Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 9 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i09.48831

PAPER

An Efficient Autism Spectrum Disorder Classification in Different Age Groups using Machine Learning Models

Ambika Rani Subhash[1,2](✉), Ashwin Kumar U Motagi[2]

[1]Department of Information Science and Engineering, BMS Institute of Technology and Management, Bengaluru, Karnataka, India

[2]School of Computer Science and Engineering, REVA University, Bengaluru, Karnataka, India

ambikasubash@bmsit.in

ABSTRACT

The current world has witnessed the emergence of various illnesses, such as autism spectrum disorder (ASD), that are not yet medically recognized. It impacts multiple behavioral domains, such as repetitive and stereotyped behavior, social competence, and linguistic skills. This condition is a severe neurodevelopmental disorder. It Identifying and classifying ASD is challenging and time-consuming due to its symptoms being remarkably similar to those of many other mental illnesses. Machine learning-based models are increasingly being used to predict a wide range of human diseases, leveraging various physiological and other characteristics. Our study aims to develop a classification model that can predict the likelihood of ASD in various age groups, such as toddlers, children, adolescents, and adults. We have utilized several machine learning (ML) algorithms, including support vector machine (SVM), Naive Bayes (NB), random forest (RF), extra trees classifier (ET), k-nearest neighbor (K-NN), decision tree (DT), Ada boost classifier (AB), and stochastic gradient descent (SGD) classifiers. These models are tested using four unique non-clinical ASD screening datasets that are publicly available from Kaggle and the UCI library. In the first dataset, there are 1054 instances and 19 features related to toddlers. The remaining ones consist of 21 traits and, for children, adolescents, and adults, 292, 104, and 704 cases, respectively. The outcomes of the experimentation have shown that the SDG, DT, and ET classifiers are the most commonly used models and have achieved results with almost 100% accuracy.

KEYWORDS

autism spectrum disorder (ASD), machine learning (ML), support vector machine (SVM), naive bayes (NB), random forest (RF), extra trees (ET), k-nearest neighbor (K-NN), decision tree (DT), Ada boost (AB), stochastic gradient descent (SGD)

1 INTRODUCTION

We have observed numerous diseases that cannot be clinically diagnosed, among which autism spectrum disorder (ASD) is one example. This condition affects many behavioral domains, such as social and communication skills, as well as stereotyped

Subhash, A.R., Motagi, A.K.U. (2024). An Efficient Autism Spectrum Disorder Classification in Different Age Groups using Machine Learning Models. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(9), pp. 17–38. https://doi.org/10.3991/ijoe.v20i09.48831

Article submitted 2024-02-28. Revision uploaded 2024-04-19. Final acceptance 2024-04-19.

and repetitive behaviors. It is a significant neurodevelopmental disorder [1]. A set of neurological conditions known as ASDs impede the brain's normal development [2]. ASD can lead to social challenges, sensory issues, repetitive behaviors, and intellectual disabilities. Psychiatric or neurological conditions like hyperactivity, attention deficit disorder, anxiety, depression, and epilepsy are also common in patients with ASD [3]. Though the exact cause of the disease is unknown, biological factors such as brain inflammation, genetic abnormalities, and unfavorable pregnancy circumstances are thought to be involved. The number of children identified with ASD is rapidly increasing, underscoring the need for further exploration of this population. Correct scientific procedures are particularly important [4].

With ASD, it affects how people act, communicate, and educate themselves [5]. The symptoms and indicators of ASD start early; 0.63% of very young children have been diagnosed with ASD, and the condition also affects adolescents and adults, according to WHO statistics [6]. An individual with ASD may experience mental health issues such as anxiety and misconceptions, which can hinder their ability to function well at different stages of life [7]. Early diagnosis and treatment are always important [8]. The behavior of the affected individual, which may include dangerous behavior influenced by movies and animations, is one of the most obvious indicators of ASD [9].

Conversely, ASD is a type of neurological disorder that significantly affects an individual's entire life. It's important to remember that both environmental and genetic factors could contribute to the development of this illness [8]. Patients with this illness cannot be completely cured, but if the symptoms are identified early on, the effects can be temporarily lessened. Because researchers have been unable to pinpoint the exact cause of ASD, they assume that human genes are to blame [9]. Even though a person with ASD frequently finds it difficult to interact with others and faces some major issues [10], such as lack of pain sensitivity, difficulty maintaining proper eye contact while communicating, inability to respond to sounds made by objects or humans, challenges in understanding and conveying proper gestures for effective interaction, lack of engagement with others, and a preference for living alone.

Moreover, early diagnosis based on a variety of physiological and health parameters seems feasible due to the increasing use of machine learning (ML)-based models in predicting various human diseases. This characteristic has piqued our curiosity about ASD prediction, diagnosis, analysis, and treatment approaches [11]. The task of diagnosing ASD is challenging because it relies on distinguishing it from other mental illnesses that exhibit symptoms closely resembling those of ASD. Furthermore, ML is the most popular field for detecting symptoms of autism by utilizing various techniques to identify the condition and determine whether the patient is affected or not [12]. Furthermore, ML has been used in a wide range of applications to solve real-time problems [13–17].

1.1 Problem definition

The severe, lifelong neurodevelopmental syndrome known as ASD is characterized by long-term or restricted impairments in thinking, behavior, activities, and socio-communication skills. The prediction and classification of ASD among individuals vary with age and other personal factors. The aim of this study is to develop a more accurate and sophisticated machine learning model for diagnosing ASD across various age groups. The ML algorithm provides the medical treatment system with accurate predictions for autism spectrum disorder.

1.2 Research contributions

Our contributions to the paper are highlighted as follows:

- Prediction of ASD by considering the possible age groups in the dataset (Toddler, Child, Adolescent, and Adult).
- Feature selection and normalization across different datasets.
- Improved machine learning-based ASD prediction with higher accuracy and enhanced performance.
- The comparison of the proposed model with related earlier works proves to be effective.

1.3 Paper organization

The rest of the paper is organized as follows: Section 2 provides a summary of the literature on ASD prediction and classification models. Section 3 presents the proposed work, including the dataset description and the machine learning model. Section 4 presents the implementation, results, and discussions. Finally, the paper concludes in Section 5.

2 LITERATURE REVIEW

This section describes previous research on the detection and prediction of ASD using ML-based techniques. The primary goal is to examine and identify certain shortcomings to suggest a new, improved, and superior ML-based method for predicting autism spectrum disorder.

One of the most popular methods for identifying functional patterns to diagnose various illnesses and determine appropriate treatments is the ML approach [11–12]. Various techniques are used in the treatment of autism patients to determine whether or not the patient is affected. Azian A. et al. [18] have proposed three methods for testing ML techniques that could be applied to regression and classification: least absolute shrinkage, Chi-square, and selection operator (LASSO). These methods include logistic regression (LR), random forest (RF), and K-nearest neighbors (K-NN). Among the different techniques used, the experimental results demonstrated that the LR had a maximum accuracy of 97.541%. The authors of [19] proposed a method for identifying autism through the use of ideal behavioral sets. To check for ASD, they employed a swarm intelligence-based binary feature selection method. There are 21 features in the examined and categorized dataset that were extracted from the ML repository. Out of the 21 characteristics examined in the ASD dataset, the authors found that only ten could be used to differentiate between patients with ASD and those without. The system was able to achieve an accuracy of 97.79%, as demonstrated by their experimental results.

An efficient method for assessing ML techniques for early ASD detection was demonstrated by Hasan et al. [20]. This system classified feature-scaled datasets using eight simple yet effective ML algorithms and four attribute scaling (AS) techniques. With the highest accuracy rates of 99.03% for adults, 97.12% for teenagers, 99.25% for toddlers, and 97.95% for children, AB and LDA successfully identified autism. In a different study, Rodrigues et al. [21] utilized functional magnetic resonance imaging and ML to identify potential markers associated with the prevalence of ASD. They measured severity using the ADOS score. With a cingulum region accuracy of 73.8%, their results suggest a functional difference between ASD subclasses.

Finding the most accurate method to calculate ASD among multiple measurements made in various classifiers, such as the Gaussian Radial Kernel and support vector machine (SVM), was recommended in [22]. With a 95% accuracy rate, the results obtained demonstrated the highest level of accuracy. Utilizing the publicly available standard ASD dataset. Raj et al. [23] presented a framework that incorporates several ML algorithms. In data for adults, adolescents, and children, the performance accuracy of predictive and classification models for ASD was 99.53%, 96.68%, and 98.30%, respectively. Hossain et al. [24] made an effort to identify the most critical features and organize early diagnosis by utilizing classification algorithms to enhance diagnostic procedures. They found that the accuracy of SVM is greater than that of all other ML methods. They demonstrated that the telief algorithm is the most effective technique for identifying the most important features in ASD datasets.

