Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 29s (2025)
https://www.jneonatalsurg.com

Multi-Scaler Fusion Framework for Pediatric Autism Prediction Using Ensemble Machine Learning and Feature Attribution

M. Prasanthi Kumari[1], Dr. D Mohan Reddy[2]

[1] M. Tech Student, Department of Computer Science and Engineering, Kakinada Institute of Technology and Science

[2] Phd, Professor & Principal, Department of Computer Science and Engineering, Kakinada Institute of Technology and Science

Cite this paper as: M. Prasanthi Kumari, Dr. D Mohan Reddy, (2025) Multi-Scaler Fusion Framework for Pediatric Autism Prediction Using Ensemble Machine Learning and Feature Attribution, *Journal of Neonatal Surgery*, 14 (29s), 368-376

ABSTRACT

In this study, we provide a machine learning framework for ASD early diagnosis that makes use of four remarkable feature scaling methods and eight main machine learning algorithms. The suggested design makes use of four outstanding feature scaling (FS) methods: quantile transformer (QT), power transformer (PT), normaliser, and max abs scaler (MAS). For the feature-scaled datasets, eight trustworthy machine learning methods are utilised: AdaBoost (AB), Random forest (RF), decision Tree (DT), k-Nearest neighbours (KNN), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), guide Vector system (SVM), and Linear Discriminant analysis (LDA). The study utilises four standard datasets on ASD: infants, young adults, children, and adults. Basic function selection approaches and primary type algorithms for every ASD dataset are observed. To get the most reliable results, use the normaliser FS on younger children and the QT FS method on older adults and teens. In order to rank the important attributes in order of importance, four feature selection techniques (FSTs) are used: information gain attribute evaluator (IGAE), gain ratio attribute evaluator (GRAE), relief F attribute evaluator (RFAE), and correlation attribute evaluator (CAE). These FSTs are used to assess the risk factors for autism spectrum disorder (ASD). Results from using the proposed paradigm for early ASD diagnosis are better than those from using current methods.

Keywords: *Autism Spectrum Disorder (ASD), Machine Learning (ML), Feature Scaling (FS), Feature Selection Techniques (FSTs), Classification Algorithms, Early Detection.”*

1. INTRODUCTION

A neurodevelopmental issue known as "Autism Spectrum disorder (ASD)" causes a person to struggle in their relationships and social interactions from a young age because it affects brain development. The symptoms of autism spectrum disorder (ASD) may range greatly in intensity and are defined by restricted and repetitive activities [1]. The lengthy and subjective behavioural evaluations used to diagnose autism spectrum disorder add another layer of complexity to the process. However, depending on severity and access to healthcare, some cases of ASD may be undiagnosed until later in life, even though this is not uncommon and is often done by the age of [2]. Early detection technologies are exist, but their practical applicability is typically limited, which causes treatments to be postponed [3].

Recently, "machine learning (ML)" has emerged as a practical method for diagnosing ASD. Diseases such as diabetes, stroke, and coronary heart failure may now be better predicted with the use of machine learning, which allows for earlier detection and treatment. Reduce reliance on subjective scientific evaluations using this paper's machine learning method to ASD analysis.

Use of four representative ASD datasets spanning many age groups (toddlers, kids, teens, and adults) is central to the study. Encoding class variables and handling missing values are both part of the preprocessing procedure. Utilising four feature scaling techniques—"Quantile Transformer (QT), power Transformer (PT), Normaliser, and Max Abs Scaler (MAS)"—to enhance type overall performance is mandated [6]. Ultimately, the most effective models for predicting autism spectrum disorder are determined using eight machine learning classifiers: "AdaBoost (AB)", "Random forest (RF), choice Tree (DT), k-Nearest neighbours (KNN), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), support Vector machine (SVM), and Linear Discriminant analysis (LDA)" [7].

Additionally, the study finds important ASD risk indicators and examines the effect of several feature selection procedures on class accuracy. Four feature selection approaches are used to find the best features in each dataset: "info gain attribute Evaluator (IGAE)," "gain Ratio attribute Evaluator (GRAE)," "relief F attribute Evaluator (RFAE)," and "Correlation characteristic Evaluator (CAE)" [3]. By combining several approaches, the suggested framework hopes to improve ASD detection and push paediatric healthcare research forward [1].

