INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Check for updates

OPEN ACCESS

**Received:** 20/09/2025

**Accepted:** 17/10/2025

**Published:** 05/11/2025

# PAASD_HPT: Predictive Analytics for Autism Spectrum Disorder Prediction using Hyperparameter Tuning of Machine Learning Models

**J Francis Julee Rajam[1]*, S Britto Ramesh Kumar[2]**

**1** Research Scholar, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli- 620002, Tamil Nadu, India.
**2** Associate Professor and Head, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli- 620002, Tamil Nadu, India.

**Citation:** Francis Julee Rajam J, Britto Ramesh Kumar S (2025) PAASD_HPT: Predictive Analytics for Autism Spectrum Disorder Prediction using Hyperparameter Tuning of Machine Learning Models. Indian Journal of Science and Technology 18(41): 3260-3269. https://doi.org/10.17485/IJST/v18i41.1613

* **Corresponding author**.

francisjuleerajam_ss1@mail.sjctni.edu

**Funding:** None

**Competing Interests:** None



Published By Indian Society for Education and Environment (iSee)

**ISSN**
Print: 0974-6846
Electronic: 0974-5645

## Abstract

**Background:** Autism Spectrum Disorder (ASD) is a neurological condition which results in substantial communication, community, and behavioral challenges for those affected. Individuals with autism experience communication challenges, social interaction issues, and exhibit repetitive behaviors. Various machine learning techniques have been employed to distinguish persons with ASD from those without. **Objectives:** To design a hyperparameter tunning algorithm for machine learning model for predicting the ASD to enhance its overall efficiency. **Methods:** This study suggests Predictive Analytics for Autism Spectrum Disorder prediction using Hyper-Parameter Tuning (PAASD_HPT) of machine learning models. The Sine Cosine Algorithm (SCA) is employed to tune the hyper-parameter of different machine learning approaches including AdaBoost, Decision Tree, KNN, Random Forest and SVM. All classifiers' performance is evaluated using the ASD dataset, both prior to and following the hyperparameter adjusting process. **Findings:** The PAASD_HPT was tested on autism prediction in adults' dataset from Kaggle. Empirical findings demonstrate that hyperparameter adjustment of machine learning models has diverse effects on predicting ASD. The proposed hyperparameter tuning approach achieves accuracy of 98.31% for AdaBoost, 96.62% for DT, 97.46% for KNN, 98.73% for RF and 96.20% for SVM. Upon hyperparameter optimization, Random Forest and AdaBoost surpasses alternative models with a notable statistical distinction. **Novelty:** The proposed PAASD_HPT utilized SCA to tune the hyperparameter. The novelty of this work is to optimize the machine learning models parameters using SCA. Thus, the suggested method helps to increase the accuracy while predicting the ASD using machine learning models.

**Keywords:** Autism Spectrum Disorder (ASD); Classification; Predictive Analytics; Hyper-parameter tuning; ML Algorithms; Sine Cosine Algorithm

## 1 Introduction

Autism spectrum disorder (ASD) is a developmental condition which affects speech and behavior, typically emerging in the infant years and enduring throughout an individual lifetime[1]. The etiology of ASD remains ambiguous, and there are yet no validated medical treatments that demonstrate efficacy. Prior study has demonstrated that the brain's capacity to modify and adapt diminishes with age. Timely intervention for children with autism can enhance their verbal and cognitive abilities prior to the emergence of behavioral issues[2]. Consequently, the timely identification of ASD is critically important.

Many researchers have used a diversity of Machine Learning (ML) approaches in recent years to assess and identify ASD. Bousidrah *et al.*[3] employed ML methods, such as KNN, SVM, and Random Forest (RF), to diagnose ASD. Generalizability necessitates validation with varied datasets. Thalor *et al.*[4] suggested an ML-based methodology for evaluating ASD utilizing varied datasets. It encompasses behavioral and neuroimaging data to construct predictive models for early detection. It improves accessibility and impartiality in educational and communal environments.

Elbattah *et al.*[5] proposed ML methods for the early and precise analysis of ASD. It encompasses eye-tracking and neuroimaging data to discern patterns and biomarkers, hence improving diagnostic precision. The obstacles encompass standardization, reproducibility, and clinical application. Subhash *et al.*[6] created a categorization model capable of predicting the probability of ASD using various ML algorithms across diverse age demographics, such as toddlers, children, adolescents, and adults.

