IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 December 2022, accepted 25 December 2022, date of publication 26 December 2022,
date of current version 16 February 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3232490

RESEARCH ARTICLE

A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders

S. M. MAHEDY HASAN[1], MD PALASH UDDIN[2,3], (Member, IEEE), MD AL MAMUN[1], (Senior Member, IEEE), MUHAMMAD IMRAN SHARIF[4], ANWAAR ULHAQ[5], AND GOVIND KRISHNAMOORTHY[6]

[1]Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh
[2]Department of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur 5200, Bangladesh
[3]School of Information Technology, Deakin University, Geelong, VIC 3220, Australia
[4]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Punjab 47040, Pakistan
[5]School of Computing, Mathematics and Engineering, Charles Sturt University, Port Macquarie, NSW 2444, Australia
[6]School of Psychology and Wellbeing, University of Southern Queensland, Ipswich, QLD 4305, Australia

Corresponding author: Anwaar Ulhaq (aulhaq@csu.edu.au)

This work was supported by the Regional Australia Mental Health Research and Training Institute, Manna Institute, NSW, Australia, under Grant 0000103935.

ABSTRACT Autism Spectrum Disorder (ASD) is a type of neurodevelopmental disorder that affects the everyday life of affected patients. Though it is considered hard to completely eradicate this disease, disease severity can be mitigated by taking early interventions. In this paper, we propose an effective framework for the evaluation of various Machine Learning (ML) techniques for the early detection of ASD. The proposed framework employs four different Feature Scaling (FS) strategies i.e., Quantile Transformer (QT), Power Transformer (PT), Normalizer, and Max Abs Scaler (MAS). Then, the feature-scaled datasets are classified through eight simple but effective ML algorithms like Ada Boost (AB), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). Our experiments are performed on four standard ASD datasets (Toddlers, Adolescents, Children, and Adults). Comparing the classification outcomes using various statistical evaluation measures (Accuracy, Receiver Operating Characteristic: ROC curve, F1-score, Precision, Recall, Mathews Correlation Coefficient: MCC, Kappa score, and Log loss), the best-performing classification methods, and the best FS techniques for each ASD dataset are identified. After analyzing the experimental outcomes of different classifiers on feature-scaled ASD datasets, it is found that AB predicted ASD with the highest accuracy of 99.25%, and 97.95% for Toddlers and Children, respectively and LDA predicted ASD with the highest accuracy of 97.12% and 99.03% for Adolescents and Adults datasets, respectively. These highest accuracies are achieved while scaling Toddlers and Children with normalizer FS and Adolescents and Adults with the QT FS method. Afterward, the ASD risk factors are calculated, and the most important attributes are ranked according to their importance values using four different Feature Selection Techniques (FSTs) i.e., Info Gain Attribute Evaluator (IGAE), Gain Ratio Attribute Evaluator (GRAE), Relief F Attribute Evaluator (RFAE), and Correlation Attribute Evaluator (CAE). These detailed experimental evaluations indicate that proper finetuning of the ML methods can play an essential role in predicting ASD in people of different ages. We argue that the detailed feature importance analysis in this paper will guide the decision-making of healthcare practitioners while screening ASD cases. The proposed framework has achieved promising results compared to existing approaches for the early detection of ASD.

INDEX TERMS Autism spectrum disorder, machine learning, classification, feature scaling, feature selection technique.

The associate editor coordinating the review of this manuscript and approving it for publication was Santosh Kumar.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition associated with brain development that starts early

stage of life, impacting a person's social relationships and interaction issues [1], [2]. ASD has restricted and repeated behavioral patterns, and the word spectrum encompasses a wide range of symptoms and intensity [3], [4], [5]. Even though there is no sustainable solution for ASD, simply early intervention and proper medical care will make a significant difference in a kid's development to focus on improving a child's behaviors and skills in communication [6], [7], [8]. Even so, the identification and diagnosis of ASD are really difficult and sophisticated, using traditional behavioral science. Usually, Autism is most commonly diagnosed at about two years of age and can also be diagnosed later, based on its severity [9], [10], [11]. A variety of treatment strategies are available to detect ASD as quickly as possible. These diagnostic procedures aren't always widely used in practice until a severe chance of developing ASD. The authors in [12] provided a short and observable checklist that can be seen at different stages of a person's life, including toddlers, children, teens, and adults. Subsequently, the authors in [13] constructed the ASDTests mobile apps system for ASD identification as fast as possible, depending on a range of questionnaire surveys, Q-CHAT, and AQ-10 methods. Consequently, they also created an open-source dataset utilizing mobile phone app information and submitted the datasets to a publicly accessible website called the University of California-Irvine (UCI) machine learning repository and Kaggle for more development in this area of study. Over the past few years, several studies have been conducted incorporating various Machine Learning (ML) approaches to analyze and diagnose ASD and also other diseases, such as diabetes, stroke, and heart failure prediction as quickly as possible [14], [15], [16]. The authors in [17] analyzed the ASD attributes utilizing Rule-based ML (RML) techniques and confirmed that RML helps classification models boost classification accuracy. The authors in [18] combined the Random Forest (RF) along with Iterative Dichotomiser 3 (ID3) algorithms and produced predictive models for children, adolescents, and adults. The authors in [19] introduced a new evaluation tool, integrating ADI-R and ADOS ML methods, and implemented different attribute encoding approaches to resolve data insufficiency, non-linearity, and inconsistency issues. Another study conducted by the authors in [13] demonstrates a feature-to-class and feature-to-feature correlation value utilizing cognitive computing and implemented Support Vector Machines (SVM), Decision Tree (DT), Logistic Regression (LR) as ASD diagnostic and prognosis classifiers [17]. In addition, the authors in [20] explored traditionally formed (TD) ($N = 19$) and ASD ($N = 11$) cases, in which a correlation-based attribute selection was used to determine the importance of the attributes. In 2015, the authors in [21] investigated ASD and TD children and recognized 15 preschool ASDs using only seven features. Besides that, they conveyed that cluster analysis might effectively analyze complex patterns to predict ASD phenotype and diversity. The authors in [22] contrasted the classifier accuracy of K-Nearest Neighbors (KNN), LR, Linear Discrimination Analysis (LDA), Classification and Regression Trees (CART), Naive Bayes (NB), and SVM for adult ASD prediction. In [23], an ML model via induction of rules was proposed for autism detection, which includes testing on only one dataset and limited comparison. The authors in [17] used LR analysis to build an ML autism classification approach, which also falls into the lack of extensive validation and comparison. The authors in [24] scrutinized autism data and observed that 5 of the overall 65 characteristics are sufficient to detect ASD through attention deficit hyperactivity disorder (ADHD). In 2019, the authors in [25] constructed an RF-based model for the prediction of ASD utilizing behavioral features. In addition, the authors in [26] used LDA and KNN methods to identify ASD Children between the ages of 4 and 11 years. In 2018, the authors in [27] suggested an ASD model based on the RF classifier for children between the ages of 4-11. The authors in [28] evaluated the predictive performance of the Deep Neural Network (DNN) in the diagnosis of ASD utilizing two distinct Adult datasets. In 2019, the authors in [18] constructed a smartphone application programming interface on RF-CART and RF-ID3 for the diagnosis of ASDs of all ages. The authors in [29] assessed the performance of multiple SVM kernels in classifying ASD data for children and explored that the polynomial kernel worked much better. The authors in [1] performed several feature selection techniques on four ASD datasets and found that the SVM classifier performed better for RIPPER-based toddler subset, correlation-based feature selection (CFS) and Boruta CFS intersect (BIC) method-based child subset and CFS-based adult subset. Furthermore, they applied Shapley Additive Explanations (SHAP) method to various feature subsets, which achieved the highest accuracy and ranked their features based on performance. The authors in [30] carried out ensemble ML approaches of Fuzzy K-Nearest Neighbor (FKNN), Kernel Support Vector Machines (KSVM), Fuzzy Convolution Neural Network (FCNN), and Random Forest (RF) to classify Parkinson's disease and ASD. Finally, the classification results are verified utilizing Leave-One-Person-Out Cross Validation (LOPOCV). The authors in [31] performed an evolutionary cultural optimization algorithm to optimize the weights of Artificial Neural Networks (ANN) in classifying three benchmark datasets of autism screening Toddlers, Children, and Adults. The authors in [32] performed an experimental analysis using 16 different ML models, among them, four bio-inspired algorithms, namely, Gray Wolf Optimization (GWO), Flower Pollination Algorithm (FPA), Bat Algorithms (BA), and Artificial Bee Colony (ABC) were employed for optimizing the wrapper feature selection method in order to select the most informative features and to increase the accuracy of the classification models on genetic and personal characteristics datasets. Another study conducted by the authors in [33] combined three benchmark datasets as Toddlers, Adolescents, and Adults and performed a Light Gradient Boosting Machine (LGBM) classifier to classify ASD. The authors in [34] utilized Extreme Learning Machines (ELM) and Random Vector Function Link (RVFL) generalization

techniques to classify the Toddlers, Adolescents, and Adults datasets.

This study gathers four standard ASD datasets (Toddlers, Children, Adolescents, and Adults) and initially preprocesses the datasets (manipulation of missing values and encoding). Then, four Feature Scaling (FS) methods including Quantile Transformer (QT), Power Transformer (PT), Normalizer, and Max Abs Scaler (MAS) are undertaken to map the datasets into an appropriate format for further assessments. Thereafter, the feature-scaled datasets are classified by eight simple but effective classification approaches (AB, RF, DT KNN, Gaussian Naive Bayes (GNB), LR, SVM, and LDA), and the best classification models are identified. Meanwhile, we also explore the significance of the FS methods on each dataset by analyzing the experimental outcomes of the transformed datasets. Afterward, four Feature Selection Techniques (FST) i.e., Info Gain Attribute Evaluator (IGAE), Gain Ratio Attribute Evaluator (GRAE), Relief F Attribute Evaluator (RFAE), and Correlation Attribute Evaluator (CAE) are implemented to calculate the risk factors of ASD and rank the most important features of these feature-scaled Toddlers, Children, Adolescents and Adults datasets. Accordingly, this study suggests that ML methods can be applied to help identify the most significant features of ASD detection based on the FST-based feature importance analysis and this will help physicians diagnose ASD cases accurately. Notice that the work presented in [35] may seem somewhat similar to ours. However, the notable differences are as follows. (i) We consider four promising FS methods (QT, PT, Normalizer, and MAS), whereas the three FS methods (Logarithmic, ZScore, and Sine) used in [35] are obsolete nowadays. (ii) After applying each FS method, we find the best FST from a list of IGAE, GRAE, RFAE, and CAE for each dataset to train the ML models, whereas [35] did not consider any such tuning of the FST methods. (iii) We consider eight simple but effective ML models for the prediction, whereas the ML models used in [35] are archaic in this domain. (iv) Finally, we compare more recent works with our proposed model in contrast to [35]. To this end, the key contributions of this paper are summarized as follows.