2. RELATED WORK

Complex non -violation of neurodevelopmental disorders known as an ASD represents impressive diagnostic obstacles, especially in the first years of the child's life. Relying on the judgment of the physician and observing behavior in traditional diagnostic techniques often leads to an inconsistent or delayed diagnosis. As a result, ASD detection focused on the use of AI and ML technologies has changed to accelerate, more accurate and reliable.

The shortcomings of current procedures and the need for more standardized, scalable instruments were brought to light in extensive review of diagnostic approaches in early childhood Alrehili et al. [1]. Scientists are exploring EHRS and predictive algorithms to solve these problems. The use of ML models for electronic data on health recording earlier to the age of one has shown that Engelhard et al. [2], providing hope for the development of early diagnostic tools for autism spectrum disorder.

We also had a success with models that are based on language. To better detect autistic symptoms in young people, Themistocleous et al. [3] They used machine learning and processing of natural language (NLP) to explore narrative speech formulas. In addition, comprehensive research ASD should include many data ways, including speech, movement and behavior, as emphasized by the systematic evaluation of diagnostic tools based on AI Joudar et al. [4].

The need for effective and expandable diagnostic frames is further demonstrated by studies of prevalence, such as Salai et al. [5] that emphasizes the worldwide range of autism. Sheldrick et al. [6] They showed that in timely intervention settings, screening approaches with AI improvements significantly increased the identification rate of ASD and reduced irregularities in the field of health, especially in disadvantaged groups.

According to Srinivasan et al. There is a high need for easily accessible and automated diagnostic tools. [7] who looked at the opinions of the carer. Megerian et al. [8] Evaluation of commercially available diagnostic equipment driven AI.

Another promising area is the analysis that relies on gestures. GOPI et al. Introduced by GOPI et al. [9]. At the same time, virtual behavior evaluation using AI platform has increased the diagnostic approach to rural and distant places through Telehealth -based techniques, as discussed by Liu and MA [10].

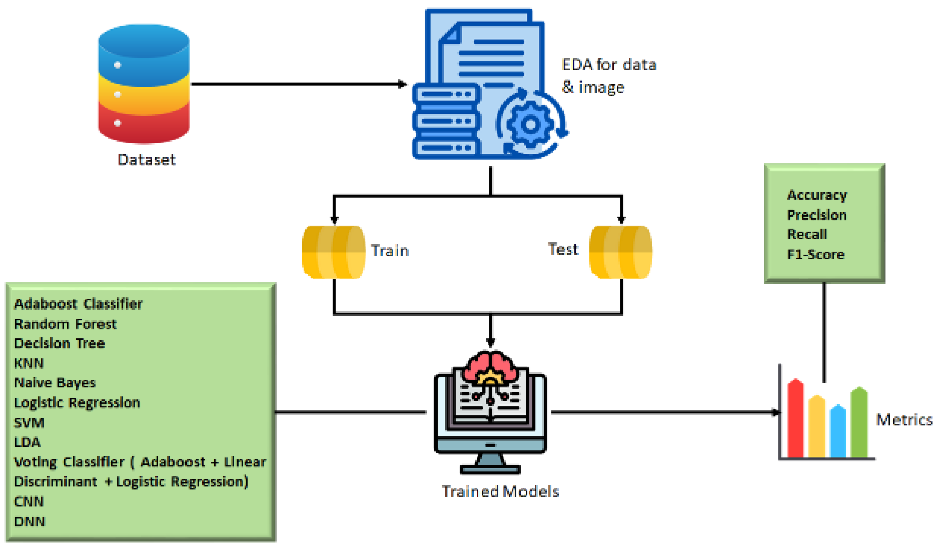
Another important factor to consider is AI. Trust and openness in the diagnosis of ASD are essential for the adoption of a clinical physician and therefore Jeon et al. [11] He built a machine learning system that includes explained artificial intelligence (XAI). Further evidence that a relatively small amount of clinical and contextual data, if operated through ML algorithms, can provide reliable predictions asd Rajagopalan et al. [12].