Abdelwahab *et al.*[7] provided a machine-learning architecture aimed at detecting ASD in persons of several age groups. The results illustrate the efficacy of prediction models utilizing ML techniques as essential instruments for achieving this objective. The only constraint of this study is the minimal size of the dataset. Ayub *et al.*[8] proposed a framework for diagnosing ASD. The findings indicate that ML techniques, especially ensemble approach, have potential for enhancing the correctness and effectiveness of ASD diagnosis. Farooq *et al.*[9] employed a federated learning for autism diagnosis by locally training two distinct ML classifiers to categorize ASD variables and identify ASD in both children and adults.

Hajjej *et al.*[10] presented a two-phase method, wherein the initial phase utilizes multiple ML techniques, including a combination of RF and XGBoost for ASD detection. It concentrates on determining suitable instructional strategies for kids with ASD by assessing their physical, linguistic, and behavioral performance. Kavadi *et al.*[11] proposed a ML based approach for the classification of medical data, designed to enhance the efficacy of ASD analysis. Reghunathan *et al.*[12] employed various ML for the precise identification of ASD in children, adolescents, and adults. Feature reduction identifies essential attributes that contribute to ASD categorization within each age group with the Cuckoo Search Algorithm.

Jeon *et al.*[13] proposed a methodology that integrates eXplainable Artificial Intelligence (XAI) approaches with a comprehensive data-preprocessing pipeline to enhance the accuracy and interpretability of ML-based diagnostic tools. Rasul *et al.*[14] utilized various machine learning techniques to discern essential ASD characteristics, with the objective of improving and automating the diagnosing procedure. It does not address feature importance in classification models and provides insufficient discourse on potential biases in dataset selection.

Ben-Sasson *et al.*[15] created a ML algorithm to forecast ASD in neonates utilizing health details from a national screening initiative. It is validated by 3-fold cross-validation, using gradient-boosting machine techniques to attain an average AUC of 0.86, identifying developmental delay and parental concern as primary predictors. Haque *et al.*[16] illustrated the efficacy of traditional ML algorithms for the early identification of ASD in toddlers and children.

Notwithstanding these advancements, previous approaches often exhibit certain shortcomings. One of the primary drawbacks is that these methods do not account for the hyper-parameter tuning process. In the absence of hyper-parameter adjustment, machine learning models may experience several deficiencies, such as inferior performance, over-fitting or under-fitting, and heightened computational expenses. These deficiencies can significantly impair the model's ability to adapt to new data, leading to incorrect forecasts.

This study introduces an effective predictive analytics method for autism spectrum disorder prediction, termed PAASD_HPT, which employs hyper-parameter adjusting of machine learning models. This work employs the Sine Cosine Algorithm (SCA) to optimize the hyperparameters of diverse ML approaches including AdaBoost, Decision Tree, KNN, RF, and SVM. The performance of all classifiers is assessed using the ASD dataset, both before and after the hyperparameter tuning process. Empirical evidence indicates that hyperparameter tuning of ML techniques yields varying impacts on the prediction of ASD.

The paper is arranged as follows: Section 2 explains the proposed methodology for ASD. Section 3 showcases experimental results and a comparison analysis with existing systems. Ultimately, Section 4 presents conclusions and delineates possibilities for further research in adult autism prediction.

## 2 Methodology

This section explains the proposed PAASD_HPT, an efficient predictive analytics for ASD prediction using hyper-parameter tuning of machine learning models. Sine Cosine Algorithm is employed to tune the hyper-parameter of different ML models such as, AdaBoost, DT, KNN, RF and SVM. The Sine Cosine Algorithm is metaheuristic optimization technique that use sine and cosine functions to investigate and employ the search space in pursuit of optimal solutions. Hyperparameter optimization is essential for machine learning algorithms to attain optimal performance. The proposed PAASD_HPT approach has the following phases: Preprocessing, Feature Selection, SCA based hyper-parameter tuning and Prediction with evaluation metrics. Figure 1 depicts the architecture of suggested PAASD_HPT.