- We develop a generalized ML framework for early-stage detection of ASD in people of different ages.
- We solve the imbalanced class distribution issue through Random Over Sampler to avoid the ML models being biased towards the majority class samples.
- We select the best Feature Scaling (FS) method to map individual ASD dataset's feature values to improve the prediction performance.
- We investigate eight simple but effective ML approaches on each feature-scaled ASD dataset, analyze their classification performances and identify the best FS techniques for each ASD dataset.
- Furthermore, we also calculate and analyze the feature importance values on each best feature-scaled ASD dataset based on four FSTs to identify the risk factors for ASD prediction.
- Finally, we also perform extensive experiments and comparisons using four different standard ASD datasets.

The remaining part of the paper is organized as follows. Section 2 demonstrates the proposed research methodology and material used in the study. Section 3 analyzes the detailed experimental outcomes while Section 4 discusses the comparative results of the progressive works in this domain. At last, Section 5 summarizes and concludes the observations and findings.

TABLE 1. Datasets description.

Dataset	No. of Instances	Positive Class	Negative Class	Male	Female
Toddlers	1054	728	326	735	319
Children	292	76	28	208	84
Adolescents	104	185	107	49	49
Adults	704	508	196	367	337

II. MATERIALS & METHODS

A. DATASET DESCRIPTION

We collect the four ASD datasets (Toddlers, Adolescents, Children, and Adults) from the publicly available repositories: Kaggle and UCI ML [36], [37], [38], [39]. The authors in [13] created the ASDTests smartphone app for Toddlers, Children, Adolescents, and Adults ASD screening using QCHAT-10 and AQ-10. The application computes a score of 0 to 10 for every individual, with which the final score is 6 out of 10 which indicates an individual has positive ASD. In addition, ASD data is obtained from the ASDTests app while open-source databases are developed in order to facilitate research in this area. The detailed description of the Toddlers, Children, Adolescents, and Adults ASD datasets are given in Table 1 and Table 2.

B. METHOD OVERVIEW

This research aims to create an effective prediction model using different types of ML methods to detect autism in people of different ages. First of all, the datasets are collected, and then the preprocessing is accomplished via the missing values imputation, feature encoding, and oversampling. The Mean Value Imputation (MVI) method is used to impute the missing values of the dataset. Then, the categorical feature values are converted to their equivalent numerical values using the One Hot Encoding (OHE) technique. Table 1 shows that all four datasets used in this work have an imbalanced class distribution problem. As such, a Random Over Sampler strategy is used to alleviate this issue. After completing the initial preprocessing, the datasets' feature values are scaled using four different FS techniques i.e., QT, PT, Normalizer, and MAS (see their detailed operations in Table 3). The feature-scaled datasets are then classified using eight different ML classification techniques i.e., AB, RF, DT, KNN, GNB, LR, SVM, and LDA. Comparing the classification outcomes of the classifiers on different feature-scaled

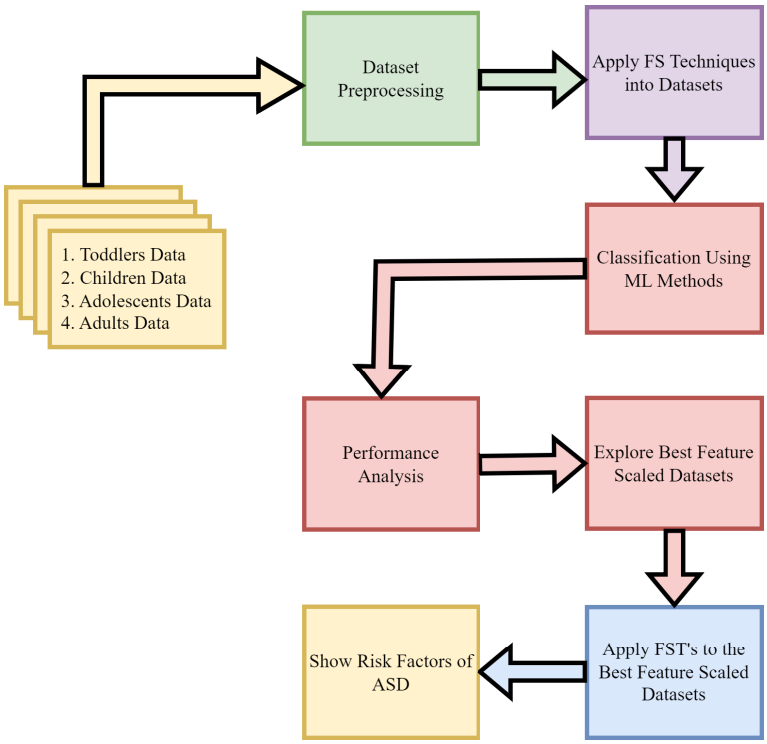


FIGURE 1. Sequential workflow for detecting ASD at an early stage.

TABLE 2. Feature description of the ASD datasets.

Attribute	Type	Description
Age	Number	Adolescents, Children, Adults (years), Toddlers (month)
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean	Whether the case was born with jaundice
Family member with PDD	Boolean	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician etc
Country of residence	String	List of countries in text format
Used the screening app before	Boolean	Whether the user has used a screening app
Screening Method Type	Integer	The type of screening methods chosen based on age category
A1: Question 1 (Q1) Answer	Binary	The answer code of the question based on the screening method used
A2: Question 2 (Q2) Answer	Binary	The answer code of the question based on the screening method used
A3: Question 3 (Q3) Answer	Binary	The answer code of the question based on the screening method used
A4: Question 4 (Q4) Answer	Binary	The answer code of the question based on the screening method used
A5: Question 5 (Q5) Answer	Binary	The answer code of the question based on the screening method used
A6: Question 6 (Q6) Answer	Binary	The answer code of the question based on the screening method used
A7: Question 7 (Q7) Answer	Binary	The answer code of the question based on the screening method used
A8: Question 8 (Q8) Answer	Binary	The answer code of the question based on the screening method used
A9: Question 9 (Q9) Answer	Binary	The answer code of the question based on the screening method used
A10: Question 10 (Q10) Answer	Binary	The answer code of the question based on the screening method used
Screening Score	Integer	The final score obtained based on the scoring algorithm of the screening method used. This was computed in an automated manner
ASD	Boolean	Toddlers, Children, adolescent or Adults diagnosed with ASD

ASD datasets, the best-performing classification methods, and the best FS techniques for each ASD dataset are identified. After those analyses, the ASD risk factors are calculated, and the most important attributes are ranked according to their importance values using four different FSTs i.e., IGAE, GRAE, RFAE, and CAE (see the detailed operations in Table 4). To this end, Fig. 1 represents the proposed research pipeline to analyze the ASD datasets and calculate the risk factors that are most responsible for ASD detection.

C. MACHINE LEARNING METHOD

1) ADA BOOST (AB)

AB is a tree-based ensemble classifier that incorporates many weak classifiers to reduce misclassification errors [41].

TABLE 3. Detailed description of the different FS methods [40].

Name	Definition	Formula
QT	It transforms the variable distribution to normal distribution	$Q(p;\lambda) = \frac{-ln(1-p)}{\lambda}$ (1) where $Q(p;\lambda)$ defines the quantile function, λ denotes intensity and p specifies quartile.
PT	It corrects the skewness of the variable by changing the distribution and making it more Gaussian	$X_i^{\lambda} = \frac{(X_i+1)^{\lambda}-1}{\lambda}$ (2) where X_i specifies the untransformed variable, λ is the boxcox parameter and X_i^{λ} represents the transformed variable.
Normalizer	It works row-wise and converts all the values between 0 and 1	$X_{scaled} = \frac{X}{Sum(X)}$ (3) where X represents the sample value in a column and $Sum(X)$ is the sum of all samples in a particular column.
MAS	It takes each column's absolute maximum value and divides each value in the column by the maximum value.	$X_{scaled} = \frac{X}{Absolute[X_{max}]}$ (4) where X represents each sample value in a column and X_{max} represents the maximum value of this column.

TABLE 4. Detailed description of the different FST methods [40].

Name	Definition	Formula
IGAE	It measures the value of information gain for each attribute concerning the output variable.	$IG(C,A) = H(C) - H(C\|A)$ (5) where H is the information entropy, IG is the information gain, C represents the class and A represents the attribute.
GRAE	It computes the gain ratio value for each attribute with respect to the class variable.	$GR(C,A) = \frac{H(C)-H(C\|A)}{H(A)}$ (6) where GR is the gain ratio.
RFAE	It calculates a feature's worthiness by repeatedly sampling an instance and considering the given attribute's value for the closest instance of the same and different classes.	$R_x = P(DiffX\|DiffClass) - P(DiffX\|SameClass)$ (7) where $P(DiffX\|DiffClass)$ represents the conditional probability value for the different class and $P(DiffX\|SameClass)$ specifies the conditional probability for the same class.
CAE	It evaluates the worth of a feature by calculating Pearson's correlation value for the class variable.	$\rho(X,Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$ (8) where $COV(X,Y)$ represents the covariance, and σ_X and σ_Y denote the standard deviation of X and Y respectively. $\rho(X,Y)$ represents the correlation between X and Y.

It selects the training set and iteratively assigns the weights depending on the previous training precision for retraining the algorithm. In order to train any weak classifier, an arbitrary subset of the full training set is used and AB assigns weights to each instance and classifier. The following equation defines the combination of several weak classifiers:

$$H(x) = Sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \tag{9}$$

where $H(x)$ defines the output of the final model through combining the weak classifiers and $h_t(x)$ represents the output of classifier t for input x and α_t specifies the weight assigned to the classifier. α_t is calculated as follows.

$$\alpha_t = \frac{0.5 * ln(1-E)}{E} \tag{10}$$

where E denote the error rate. The following equation is utilized to update the weights of each training sample-label pair (x_i, y_i).

$$D_{t+1}(i) = \frac{D_t(i) exp(-\alpha_t y_i h_t(x_i))}{Z_t} \tag{11}$$

where D_{t+1} denotes the updated weight, D_t specifies the weight of previous level, and Z_t sum of all weights.

2) RANDOM FOREST (RF)

RF is a decision tree-based ensemble classification method and follows the split and conquer technique in the input dataset to create multiple decision-making trees (known as the forest) [42]. It works in two phases. At first, it creates a forest by combining the 'N' number of decision trees and in the second phase, it makes predictions for each tree generated in the first phase. The working process of the RF algorithm is illustrated below:

1) Select random samples from the training dataset.
2) Construct decision trees for each training sample.
3) Select the value of 'N' to define the number of decision trees.
4) Repeat Steps 1 and 2.
5) For each test sample, find the predictions of each decision tree, and assign the test sample a class value based on majority voting.