Kumar and Bhattacharya [13] examined regional differences in ASD diagnostics and emphasized the need for regionally and cultural -specific AI models. It is necessary to have multidimensional models that can distinguish ASD from similar diseases, because Mr. et al. [14] They emphasized neurological co -morbidity that can make it difficult to diagnose. Finally, a study of Lucas et al. [15] He was concerned with how parents feel about genetic testing and how important it is to make diagnostic systems driven by AI to be transparent and secure data.

These findings as a whole show that AI and ML change the game in terms of identifying ASD. They make it easier, more accurate and previously interfered and are also more accessible, especially when they are built with inclusivity, explanability and clinical significance.

3. MATERIALS AND METHODS

Adaboost, Random forest, selection Tree, KNN, Naïve Bayes, Logistic Regression, SVM, LDA, voting Classifier, CNN, and DNN are some of the machine learning algorithms that are part of the proposed approach to early ASD diagnosis. The goal is to have a green framework. It applies function scaling algorithms while evaluating their effectiveness on four ASD datasets: toddlers, adolescents, youngsters, and adults. Furthermore, in order to prioritise exceptional improvements, four unique selection procedures are utilised: "IGAE, GRAE, RFAE, and CAE." Research has shown that ASD diagnosis powered by AI may improve diagnostic accuracy and make early intervention easier [1] [3]. Clinical applications benefit from enhanced interpretability when machine learning and explainable AI are combined [4] [11]. Improving the accessibility of ASD diagnosis, the technique also aids distant assessment [10]. The goal of this strategy is to use AI-driven models to improve ASD identification and reduce the expenses associated with misdiagnosis.



"Fig.1 Proposed Architecture"

A (Fig.1)machine learning pipeline is shown in this diagram. The process starts with a dataset and continues with EDA for the data and photos. Separate sets, one for training and one for testing, are made from the preprocessed data. For model training, a number of machine learning algorithms are used, such as "Adaboost, Random Forest, Decision Tree, KNN, Naive Bayes, Logistic Regression, SVM, LDA, CNN, and DNN". Lastly, measures like "Accuracy, Precision, Recall, and F1-Score" are used to assess the efficacy of these trained models.

i) Dataset Collection:

The process of acquiring datasets involves importing both raw and image records that are relevant to autism spectrum disorder identification. Raw data comes from popular ASD datasets "(toddlers, adolescents, children, Adults)" and includes patient information, behavioural styles, and diagnostic functions. In order to facilitate AI-driven judgement, image facts include expressions on the face, hand gestures, and movement patterns. In preparation for using them in machine learning-based ASD detection, the datasets undergo preprocessing to facilitate feature extraction and model training.

ii) EDA for Raw Data:

The term "exploratory data analysis (EDA)" refers to a variety of methods used to identify and prepare raw data for use in machine learning models. To guarantee the accuracy of the data, we first check for null values and fill them in using imputation methods. Encoding techniques, like One-hot Encoding or Label Encoding, convert specific values into numerical representations. Data visualisation techniques, such as scatter plots, field plots, and histograms, make it easier to examine patterns, outliers, and the distribution of data. Lastly, the use of heatmaps and statistical approaches for correlation analysis helps to clarify the linkages between features. This, in turn, makes it easier to choose the most relevant qualities for ASD detection, which improves the performance and interpretability of the model.

iii) EDA for image data

To maximise the effectiveness of versions, "exploratory data analysis (EDA)" for picture data includes pre-processing and showing photographs. It all starts with picture scaling, which ensures uniform dimensions across the board and boosts computational efficiency and uniformity. Statistical augmentation methods, such as flipping, rotating, scaling, and changing the brightness, enhance generalisability by increasing dataset heterogeneity and reducing the likelihood of overfitting. The distribution of pixels, comparison levels, and class imbalances may be better understood with the use of visualisation tools like histograms and sample picture displays. These steps guarantee top-notch advanced input data, which improves feature extraction for ML models, particularly for ASD detection using CNN and DNN deep learning frameworks. Predictions are more accurate and reliable when exploratory information analysis is done well.

iv) Training and Testing:

