



Classification And Diagnosis Of Autism Spectrum Disorder Using Machine Learning Approach

¹Xavier Mary A, ²Mohanapriya K, ³Monika I, ⁴Pooja shri V A,

¹Assistant. Professor, ²Student, ³Student, ⁴Student, ^{1,2,3,4}Computer Science And Engineering,
^{1,2,3,4}Adhiyamaan College of Engineering, Hosur, India

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.

1. 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 californiairvine (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- toclass 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-oneperson-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 fst 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.

2. MATERIALS & METHODS

2.1 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.

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

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

2.2. 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. 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.

2.3.MACHINE LEARNING METHOD

1) **ADA BOOST (AB):** AB is a tree-based ensemble classifier that incorporates many weak classifiers to reduce misclassification errors [41]. 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:

$$H(x) = \text{Sign}(\sum^T \alpha_t h_t(x)) \quad (9)$$

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.

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

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

$$D_{t+1}(i) = \frac{D_t(i)}{\exp(-\alpha y_i n_t(x_i))} \quad \text{..... (11)}$$

where D_{t+1} denotes the updated weight, D_t specifies the weight of $\mathbb{Z}_t$ 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 $x_i \in D$, exists to a class C_i , and is predicted by $|C_i \cap D|/|D|$. To classify instances in the dataset D , the required information is needed, and the following equation calculates it: $\text{Info}(D) = -\sum_{i=1}^m P_i \log_2(P_i)$ (12) where $\text{Info}(D)$ is the average amount of information needed to identify C_i of an instance, $x_i \in D$ and the objective of DT is to divide repeatedly, D , into sub datasets $D_1, D_2, \dots, D_n$. The following equation estimates the $\text{InfoA}(D)$: $\text{InfoA}(D) = \sum_{j=1}^n |D_j|/|D| * \text{Info}(D_j)$ (13) Finally, the following equation calculates the information gain value: $\text{Gain}(A) = \text{Info}(D) - \text{InfoA}(D)$ (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: $D_e = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$ (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: $\mu = \frac{1}{n} \sum_{i=1}^n x_i$ (16) $\delta = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}$ (17) $f(x) = \frac{1}{\sqrt{2\pi} * \delta} * e^{-\frac{(x-\mu)^2}{2\delta^2}}$ (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]: $p = \frac{1}{1 + e^{-x}}$, (19) where p denotes the probability of instance x . At the time of model training, for each instance $x_1, x_2, x_3, \dots, x_n$ the logistic coefficients will be $b_0, b_1, b_2, \dots, b_n$. The stochastic gradient descent method estimates and updates the values of the coefficients. $v = b_0x_0 + b_1x_1 + \dots + b_nx_n$ (20) $p = \frac{1}{1 + e^{-v}}$ (21) Now, the following equation is used to update the values of the coefficients: $b = b + l * (y - p) * (1 - p) * x$ (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 nonlinear 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: $K(x, x_0) = \exp(-(\|x - x_0\|)^2 / 2\delta^2)$ (23) where $(\|x - x_0\|)^2$ 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 $\{x_1, x_2, \dots, x_n\}$ with classes $z_i \in \{1, \dots, k\}$. The prior probability is assumed to display as Gaussian distribution $\phi(x|\mu_k, \Sigma_k)$ in each class. The model estimation is defined as follows: $a_k = \frac{1}{n} \sum_{i=1}^n I(z_i = k)$ (24) $\mu_k = \frac{1}{n_k} \sum_{i=1}^n x_i I(z_i = k)$ (25) $\Sigma_k = \frac{1}{n_k} \sum_{i=1}^n (x_i - \mu_k)(x_i - \mu_k)^T$ (26) where a_k denotes the prior probability, μ_k defines mean of all classes, Σ_k indicates the sample covariance of the class means.

3. Proposed Methodology

Figure 1 shows the steps in the proposed workflow which involves the pre-processing of data, training, and testing with specified models, evaluation of results and prediction of ASD. This work is implemented in Python 3.