The authors of the paper [25] offered several strategies and approaches for recognizing and classifying ASD. They employed a variety of ML techniques, including classifiers and neural network-based classifiers. To determine the effectiveness of the proposed system in detecting ASD, a comprehensive test was conducted. Three datasets were used in the experiments: adult, adolescent, and child. The experimental results demonstrated that when each ML classifier was tested using precision, F-beta score, and recall methods, some of them performed better in terms of accuracy than the other studied classifiers. To process the first identified autism datasets, Akter et al. [26] collected data on infants, children, teenagers, and adults. They then processed these datasets using a variety of feature transformation methods. Using these modified ASD datasets, various classification techniques were then assessed for effectiveness. The toddler dataset produced the best SVM results; the children's dataset produced the best Ada boost (AB) results; the teenage dataset produced the best Glmboost results; and the adult dataset produced the best AB results. Using ML techniques [27] such as RF, gradient boosting machine (GBM), and SVM, researchers were able to identify critical traits predictive of ASD with 98.77% accuracy in the studies mentioned. Using the previously mentioned algorithms, a second study examined the relationship between gut microbiota and ASD and discovered that Parasutterella and Alloprevotella were important genera associated with autism spectrum disorder.

Additionally, a different study suggested [28] employing ML to pinpoint the distinctive traits of individuals with ASD and differentiate autism subgroups from neurotypical groups. To accomplish this, k-means clustering was utilized to identify subcategories within autism records. Furthermore, a classification model [29] for all ages and a machine learning-based ASD prediction model were presented, combining random forest with iterative classification and a regression tree. Based on evaluation metrics, the recommended prediction model outperformed the existing datasets.

Some of the authors also worked on utilizing deep learning mechanisms. For instance, in the context of diagnosing ASD, Yin et al. [30] developed deep learning techniques using functional magnetic resonance imaging (fMRI) data from brain systems. A graph-based classification method was employed in another study [31]. While this approach yields superior results, it does not address missing values or apply data normalization. An earlier study [32] examined intrinsic brain networks. It is inferred that aberrant mechanisms could be the cause of ASD. Dysfunction in the SN and visual systems, as well as related processes, may underlie individual differences in the severity of ASD symptoms. According to Smith et al. [33], there is an association between symptoms of ASD disorder and weaker communication with RSN chronological features, as well as a higher severity of symptoms in individuals with ASD. The results imply that entropy and FC provide more information about the brain's temporal-spatial organization.

3 PROPOSED WORK

The proposed method, as illustrated in Figure 1, introduces a system comprising concepts that are employed through optimal feature selection to aid individuals in understanding, comprehending, or estimating the probability of having autism spectrum disorder. The primary goal of the proposed abstract model is to convey the fundamental ideas and characteristics of the system. The purpose of the model is to provide software users with an interpretive understanding of the framework. The six main steps in the proposed model are: (i) Data collection, which involves collecting data from various sources with different parameters. (ii) Data preprocessing involves imputing missing values rather than deleting them. Also, remove the outliers from the dataset. (iii) Data splitting is used to divide the training, testing, and validation processes. (iv) Implement a classification model using ML classifiers such as SVM, Naive Bayes (NB), RF, extra trees (ET), KNN, decision tree (DT), AB, and stochastic gradient descent (SGD) to validate the classification results. (v) Model evaluation involves assessing performance using metrics such as accuracy, precision, recall, sensitivity, specificity, mean score, and AUC-ROC. (vi) Model validation is carried out using the k-fold mechanism.

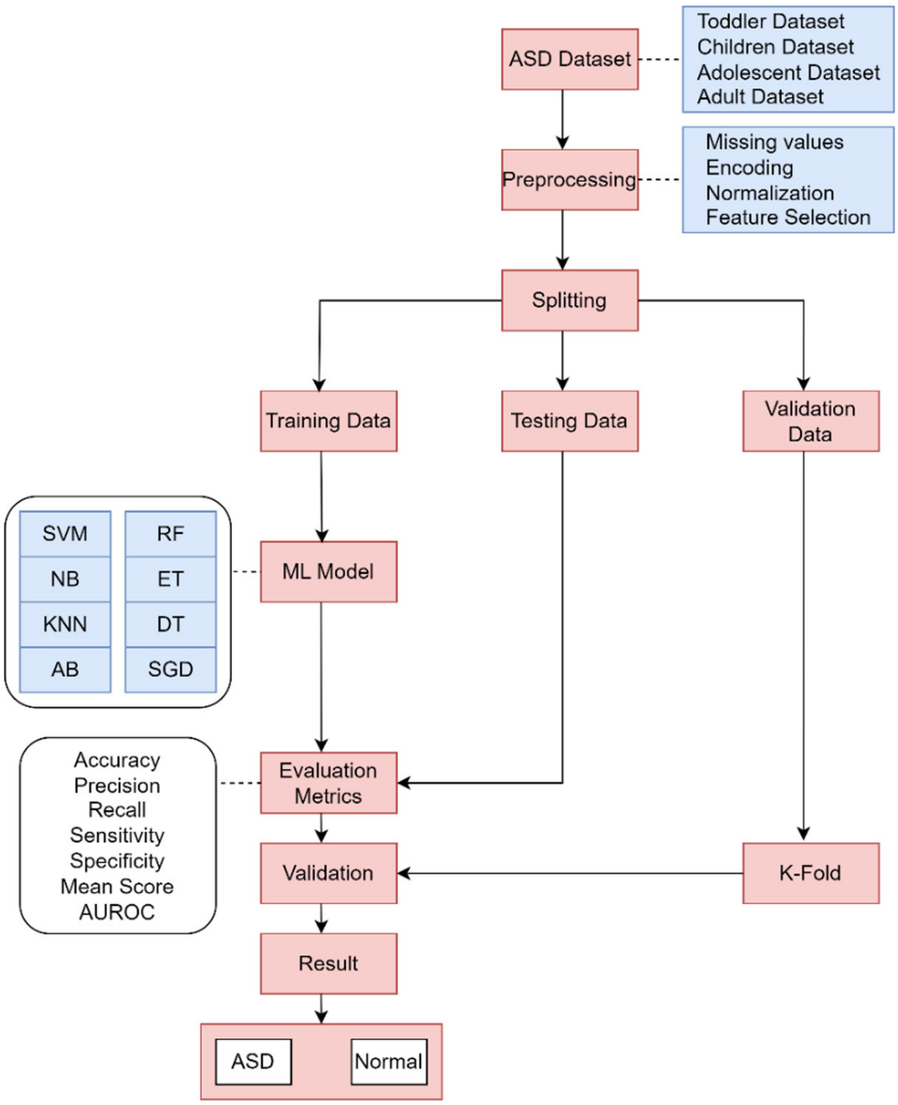

Fig. 1. Proposed system for ASD detection

3.1 Dataset description

We have utilized four ASD datasets for experimentation and analysis. We have used datasets of toddlers, children, adolescents, and adults. The overview of the datasets used is provided in Table 1. The datasets under consideration consist of 21 common features but vary in sample sizes, with the largest containing up to 1054 samples. The description of 21 features is provided in Table 2, but the toddler's dataset only includes 19 features. Further, the descriptions of the ten questionnaires (Q-CHAT-10) are provided in Table 3. The term "person" in the table refers to either a toddler, child, adolescent, or adult. The behavior of a person, whether affected by ASD or not, is classified based on all the listed features.

Table 1. Description of used ASD dataset

No	Source	Category	Attribute Type	# Features	# Samples
1	[34]	Toddler	Categorical, continuous and binary	19	1054
2	[35]	Children	Categorical, continuous and binary	21	292
3	[36]	Adolescent	Categorical, continuous and binary	21	104
4	[37]	Adult	Categorical, continuous and binary	21	704

Table 2. Description of features in ASD dataset

Feature	Feature Description
A1–A10	Screening questions
'age'	Age of candidate
'gender'	Gender of candidate
'ethnicity'	Ethnicity
'jundice'	Born with jaundice
'austim'	Family member with same problem
'contry_of_res'	Country of residence
'used_app_before'	User is familiar with application or not
'relation'	Who is responsible for success of test?
'age_desc'	Kind of screening test
'result'	Screening result
'Class/ASD'	Class/ASD

Table 3. Screening questionnaires used in ASD dataset

Number	Question
A1	What is the response when you call a person by name?
A2	How easy to get an eye contact with person?
A3	Does the person ask for anything by pointing to it?
A4	Does the person try to share anything with you?
A5	Does the person pretends to take care of his own things?
A6	Does the person follow you where you are looking and pointing to?

(Continued)

Table 3. Screening questionnaires used in ASD dataset *(Continued)*

Number	Question
A7	Does the person identify and try to comfort his family members when they are upset?
A8	What are the possible first words the person uttered all about?
A9	Does the person uses or makes any gestures?
A10	Does the person keep staring at anything without proper purpose?

3.2 Data pre-processing

It is one of the methods used to transform raw data into a comprehensible and useful format. Real-world data often contains a significant number of null variables and errors, which can render it incomplete and unstructured. We have considered pre-processing methods including handling missing values, data normalization, encoding, feature selection, and dataset splitting.

Missing values: To identify and handle outliers, this step involves an exploratory data analysis process. An iterative imputer was used to handle the missing values. Using the iterative imputation method, each feature is imputed based on the other features.

Data normalization: The data fields in the dataset may be of different data types. For effective classification, the data values must be encoded with the same data type. The parameter values are scaled between 0 and 1 to obtain an accurate result. The numeric column values in the dataset are adjusted before they can be standardized, without altering the possible values or deleting any data.