A crucial way to evaluate the effectiveness of versions in machine learning is to divide data into training and testing units. In most cases, academic records make up 70–80% of the total, while recreational records make up 20–30%. The training set helps the model see patterns, while the test set assesses its ability to generalise using new data. The fair distribution of lessons in picture-based datasets is ensured by this department. Appropriate records division prevents overfitting and guarantees the best model performance on real-world datasets for "Autism Spectrum disorder (ASD)" identification. Class proportions are maintained by stratified splitting procedures, which improve dependability and robustness in categorisation and prediction duties.

v) Algorithms:

“AdaBoost:” Adaptive Boosting (AdaBoost) augments vulnerable classifiers, usually selection stumps, by using iteratively focusing on misclassified instances to elevate accuracy. it is extensively applied for autism prognosis owing to its proficiency in handling unbalanced data successfully [1][4].

“Random Forest:” Random forest is an ensemble learning method that generates several decision trees to beautify category precision. It is utilized in autism identification with the aid of the analysis of behavioral and genetic data for early prognosis with significant reliability [3] [7].

“Decision Tree:” decision trees are hierarchical structures hired for class purposes. Within the identification of "autism spectrum disorder (ASD)", they facilitate the recognition of critical diagnostic factors thru the evaluation of organized medical data, hence improving early detection methodologies. [3][6].

“K-Nearest Neighbors (KNN):” KNN is a non-parametric method that categorizes data points primarily based on their closeness. It’s far applied in ASD research to identify trends in behavioral and digital health record data for early autism diagnosis [2][5].

“Naïve Bayes:” Naïve Bayes is a probabilistic classifier derived from Bayes' theorem. It is efficacious for ASD identity, examining linguistic patterns and clinical characteristics to forecast autism traits in children [3][6].

“Logistic Regression:” Logistic Regression forecasts binary results, which include the presence or absence of ASD, via modeling the relationships among risk variables. Its miles regularly applied in ASD screening to discover persons at high threat [1] [4].

“Support Vector Machine (SVM):” "support Vector machine (SVM)" is proficient in diagnosing "Autism Spectrum disorder (ASD)" instances by using figuring out an excellent hyperplane. It’s miles utilized in the evaluation of speech and behavioral facts for specific autism prediction [3][9].

“Linear Discriminant Analysis (LDA):” LDA is a supervised classification method that identifies linear combinations of capabilities to beautify class separation, subsequently increasing the accuracy of ASD diagnosis through behavioral and genetic data [3] [6].

“Voting Classifier (AdaBoost + LDA + Logistic Regression):” A voting Classifier amalgamates AdaBoost, LDA, and Logistic Regression to enhance ASD identity. This ensemble method improves prediction accuracy by using many methods. [3][8].

“Convolutional Neural Network (CNN):” "Convolutional Neural Networks (CNNs)" analyze picture statistics by examining facial expressions and movement styles for the detection of "Autism Spectrum disorder (ASD)". They derive profound features for precise category of autism-related characteristics. [7][10].

“Deep Neural Network (DNN):” "Deep Neural Networks (DNNs)" are multi-layered architectures that figure tricky styles from behavioral and speech facts, hence improving autism diagnosis via the identity of nuanced variations in "Autism Spectrum disorder (ASD)" characteristics [8][11].

4. RESULTS & DISCUSSION

Accuracy: A test capacity towards create a proper difference between healthy & sick cases is a measure of accuracy. We can determine accuracy of a test through calculating proportion of cases undergoing proper positivity & genuine negative. It is possible towards express this mathematically:

$$"Accuracy" = \frac{"\text{TP} + \text{TN}"}{"\text{TP} + \text{FP} + \text{TN} + \text{FN}"} (1)$$

Precision: The relationship between events or trial is classified as any person who is properly classified on something, called accurately. Therefore, it is a formula towards consider determining the precision:

$$"Precision" = \frac{"\text{True Positive}"}{"\text{True Positive } + \text{False Positive}"} (2)$$

Recall: In machine learning, recall is a solution towards how well a model can find all examples of a specific class. ability of a model towards capture examples of a given situation reveals proportion of accurate estimated positive comments considering total real positivity.

$$"Recall" = \frac{"\text{TP}"}{"\text{TP } + \text{ FN}"} (3)$$

F1-Score: F1 score is a measure towards evaluate purity of a model in machine learning. It takes recall & precision of a model & mixes them. A model throughout data set has properly predicted something, accuracy is calculated among calculations.