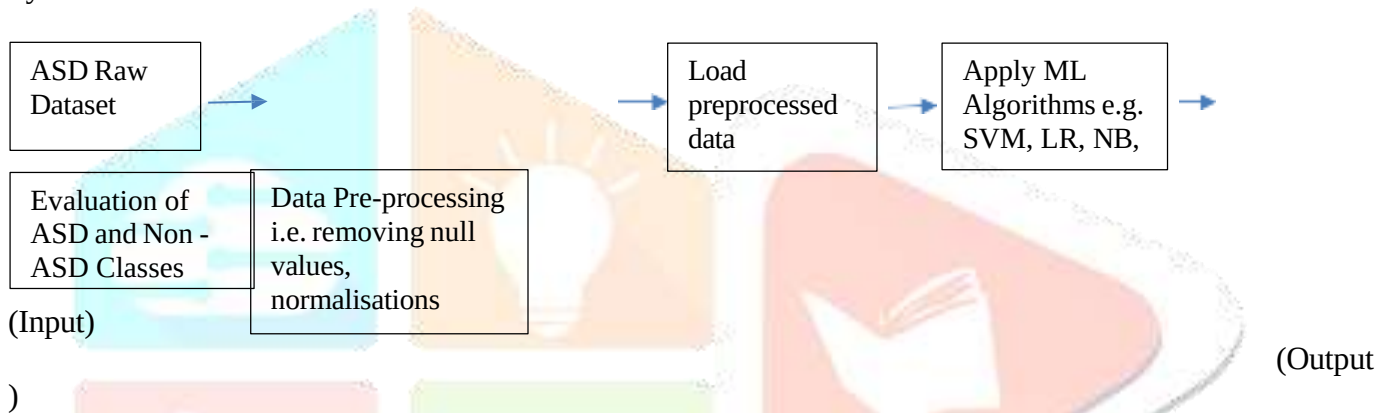


Figure. 1. Steps in the proposed ASD detection solution

3.1. Data Pre-Processing

Data pre-processing is a technique in which transform the raw data into a meaningful and understandable format. Real-world data is commonly incomplete and inconsistent because it contains lots of errors and null values. A good pre-processed data always yields to a good result. Various Data pre-processing methods are used to handle incomplete and inconsistent data like as handling missing values, outlier detection, data discretization, data reduction (dimension and numerosity reduction), etc. The problems of missing values in these dataset has been handled by imputation method.

3.2. Training and Testing Model

The whole dataset has been split into two parts i.e. one part is training the dataset and the other one is testing dataset with a ratio of 80:20 respectively. For cross-validation purposes again training data has been split into two parts. One part is the training dataset and another part is the validation dataset into an 80:20 ratio respectively. Figure 2 shows the final training, testing and validation sets on which classification has been performed.