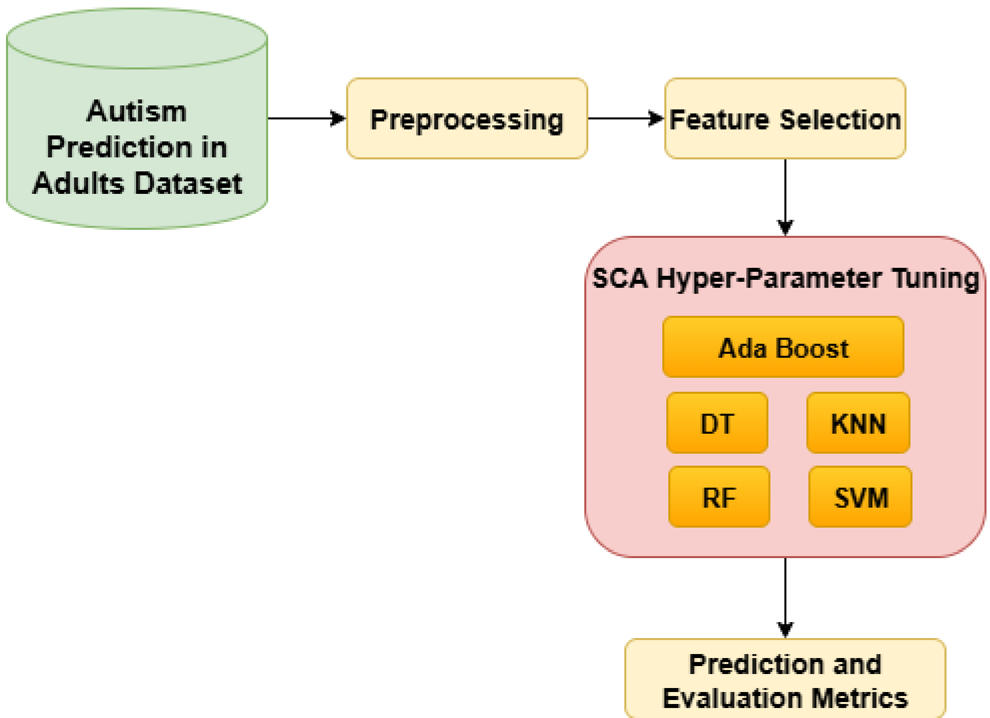


**Fig 1.** Proposed PAASD_HPT Architecture

### 2.1 Dataset Description

A dataset sourced from Kaggle, titled "Autism Prediction in Adults[17], encompasses various attributes, including patient ID, scores derived from the Autism Spectrum Quotient (AQ) 10-item screening tool (A1_Score to A10_Score), age, gender, ethnicity, history of neonatal jaundice, familial history of autism diagnosis, country of residence, prior screening test history, AQ1-10 screening test outcomes, patient age description, relationship of the individual who completed the test, and the target variable designated as "Class/ASD" (0 for No and 1 for Yes). The dataset utilized in this research comprises a diverse array of factors that may affect ASD diagnosis. It encompasses demographic data, including age and gender, with characteristics such as familial history and neonatal jaundice, which may be pertinent for diagnosis. The AQ scores offer significant screening information; the goal column "Class/ASD" functions as the experiment's definitive reference. This dataset offers a comprehensive resource for the training and assessment of ASD prediction models.

### 2.2 Data Preprocessing

For preprocessing, this paper uses MAPLE (**M**issing data imputation and **A**nomaly removal for **P**reprocessing with **L**earning **E**nhancement)[18]. It combines advanced data imputation and outlier handling techniques to improve predictive accuracy and data quality. Feature selection is performed using FlexiFeat[19], an ensemble method that combines filter-based (CfsSubsetEval with BestFirst), wrapper-based (WrapperSubsetEval with GreedyStepwise), and embedding (ReliefF with Ranker) strategies to retain exclusively the most relevant features.

## 2.3 Hyper parameter tuning

This paper uses SCA to tune the hyper-parameter of AdaBoost, Decision Tree, KNN, RF and SVM machine learning approaches. The Sine Cosine Algorithm developed by Mirjalili in 2016 [20], is a population-based optimization method designed to address various optimization challenges. The SCA produces many initial random solutions and directs them to converge towards the optimal answer with a mathematical model founded on sine and cosine functions. It combines numerous arbitrary and adaptive variables to sustain exploration and exploitation of the search space at diverse optimization milestones.

The optimization procedures are categorized into two primary phases: exploration (diversification, serving as a global search) and exploitation (intensification, functioning as a local search). During the exploration phase, the optimization algorithms abruptly combine random solutions from the solution set with a significant degree of randomization to identify potential regions inside the search space. During the exploitation phase, progress is modest, with random solutions advancing at a slower pace, and random alterations are significantly less compared to the exploration phase.