Feature selection: Given the significance of feature selection, sequential forward selection (SFS) is utilized. Several features with low significance have been excluded from the dataset using this method. Among the features listed in Table 2, we have excluded 'country_of_res,' 'used_app_before,' 'age_desc,' and 'relation' as they do not significantly impact the classification.

3.3 Dataset splitting

The entire dataset of autism patients is currently divided into two parts for testing and training. According to the suggested model, 70% of the data was used for the training partition, and the remaining 30% was used for testing. The number of instances for training and testing varies across different datasets: toddlers (1054), children (292), adolescents (104), and adults (704). Training data is further divided into two subsections for cross-validation, with a ratio of 80:20 for the training and validation datasets.

3.4 Classification model

We have utilized familiar ML algorithms for screening and classifying ASD. We have used SVM, NB, RF, ET, KNN, DT, AB, and SGD to validate the classification results.

i) **Support vector machine:** It is used to solve regression and classification problems. Due to its effectiveness and capacity to achieve exceptional accuracy in the majority of data, it is typically utilized in classification problems. The goal of this algorithm is to maximize the margin to determine the optimal degree of separation between the classes. The primary concept of the SVM model is to divide classes based on the establishment of decision boundaries [38].

ii) **Naïve Bayes:** This method, which is a supervised ML technique, is based on the principle of probability. The efficiency of forecasting and processing speed define this approach. When compared to SVM and other ML models, Naïve Bayes (NB) requires less training time. This is primarily due to its statistical concept of determining the probability of a desired outcome [39].

iii) **Decision tree classifier:** It can be applied to regression and classification problems. Still, the most common application is resolving classification problems. Dataset attributes are present in the internal nodes of a tree. Each leaf node in the tree structure represents a conclusion, while the branches represent decision rules. Additionally, DT poses a query and creates subtrees within the tree based on the response (Yes or No) [40].

iv) **Random forest classifier:** The DT is based on the RF approach. DTs utilize the related questions and their respective answers to narrow down the range in the tree with a high confidence level in order to generate a single forecast. Some DT predictions may be inaccurate. Nonetheless, forecast accuracy is increased when multiple DTs are combined into a single model. The combination of multiple DTs alludes to RF, an ML model that is applied to regression and classification [41].

v) **Extra trees classifier:** The extremely randomized tree classifier, or ET classifier for short, is an ensemble ML algorithm used for classification tasks. It is comparable to and an extension of the RF algorithm. This algorithm is robust and versatile, employing randomization and ensemble learning to generate dependable predictions in classification tasks [42].

vi) **K-nearest neighbor:** K-NN, or K-nearest neighbors, is a fundamental ML technique that is user-friendly and highly efficient. It doesn't require complicated mathematical equations, unlike other models. It can be used for tasks involving both classification and regression. The main idea is based on finding similar data nearby by measuring the distances between data points. The 'K' component denotes the quantity of nearby data points taken into account, and it is essential to carefully choose it to reduce the possibility of prediction errors [43].

vii) **Ada boost classifier:** A popular ensemble learning algorithm in ML is called AB, short for Adaptive Boosting. To build a robust predictive model, it combines the outputs of multiple weak learners, typically decision trees. AB assigns a weight to each data point, emphasizing incorrectly classified data points in subsequent iterations. This allows the algorithm to focus on improving its performance in those instances. The final prediction is the weighted total of the predictions made by the weak learners. With its reputation for managing intricate datasets and enhancing overall model accuracy, AB is a flexible and potent algorithm for classification applications [44].

viii) **SGD classifier:** For classification tasks, the stochastic gradient descent classifier, or SGD classifier, is a ML algorithm. It is a member of the linear classifier family and trains its model using the stochastic gradient descent optimization technique. The SGD classifier is computationally efficient and suitable for large datasets because it updates the model parameters based on a randomly chosen subset of the data instead of processing the entire dataset in each iteration. In online learning scenarios where the model can instantly adjust to new data, it is especially helpful. Because of its adaptability and ability to address a wide range of classification issues, the algorithm offers a scalable and effective method for training models [45].

3.5 Model evaluation metrics

The evaluation metrics are explained and presented in this section. Determining how well a predictive model performs in achieving an objective requires evaluating the model's performance. The performance and effectiveness of the classification model are evaluated using performance assessment metrics on the test dataset. Using the appropriate metrics is essential for evaluating performance, including sensitivity, specificity, accuracy, precision, recall, mean score, and AUC-ROC. The primary parameters that determine these metrics are the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Accuracy is the ratio of correctly classified samples to all samples, and it is one of the most commonly used metrics to assess classification performance. It is calculated as:

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

Precision is defined as the ratio of correctly identified positives to all predicted positives. Mathematically speaking, it is calculated as:

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

Recall is the total number of accurate predictions made across all valid samples. It is computed as:

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

Mean score: Generally speaking, a "mean score" is the average of a group of scores or numerical values. It is a central tendency that provides a general representation of the data. Depending on the specific context, the term "mean score" in ML can have various interpretations. Here are two typical interpretations: cross-validation mean score and model evaluation mean score.

F1 score: A popular metric in statistics and ML for assessing a classification model's performance is the F1 score. When working with imbalanced datasets-meaning when one class significantly outnumbers the other—it is particularly beneficial. It is calculated as:

$$F1\,score = \frac{2(Precison * Recall)}{Precison + Recall}$$

Sensitivity: Often referred to as the hit rate, recall rate or true positive rate, this crucial performance indicator is used in binary classification. It assesses a model's ability to accurately identify positive instances among all the real positive instances in the dataset. It is calculated as:

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

Specificity: A key performance indicator in binary classification is specificity, which assesses a model's ability to accurately identify negative instances among all actual negative instances in the dataset. It is calculated as:

$$Specificity = \frac{TN}{TN + FP}$$

Area under curve: One metric used to assess the performance of a binary classification model is the receiver operating characteristic (ROC)'s area under the curve (AUC). The trade-off between the true positive rate (sensitivity) and the false positive rate (specificity) at different threshold settings is graphically represented by the ROC curve.

3.6 Model validation

A mathematical technique for evaluating mastery-learning capacities is cross-validation. The validation process uses the K-fold validation approach. The entire dataset is utilized for both testing and training in the K-fold method. In this manner, 70% of the dataset is used for training, while 30% is allocated for testing, considering the relevant test cases. The results are then validated and verified with the entire dataset.

4 RESULTS AND DISCUSSION

The experimentation was conducted using Google CoLab. Pandas is used to load the dataset, and Matplotlib is used to create plots using Python packages. Pre-processing in the Jupyter Notebook involves using Python to program subsets, selecting the best features, and handling missing values. Python is also used in the implementation of the ML steps. A Windows 10 PC with the following specifications was used for the experiments to successfully run and validate the proposed model: a 2.9 GHz Intel Core i7 CPU, 8 GB of RAM, an Intel HD Graphics 620 GPU, and a 5 GB disk.

The model evaluation results for the selected dataset are presented and discussed in this section. The way in which the models are measured and trained affects the evaluation results. The assessment of ASD was conducted using a conventional ML methodology that included SVM, NB, RF, ET, KNN, DT, AB, and SGD. As defined in Section 3.1, we conducted experiments on four different datasets, and the results were recorded after 50 iterations. The empirical performance evaluation of classifiers based on conventional ML algorithms is summarized in Tables 4 to 7.

Table 4. Evaluation results for toddler dataset

Model	Accuracy	Sensitivity	Specificity	Mean	Precision	Recall	F1_Score	AUROC
DT	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ET	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AB	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SGD	1.000	1.000	1.000	1.000	1.000	1.000	1.000	NaN
SVM	0.997	0.995	1.000	0.997	1.000	0.995	0.998	1.000
KNN	0.972	0.963	0.990	0.975	0.995	0.963	0.979	0.991
NB	0.959	0.950	0.980	0.963	0.990	0.950	0.970	0.998

Table 5. Evaluation results for children dataset

Model	Accuracy	Sensitivity	Specificity	Mean	Precision	Recall	F1_Score	AUROC
SGD	0.989	0.973	1.000	0.987	1.000	0.973	0.986	NaN
SVM	0.955	1.000	0.922	0.959	0.902	1.000	0.949	1.000
ET	0.932	0.973	0.902	0.936	0.878	0.973	0.923	0.988
RF	0.898	0.973	0.843	0.905	0.818	0.973	0.889	0.976
NB	0.875	0.865	0.882	0.874	0.842	0.865	0.853	0.975
AB	0.852	0.919	0.804	0.858	0.773	0.919	0.840	0.861
KNN	0.841	1.000	0.725	0.855	0.725	1.000	0.841	0.946
DT	0.818	0.892	0.765	0.825	0.733	0.892	0.805	0.828