3) DECISION TREE (DT)

DT follows a top-down approach to build a predictive model for class values using training data-inducing decision-making rules [43]. This research utilized the information gain method to select the best attribute. Assuming P_i, the probability such that $x_i \epsilon D$, exists to a class C_i, and is predicted by $|Ci, D|/|D|$. To classify instances in the dataset D, the required information is needed, and the following equation calculates it:

$$Info(D) = -\sum_{i=1}^{m} P_i \log_2(P_i) \tag{12}$$

where $Info(D)$ is the average amount of information needed to identify C_i of an instance, $x_i \epsilon D$ and the objective of DT is to divide repeatedly, D, into sub datasets $D_1, D_2 \ldots\ldots\ldots D_n$. The following equation estimates the $Info_A(D)$:

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} * Info(D_j) \tag{13}$$

Finally, the following equation calculates the information gain value:

$$Gain(A) = Info(D) - Info_A(D) \tag{14}$$

4) K-NEAREST NEIGHBORS (KNN)

KNN classifies the test data by utilizing the training data directly by calculating the K value, indicating the number of KNN [43]. For each instance, it computes the distance between all the training instances and sorts the distance. Furthermore, a majority voting technique is employed to assign the final class label to the test data. This research applies Euclidean distance to calculate the distances among instances. The following equation represents the Euclidean distance calculation:

$$D_e = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \tag{15}$$

where D_e indicates the euclidean distance, X_i denotes the testing sample values, Y_i specifies the training sample values and n represents the total number of sample values.

5) GAUSSIAN NaïVE BAYES (GNB)

GNB algorithm follows a normal distribution and is used for classification when all the data values of a dataset are numeric [43]. To compute the probability values of any instance with respect to the class value mean and standard deviation are calculated for each attribute of the dataset. Consequently, for testing, when any instance comes, it utilizes the mean and standard deviation values to calculate the probability of the test instance. The necessary equations are given below:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{16}$$

$$\delta = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{17}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} * \frac{1}{\delta} * e^{-(x-\mu)2} \tag{18}$$

where μ indicates the mean, δ represents standard deviation, x_i denotes all samples in a particular column, n indicates the total number of samples and f_x presents the conditional probability of class value.

6) LOGISTIC REGRESSION (LR)

Based on a given dataset of independent variables, logistic regression calculates the likelihood that an event will occur, such as voting or not voting. Given that the result is a probability, the dependent variable's range is 0 to 1. In logistic regression, the odds—that is, the likelihood of success divided by the probability of failure-are transformed using the logit formula. The following formulae are used to express this logistic function, which is sometimes referred to as the log odds or the natural logarithm of odds [43]:

$$p = \frac{1}{1 + e^{-x}}, \tag{19}$$

where p denotes the probability of instance x. At the time of model training, for each instance $x_1, x_2, x_3 \ldots\ldots x_n$ the logistic coefficients will be $b_0, b_1, b_2 \ldots\ldots b_n$. The stochastic gradient descent method estimates and updates the values of the coefficients.

$$v = b_0 x_0 + b_1 x_1 + \ldots\ldots\ldots + b_n x_n \tag{20}$$

$$p = \frac{1}{1 + e^{-v}} \tag{21}$$

Now, the following equation is used to update the values of the coefficients:

$$b = b + l * (y - p) * (1 - p) * p * x \tag{22}$$

Initially, all the coefficient values are 0 and y is the output value for each training sample, where l denotes learning rate, x represents biased input for b_0 and is always 1. It updates the

values of the coefficients until it predicts the correct output at the training stage.

7) SUPPORT VECTOR MACHINE (SVM)

SVM is used to classify both linear and non-linear data and mostly works well for high-dimensional data with non-linear mapping. It explores the decision boundary or optimal hyperplane to separate one class from another. This study used Radial Basis Function (RBF) as a kernel function and SVM automatically defines centers, weights, and thresholds and reduces an upper bound of expected test error [29], [44]. The following equation represents the RBF function:

$$K(x, x') = exp(-\frac{(||x - x'||)^2}{2\delta^2}) \tag{23}$$

where $(||x - x'||)^2$ defines the squared Euclidean distance between the two feature samples and δ is a free parameter.

8) LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA is a dimensionality reduction technique but can be used for classification by exploring the linear combination of features [45]. LDA uses the Bayes theorem to estimate the probability. Let us, consider k classes and n training samples that are defined as $\{x_1, x_2 \ldots\ldots\ldots x_n\}$ with classes $z_i \epsilon \{1 \ldots . k\}$. The prior probability is assumed to display as Gaussian distribution $\phi(x|\mu_k, \Sigma)$ in each class. The model estimation is defined as follows:

$$a_k = \frac{\sum_{i=1}^{n} l * (z_i = k)}{n} \tag{24}$$

$$\mu_k = \frac{\sum_{i=1}^{n} x_i * l * (z_i = k)}{\sum_{i=1}^{n} l * (z_i = k)} \tag{25}$$

$$\Sigma = \frac{\sum_{i=1}^{n} (x_i - \mu_{zi})(x_i - \mu_{zi})^T}{n}, \tag{26}$$

where a_k denotes the prior probability, μ_k defines mean of all classes, Σ indicates the sample covariance of the class means.

III. EXPERIMENTAL RESULTS ANALYSIS

A. EXPERIMENTAL SETUP

In order to conduct the experiment, an open-source cloud-based service named Google Collaboratory provided by Google is utilized. The scikit-learn package of Python programming language is used to complete the data preprocessing, feature scaling, feature selection, and classification tasks. In this work, a 10-fold cross-validation technique [46], [47], [48] is utilized to construct prediction models using four different ASD (Toddlers, Children, Adolescents, and Adults) datasets. In 10-fold cross-validation, during training, the datasets are randomly divided into equal 10 folds. During model building, 9 folds are used and training and the remaining one is used for testing. Hence, this procedure is repeated 10 times, and finally, average the results. Here, due to the lack of enough samples in the datasets, 10-fold cross-validation is used to prevent the model from overfitting and reducing the variance during model building and generalize the model with a small amount of data. If we perform hold-out validation with a fixed test set, then, there would have a possibility of potential overfitting during model building and it will increase the variance and thus cannot generalize the prediction model for unseen test data. Various statistical evaluation measures including accuracy, Receiver Operating Characteristics (ROC) curve, F1-score, precision, recall, Mathews Correlation Coefficient (MCC), Kappa score, and Log loss are considered to justify the experimental outcomes. The evaluation measures are calculated using the following formulae.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{27}$$

$$Precision = \frac{TP}{FP + TP} \tag{28}$$

$$Recall = \frac{TP}{FN + TP} \tag{29}$$

$$F1 - Score = \frac{2TP}{FN + FP + 2TP} \tag{30}$$

$$MCC = \frac{(TP * TN - FP * FN)}{((TP+FP)(TP+FN)(TN+FP)(TN+FN))^{\frac{1}{2}}} \tag{31}$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{32}$$

$$LogLoss = -1.0 * (y \log(y')) + (1 - y) * \log(1 - y') \tag{33}$$

The following terms represent the above equations. TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative; p_o is the relative observed agreement among raters; and p_e is the hypothetical probability of chance agreement; y is the actual/true value and y' is the prediction probability of each observation.

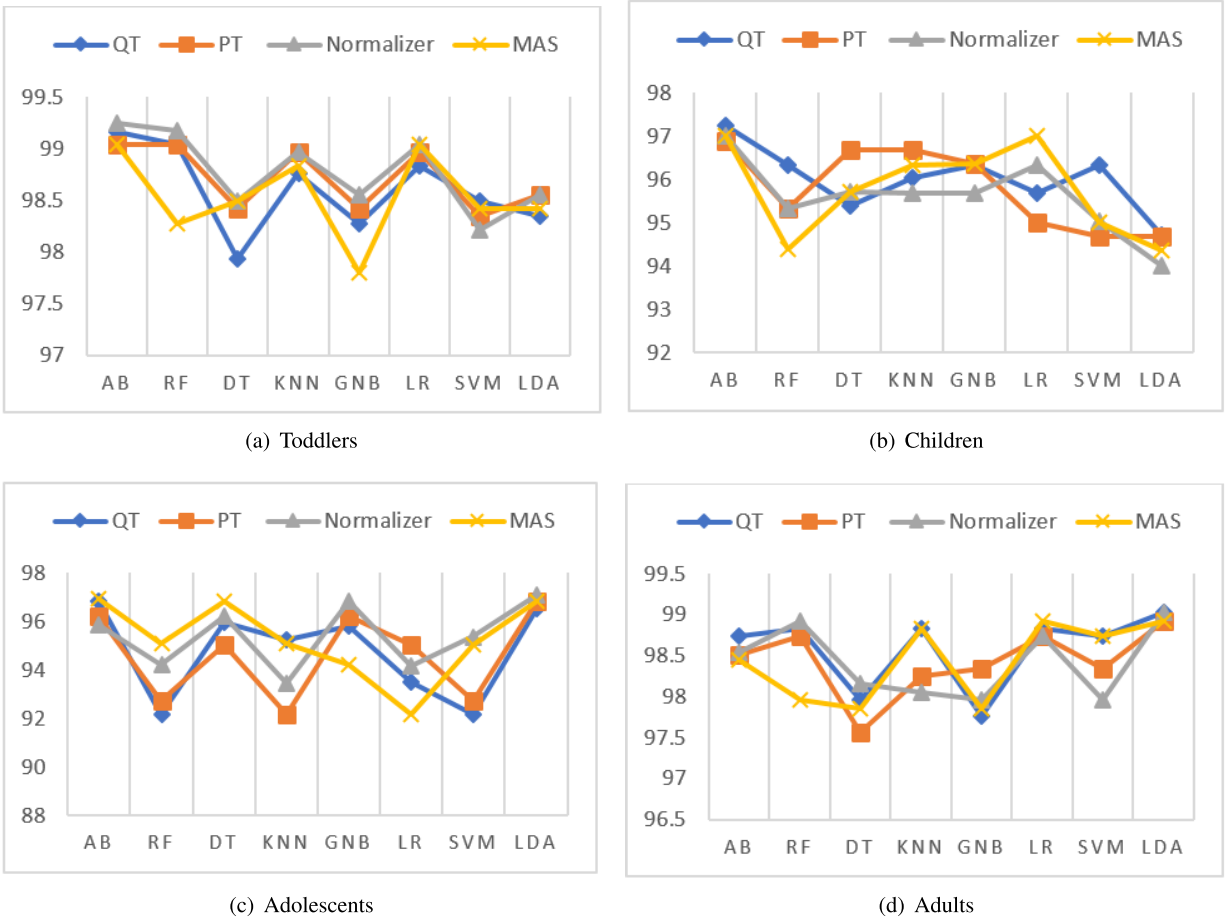
B. ANALYSIS ON ACCURACY

Accuracy represents the actual prediction performance of any classifier. The higher the value of accuracy indicates better prediction and lower the miss-classification. The accuracy values of various classifiers on different feature-scaled datasets are presented in Table 5.

In this case, LDA delivers the best accuracy of 97.12% for the normalizer-scaled Adolescent dataset. Moreover, while investigating the results of the feature-scaled Adult dataset, it is seen that both the QT and normalizer-scaled datasets perform better than the other FS methods. In both of the cases, LDA achieves the best accuracy value of 99.03%. Additionally, the accuracy values of various ML classifiers on feature-scaled Toddlers, Children, Adolescents, and Adult datasets are contrasted in Fig. 2.