The mathematical formulas for changing positions in SCA are provided below.

$$X_i^{t+1} = \begin{cases} X_i^t + r_1 \times \sin(r_2) \times |r_3 P_i^t - X_i^t| & r_4 < 0.5 \\ X_i^t + r_1 \times \cos(r_2) \times |r_3 P_i^t - X_i^t| & r_4 \geq 0.5 \end{cases} \tag{1}$$

$X_i^{t+1}$ denotes the configuration of the current choice in the $i^{th}$ measurement at the $t^{th}$ model, whereas $r_1$, $r_2$, $r_3$, and $r_4$ are arbitrary numbers. $P_i$ signifies the positioning of the location factor in the $i^{th}$ dimension. The input $r_1$ designates the subsequent placement regions, which may be inside or beyond the space separating the solution from the goal. The $r_2$ option delineates the extent of movement towards or away from the target. The parameter $r_3$ allocates arbitrary weights to the destination to stochastically amplify ($r_3 > 1$) or diminish ($r_3 < 1$) the impact of desalination on distance determination. The parameter r4 is a stochastic variable within the interval [0,1] that signifies equal distribution between the sine and cosine components. The parameter $r_1$ can be calculated as,

$$r_1 = \pi * \left(1 - \frac{t}{T}\right) \tag{2}$$

Where t and T denotes the current and maximum number of iteration.

The SCA is likely capable of identifying the global optimum of optimization issues for the following reasons:

- SCA produces and cultivates a collection of random initial solutions for a certain optimization issue, hence primarily using robust exploration and avoidance of local optima, which is aligned with individual-based optimization methods.
- The SCA seamlessly transitions from exploration to exploitation search tactics by employing adjustable range values in sine and cosine functions.
- The globally optimal solution is stored in a variable as the goal point and will remain intact during the optimization process.
- As the solutions adjust their positions around the currently optimal solution, there is a propensity to identify the most advantageous regions of the search spaces during the optimization process.

SCA is efficiently employed to address diverse optimization challenges across multiple domains, including engineering design, machine learning, image processing, network applications, and scheduling. This paper employs SCA for the parameter tuning of machine learning models. Algorithm-1 shows the proposed PAASD_HPT.

**Algorithm-1: PAASD_HPT**
***Input:*** Preprocessed Autism Dataset
***Output:*** Predicted Results and Evaluation Metrics
Step1: Load Preprocessed Dataset
Step2: Divide the dataset into 60% trainData, 20% validationData and 20% testData
Step3: Train Machine Learning Models (AdaBoost, DT, KNN, RF, SVM) on trainData.
Step4: Trained Model is evaluated on validationData
Step5: Apply SCA hyper-parameter tuning of ML models using Algorithm-2
Step6: Use fined tuned model to test testData
Step7: Predict ASD
Step8: Evaluate Performance using metrics

In **algorithm-1**, the preprocessed data set is separated into train/test/validation split, which is an essential method in machine learning for assessing model efficacy and mitigating overfitting. The process entails partitioning a dataset into three subsets: a training (60%) for model development, a validation (20%) for hyperparameter optimization, and a test (20%) for the final evaluation of the model on new data.

The training dataset is trained using five ML approaches: AdaBoost, DT, KNN, RF and SVM. The model is assessed on validation dataset. The Sine Cosine Algorithm is applied for hyper-parameter tuning of ML models (Algorithm-2). Finally, the fine-tuned model is used to test the test data for ASD prediction. The metrics accuracy, precision, recall, F1-score and specificity were employed to analyze the efficiency. Figure 2 shows the workflow of PAASD_HPT.

**Algorithm-2: SCA – Hyper-parameter tuning**
***Input:*** TrainData, ValidationData, TestData, List of ML Model = {AdaBoost, DT, KNN, RF, SVM}
***Output:***Fine tuned ML Model

```
Initialize pop_size = 10, MaxIter = 50
For i = 1 to pop_size
 Randomly Select the ML Model
 Evaluate the ML Model using TrainData and ValidationData
 Compute Accuracy
 Find the best ML (Highest Accuracy)
End For
For iter = 1 to MaxIter
 Compute r1 using Eq. (2)
 For i = 1 to pop_size
  Randomly select ML model
  For j = 1 to number of parameters in ML model
   xj = parameter_value
   xp = parameter_value (best ML)
   r2 = 2 * π * rand()
   r3 = 2 * rand()
   r4 = rand()
   Compute new parameter_value based on Eq. (1)
   Set new parameter_value
  EndFor
  Evaluate the ML Model using TrainData and ValidationData
  Find the best ML
 EndFor
EndFor
Return Fine Tuned ML Model
```

In **algorithm-2**, the initial population are randomly created, assign default parameters for ML models and assess the technique using training and validation dataset. Select and assign the best model based on highest accuracy. The new parameter value for ML model is computed using **Eq. (1)**. The ML model is evaluated with the new parameters using TrainData and