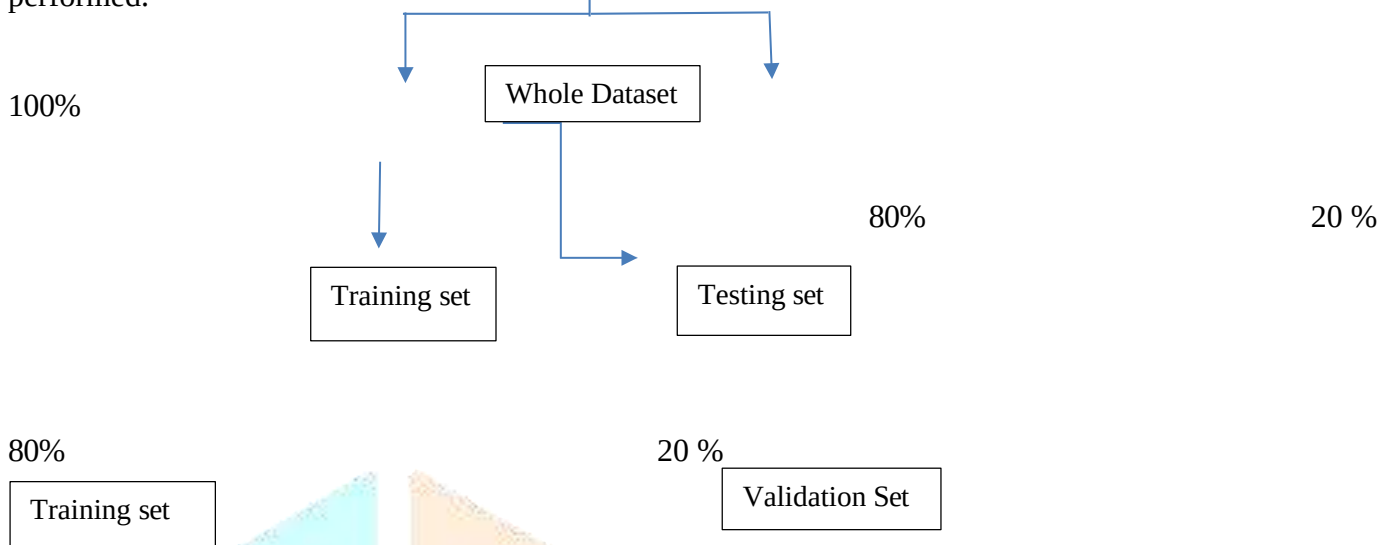


Figure. 2. Final Training, Testing and Validation Sets

4. Support Vector Machine (SVM)

SVM is a linear supervised machine learning approach that is used for classification and regression. It is a pattern recognition problem solver. It does not cause the problem of overfitting. SVM separates the classes by defining a decision boundary [19].

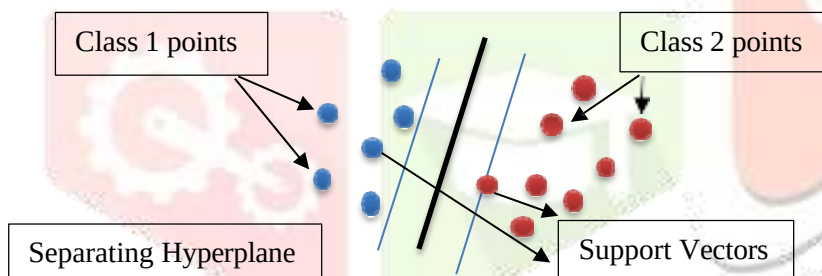


Figure. 3. An SVM classifier

4.1. Naïve Bayes (NB) :

A naive Bayes classifier is a supervised learning algorithm. It is a generative model and is based on joint probability distribution. The Naive Bayes concept based on independence assumptions. It exhibits less training time as compared to SVM and ME model. It calculates the posterior probability for a dataset using the prior probability and likelihood [17].

4.2. Convolutional Neural Network (CNN):

CNN is one of the deep learning techniques known to build models for various problems [24] or [25] or [26]. It is a feed- forward neural network that is inspired by the human brain. A CNN model contains one input layer, one output layer, and many other different layers i.e. convolution layers, max pooling, fully connected layers, and normalization layers. Their activation functions can be computed with Matrix Multiplication, which is followed by a bias offset.