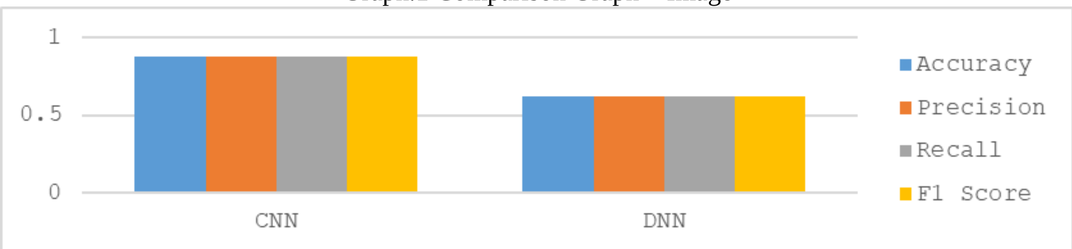
$$"F1\ Score" = "2" * \frac{"Recall \text{ X } Precision"}{"Recall + Precision"} * "100" (1)$$

Each method is assessed using four performance measures: accuracy, precision, recall, and F1-score. These metrics are shown in Tables 1 through 5. All measures show that the CNN & Voting Classifier is the best. Additionally, tables are provided for the purpose of comparing the metrics of other algorithms.

“Table.1 Performance Evaluation – Image”

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
CNN	**0.8788**	**0.8789**	**0.8789**	**0.8789**
DNN	0.6173	0.6173	0.6173	0.6173

“Graph.1 Comparison Graph – Image”



“Table.2 Performance Evaluation Table - Data (Quantile Transformer)”

ML Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
AdaBoost Classifier	1.000	1.00	1.000	1.000
Random Forest Classifier	0.949	1.00	0.870	0.930
Decision Tree Classifier	0.898	0.95	0.792	0.864
KNN	0.831	0.95	0.679	0.792
Naive Bayes	0.864	0.80	0.800	0.800
Logistic Regression	1.000	1.00	1.000	1.000
SVC	0.983	1.00	0.952	0.976
LDA	0.983	1.00	0.952	0.976
Voting Classifier	1.000	1.00	1.000	1.000

“Table.3 Performance Evaluation Table – Data (Power Transformer)”

ML Model	Accuracy	Precision	Recall	F1 Score
AdaBoost	1.000	1.00	1.000	1.000
Random Forest	0.949	1.00	0.870	0.930
Decision Tree	0.881	0.90	0.783	0.837
KNN	0.831	0.95	0.679	0.792
Naive Bayes	0.864	0.85	0.773	0.810
Logistic Regression	1.000	1.00	1.000	1.000
SVM	0.966	1.00	0.909	0.952
LDA	1.000	1.00	1.000	1.000
Voting Classifier	1.000	1.00	1.000	1.000

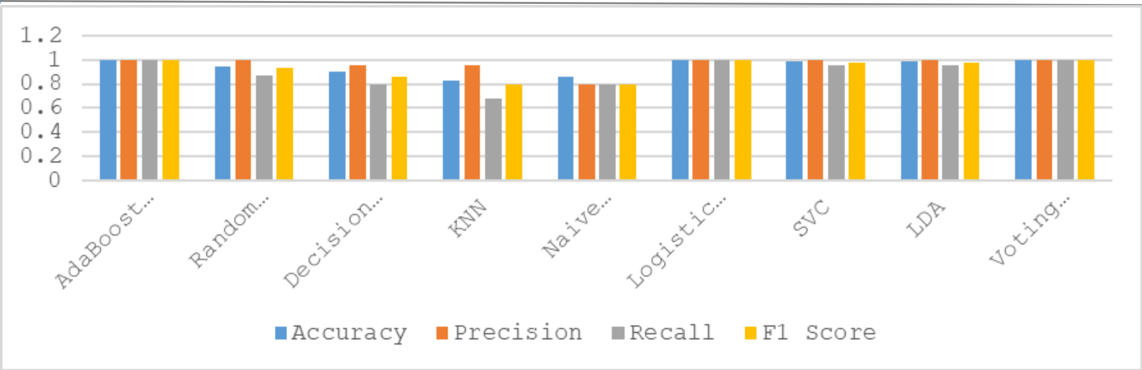
“Table.4 Performance Evaluation Table – Data (MaxAbsScaler - MAS)”