Table 6. Evaluation results for adolescent dataset

Model	Accuracy	Sensitivity	Specificity	Mean	Precision	Recall	F1_Score	AUROC
ET	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SVM	0.938	0.900	1.000	0.946	1.000	0.900	0.947	0.992
SGD	0.938	0.900	1.000	0.946	1.000	0.900	0.947	NaN
RF	0.938	100.000	0.833	0.924	0.909	100.000	0.952	0.992
KNN	0.875	100.000	0.667	0.847	0.833	100.000	0.909	0.975
NB	0.750	0.650	0.917	0.772	0.929	0.650	0.765	0.900
AB	0.750	0.750	0.750	0.750	0.833	0.750	0.789	0.750
DT	0.688	0.700	0.667	0.685	0.778	0.700	0.737	0.683

Table 7. Evaluation results for adult dataset

Model	Accuracy	Sensitivity	Specificity	Mean	Precision	Recall	F1_Score	AUROC
SGD	1.000	1.000	1.000	1.000	1.000	1.000	1.000	NaN
SVM	0.986	0.947	1.000	0.978	1.000	0.947	0.973	1.000
ET	0.981	0.930	1.000	0.970	1.000	0.930	0.964	0.999
NB	0.976	0.930	0.994	0.967	0.981	0.930	0.955	0.998
RF	0.972	0.912	0.994	0.959	0.981	0.912	0.945	0.998
KNN	0.958	0.912	0.974	0.948	0.929	0.912	0.920	0.993
DT	0.929	0.877	0.948	0.918	0.862	0.877	0.870	0.913
AB	0.929	0.877	0.948	0.918	0.862	0.877	0.870	0.913

We have carefully measured several performance metrics in our extensive analysis, such as the F1 score, accuracy, precision, mean score, sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC). Together, these metrics provide a comprehensive overview of the ML model's performance across various domains. The results do, however, highlight one important finding: there is a noticeable difference in the model's performance between datasets. Our deliberate choice to utilize authentic datasets without balancing, which mirrors the inherent complexities and imbalances present in real-world data, is the reason behind this inconsistency. This emphasizes the importance of considering dataset characteristics when evaluating models. It underscores the need for meticulous model selection and fine-tuning to achieve optimal results across various scenarios. A comprehensive understanding of the effectiveness of classifiers is enhanced by incorporating a variety of performance metrics. This, in turn, encourages a data-driven approach to selecting and enhancing models.

From Figure 2, it is clear that the remarkable 100% accuracy achieved by DT, RF, ET, AB, and SGD classifiers on a toddler dataset highlights the potential effectiveness of these models in identifying complex patterns in the data. The remarkable accuracy indicates that these algorithms have successfully captured the fundamental structure of the features associated with toddlers, demonstrating their proficiency in classification tasks for this specific dataset. The accompanying figure visually validates the strong performance of these classifiers and demonstrates their ability to predict outcomes effectively. On the other hand, while still achieving remarkable accuracy levels between 95% and 99%, SVM, KNN, and NB classifiers appear to perform slightly less accurately than their counterparts. This subtle variation in how well different algorithms perform encourages more research into the specific details of the toddler dataset, which may shed light on the types of data that each algorithm is better at handling.

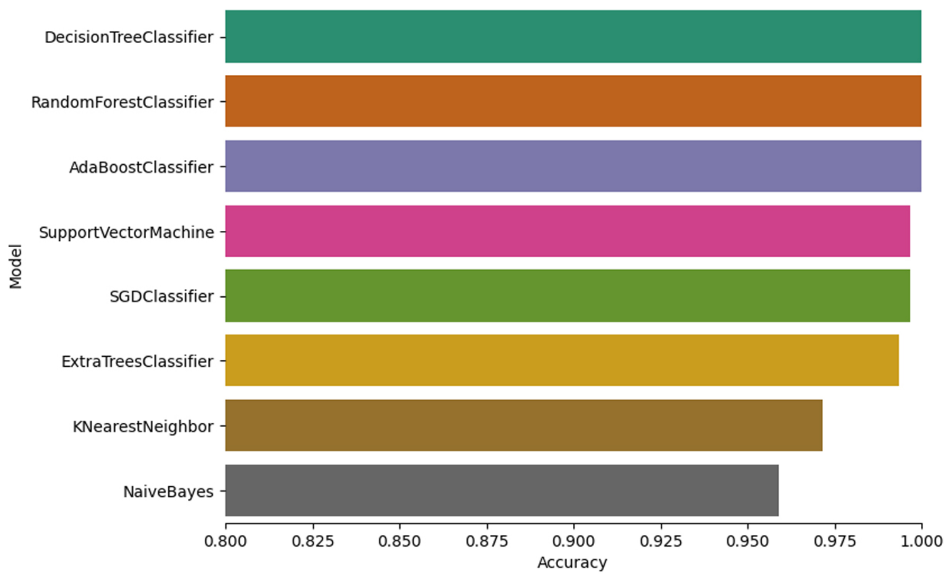

Fig. 2. Accuracy comparison of machine learning algorithms on toddler dataset

Based on the results shown in Figure 3, we compared classifiers using a children's dataset and found significant differences in performance measures. In particular, the SVM and SGD classifiers have performed better, achieving accuracy rates between 95% and 99%. This excellent performance indicates that these models are proficient at identifying patterns among the intricate features of the children dataset, showcasing their effectiveness in classification tasks specific to this scenario. Figure 3 that accompanies these results visually emphasizes the consistent high accuracy of SGD and SVM classifiers, showcasing their reliability in various scenarios. On the other hand, the alternative classifiers in our study have shown somewhat inconsistent results, with accuracy ranging from 80% to 93%. Although this range is still impressive, it shows that the predictive abilities of these classifiers differ noticeably from those of the SGD and SVM models. The observed discrepancies in accuracy highlight how crucial it is to carefully select models based on the unique characteristics and nuances of the dataset in question.

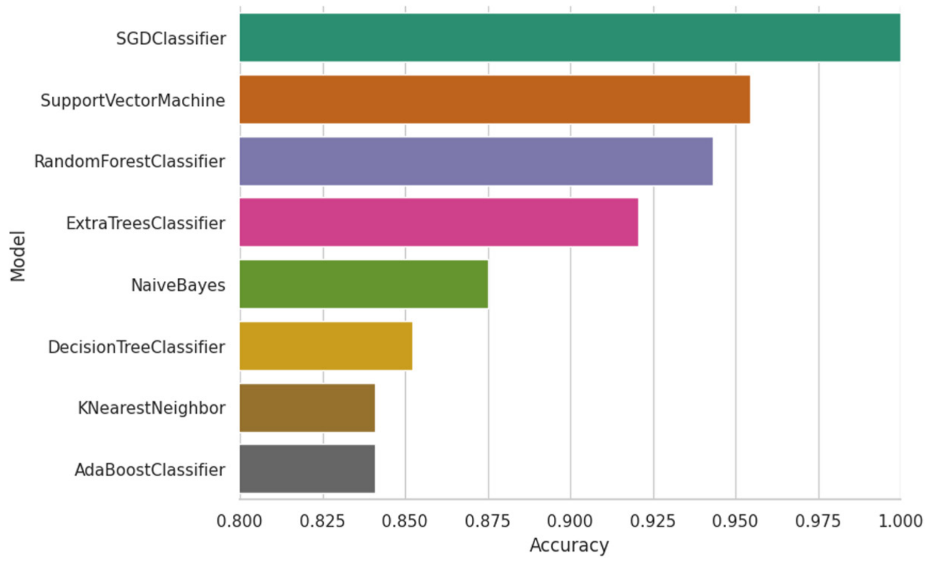

Fig. 3. Accuracy comparison of machine learning algorithms on child dataset

Surprisingly, the ET classifier has proven to be incredibly accurate, scoring a perfect 100% on the teenage dataset. This remarkable performance indicates that ET, with its randomization techniques and ensemble learning approach, has successfully captured the complex patterns present in the teenage dataset, demonstrating its resilience in classification tasks designed for this age group. Accuracy rates ranging from 90% to 94% have also been achieved by the RF, SVM, and SGD classifiers collaborating to yield impressive outcomes. The accompanying Figure 4 highlights the robust and consistent performance of these models in a variety of scenarios across the adolescent dataset. It is noteworthy, though, that other classifiers in our study have shown a slightly wider accuracy range, varying from 68% to 88%. This variability suggests that while ET, RF, SVM, and SGD models perform exceptionally well, other classifiers might struggle to match the intricacies of the teenage dataset. The subtle differences in accuracy emphasize the importance of using a customized strategy when selecting classifiers based on the distinctive features of the dataset.