4.3. Logistic Regression (LR):

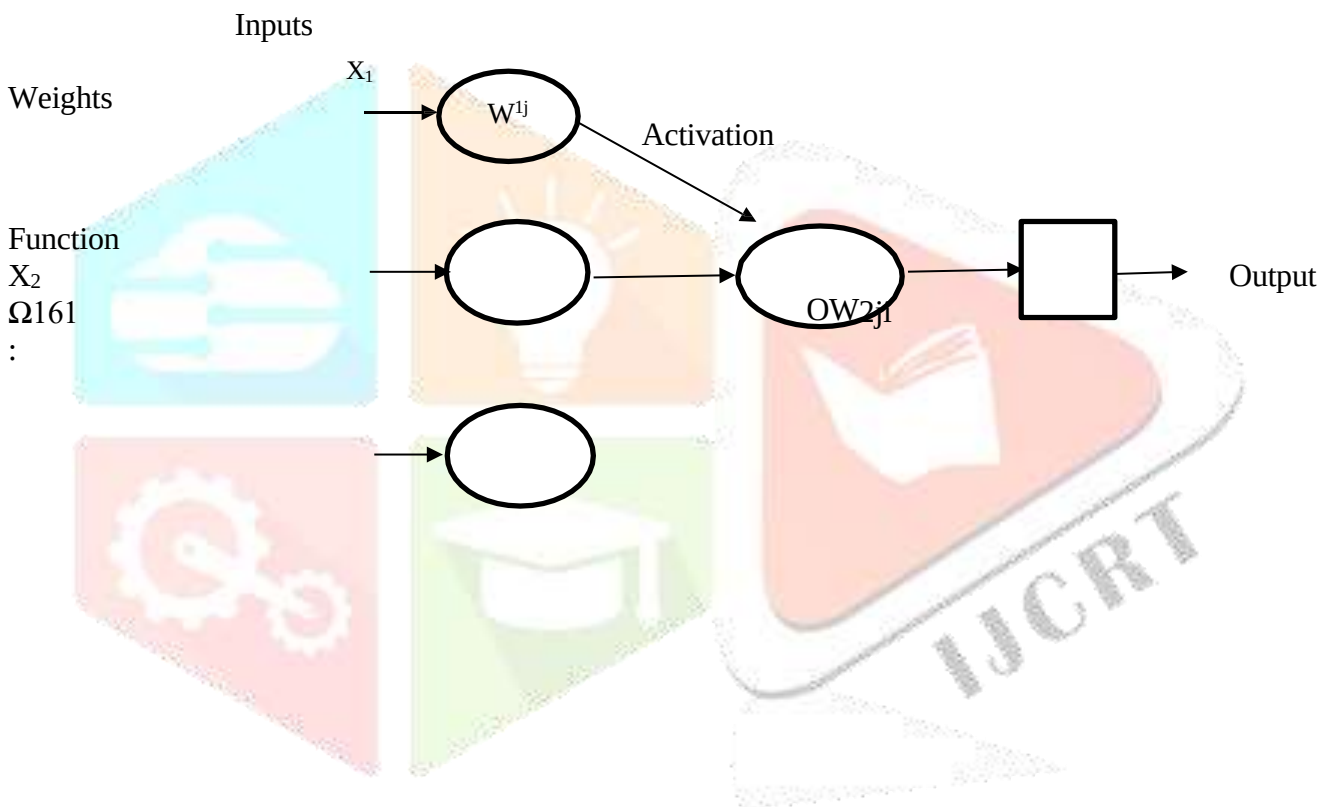
LR is a regression tool that is used to analyse the binary dependent variables. Its output value lies in either the 0 or 1 form. It is used for the continuous value dataset. It tells the relationship between one dependent binary variable, and one nominal or ordinary variable. It can be represented by the sigmoidal function.

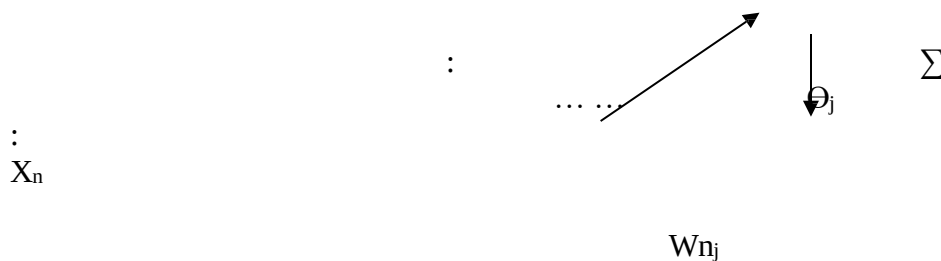
4.4. K- Nearest Neighbour (KNN):

KNN is a supervised learning approach and is the simplest of all. It is used for classification as well as regression problems. It assumes that similar data exist nearby. The 'K' part indicates the number of seed point that is to be selected. It should be chosen carefully to reduce the error. Thus it is based on the idea of similarity which can be in terms of distance, closeness or proximity. The most common distance measure is Euclidean distance.