C. ANALYSIS ON PRECISION

Precision represents positive predictive value and a higher value of precision means the true positive value is high and the false positive value is low. The precision values of various classifiers on different feature-scaled datasets are presented



(a) Toddlers (b) Children

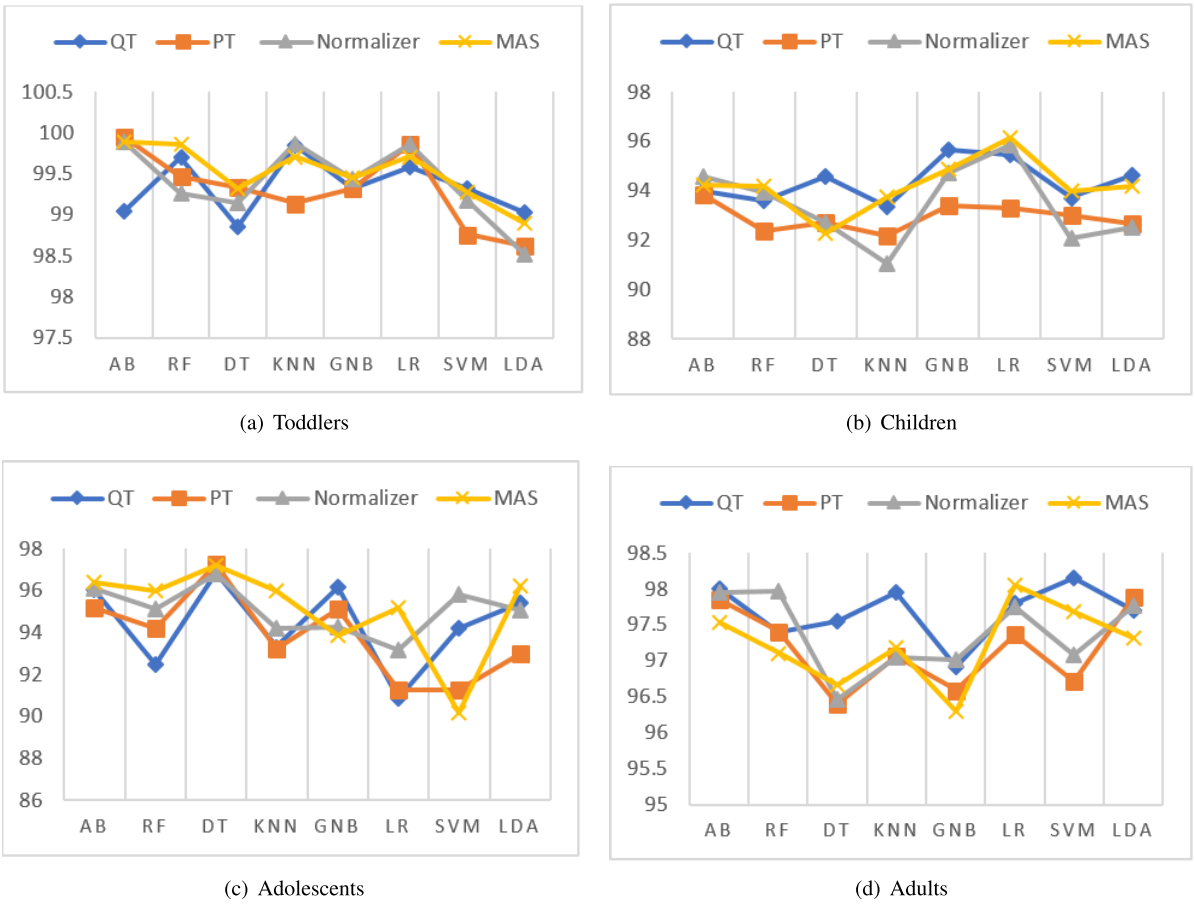
(c) Adolescents (d) Adults

FIGURE 2. Accuracy of the classifiers on different feature-scaled datasets.

TABLE 5. Accuracy of different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	99.17	99.04	97.94	98.76	98.28	98.83	98.49	98.35
	PT	99.04	99.04	98.42	98.97	98.42	98.97	98.35	98.56
	Normalizer	99.25	99.18	98.49	98.97	98.56	99.04	98.21	98.56
	MAS	99.04	98.28	98.49	98.83	97.80	99.04	98.42	98.42
Children	QT	97.95	96.35	95.39	96.03	96.34	95.68	96.35	94.72
	PT	96.89	95.35	96.68	96.68	96.37	95.02	94.69	94.70
	Normalizer	97.02	95.35	95.71	95.7	95.7	96.33	95.03	94.02
	MAS	97.02	94.4	95.71	96.35	96.37	97.02	95.01	94.37
Adolescents	QT	96.86	92.18	95.96	95.26	95.8	93.5	92.18	96.52
	PT	96.23	92.76	95.06	92.18	96.23	95.06	92.76	96.86
	Normalizer	95.86	94.25	96.23	93.45	96.82	94.18	95.38	97.12
	MAS	96.93	95.10	96.86	95.12	94.23	92.18	95.06	96.86
Adults	QT	98.74	98.83	97.96	98.83	97.77	98.83	98.74	99.03
	PT	98.52	98.74	97.57	98.25	98.35	98.74	98.35	98.93
	Normalizer	98.54	98.93	98.16	98.06	97.96	98.74	97.96	99.03
	MAS	98.45	97.96	97.86	98.83	97.86	98.93	98.74	98.93

in Table 6. Analyzing the precision values of the Toddler dataset, it is found that the AB classifier provides the best precision of 99.95% while PT is used as the FS method. While reviewing the feature-scaled Children dataset, it is noticed that the LR classifier obtains the highest precision of 96.16% for MAS in classifying ASD. Furthermore, inspecting the feature-scaled Adolescent dataset, we observe that DT delivers the best precision of 97.25% while using PT as FS method. Moreover, while investigating the results of the feature-scaled Adult dataset, it is seen that both the QT-transformed datasets perform better than the other FS methods. In that case, SVM achieves the best precision value of 98.16%. Additionally,



(a) Toddlers (b) Children

(c) Adolescents (d) Adults

FIGURE 3. Precision of the classifiers on different feature-scaled datasets.

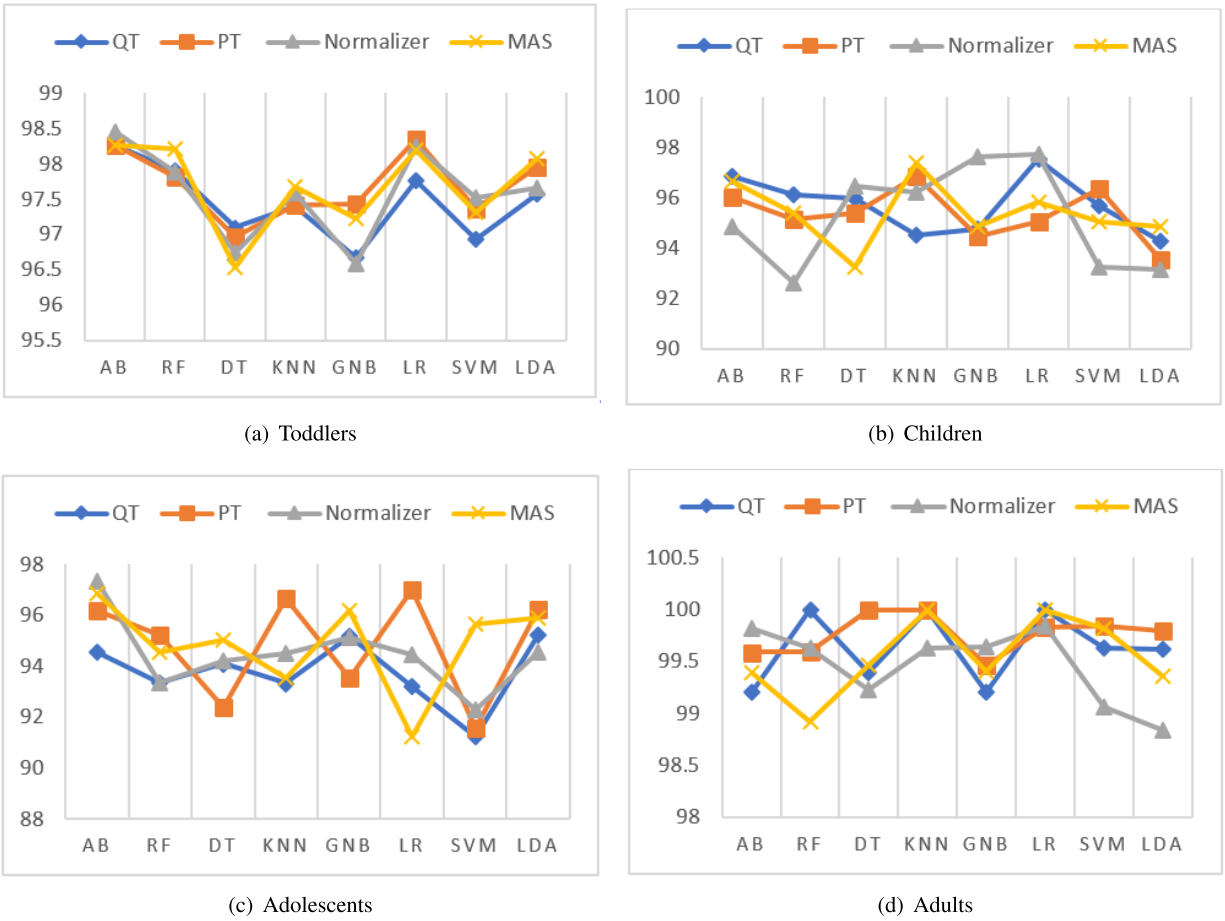
TABLE 6. Precision of the different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	99.05	99.70	98.85	99.85	99.32	99.59	99.32	99.03
	PT	99.95	99.47	99.34	99.14	99.32	99.86	98.76	98.62
	Normalizer	99.89	99.27	99.15	99.87	99.44	99.86	99.17	98.52
	MAS	99.89	99.86	99.32	99.71	99.45	99.71	99.28	98.90
Children	QT	94.02	93.63	94.60	93.39	95.68	95.45	93.73	94.66
	PT	93.86	92.37	92.71	92.17	93.41	93.31	93.00	92.69
	Normalizer	94.57	93.94	92.71	91.06	94.73	95.86	92.10	92.55
	MAS	94.23	94.2	92.31	93.77	94.88	96.16	94.01	94.22
Adolescents	QT	96.07	92.50	96.90	93.32	96.2	90.89	94.20	95.45
	PT	95.20	94.20	97.25	93.25	95.14	91.25	91.25	93.02
	Normalizer	96.12	95.12	96.8	94.2	94.25	93.2	95.84	95.1
	MAS	96.42	96.02	97.2	96.02	93.89	95.2	90.2	96.25
Adults	QT	98.0	97.40	97.56	97.95	96.91	97.80	98.16	97.71
	PT	97.86	97.41	96.40	97.07	96.59	97.37	96.71	97.89
	Normalizer	97.95	97.97	96.47	97.05	97.01	97.76	97.09	97.77
	MAS	97.54	97.11	96.67	97.19	96.30	98.05	97.68	97.32

the precision values of various ML classifiers on feature-scaled Toddlers, Children, Adolescents, and Adult datasets are contrasted in Fig. 3.

