uci.edu/ml

22. Thabtah, F., Abdelhamid, N., Peebles, D.: A machine learning autism classification based on logistic regression analysis. Health Information Science and Systems 7(1), 1–11 (2019)

23. Thabtah, F., Peebles, D.: A new machine learning model based on induction of rules for autism detection. Health informatics journal 26(1), 264–286 (2020)

24. Wiggins, L.D., Reynolds, A., Rice, C.E., Moody, E.J., Bernal, P., Blaskey, L., Rosenberg, S.A., Lee, L.C., Levy, S.E.: Using Standardized Diagnostic Instruments to Classify Children with Autism in the Study to Explore Early Development. Journal of Autism and Developmental Disorders 45(5), 1271–1280 (2015)

25. Zunino, A., Morerio, P., Cavallo, A., Ansuini, C., Podda, J., Battaglia, F., Veneselli, E., Becchio, C., Murino, V.: Video Gesture Analysis for Autism Spectrum Disorder Detection. In: 24th International Conference on Pattern Recognition (ICPR), vol. 2018-August, pp. 3421–3426. IEEE (2018)