4.5. Artificial Neural Network:

ANN is a neural network that has a connection with multiple neurons. Each neuron cell having a group of input values and associated weights. The most common artificial 0 neural network feeds forward neural network. In this network, the flow of information moves in the only forward direction. This type of network contains three main layers, first is the input layer, the second is a hidden layer and last is the output layer. There is no cycle or loop in the network [21].





5. Model Evaluation:

Model evaluation is a crucial step in assessing the effectiveness, accuracy, and reliability of machine learning models used for ASD classification and diagnosis. It ensures that the model performs well on real-world data and can generalize beyond the training dataset.

Key Evaluation Metrics:

- Accuracy – Measures the overall correctness of the model in predicting ASD and non-ASD cases.
- Precision – Indicates how many of the predicted ASD cases are actually correct, reducing false positives.
- Recall (Sensitivity) – Shows how well the model identifies actual ASD cases, minimizing false negatives.
- F1-Score – A balanced metric that combines precision and recall, useful for imbalanced datasets.
- ROC Curve & AUC (Area Under Curve) – Measures the model's ability to differentiate between ASD and non-ASD cases. A higher AUC value indicates better performance.

Evaluation Techniques:

- Confusion Matrix: A table that visualizes true positives, true negatives, false positives, and false negatives.
- Cross-Validation (K-Fold): Helps ensure the model's performance is stable across different subsets of the dataset.
- Overfitting Detection: Regularization techniques and validation sets help prevent the model from memorizing the training data instead of learning patterns.

A well-evaluated model ensures accurate and early diagnosis of ASD, making it a valuable tool in clinical and healthcare applications.

6. Literature Survey

Recent studies have explored various machine learning (ML) approaches for the early detection of Autism Spectrum Disorder (ASD). Bala et al. (2022) developed an ML model incorporating feature selection techniques and classifiers like Support Vector Machine (SVM), achieving up to 99.61% accuracy in child datasets.

Another study by Liao et al. (2022) proposed a fusion of physiological data (EEG) and behavioral data (eye fixation and facial expression) using a weighted naive Bayes algorithm, resulting in an 87.50% classification accuracy.

In 2023, a study integrated multiple behavioral and developmental assessments with ML classifiers, including decision trees and eXtreme gradient boosting (XGB), achieving an accuracy of 78.3%, surpassing senior pediatricians' 66.7% accuracy.

Additionally, a 2024 systematic review highlighted the effectiveness of ML in analyzing motion patterns for early ASD detection, though challenges remain in clinical application. These studies underscore the potential of ML techniques in enhancing early ASD diagnosis through diverse data sources and methodologies.

Vaishali R, Sasikala R. et al. [3] have proposed a method to identify Autism with optimum behavior sets. In this work, an ASD diagnosis dataset with 21 features obtained from the UCI machine learning repository experimented with swarm intelligence based binary firefly feature selection wrapper.

The alternative hypothesis of the experiment claims that it is possible for a machine learning model to achieve a better classification accuracy with minimum feature subsets. Using Swarm intelligence based single-objective binary firefly feature selection wrapper it is found that 10 features among 21 features of ASD dataset are sufficient to distinguish between ASD and non-ASD patients. The results obtained with this approach justifies the hypothesis by producing an average accuracy in the range of 92.12%–97.95% with optimum feature subsets which are approximately equal to the average accuracy produced by the entire ASD diagnosis dataset.