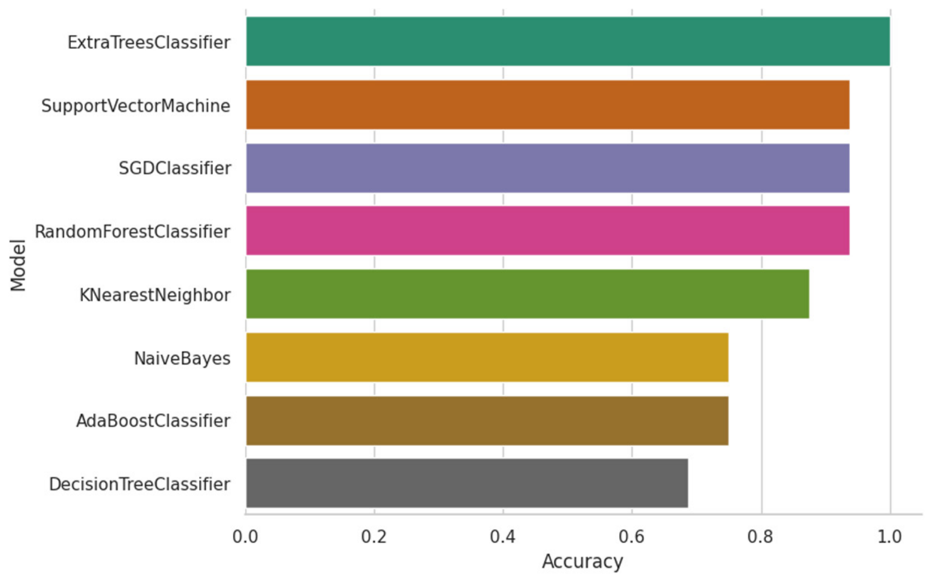

Fig. 4. Accuracy comparison of machine learning algorithms on adolescent dataset

In our analysis of the adult dataset, the SGD classifier has proven to be an outstanding performer, achieving a flawless accuracy rate of 100%. This impressive outcome highlights the effectiveness of the SGD classifier in identifying intricate patterns specific to the adult population. It demonstrates strong performance in classification tasks tailored for this age group. Concurrently, other classifiers that we used in this investigation have also yielded excellent results, with accuracy rates ranging between 92% and 98%. These results are graphically represented in the accompanying Figure 5, which illustrates the consistent performance of the SGD classifier and other models across various instances in the adult dataset. Although the SGD classifier is exceptional for its flawless accuracy, the competitive performance of other classifiers suggests a diverse range of models that are suitable for this specific dataset.

A comprehensive examination of the confusion matrices for each classifier model across various datasets provide valuable insights into the distinct performance characteristics and challenges linked to each dataset. Interestingly, Table 8 shows indications of an uneven distribution of classes in both the Toddler and Adult datasets. This imbalance may make it difficult for classifiers to predict minority classes accurately, potentially leading to biased outcomes. The distribution in the children and adolescent datasets, on the other hand, is noticeably more balanced, which enhances the performance of the classification models for these age groups. The KNN classifier is particularly adept at generating a well-balanced distribution of results in the children dataset, as indicated by the confusion matrix analysis. This implies that KNN effectively negotiates the nuances of class disparities, contributing to a more equitable representation of predictions for this age group across different classes. Furthermore, the confusion matrix shows that the accuracy expectations for the adolescent dataset closely align with the NB classifier. NB's balanced distribution in this case indicates that the model manages class proportions well, which helps to produce reliable predictions in the adolescent dataset. These results highlight the significance of considering both the distribution of predictions across classes and overall accuracy, especially in datasets with existing class imbalances. More accurate and dependable predictions are made possible by customizing classifier selection to the unique features of each dataset, as demonstrated by the nuanced performance of KNN and NB in the children and adolescent datasets. Understanding the strengths and weaknesses of each classifier model in addressing the challenges posed by different class distributions in datasets across various age groups is facilitated by utilizing confusion matrices.

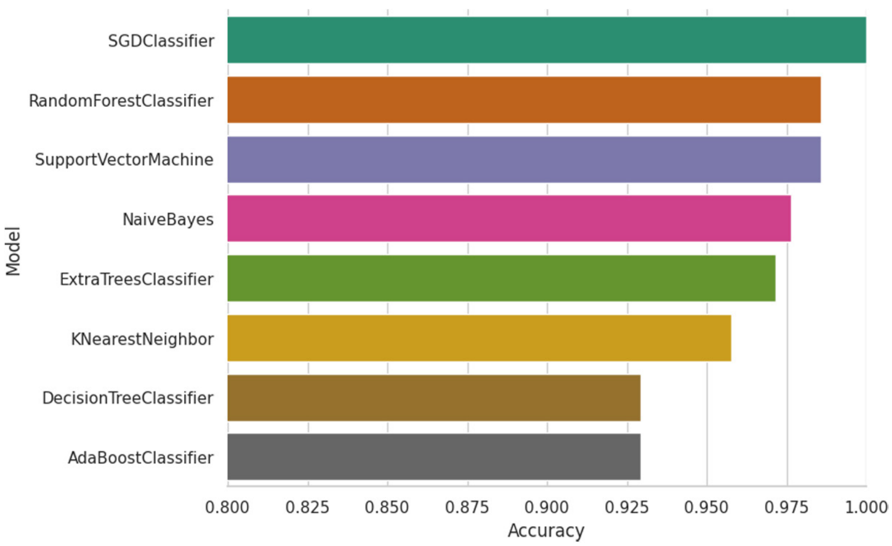

Fig. 5. Accuracy comparison of machine learning algorithms on adult dataset

Table 8. Confusion matrix of machine learning algorithms on different datasets

Model	Datasets			
	Toddler	Child	Adolescent	Adult
DT	$\begin{bmatrix} 98 & 0 \\ 0 & 219 \end{bmatrix}$	$\begin{bmatrix} 43 & 8 \\ 3 & 34 \end{bmatrix}$	$\begin{bmatrix} 9 & 3 \\ 6 & 14 \end{bmatrix}$	$\begin{bmatrix} 147 & 8 \\ 7 & 50 \end{bmatrix}$
RF	$\begin{bmatrix} 98 & 0 \\ 0 & 219 \end{bmatrix}$	$\begin{bmatrix} 45 & 6 \\ 1 & 36 \end{bmatrix}$	$\begin{bmatrix} 12 & 0 \\ 3 & 17 \end{bmatrix}$	$\begin{bmatrix} 155 & 0 \\ 3 & 54 \end{bmatrix}$
ET	$\begin{bmatrix} 98 & 0 \\ 0 & 219 \end{bmatrix}$	$\begin{bmatrix} 44 & 7 \\ 0 & 37 \end{bmatrix}$	$\begin{bmatrix} 12 & 0 \\ 2 & 18 \end{bmatrix}$	$\begin{bmatrix} 155 & 0 \\ 4 & 53 \end{bmatrix}$
SVM	$\begin{bmatrix} 98 & 0 \\ 1 & 218 \end{bmatrix}$	$\begin{bmatrix} 47 & 4 \\ 0 & 37 \end{bmatrix}$	$\begin{bmatrix} 12 & 0 \\ 2 & 18 \end{bmatrix}$	$\begin{bmatrix} 155 & 0 \\ 3 & 54 \end{bmatrix}$
AB	$\begin{bmatrix} 98 & 0 \\ 0 & 219 \end{bmatrix}$	$\begin{bmatrix} 41 & 10 \\ 4 & 33 \end{bmatrix}$	$\begin{bmatrix} 9 & 3 \\ 7 & 13 \end{bmatrix}$	$\begin{bmatrix} 147 & 8 \\ 5 & 52 \end{bmatrix}$
KNN	$\begin{bmatrix} 97 & 1 \\ 8 & 211 \end{bmatrix}$	$\begin{bmatrix} 37 & 14 \\ 0 & 37 \end{bmatrix}$	$\begin{bmatrix} 8 & 4 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 151 & 4 \\ 5 & 52 \end{bmatrix}$
NB	$\begin{bmatrix} 96 & 2 \\ 11 & 208 \end{bmatrix}$	$\begin{bmatrix} 42 & 6 \\ 5 & 32 \end{bmatrix}$	$\begin{bmatrix} 11 & 1 \\ 7 & 13 \end{bmatrix}$	$\begin{bmatrix} 154 & 1 \\ 4 & 53 \end{bmatrix}$
SGD	$\begin{bmatrix} 98 & 0 \\ 0 & 219 \end{bmatrix}$	$\begin{bmatrix} 51 & 0 \\ 0 & 37 \end{bmatrix}$	$\begin{bmatrix} 12 & 0 \\ 6 & 14 \end{bmatrix}$	$\begin{bmatrix} 155 & 0 \\ 0 & 57 \end{bmatrix}$

The receiver operating characteristic (ROC) curves are a comprehensive metric that can be used to visualize the performance of ML models. This graph illustrates the trade-off between true positive rates and false positive rates at various thresholds. These ROC curves are displayed in the accompanying Figures 6 to 9, which offer a dynamic representation of the performance of various classifiers on the toddler dataset. Surprisingly, classifiers with AUC values of 1 include DT, RF, ET, AB, and SVM. An ideal classification scenario is one in which the model achieves perfect discrimination between positive and negative instances, as indicated by a perfect AUC score. Concurrently, Figure 6 shows that NB and KNN classifiers are not far behind, producing excellent AUC values of 0.99. This nearly flawless score indicates that these classifiers are still highly effective at accurately classifying data, even though they may not be as superior as their counterparts. The robust discriminative abilities of the classifiers are further supported by the ROC analysis, which highlights their capacity to capture subtle patterns within the toddler dataset. DT, RF, ET, AB, SVM, KNN, and NB all achieved remarkable AUC values, which attest to their effectiveness in making precise predictions for various instances in the toddler dataset. This detailed visualization enhances our comprehension of the classifiers' abilities to process data related to toddlers by offering a nuanced perspective on the subtle variations in their performance. It also reaffirms the previously mentioned high accuracy.