D. ANALYSIS ON RECALL

Recall represents a true positive rate and a higher value of recall means the true positive value is high and the false negative value is low. When the true positive is high and the false negative is low that means better prediction. The recall values of various ML classifiers on different feature-scaled datasets are presented in Table 7. While reviewing the recall results of the feature-scaled Toddler dataset, it is observed that AB obtains the highest recall of 98.45% for the normalizer-scaled Toddler dataset.



(a) Toddlers (b) Children

(c) Adolescents (d) Adults

FIGURE 4. Recall of the classifiers on different feature-scaled datasets.

TABLE 7. Recall of the different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	98.27	97.90	97.10	97.41	96.68	97.77	96.93	97.57
	PT	98.27	97.82	96.97	97.42	97.44	98.35	97.37	97.96
	Normalizer	98.45	97.89	96.74	97.59	96.59	98.25	97.53	97.66
	MAS	98.27	98.21	96.54	97.68	97.23	98.19	97.31	98.08
Children	QT	96.84	96.11	96.00	94.51	94.78	97.54	95.70	94.29
	PT	96.05	95.17	95.39	96.87	94.49	95.03	96.39	93.56
	Normalizer	94.88	92.63	96.48	96.21	97.63	97.72	93.25	93.14
	MAS	96.64	95.4	93.25	97.38	94.86	95.81	95.03	94.84
Adolescents	QT	94.58	93.35	94.10	93.33	95.2	93.21	91.24	95.24
	PT	96.21	95.24	92.38	96.69	93.54	97.03	91.58	96.23
	Normalizer	97.36	93.35	94.23	94.52	95.12	94.45	92.3	94.58
	MAS	96.89	94.56	95.02	93.56	96.2	91.25	95.65	95.89
Adults	QT	99.21	100.00	99.40	100.00	99.21	100.00	99.63	99.62
	PT	99.59	99.60	100.00	100.00	99.47	99.83	99.84	99.80
	Normalizer	99.82	99.63	99.23	99.63	99.64	99.84	99.06	98.84
	MAS	99.4	98.92	99.47	100.00	99.41	100.00	99.82	99.36

Investigating the feature-scaled Children datasets, we find that LR delivers the best recall value of 97.72% while normalizer as FS method. Moreover, inspecting the recall results of feature-scaled adolescent datasets, it is noticed that AB achieves the highest recall of 97.36% for normalizer-scaled Adolescent datasets. Finally, we analyze the outcomes of feature-scaled Adult datasets and find that RF, KNN, and LR deliver the highest recall of 100.00% for PT, while DT and KNN obtain the best recall of 100.00% for PT and KNN, LR also obtains 100.00% recall value for MAS-scaled

(a) Toddlers (b) Children

(c) Adolescents (d) Adults

FIGURE 5. ROC of the classifiers on different feature-scaled datasets.

TABLE 8. ROC of the different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	99.97	99.94	99.93	99.97	99.94	99.99	99.95	99.91
	PT	99.96	99.96	99.92	99.94	99.94	99.99	99.89	99.94
	Normalizer	99.99	99.96	99.96	99.96	99.94	99.98	99.96	99.96
	MAS	99.96	99.96	99.96	99.96	99.35	99.98	99.92	99.92
Children	QT	99.48	98.73	98.82	99.30	99.46	99.41	98.68	98.47
	PT	99.35	99.09	99.38	98.9	99.23	99.31	98.94	98.48
	Normalizer	99.41	99.02	99.20	99.19	99.73	99.59	99.34	98.64
	MAS	99.19	98.99	99.69	99.06	99.33	99.58	98.61	98.13
Adolescents	QT	99.72	97.46	98.92	98.97	98.87	99.00	99.12	99.45
	PT	99.06	98.25	98.99	98.88	98.78	98.87	99.36	99.02
	Normalizer	99.45	97.98	99.05	99.02	99.20	97.99	99.12	99.61
	MAS	99.21	99.01	99.06	99.04	98.88	99.02	99.61	99.72
Adults	QT	99.94	99.88	99.94	99.87	99.95	99.96	99.92	99.95
	PT	99.92	99.92	99.90	99.92	99.98	99.96	99.96	99.99
	Normalizer	99.92	99.92	99.92	99.86	99.90	99.97	99.92	99.99
	MAS	99.91	99.81	99.92	99.94	99.92	99.98	99.94	99.95

adult's datasets. Besides, we also compare the recall values of various ML classifiers on feature-scaled Toddlers, Children, Adolescents, and Adult datasets in Fig. 4.

E. ANALYSIS ON ROC

The ROC value indicates the ability of any classifier to distinguish between positive and negative classes. The ROC values of various ML classifiers on different feature-scaled datasets are presented in Table 8. While reviewing the ROC results of the feature-scaled Toddler dataset, it is observed that LR obtains the highest ROC of 99.99% for both QT and PT and AB achieves 99.99% for the normalizer method. Investigating the feature-scaled Children dataset, it is found that GNB delivers the best ROC value of 99.73% using normalizer as

(a) Toddlers (b) Children

(c) Adolescents (d) Adults

FIGURE 6. F1-score of the classifiers on different feature-scaled datasets.

TABLE 9. F1-score of the different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	98.99	98.80	97.94	98.51	97.87	99.09	98.16	98.29
	PT	99.12	98.04	98.45	98.85	98.56	98.82	98.16	98.48
	Normalizer	99.14	98.65	98.32	98.62	98.24	98.67	97.92	97.92
	MAS	98.99	98.96	97.69	98.91	98.19	99.02	98.28	98.19
Children	QT	97.02	93.15	95.38	95.37	95.17	96.29	93.06	92.90
	PT	96.27	94.69	93.06	93.26	95.18	95.58	94.27	94.72
	Normalizer	97.02	94.00	95.07	94.10	96.43	96.74	93.21	93.45
	MAS	96.78	95.38	94.51	94.12	93.64	95.38	94.11	94.41
Adolescents	QT	97.69	92.58	95.19	93.69	94.43	94.35	93.6	95.12
	PT	95.93	95.12	93.45	95.02	95.68	92.12	92.6	92.88
	Normalizer	97.52	94.87	94.58	94.88	92.35	92.18	94.2	93.86
	MAS	97.41	93.25	95.87	92.88	94.12	95.02	93.65	94.88
Adults	QT	98.86	98.38	97.85	98.80	98.07	98.71	98.19	99.18
	PT	98.74	98.36	97.88	98.57	97.98	98.80	98.62	99.11
	Normalizer	98.92	98.89	97.91	98.54	97.99	98.70	98.40	98.62
	MAS	98.65	98.27	98.06	98.83	98.48	99.01	98.63	98.96

the FS method. Moreover, inspecting the ROC results of the feature-scaled Adolescent dataset, we notice that both AB and LDA achieve the highest ROC of 99.72% for QT and MAS-scaled datasets. Finally, we analyze the outcomes of feature-scaled Adult datasets and find that LDA delivers the highest ROC value of 99.99% while using PT and normalizer as the FS methods. We compare the ROC values of various ML classifiers on feature-scaled Toddlers, Children, Adolescents, and Adult datasets in Fig. 5.

F. ANALYSIS ON THE F1-SCORE

F1-score takes the harmonic mean of the precision and recall values and a higher value of it indicates better prediction.

(a) Toddlers (b) Children

(c) Adolescents (d) Adults

FIGURE 7. Kappa of the classifiers on different feature-scaled datasets.

TABLE 10. Kappa of the different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	98.88	95.84	94.46	96.53	94.46	96.53	93.06	93.79
	PT	98.89	93.14	94.51	93.14	94.51	94.50	91.78	93.82
	Normalizer	98.97	98.62	98.62	97.93	97.94	99.31	98.62	95.87
	MAS	98.87	97.93	98.62	97.93	98.62	97.93	98.62	99.31
Children	QT	92.95	91.46	90.09	90.01	93.35	89.99	93.29	89.77
	PT	92.97	90.01	93.29	93.44	89.77	90.07	90.16	89.77
	Normalizer	93.78	90.09	90.07	90.15	93.44	90.15	89.25	86.81
	MAS	90.26	86.81	89.99	90.07	90.15	93.42	90.12	90.15
Adolescents	QT	92.32	84.61	92.21	92.02	91.88	86.83	88.95	94.02
	PT	91.89	87.12	91.82	88.32	92.01	85.56	90.02	94.02
	Normalizer	92.25	89.26	92.76	90.12	92.03	92.12	89.82	91.98
	MAS	92.01	92.00	91.85	90.25	92.09	88.78	86.67	92.12
Adults	QT	97.06	97.07	95.08	97.07	96.11	99.02	95.14	99.02
	PT	98.00	98.05	96.11	97.08	96.11	97.06	96.09	97.07
	Normalizer	96.99	97.08	94.17	95.13	97.08	97.05	95.13	99.02
	MAS	96.83	96.10	94.17	98.05	96.09	98.05	97.08	97.08

The F1-score values of various ML classifiers on different feature-scaled datasets are presented in Table 9. While reviewing the F1-score results of the feature-scaled Toddler dataset, we observe that AB obtains the highest F1-score of 99.14% for the normalizer-scaled Toddler dataset. Investigating the feature-scaled Children dataset, it is found that AB delivers the best F1-score value of 97.02% while using QT and normalizer as FS methods. Moreover, inspecting the F1-score results of feature-scaled Adolescent datasets, we notice that AB achieves the highest F1-score of 97.69% for the QT-scaled Adolescent dataset. Finally, we analyze the outcomes of the feature-scaled Adult dataset and notice

(a) Toddlers (b) Children

(c) Adolescents (d) Adults

FIGURE 8. Log loss of the classifiers on different feature-scaled datasets.