ML Model	Accuracy	Precision	Recall	F1 Score
AdaBoost	1.000	1.00	1.000	1.000
Random Forest	0.949	1.00	0.870	0.930
Decision Tree	0.864	0.90	0.750	0.818
KNN	0.847	1.00	0.690	0.816
Naive Bayes	0.864	0.85	0.773	0.810
Logistic Regression	1.000	1.00	1.000	1.000
SVC	0.983	1.00	0.952	0.976
LDA	1.000	1.00	1.000	1.000
Voting Classifier	1.000	1.00	1.000	1.000

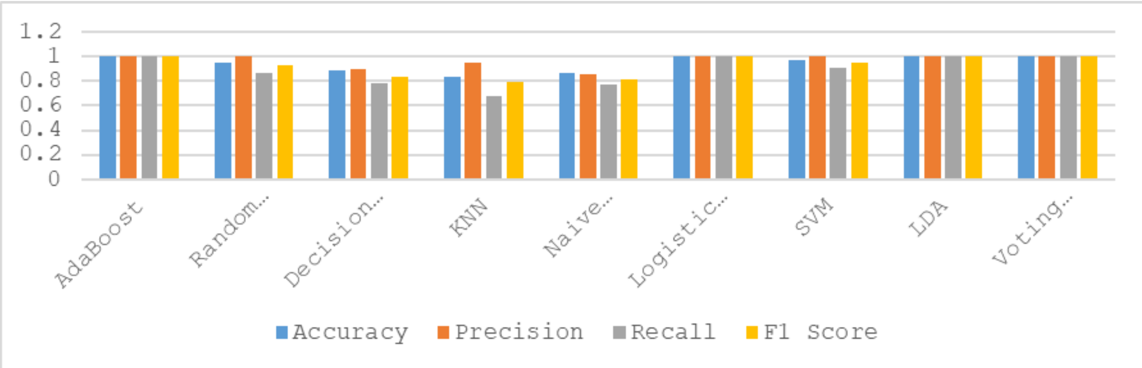
“Table.5 Performance Evaluation Table – Data (Normalizer)”

ML Model	Accuracy	Precision	Recall	F1 Score
AdaBoost	0.966	1.00	0.909	0.952
Random Forest	0.915	0.95	0.826	0.884
Decision Tree	0.780	0.75	0.652	0.698
KNN	0.678	0.95	0.514	0.667
Naive Bayes	0.610	0.45	0.429	0.439
Logistic Regression	0.576	0.65	0.419	0.510
SVC	0.661	0.25	0.500	0.333
LDA	0.797	0.70	0.700	0.700
Voting Classifier	0.864	0.90	0.750	0.818

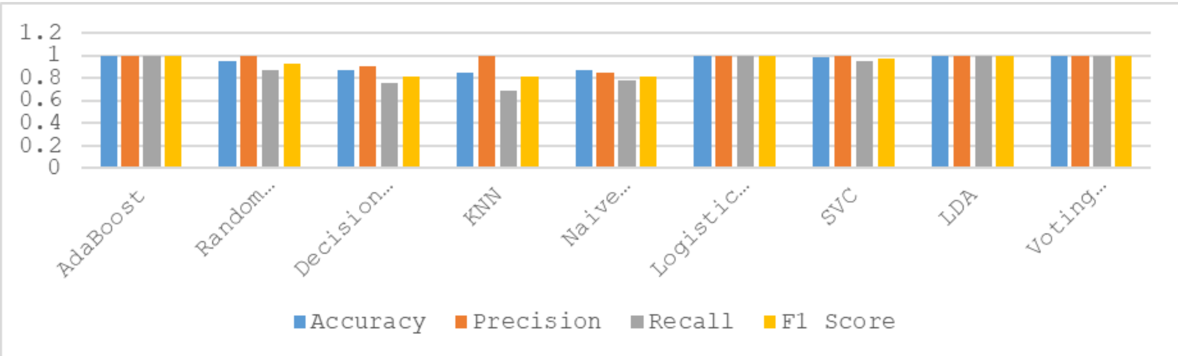
“Graph.2 Comparison Graph - Data (Quantile Transformer)”