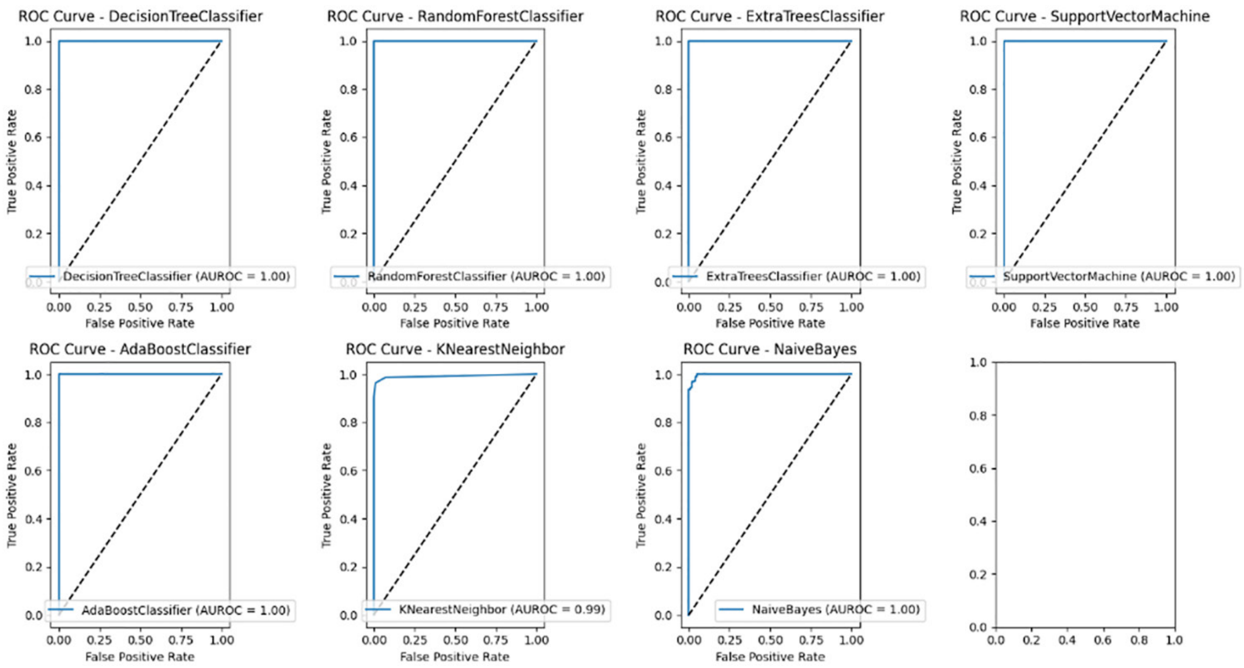

Fig. 6. Measurement of AUROC of machine learning algorithms on toddler dataset

Figure 7, which accompanies the SVM classifier, stands out among the others. The SVM classifier's exceptional ability to correctly classify instances within the children's dataset is highlighted by its perfect AUC score, which strikes a balance between true positive rates and false positive rates. Other classifiers depicted in the figures, on the other hand, exhibit AUC values ranging from 0.82 to 0.98. This variability highlights subtle differences in how well the classifiers perform, with each model showcasing its own advantages and characteristics in navigating the complexities of the children's dataset. Even though some classifiers achieve an AUC value close to 1.0, they still do not quite reach the level of perfect discrimination exhibited by the SVM classifier.

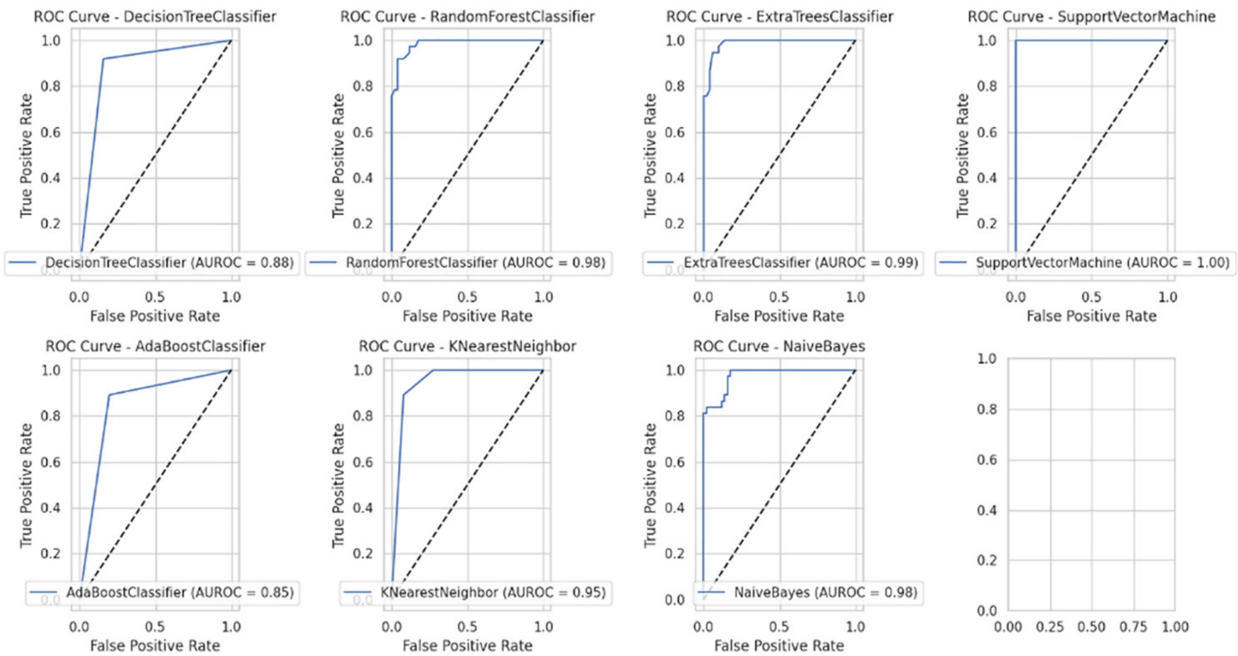

Fig. 7. Measurement of AUROC of machine learning algorithms on child dataset

The classifier named ET is the most effective on the adolescent dataset, exhibiting a maximal AUC value of 1. This flawless AUC score highlights the robustness of ET in identifying complex patterns within the adolescent dataset and indicates an excellent ability to distinguish between positive and negative instances. DT, on the other hand, has the least favorable AUC value in this situation, scoring 0.68. The lower AUC value suggests potential challenges in capturing the underlying complexities, indicating that DT may be less effective at distinguishing between classes within the adolescent dataset. The AUC values of the remaining classifiers in Figure 8 range from 0.75 to 0.99. The variation in AUC scores reveals subtle differences in the performance of these models, with each one exhibiting varying degrees of accuracy in correctly classifying instances from the adolescent dataset. As some classifiers approach the level of perfection achieved by ET, others demonstrate their effectiveness in handling the complexities of the dataset by achieving AUC values in the mid-to-high range.

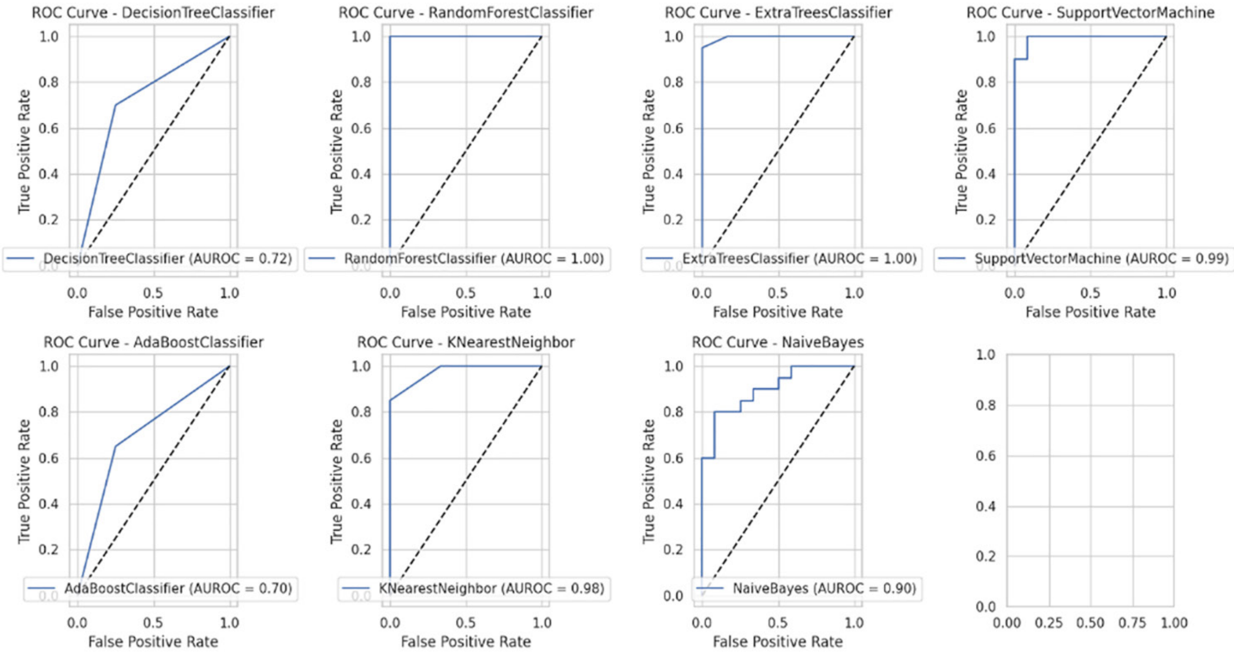

Fig. 8. Measurement of AUROC of ML algorithms on adolescent dataset

The SVM classifier achieves a perfect AUC value of 1.0, making it the clear winner on the adult dataset. This exemplary AUC score underscores the SVM classifier's robust performance in recognizing intricate patterns and confirms its exceptional ability to precisely differentiate between positive and negative instances within the adult dataset. Some of the other classifiers displayed in Figure 9, have AUC values ranging from 0.91 to 0.99. These AUC scores indicate very good discriminatory abilities, although they are slightly less optimal than those of the SVM classifier. All of the classifiers in this range demonstrate excellent accuracy in distinguishing between the classes in the adult dataset.