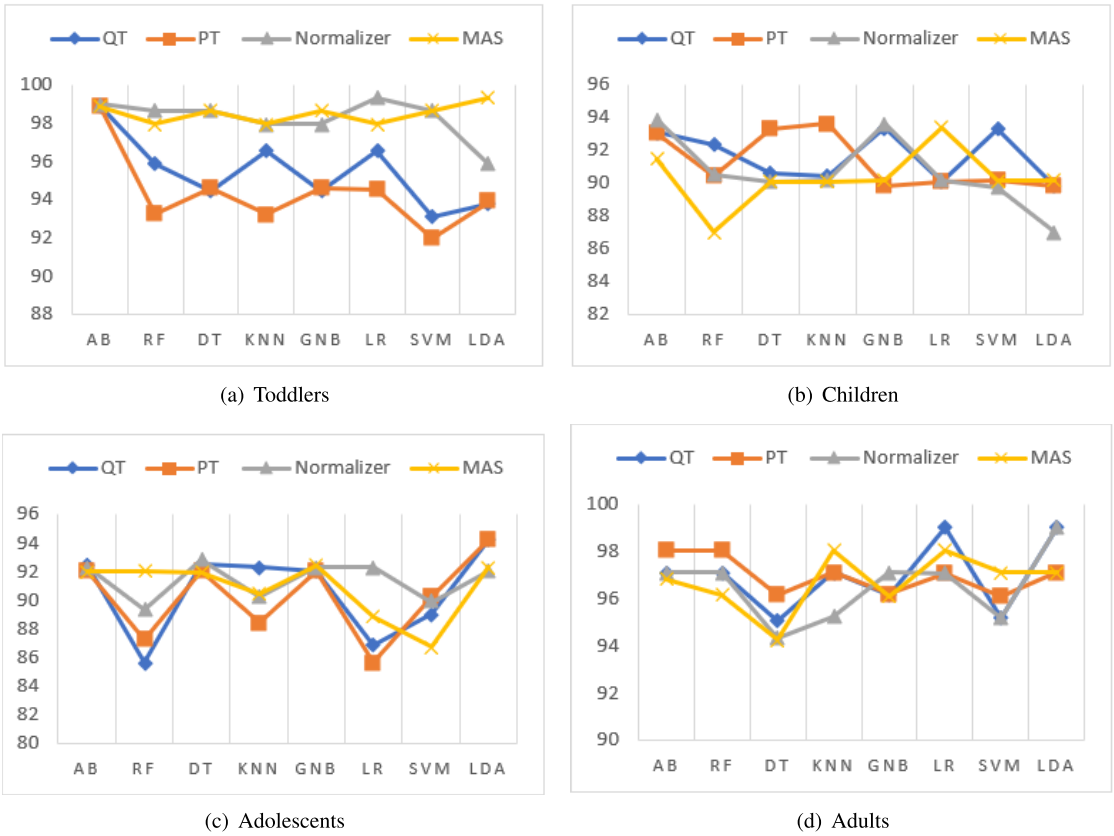
TABLE 11. Log loss of the different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	0.0935	0.7097	0.9462	0.5914	0.9462	0.5914	1.18	1.06
	PT	0.1098	1.18	0.9462	1.18	0.9462	0.9462	1.41	1.06
	Normalizer	0.0802	0.2356	0.2365	0.3548	0.3548	0.118	0.2365	0.7097
	MAS	0.1098	0.3548	0.2365	0.3548	0.2365	0.3548	0.2365	0.1182
Children	QT	0.98	1.32	1.69	1.69	1.13	1.69	1.13	1.69
	PT	0.98	1.69	1.13	1.13	1.69	1.69	1.69	1.69
	Normalizer	1.10	1.69	1.69	1.69	1.13	1.69	1.69	2.26
	MAS	1.13	2.26	1.69	1.69	1.69	1.13	1.69	1.69
Adolescents	QT	1.32	2.65	1.32	1.33	1.32	2.65	1.32	1.12
	PT	1.32	2.65	1.32	2.65	1.32	1.32	2.65	1.12
	Normalizer	1.32	1.32	1.32	1.32	1.32	1.32	2.65	1.32
	MAS	1.32	2.65	1.32	1.33	1.32	2.65	1.32	1.12
Adults	QT	0.33	0.50	0.83	0.50	0.67	0.16	0.83	0.16
	PT	0.67	0.33	0.67	0.50	0.67	0.50	0.67	0.50
	Normalizer	0.33	0.50	1.00	0.83	0.50	0.50	0.83	0.16
	MAS	0.46	0.67	1.00	0.33	0.67	0.33	0.50	0.50

that LDA delivers the highest F1-score value of 99.11% while using PT as the FS method. We compare the F1-score values of various ML classifiers on feature-scaled Toddlers, Children, Adolescents, and Adult datasets in Fig. 6.

G. ANALYSIS ON KAPPA

Kappa score measures the degree of agreement between true class and predicted class. The higher value of kappa means a better prediction which indicates a higher degree of agreement between actual and predicted values. The kappa values of various ML classifiers on different feature-scaled datasets are presented in Table 10. While reviewing the kappa results of the feature-scaled Toddler dataset, it is observed that both the normalizer and MAS-scaled datasets provide the best kappa value and outperform the other FS methods. Consequently, both LR and LDA obtain the highest



(a) Toddlers (b) Children

(c) Adolescents (d) Adults

FIGURE 9. MCC of the classifiers on different feature-scaled datasets.

TABLE 12. MCC of the different ML classifiers on ASD datasets.

Dataset	FS	AB	RF	DT	KNN	GNB	LR	SVM	LDA
Toddlers	QT	98.91	95.87	94.47	96.55	94.47	96.55	93.10	93.80
	PT	98.93	93.28	94.59	93.22	94.59	94.54	92.00	93.88
	Normalizer	98.99	98.63	98.62	97.95	97.94	99.31	98.63	95.89
	MAS	98.89	97.96	98.63	97.96	98.63	97.96	98.63	99.31
Children	QT	93.10	92.33	90.63	90.46	93.35	90.04	93.29	89.82
	PT	93.08	90.46	93.29	93.64	89.82	90.11	90.21	89.82
	Normalizer	93.88	90.53	90.11	90.20	93.64	90.20	89.77	87.00
	MAS	91.50	87.00	90.04	90.11	90.20	93.42	90.17	90.20
Adolescents	QT	92.45	85.63	92.49	92.31	92.03	86.85	89.02	94.25
	PT	92.02	87.25	92.03	88.42	92.03	85.63	90.23	94.25
	Normalizer	92.31	89.36	92.85	90.23	92.31	92.31	89.88	92.02
	MAS	92.01	92.02	91.89	90.45	92.42	88.86	86.71	92.31
Adults	QT	97.12	97.12	95.08	97.11	96.18	99.03	95.18	99.03
	PT	98.07	98.07	96.18	97.12	96.19	97.10	96.11	97.11
	Normalizer	97.11	97.12	94.33	95.25	97.12	97.06	95.17	99.02
	MAS	96.85	96.18	94.24	98.07	96.11	98.07	97.12	97.12

MCC of 99.31% for normalizer and MAS-scaled Toddler datasets. Investigating the feature-scaled Children datasets, it is found that AB delivers the best kappa value of 93.78% using normalizer as FS method. Moreover, inspecting the kappa results of feature-scaled Adolescent datasets, we notice that LDA achieves the highest MCC of 94.02% for both QT and PT-scaled datasets respectively. Finally, we analyze the outcomes of feature-scaled Adult datasets and see that both LR and LDA deliver the highest kappa value of 99.02% while using QT and normalizer as the feature scaling methods. Besides, we also compare the kappa values of various ML classifiers on feature-scaled Toddlers, Children, Adolescents, and Adult datasets in Fig. 7.

H. ANALYSIS ON LOG LOSS

The log loss value indicates how close the prediction probability is to the true values. The lower the log loss value, the better the prediction. The log loss values of various ML

TABLE 13. Feature importance for the normalizer-scaled toddlers.

Feature Name	IGAE	GRAE	RFAE	CAE
A1Score	0.190789	0.193046	0.252372	0.5038
A2Score	0.173791	0.175119	0.298482	0.4635
A3Score	0.137306	0.141301	0.21907	0.4097
A4Score	0.199316	0.199403	0.264516	0.5052
A5Score	0.25161	0.252053	0.305787	0.5633
A6Score	0.25161	0.252053	0.305787	0.5633
A7Score	0.229182	0.245337	0.250095	0.5632
A8Score	0.14363	0.144324	0.258539	0.4272
A9Score	0.277909	0.277997	0.350664	0.5773
A10Score	0.023146	0.023657	0.138235	0.1798
Age	0.016184	0.040015	0.021652	0.0668
Gender	0.00977	0.011045	0.018975	0.1177
Ethnicity	0.029538	0.01162	0.056357	0.0737
Jaundice	0.004052	0.004789	0.021652	0.0741
Family member with autism	0.000131	0.000205	0.01129	0.0135
Relation	0.003077	0.011662	-0.000759	0.0132

TABLE 14. Feature importance for the normalizer-scaled children.

Feature Name	IGAE	GRAE	RFAE	CAE
A1Score	0.116637	0.123048	0.19007	0.3935
A2Score	0.038218	0.038348	0.07637	0.229
A3Score	0.122714	0.149296	0.11062	0.3955
A4Score	0.249648	0.251567	0.28185	0.5685
A5Score	0.112333	0.136666	0.15582	0.3799
A6Score	0.135612	0.156652	0.12295	0.4173
A7Score	0.055143	0.057011	0.07945	0.2739
A8Score	0.143457	0.143462	0.18459	0.4384
A9Score	0.177918	0.177942	0.16575	0.4862
A10Score	0.153613	0.181338	0.18596	0.4399
Age	0.00000	0.000000	0.00871	0.0753
Gender	0.001086	0.001255	0.00274	0.0388
Ethnicity	0.036504	0.014577	0.01757	0.0556
Jaundice	0.000453	0.000535	-0.00308	0.025
Family member with autism	0.001724	0.002642	-0.00822	0.0488
Country of Res	0.235509	0.054281	0.02705	0.0752
Used app before	0.001632	0.007048	-0.00205	0.0472
Age desc	0.00000	0.000000	0.00000	0.00000
Relation	0.011968	0.014917	0.00671	0.0614

classifiers on different feature-scaled datasets are presented in Table 11. While reviewing the log loss results of the feature-scaled Toddler and children datasets, we observe that AB obtains the lowest log loss of 0.0802% and 0.98% for the normalizer-scaled toddler and QT and PT-scaled children. Furthemore, it is noticed that LDA achieves the lowest log loss of 1.12% for QT, PT, and MAS-scaled adolescents datasets. Finally, we analyze the outcomes of feature-scaled Adult datasets and see that both LR and LDA deliver the highest log loss value of 0.16% while using QT and normalizer as the feature scaling methods. Besides, we also compare the log loss values of various ML classifiers on feature-scaled Toddlers, Children, Adolescents, and Adult datasets in Fig. 8.

I. ANALYSIS ON MCC

MCC takes all the coefficient of confusion matrix such as TP, TN, FN and FP into consideration to calculate the degree of correlation. The higher value of MCC represents better prediction and strong correlation between actual and predicted class. While reviewing the MCC results of the feature-scaled Toddler dataset, we observe that both LR and LDA obtain the highest MCC of 99.31% for normalizer and MAS-scaled Toddler datasets. Investigating the feature-scaled children datasets, it is found that AB delivers the best MCC value of 93.88% using normalizer as the FS method. Moreover, inspecting the MCC results of feature-scaled Adolescent datasets, we notice that LDA achieves the highest MCC of 94.25% for both QT and PT-scaled datasets respectively. Finally, we analyze the outcomes of feature-scaled Adult datasets and find that both LR and LDA deliver the highest MCC value of 99.03% while using QT as the feature scaling method. Besi'des, we also compare the MCC values of various ML classifiers on feature-scaled toddlers, children, adolescents, and adults datasets in Fig. 9.