Fadi Thabtah et al. [8] have proposed an ASD screening model using Machine Learning Adaption and DSM-5. A screening tool has been used to realize one or more goals in ASD screening. In this paper, the researcher discussed the ASD Machine Learning classification with their pros and cons. The researcher tried to highlight the problem accompanying with existing ASD screening tools and the consistency of such tools using the DSM-IV instead of the DSM-5 manual.

M. S. Mythili, A. R. Mohamed Shanavas et al. [13] have a study on ASD using Classification Techniques. The main aim of this paper was to detect the autism problem and the levels of autism. In this Neural Network, SVM and Fuzzy techniques with WEKA tools are used to analyze the student's behavior and their social interaction.

J. A. Kosmicki¹, V. Sochat, M. Duda and D.P. Wall Et al. [14] have supposed a searching method for a least set of traits for autism detection. In this, the authors used a machine learning approach to evaluate the clinical assessment of ASD. The ADOS was performed on the subset of behaviors of children based on the autism spectrum. ADOS having four modules. In this work, 8 different machine learning algorithms were employed, involving stepwise backward feature identification on score sheets from 4540 individuals. It uses 9 out of the 28 behaviors from module 2 and 12 out of the 28 behaviors from module 3 to identify an ASD risk with an overall accuracy of 98.27% and 97.66% respectively.

Li B, A. Sharma, J Meng, S. Purushwalkam, E. Gowen (2017) et al. [11] have used machine learning classifiers to detect autistic adults by imitation method. The goal of this study was to investigate the fundamental problem related to discriminative test conditions and kinematic parameters. The dataset contains 16 ASC participants who have a series of hand movements. In this 40 kinematic constraints from 08 imitation conditions has been extracted by using machine learning methods. This research shows that for a small sample, there is a feasibility of applying machine learning methods to analyze high-dimensional data and the diagnostic classification of autism. The sensitivity rates achieved by RIPPER which have the features Va (87.30 %), CHI (80.95%), IG (80.95%), Correlation (84.13%), CFS (84.13%), and “no feature selection” (80.00%) on the AQ-Adolescent dataset.

It is evident from the above discussed section that there is definitely a need to explore the possibility of applying deep learning based models for the detection of ASD in human population. Most of the work discussed above use conventional machine learning approaches and hence are limited in their performance. In this work, performance of several machine learning models have been compared to that of the deep learning model for this purpose. Separate models have been prepared for separate population set (discussed in section below) and compared individually.

7. Workflow:

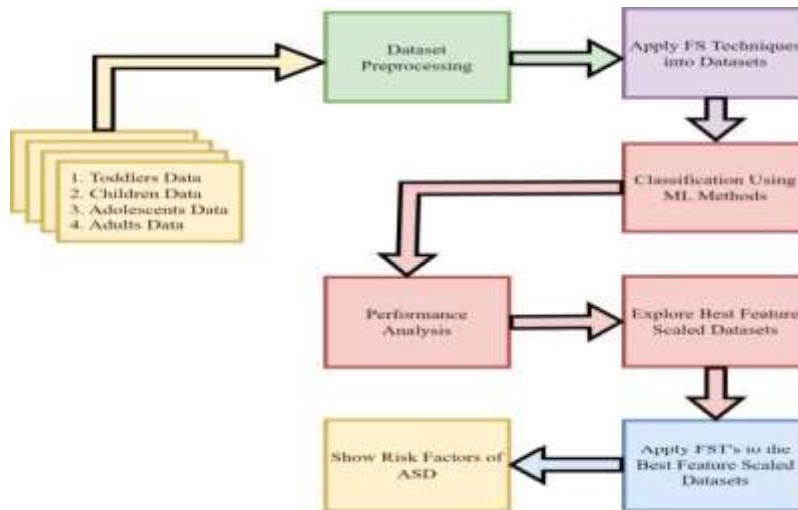


FIGURE 1. Sequential workflow for detecting ASD at an early stage.

8. 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 featurescaled 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.

9. 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," *Biomedicine*, 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

spirometry 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 multiobjective 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:



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.



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.



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.



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.



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.