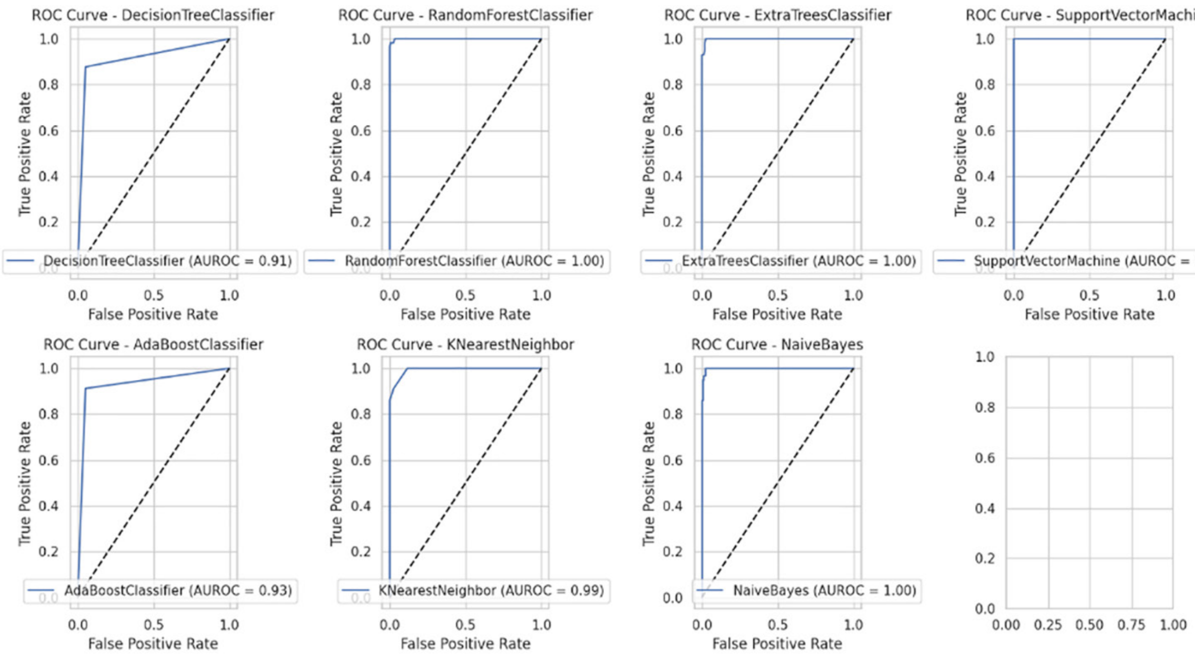

Fig. 9. Measurement of AUROC of machine learning algorithms on adult dataset

In order to enhance the accuracy of predicting ASD, datasets pertaining to ASD were collected and analyzed as part of research projects. In the study, several ML classifiers were utilized, including AB, SVM, NB, RF, ET, KNN, DT, and SGD. The accuracy of the suggested model was found to be comparable when the results were compared with those of other recent studies [28] and [29], as shown in Table 9. Crucially, the model demonstrated efficacy in diagnosing ASD in patients of various ages, from toddlers to adults. The results suggest potential applications for the developed model in the field of diagnosing ASD.

Table 9. Comparison of performance results

	Toddler			Children			Adolescent			Adult		
	[28]	[29]	Proposed Work	[28]	[29]	Proposed Work	[28]	[29]	Proposed Work	[28]	[29]	Proposed Work
SVM	92.7	93.8	99.6	89.9	93.8	95.4	84.9	85.2	93.7	93.8	95.2	98.5
NB	88.9	94.1	95.8	73.4	84.6	87.5	75.3	70.8	75.0	96.9	97.6	97.6
DT	–	100	100	–	80.1	81.2	–	96.8	68.7	–	91.8	92.9
RF	81.5	98.5	100	85.4	88.7	89.7	90.5	91.8	93.7	89.7	96.4	97.1
ETC	–	–	100	–	–	93.2	–	–	100	–	–	98.1
KNN	90.5	95.7	97.1	79.8	81.9	84.1	79.8	80.9	87.5	91.8	93.8	95.7
AB	–	–	100	–	–	85.2	–	–	75.0	–	–	92.9
SGD	–	–	100	–	–	98.8	–	–	93.7	–	–	100

5 CONCLUSION

Diagnosing ASD can be challenging due to the complexity of associated disorders, which involve behavioural, emotional, structural, and mental components.

As there are currently no conclusive medical tests for all relevant features, the diagnosis of ASD is based on laborious and intricate procedures involving psychological evaluations and observation of responses. Accurate detection of ASD is further complicated by the absence of effective screening methods. A recent and promising advancement in enhancing the accuracy and effectiveness of ASD prediction is ML. The effectiveness of diagnosing ASD across various age groups was demonstrated by applying different machine learning models to a dataset that included toddlers and adults. The findings suggest potential applications for the model in diagnosing ASD. Prospects for the future include exploring transfer-learning models such as ResNet and MobileNet, which utilize image datasets to improve the accuracy of detecting ASD in autistic children at a young age. Furthermore, deep learning techniques might be useful in the future for determining the severity of the disorder.

6 REFERENCES

[1] J. Kang, X. Han, J. Song, Z. Niu, and X. Li, "The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data," *Computers in Biology and Medicine*, vol. 120, p. 103722, 2020. https://doi.org/10.1016/j.compbiomed.2020.103722

[2] L. Crane, R. Batty, H. Adeyinka, L. Goddard, L. A. Henry, and E. L. Hill, "Autism diagnosis in the United Kingdom: Perspectives of autistic adults, parents and professionals," *Journal of Autism and Developmental Disorders*, vol. 48, no. 11, pp. 3761–3772, 2018. https://doi.org/10.1007/s10803-018-3639-1

[3] F. Thabtah, R. Spencer, N. Abdelhamid, F. Kamalov, C. Wentzel, Y. Ye, and T. Dayara, "Autism screening: An unsupervised machine learning approach," *Health Information Science and Systems*, vol. 10, no. 1, p. 26, 2022. https://doi.org/10.1007/s13755-022-00191-x

[4] D. D. Khudhur and S. D. Khudhur, "The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups," *Measurement: Sensors*, vol. 27, p. 100774, 2023. https://doi.org/10.1016/j.measen.2023.100774

[5] K. D. Cantin-Garside, Z. Kong, S. W. White, L. Antezana, S. Kim, and M. A. Nussbaum, "Detecting and classifying self-injurious behaviour in autism spectrum disorder using machine learning techniques," *Journal of Autism and Developmental Disorders*, vol. 50, no. 11, pp. 4039–4052, 2020. https://doi.org/10.1007/s10803-020-04463-x

[6] N. A. Mashudi, N. Ahmad, and N. M. Noor, "Classification of adult autistic spectrum disorder using machine learning approach," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, pp. 743–751, 2021. https://doi.org/10.11591/ijai.v10.i3.pp743-751

[7] U. Erkan and D. N. H. Thanh, "Autism spectrum disorder detection with machine learning methods," *Current Psychiatry Research and Reviews*, vol. 15, no. 4, pp. 297–308, 2020. https://doi.org/10.2174/2666082215666191111121115

[8] M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkuti, "Prediction of diabetes disease using machine learning algorithms," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 1, pp. 284–290, 2022. https://doi.org/10.11591/ijai.v11.i1.pp284-290

[9] M. K. Hanif, N. Ashraf, M. U. Sarwar, D. M. Adinew, and R. Yaqoob, "Employing machine learning-based predictive analytical approaches to classify autism spectrum disorder types," *Complexity and Robustness Trade-Off for Traditional and Deep Models*, vol. 2022, pp. 1–10, 2022. https://doi.org/10.1155/2022/8134018