TABLE 15. Feature importance for the QT-scaled adolescents.

A1 Score	0.021946	0.026115	0.04519	0.1757
A2 Score	0.029069	0.029193	0.09615	0.2004
A3 Score	0.174678	0.187708	0.20192	0.4883
A4 Score	0.188168	0.214089	0.17404	0.5067
A5 Score	0.211911	0.261206	0.24904	0.5339
A6 Score	0.144373	0.185248	0.14808	0.4454
A7 Score	0.078265	0.078348	0.12212	0.3264
A8 Score	0.124797	0.12983	0.17404	0.4138
A9 Score	0.116242	0.156153	0.05673	0.4012
A10 Score	0.174678	0.187708	0.20192	0.4883
Age	0.000000	0.000000	0.00192	0.1068
Gender	0.012121	0.012134	0.04808	0.1295
Ethnicity	0.064442	0.025481	0.03005	0.1096
Jaundice	0.001019	0.001645	0.04712	0.0978
Family member with autism	0.00055	0.000965	0.02885	0.0277
County of Res	0.385831	0.095969	0.08654	0.1474
Used app before	0.015078	0.06411	0.0192	0.1456
Age desc	0.021888	0.061519	-0.00288	0.1759
Relation	0.018134	0.0104097	0.01423	0.0553

TABLE 16. Feature importance for the QT-scaled adults.

A1Score	0.07816	0.0916	0.152699	0.29763
A2Score	0.07095	0.0714	0.189631	0.31138
A3Score	0.1473	0.14808	0.197017	0.44107
A4Score	0.17514	0.17515	0.265483	0.46995
A5Score	0.24037	0.24037	0.402415	0.537
A6Score	0.24099	0.27991	0.231534	0.59209
A7Score	0.08937	0.09117	0.21108	0.35143
A8Score	0.04421	0.04729	0.138494	0.23716
A9Score	0.28855	0.3176	0.295597	0.63558
A10Score	0.12124	0.12319	0.187358	0.38592
Age	0.00000	0.00000	-0.000909	0.05917
Gender	0.00466	0.00467	0.01321	0.08038
Ethnicity	0.08855	0.03397	0.104378	0.15073
Jaundice	0.00698	0.01509	0.010511	0.10215
Family member with autism	0.02066	0.0372	0.00142	0.177741
County of Res	0.21054	0.04729	0.15071	0.1161
Used app before	0.00128	0.01028	0.003977	0.04404
Age desc	0.00000	0.00000	0.000000	0.00000
Relation	0.00113	0.00143	-0.004773	0.00538

IV. DISCUSSION AND EXTENDED COMPARISON

In the previous section, we analyzed four different ASD datasets to build prediction models for different stages of people. In order to do this, we applied various FS methods to those ASD datasets and classified them utilizing eight different simple but effective ML classifiers and also determined how the FS methods affect the classification performance. Furthermore, we also employed four different FSTs to compute the importance of the features which are more responsible for ASD prediction. Inspecting the experimental findings, the best performing classifiers model predicted ASD with AB (99.25%), AB (97.95%), LDA (97.12%), LDA (99.03%) accuracy; AB, LR (99.99%), GNB (99.73%), AB, LDA (99.72%), LDA (99.99%) ROC; AB (99.14%), AB (97.02%), AB (97.69%), LDA (99.11%) F1-score; AB (99.95%), LR (96.16%), DT (97.25%), SVM (98.16%) precision; AB (98.45%), LR (97.72%), AB (97.36%), RF, DT, KNN, LR (100%) recall; LR, LDA (99.31%), AB (93.88%), LDA (94.25%), LR, LDA (99.03%) MCC; LR, LDA (99.31%), AB (93.78%), LDA (94.02%), LR, LDA (99.02%) kappa; AB (0.0802%), AB (0.98%), LDA (1.12%), LR, LDA (0.16%) log loss for Toddlers, Children, Adolescents, Adults datasets respectively. After analyzing the experimental outcomes of different classifiers on feature-scaled ASD datasets, it is found that AB for Toddlers and Children, and LDA for Adolescents and Adults outperformed the other ML classifiers in terms of classification performance. Besides, the experimental outcomes implied that the normalizer FS method for Toddlers, normalizer FS method for Children, QT FS method for Adolescents, and QT FS method for Adults showed better performance. Additionally, we calculated the feature importance using the IGAE, GRAE, RFAE, and CAE FST methods

TABLE 17. Comparison with other works.

Dataset	Reference	Accuracy	ROC	F1	Precision	Recall	MCC	Kappa	Log Loss
Toddlers	Mousumi *et al.* [1]	97.82	99.70	97.80	-	-	-	94.87	-
	Proposed Model	99.25	99.99	99.14	99.89	98.45	98.99	98.97	0.0802
Children	Omar *et al.* [18]	92.26	-	-	-	-	-	-	-
	Thabtah *et al.* [17]	97.80	-	-	-	98.00	-	-	-
	Talabani *et al.* [49]	92.26	-	-	88.09	96.52	-	-	-
	Mousumi *et al.* [1]	99.61	99.60	99.60	-	-	-	99.21	-
	Haroon *et al.* [30]	95.5	-	96.00	97.00	98.00	90.10	-	-
	Abitha *et al.* [31]	94.1	-	-	-	-	-	-	-
	Kamma *et al.* [33]	95.82	-	-	-	-	-	-	-
	Gupta *et al.* [34]	-	-	94.71	92.59	97.09	89.32	-	-
	Proposed Model	97.95	99.73	97.02	96.16	97.72	93.88	93.78	0.98
Adolescents	Omar *et al.* [18]	93.78	-	-	-	-	-	-	-
	Thabtah *et al.* [17]	94.23	-	-	-	92.20	-	-	-
	Talabani *et al.* [49]	93.78	-	-	89.85	98.4	-	-	-
	Mousumi *et al.* [1]	95.87	99.00	95.90	-	-	-	91.74	-
	Kamma *et al.* [33]	95.82	-	-	-	-	-	-	-
	Gupta *et al.* [34]	-	-	84.21	93.25	74.15	65.53	-	-
	Proposed Model	97.12	99.72	97.69	97.25	97.36	94.25	94.02	1.12
Adults	Omar *et al.* [18]	97.10	-	-	-	-	-	-	-
	Thabtah *et al.* [17]	99.85	-	-	-	99.90	-	-	-
	Shuvo *et al.* [25]	95.71	-	-	-	85.71	-	-	-
	Talabani *et al.* [49]	96.91	-	-	90.07	96.87	-	-	-
	Mousumi *et al.* [1]	99.82	99.80	99.90	-	-	-	99.59	-
	Abitha *et al.* [31]	98.00	-	-	-	-	-	-	-
	Kamma *et al.* [33]	95.82	-	-	-	-	-	-	-
	Gupta *et al.* [34]	-	-	94.26	97.46	91.27	92.46	-	-
	Proposed Model	99.03	99.99	99.11	98.16	100.00	99.03	99.02	0.16

on the normalizer-scaled Toddlers, normalizer-scaled Children, QT-scaled Adolescents, and QT-scaled Adults to enumerate the risk factors for ASD prediction. The quantitative results are provided in Table 13, Table 14, Table 15 and Table 16. This feature importance analysis helps healthcare practitioners decide the most important features while screening ASD cases. To this end, we provide the comparative results of our work with other recent studies in Table 17.

V. CONCLUSION

In this work, we proposed a machine-learning framework for ASD detection in people of different ages (Toddlers, Children, Adolescents, and Adults). We show that predictive models based on ML techniques are useful tools for this task. After completing the initial data processing, those ASD datasets were scaled using four different types of feature scaling (QT, PT, normalizer, MAS) techniques, classified using eight different ML classifiers (AB, RF, DT, KNN, GNB, LR, SVM, LDA). We then analyzed each feature-scaled dataset's classification performance and identified the best-performing FS and classification approaches. We considered different statistical evaluation measures such as accuracy, ROC, F1-Score, precision, recall, Mathews correlation coefficient (MCC), kappa score, and Log loss to justify the experimental findings. Consequently, our proposed prediction models based on ML techniques can be utilized as an alternative or even a helpful tool for physicians to accurately identify ASD cases for people of different ages. Additionally, the feature importance values were calculated to identify the most prominent features for ASD prediction by employing four different FSTs (IGAE, GRAE, RFAE, and CAE). Therefore, the experimental analysis of this research will allow healthcare practitioners to take into account the most important features while screening ASD cases. The limitation of our research work is that the amount of data was not sufficient enough to build a generalized model for people of all stages. In the future, we intend to collect more data related to ASD and construct a more generalized prediction model for people of any age to improve ASD detection and other neuro-developmental disorders.

REFERENCES

[1] M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, "Efficient machine learning models for early stage detection of autism spectrum disorder," *Algorithms*, vol. 15, no. 5, p. 166, May 2022.

[2] D. Pietrucci, A. Teofani, M. Milanesi, B. Fosso, L. Putignani, F. Messina, G. Pesole, A. Desideri, and G. Chillemi, "Machine learning data analysis highlights the role of parasutterella and alloprevotella in autism spectrum disorders," *Biomedicines*, vol. 10, no. 8, p. 2028, Aug. 2022.

[3] R. Sreedasyam, A. Rao, N. Sachidanandan, N. Sampath, and S. K. Vasudevan, "Aarya—A kinesthetic companion for children with autism spectrum disorder," *J. Intell. Fuzzy Syst.*, vol. 32, no. 4, pp. 2971–2976, Mar. 2017.

[4] J. Amudha and H. Nandakumar, "A fuzzy based eye gaze point estimation approach to study the task behavior in autism spectrum disorder," *J. Intell. Fuzzy Syst.*, vol. 35, no. 2, pp. 1459–1469, Aug. 2018.

[5] H. Chahkandi Nejad, O. Khayat, and J. Razjouyan, "Software development of an intelligent spirography test system for neurological disorder detection and quantification," *J. Intell. Fuzzy Syst.*, vol. 28, no. 5, pp. 2149–2157, Jun. 2015.

[6] F. Z. Subah, K. Deb, P. K. Dhar, and T. Koshiba, "A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI," *Appl. Sci.*, vol. 11, no. 8, p. 3636, Apr. 2021.