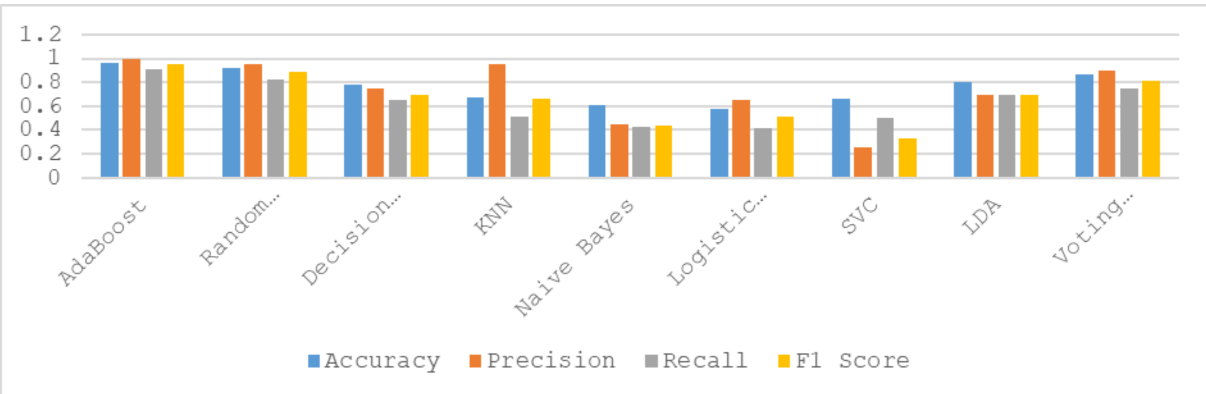
"Graph.3 Comparison Graph – Data (Power Transformer)"



"Graph.4 Comparison Graph – Data (MaxAbsScaler - MAS)"



"Graph.5 Comparison Graph – Data (Normalizer)"



Blue represents accuracy, orange recall, grey represents precision, and bright yellow represents the F1-Score in graphs 1–5. When compared to the other models, the CNN & Voting Classifier consistently achieves the best results across all criteria. The aforementioned graphs provide a visual representation of these facts.

5. CONCLUSION

In this study, we provide a machine learning framework for detecting ASD in a wide range of age groups, including newborns, children, adolescents, and adults. Our research shows that prediction models trained using ML are useful tools for this task. The ASD datasets were scaled using four wonderful characteristic scaling approaches (QT, PT, normaliser, MAS) after the initial data processing was finished, and then eight specific machine learning classifiers "(AB, RF, DT, KNN, GNB, LR, SVM, LDA)" were used for classification. Our next step was to compare the category performance of all the feature-scaled datasets and choose the most effective methods for selecting characteristics and classes. To back up the results of the experiment, we checked a number of statistical measures, including precision, recall, "Matthews's correlation coefficient (MCC)," kappa score, log loss, ROC, and accuracy. Consequently, our recommended prediction models using machine learning techniques may provide doctors with a valuable tool to accurately identify individuals with ASD, regardless of their age group. In addition, using four separate "FSTs (IGAE, GRAE, RFAE, and CAE)", function significance values were calculated to investigate the most remarkable functions for ASD prediction. Consequently, medical professionals will be able to take into account the most important factors when screening for ASD thanks to the experimental assessment of these investigations. We want to improve the diagnosis of ASD and other neurodevelopmental illnesses in the future by collecting more data on the disorder and developing a more comprehensive prediction model that can be used with people of all ages.