[10] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of autism spectrum disorder in children using machine learning techniques," *SN Computer Science*, vol. 2, no. 5, p. 386, 2021. https://doi.org/10.1007/s42979-021-00776-5

[11] S. D. Khudhur and D. D. Khudhur, "IgG-IgM antibodies-based infection time detection of COVID-19 using machine learning models," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 2, p. 340, 2022. https://doi.org/10.12928/telkomnika.v20i2.21649

[12] N. A. Ali, "Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 1, pp. 91–99, 2020. https://doi.org/10.11591/ijai.v9.i1.pp91-99

[13] L. Shalini, S. S. Manvi, N. C. Gowda, and K. N. Manasa, "Detection of phishing emails using machine learning and deep learning," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, 2022, pp. 1237–1243. https://doi.org/10.1109/ICCES54183.2022.9835846

[14] N. C. Gowda and A. Bharathi Malakreddy, "A trust prediction mechanism in edge communications using optimized support vector regression," in *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, 2023. https://doi.org/10.1109/ICCMC56507.2023.10083686

[15] K. B. Rekha and N. C. Gowda, "A framework for sentiment analysis in customer product reviews using machine learning," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 2020. https://doi.org/10.1109/ICSTCEE49637.2020.9276877

[16] L. Shalini, S. S. Manvi, B. Gardiner, and N. C. Gowda, "Image based classification of COVID-19 infection using ensemble of machine learning classifiers and deep learning techniques," in *2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, 2022. https://doi.org/10.1109/ICDSAAI55433.2022.10028859

[17] M. Rahul, P. Ravichandra, M. M. Yakoobi, and N. C. Gowda, "Deep learning-based solution for differently-abled persons in the society," in *2023 4th International Conference for Emerging Technology (INCET)*, 2023. https://doi.org/10.1109/INCET57972.2023.10170230

[18] A. A. Abdullah, S. Rijal, and S. R. Dash, "Evaluation on machine learning algorithms for classification of autism spectrum disorder (ASD)," *Journal of Physics: Conference Series*, vol. 1372, no. 1, p. 012052, 2019. https://doi.org/10.1088/1742-6596/1372/1/012052

[19] M. Bala, M. H. Ali, Md. 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, 2022. https://doi.org/10.3390/a15050166

[20] S. M. Mahedy Hasan, M. P. Uddin, M. A. Mamun, M. I. Sharif, A. Ulhaq, and G. Krishnamoorthy, "A machine learning framework for early-stage detection of autism spectrum disorders," *Institute of Electrical and Electronics Engineers (IEEE), Access*, vol. 11, pp. 15038–15057, 2023. https://doi.org/10.1109/ACCESS.2022.3232490

[21] I. D. Rodrigues, E. A. de Carvalho, C. P. Santana, and G. S. Bastos, "Machine learning and rs-fMRI to identify potential brain regions associated with autism severity," *Algorithms*, vol. 15, no. 6, p. 195, 2022. https://doi.org/10.3390/a15060195

[22] K. Chowdhury and M. A. Iraj, "Predicting autism spectrum disorder using machine learning classifiers," *arXi*, 2021. https://doi.org/10.48550/ARXIV.2101.09279

[23] S. Raj and S. Masood, "Analysis and detection of autism spectrum disorder using machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 994–1004, 2021. https://doi.org/10.1016/j.procs.2020.03.399

[24] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning techniques," in *Health Information Science and Systems*, vol. 9, no. 1, 2021. https://doi.org/10.1007/s13755-021-00145-9

[25] J. Talukdar, D. K. Gogoi, and T. P. Singh, "A comparative assessment of most widely used machine learning classifiers for analysing and classifying autism spectrum disorder in toddlers and adolescents," *Healthcare Analytics*, vol. 3, p. 100178, 2023. https://doi.org/10.1016/j.health.2023.100178

[26] T. Akter, M. S. 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. https://doi.org/10.1109/ACCESS.2019.2952609

[27] 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, 2022. https://doi.org/10.3390/biomedicines10082028

[28] M. Masum, I. Nur, M. H. Faruk, M. Adnan, and H. Shahriar, "A comparative study of machine learning-based autism spectrum disorder detection with feature importance analysis," in *COMPSAC 2022: Computer Software and Applications Conference*, vol. 3, 2022.

[29] D. D. Khudhur and S. D. Khudhur, "The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups," *Measurement: Sensors*, vol. 27, p. 100774, 2023. https://doi.org/10.1016/j.measen.2023.100774

[30] W. Yin, S. Mostafa, and F. Wu, "Diagnosis of autism spectrum disorder based on functional brain networks with deep learning," *Journal of Computational Biology*, vol. 28, no. 2, pp. 146–165, 2021. https://doi.org/10.1089/cmb.2020.0252

[31] Z. A. Huang, Z. Zhu, C. H. Yau, and K. C. Tan, "Identifying autism spectrum disorder from resting-state fMRI using deep belief network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 2847–2861, 2021. https://doi.org/10.1109/TNNLS.2020.3007943

[32] E. P. K. Pua, P. Thomson, J. Y. M. Yang, J. M. Craig, G. Ball, and M. Seal, "Individual differences in intrinsic brain networks predict symptom severity in autism spectrum disorders," *Cerebral Cortex*, vol. 31, no. 1, pp. 681–693, 2020. https://doi.org/10.1093/cercor/bhaa252

[33] R. X. Smith, K. Jann, M. Dapretto, and D. J. J. Wang, "Imbalance of functional connectivity and temporal entropy in resting-state networks in autism spectrum disorder: A machine learning approach," *Frontiers in Neuroscience*, vol. 12, 2018. https://doi.org/10.3389/fnins.2018.00869

[34] Fadi, *Autism Screening for Toddlers*, Kaggle, 2018. [Online]. Available: https://www.kaggle.com/fabdelja/autism-screening-for-toddlers.

[35] Fadi Fayez Thabtah, "Autistic spectrum disorder screening data for children," 2017. https://archive.ics.uci.edu/ml/machine-learningdatabases/00419/

[36] Fadi Fayez Thabtah, "Autistic spectrum disorder screening data for adolescent," 2017. https://archive.ics.uci.edu/ml/machine-learningdatabases/00420/

[37] Fadi Fayez Thabtah, "Autistic spectrum disorder screening data for adult," 2017. https://archive.ics.uci.edu/ml/machine-learningdatabases/00426/

[38] D. Mustafa Abdullah and A. Mohsin Abdulazeez, "Machine learning applications based on SVM classification a review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81–90, 2021. https://doi.org/10.48161/qaj.v1n2a50

[39] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Variable selection for naïve bayes classification," *Computers & Operations Research*, vol. 135, p. 105456, 2021. https://doi.org/10.1016/j.cor.2021.105456

[40] J. Chen, "PE teaching activities in colleges and universities based on decision tree," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 8, pp. 38–50, 2018. https://doi.org/10.3991/ijet.v13i08.8693

[41] J. Yu, "Academic performance prediction method of online education using random forest algorithm and artificial intelligence methods," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 5, pp. 45–57, 2021. https://doi.org/10.3991/ijet.v16i05.20297

[42] N. Wahid, A. Zaidi, G. Dhiman, M. Manwal, D. Soni, and R. R. Maaliw, "Identification of coronary artery disease using extra tree classification," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, 2023. https://doi.org/10.1109/ICICT57646.2023.10134338
[43] Y. An, M. Xu, and C. Shen, "Classification method of teaching resources based on improved KNN algorithm," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 4, pp. 73–88, 2019. https://doi.org/10.3991/ijet.v14i04.10131
[44] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference,* Berlin, Heidelberg: Springer. 2013, pp. 37–52. https://doi.org/10.1007/978-3-642-41136-6_5
[45] E. Yu, D. Dong, Y. Xu, S. Ouyang, and X. Liao, "CP-SGD: Distributed stochastic gradient descent with compression and periodic compensation," *Journal of Parallel and Distributed Computing*, vol. 169, pp. 42–57, 2022. https://doi.org/10.1016/j.jpdc.2022.05.014

7 AUTHORS

Ambika Rani Subhash received her M.Tech degree in Computer Science Engineering from REVA University. She has more than 11 years of teaching and research experience. Presently, she is pursuing her Ph.D. research at REVA University and working as an Assistant Professor in the Department of Information Science and Engineering. Additionally, she serves as the Training and Placement Officer at BMS Institute of Technology and Management in Bengaluru. Her research interests include artificial intelligence, machine learning, IoT, Big Data, and computer networks. She is a member of the International Association of Engineers (IAENG) and IEEE (E-mail: ambikasubash@bmsit.in).

Ashwin Kumar U Motagi received his Ph.D. degree in Computer Science and Engineering and has extensive teaching and research experience spanning 27 years. Presently, he is the Director and Professor of the School of Computer Science and Engineering at REVA University in Bangalore. He has published numerous articles in international journals and conferences and has had many patents granted to him. He has received the IEEE Conference ACCT-12 Best Paper Award and has obtained numerous research grants from government organizations such as DRDO. His research interests include artificial intelligence, machine learning, IoT, Bigdata and Data Analytics. He is an active member of technical societies such as CSI and IEEE.