[7] K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, "The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review," *Electronics*, vol. 10, no. 23, p. 2982, Nov. 2021.
[8] I. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. H. Ali, H. S. A. Shatnawi, S. M. Alwazer, and M. Alshahrani, "Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques," *Electronics*, vol. 11, no. 4, p. 530, Feb. 2022.
[9] P. Sukumaran and K. Govardhanan, "Towards voice based prediction and analysis of emotions in ASD children," *J. Intell. Fuzzy Syst.*, vol. 41, no. 5, pp. 5317–5326, 2021.
[10] S. P. Abirami, G. Kousalya, and R. Karthick, "Identification and exploration of facial expression in children with ASD in a contact less environment," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 2033–2042, Mar. 2019.
[11] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning techniques," *Health Inf. Sci. Syst.*, vol. 9, no. 1, pp. 1–13, Dec. 2021.
[12] C. Allison, B. Auyeung, and S. Baron-Cohen, "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," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 51, no. 2, pp. 202–212, 2012.
[13] F. Thabtah, F. Kamalov, and K. Rajab, "A new computational intelligence approach to detect autistic features for autism screening," *Int. J. Med. Inform.*, vol. 117, pp. 112–124, Sep. 2018.
[14] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104672.
[15] E. Dritsas and M. Trigka, "Stroke risk prediction with machine learning techniques," *Sensors*, vol. 22, no. 13, p. 4670, Jun. 2022.
[16] V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, "Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms," *Neural Comput. Appl.*, early access, pp. 1–17, Mar. 2022.
[17] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Inform. Health Social Care*, vol. 44, no. 3, pp. 278–297, 2018.
[18] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, "A machine learning approach to predict autism spectrum disorder," in *Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Feb. 2019, pp. 1–6.
[19] H. Abbas, F. Garberson, E. Glover, and D. P. Wall, "Machine learning approach for early detection of autism by combining questionnaire and home video screening," *J. Amer. Med. Informat. Assoc.*, vol. 25, no. 8, pp. 1000–1007, 2018.
[20] K. L. Goh, S. Morris, S. Rosalie, C. Foster, T. Falkmer, and T. Tan, "Typically developed adults and adults with autism spectrum disorder classification using centre of pressure measurements," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 844–848.
[21] A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, and I. Castiglioni, "Use of machine learning to identify children with autism and their motor abnormalities," *J. Autism Develop. Disorders*, vol. 45, no. 7, pp. 2146–2156, 2015.
[22] B. Tyagi, R. Mishra, and N. Bajpai, "Machine learning techniques to predict autism spectrum disorder," in *Proc. IEEE Punecon*, Jun. 2019, pp. 1–5.
[23] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," *Health Informat. J.*, vol. 26, no. 1, pp. 264–286, Mar. 2020.
[24] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioral distinction of autism and ADHD," *Transl. Psychiatry*, vol. 6, no. 2, pp. e732–e732, Feb. 2016.
[25] S. B. Shuvo, J. Ghosh, and A. S. Oyshi, "A data mining based approach to predict autism spectrum disorder considering behavioral attributes," in *Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2019, pp. 1–5.
[26] O. Altay and M. Ulas, "Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children," in *Proc. 6th Int. Symp. Digit. Forensic Secur. (ISDFS)*, Mar. 2018, pp. 1–4.
[27] F. N. Buyukoflaz and A. Ozturk, "Early autism diagnosis of children with machine learning algorithms," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, May 2018, pp. 1–4.
[28] M. F. Misman, A. A. Samah, F. A. Ezudin, H. A. Majid, Z. A. Shah, H. Hashim, and M. F. Harun, "Classification of adults with autism spectrum disorder using deep neural network," in *Proc. 1st Int. Conf. Artif. Intell. Data Sci. (AiDAS)*, Sep. 2019, pp. 29–34.
[29] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, pp. 41–51, Jan./Feb. 2018.
[30] A. S. Haroon and T. Padma, "An ensemble classification and binomial cumulative based PCA for diagnosis of Parkinson's disease and autism spectrum disorder," *Int. J. Syst. Assurance Eng. Manage.*, early access, pp. 1–16, Jul. 2022.
[31] R. Abitha, S. M. Vennila, and I. M. Zaheer, "Evolutionary multi-objective optimization of artificial neural network for classification of autism spectrum disorder screening," *J. Supercomput.*, vol. 78, no. 9, pp. 11640–11656, Jun. 2022.
[32] M. Alsuliman and H. H. Al-Baity, "Efficient diagnosis of autism with optimized machine learning models: An experimental analysis on genetic and personal characteristic datasets," *Appl. Sci.*, vol. 12, no. 8, p. 3812, Apr. 2022.
[33] S. P. Kamma, S. Bano, G. L. Niharika, G. S. Chilukuri, and D. Ghanta, "Cost-effective and efficient detection of autism from screening test data using light gradient boosting machine," in *Intelligent Sustainable Systems*. Singapore: Springer, pp. 777–789, 2022.
[34] U. Gupta, D. Gupta, and U. Agarwal, "Analysis of randomization-based approaches for autism spectrum disorder," in *Pattern Recognition and Data Analysis with Applications*. Singapore: Springer, pp. 701–713, 2022.
[35] T. Akter, M. Shahriare Satu, M. I. Khan, M. H. Ali, S. Uddin, P. Lio, J. M. W. Quinn, and M. A. Moni, "Machine learning-based models for early stage detection of autism spectrum disorders," *IEEE Access*, vol. 7, pp. 166509–166527, 2019.
[36] Kaggle. (2022). *Autism Spectrum Disorder Detection Dataset for Toddlers*. [Online]. Available: https://www.kaggle.com/fabdelja/autism-screening-for-toddlers
[37] UCI. (2022). *UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Adolescent Data Set*. [Online]. Available: https://shorturl.at/fhxCZ
[38] UCI. (2022). *UCI Machine Learning Repository: Autism Screening Adult Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult
[39] UCI. (2022). *UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Children Data Set*. [Online]. Available: https://shorturl.at/fiwLU
[40] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105524.
[41] D. Mease, A. J. Wyner, and A. Buja, "Boosted classification trees and class probability/quantile estimation," *J. Mach. Learn. Res.*, vol. 8, no. 3, pp. 409–439, 2007.
[42] Q. Wang, W. Cao, J. Guo, J. Ren, Y. Cheng, and D. N. Davis, "DMP_MI: An effective diabetes mellitus classification algorithm on imbalanced data with missing values," *IEEE Access*, vol. 7, pp. 102232–102238, 2019.
[43] S. M. M. Hasan, M. A. Mamun, M. P. Uddin, and M. A. Hossain, "Comparative analysis of classification approaches for heart disease prediction," in *Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (ICME)*, Feb. 2018, pp. 1–4.
[44] D. Ramesh and Y. S. Katheria, "Ensemble method based predictive model for analyzing disease datasets: A predictive analysis approach," *Health Technol.*, vol. 9, no. 4, pp. 533–545, Aug. 2019.
[45] A. Arabameri and H. R. Pourghasemi, "Spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R," in *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Amsterdam, The Netherlands: Elsevier, pp. 299–321, 2019.
[46] J. Stirling, T. Chen, and M. Adamou, "Autism spectrum disorder classification using a self-organising fuzzy classifier," in *Fuzzy Logic*. Cham, Switzerland: Springer, pp. 83–94, 2021.
[47] M. M. Rahman, O. L. Usman, R. C. Muniyandi, S. Sahran, S. Mohamed, and R. A. Razak, "A review of machine learning methods of feature selection and classification for autism spectrum disorder," *Brain Sci.*, vol. 10, no. 12, p. 949, Dec. 2020.
[48] M. Hasan, M. M. Ahamad, S. Aktar, and M. A. Moni, "Early stage autism spectrum disorder detection of adults and toddlers using machine learning models," in *Proc. 5th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2021, pp. 1–6.

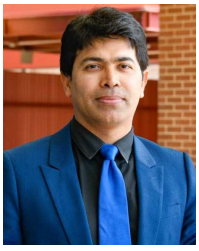
[49] H. Talabani and E. Avci, "Performance comparison of SVM kernel types on child autism disease database," in *Proc. Int. Conf. Artif. Intell. Data Process. (IDAP)*, Sep. 2018, pp. 1–5.

S. M. MAHEDY HASAN received the B.Sc. degree in computer science and engineering from the Rajshahi University of Engineering and Technology (RUET), Bangladesh. He is currently serving as an Assistant Professor with the Department of Computer Science and Engineering, RUET. Before joining RUET, he was a Lecturer at the Department of Computer Science and Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University (BSMRSTU), Bangladesh, in 2019. His research interests include computer vision, pattern recognition, machine learning, deep learning, transfer learning, biomedical engineering, bioinformatics, natural language processing, text mining, and pedagogy.

MD PALASH UDDIN (Member, IEEE) received the B.Sc. degree in computer science and engineering from Hajee Mohammad Danesh Science and Technology University (HSTU), Bangladesh, and the M.Sc. degree in computer science and engineering from the Rajshahi University of Engineering and Technology, Bangladesh. He is currently pursuing the Ph.D. degree with the School of Information Technology, Deakin University, Australia. He is also an Academic Faculty Member with HSTU. His research interests include machine learning, federated learning, blockchain, and remote sensing image analysis.

MD AL MAMUN (Senior Member, IEEE) received the B.Sc. degree in computer science and engineering from the Rajshahi University of Engineering and Technology (RUET), Bangladesh, in 2005, and the Ph.D. degree in computer science from the University of New South Wales (UNSW), Canberra, Australia, in 2011. He is currently working as a Professor with the Department of Computer Science and Engineering, RUET. He has published more than 50 publications in several international journals and conferences. His area of research interests include satellite image mining (image compression, change detection, prediction and forecasting, adaptive linear and non-linear modeling), computer vision (pattern recognition and image classification, objects recognition, feature extraction, and nonlinear image classification), machine learning, and data mining. He has served as a Reviewer for various IEEE-sponsored conferences and journals, such as IEEE Transactions on Geoscience and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, and IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. He attended several national and international conferences and served as a member of the organizing committee or technical program committee (TPC). He is also serving as the Director of the ICT Cell, RUET; an Executive Member of the IEEE Computer Society Bangladesh Chapter; an Adviser of the IEEE Computer Society Student Branch RUET; and an Executive Member of the Robotic Foundation, Eastern Region, Bangladesh.

MUHAMMAD IMRAN SHARIF received the B.S. and M.S. degrees in computer science from COMSATS University Islamabad, Wah Campus, Pakistan, in 2019 and 2021, respectively. His research interests include medical imaging, machine learning, computer vision, artificial intelligence, and pattern recognition.



ANWAAR ULHAQ received the Ph.D. degree in artificial intelligence from Monash University, Australia. He is currently working as a Senior Lecturer (AI) with the School of Computing, Mathematics, and Engineering, Charles Sturt University, Australia. He has developed national and international recognition in computer vision and image processing. His research has been featured 16 times in national and international news venues, including ABC News and IFIP (UNESCO). He is an Active Member of IEEE, ACS, and the Australian Academy of Sciences. As the Deputy Leader of the Machine Vision and Digital Health Research Group (MaViDH), he provides leadership in artificial intelligence research and leverages his leadership vision and strategy to promote AI research by mentoring junior researchers in AI and supervising HDR students devising plans to increase research impact.

GOVIND KRISHNAMOORTHY is currently a Clinical Psychologist and a Senior Lecturer with the School of Psychology and Wellbeing, University of Southern Queensland, Australia. His research and clinical practice focus on improving mental health and educational outcomes for children and adolescents. He has collaborated with health services, schools, and community services in implementing place-based and systems approaches to support developmental disorders and mental health concerns in children, adolescents, and their families.