We want to improve ASD detection in the future by acquiring more diverse and comprehensive datasets to enhance model generalisability across all age groups. Prediction accuracy and interpretability may be enhanced with the use of deep learning models and explainable AI approaches. Enhancing healthcare professionals' abilities for early and accurate ASD diagnosis may be achieved by expanding the framework to become aware of new neurodevelopmental illnesses and establishing a real-time screening tool for practical software.

REFERENCES

[1] Alrehaili, R. A., ElKady, R. M., Alrehaili, J. A., & Alreefi, R. M. (2023). Exploring early childhood autism spectrum disorders: A comprehensive review of diagnostic approaches in young children. Cureus, 15(12).

[2] Engelhard, M. M., Henao, R., Berchuck, S. I., Chen, J., Eichner, B., Herkert, D., ... & Dawson, G. (2023). Predictive value of early autism detection models based on electronic health record data collected before age 1 year. JAMA network open, 6(2), e2254303-e2254303.

[3] Themistocleous, C. K., Andreou, M., & Peristeri, E. (2024). Autism Detection in Children: Integrating Machine Learning and Natural Language Processing in Narrative Analysis. Behavioral Sciences, 14(6), 459.

[4] Joudar, S. S., Albahri, A. S., Hamid, R. A., Zahid, I. A., Alqaysi, M. E., Albahri, O. S., & Alamoodi, A. H. (2023). Artificial intelligence-based approaches for improving the diagnosis, triage, and prioritization of autism spectrum disorder: a systematic review of current trends and open issues. Artificial Intelligence Review, 56(Suppl 1), 53-117.

[5] Salari, N., Rasoulpoor, S., Rasoulpoor, S., Shohaimi, S., Jafarpour, S., Abdoli, N., ... & Mohammadi, M. (2022). The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis. Italian Journal of Pediatrics, 48(1), 112.

[6] Sheldrick, R. C., Carter, A. S., Eisenhower, A., Mackie, T. I., Cole, M. B., Hoch, N., ... & Pedraza, F. M. (2022). Effectiveness of screening in early intervention settings to improve diagnosis of autism and reduce health disparities. JAMA pediatrics, 176(3), 262-269.

[7] Srinivasan, S., Ekbladh, A., Freedman, B., & Bhat, A. (2021). Needs assessment in unmet healthcare and family support services: A survey of caregivers of children and youth with autism spectrum disorder in Delaware. Autism Research, 14(8), 1736-1758.

[8] Megerian, J. T., Dey, S., Melmed, R. D., Coury, D. L., Lerner, M., Nicholls, C. J., ... & Taraman, S. (2022). Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. NPJ digital medicine, 5(1), 57.

[9] Gopi, V. P., Francis, B., & Thomas, A. (2024). Early-stage identification of autism in children using gesture

monitoring based on artificial intelligence. In Advances in Artificial Intelligence (pp. 491-522). Academic Press.

[10] Liu, M., & Ma, Z. (2022). A systematic review of telehealth screening, assessment, and diagnosis of autism spectrum disorder. Child and adolescent psychiatry and mental health, 16(1), 79.

[11] Jeon, I., Kim, M., So, D., Kim, E. Y., Nam, Y., Kim, S., ... & Moon, J. (2024). Reliable Autism Spectrum Disorder Diagnosis for Pediatrics Using Machine Learning and Explainable AI. Diagnostics, 14(22), 2504.

[12] Rajagopalan, S. S., Zhang, Y., Yahia, A., & Tammimies, K. (2024). Machine Learning Prediction of Autism Spectrum Disorder From a Minimal Set of Medical and Background Information. JAMA Network Open, 7(8), e2429229-e2429229.

[13] Kumar, A., & Bhattacharya, S. (2024). Unveiling autism spectrum disorder in South East Asia through a public health Lens. Frontiers in Child and Adolescent Psychiatry, 3, 1489269.

[14] Pan, P. Y., Bölte, S., Kaur, P., Jamil, S., & Jonsson, U. (2021). Neurological disorders in autism: A systematic review and meta-analysis. Autism, 25(3), 812-830.

[15] Lucas, H. M., Lewis, A. M., Lupo, P. J., & Schaaf, C. P. (2022). Parental perceptions of genetic testing for children with autism spectrum disorders. American Journal of Medical Genetics Part A, 188(1), 178-186.