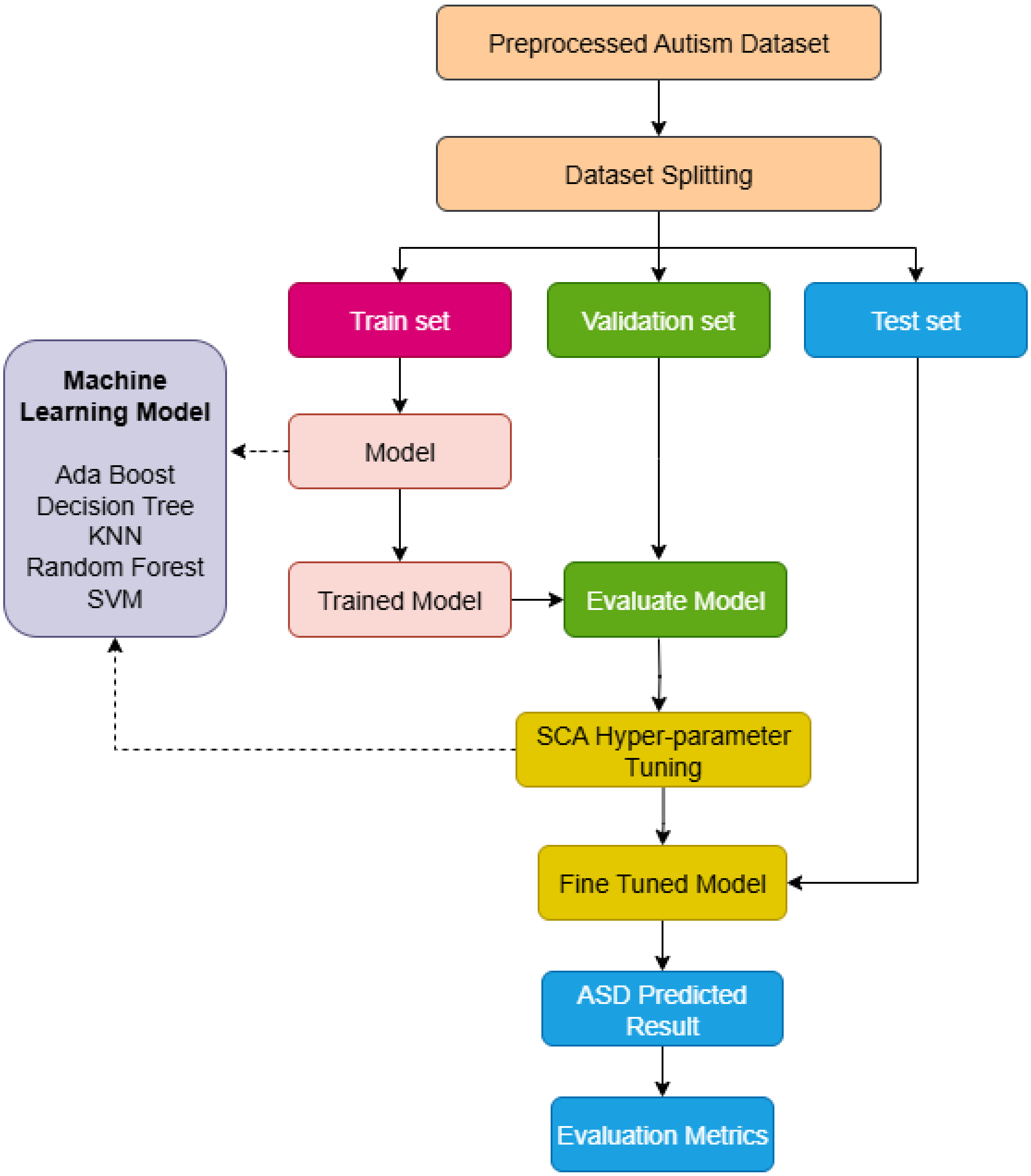


**Fig 2.** PAASD_HPT Workflow

ValidationData to find the best model based on highest accuracy. These steps are repeatedly executed number of times until it reaches maximum iterations.

## 3 Results and Discussion

This section presents an assessment of the proposed PAASD_HPT algorithm for identifying ASD in adults. All studies were conducted on a Laptop equipped with an Intel Core i7-1260P CPU (12 cores, 2.1 GHz, 18 MB L3 cache), 64 GB of RAM, and Windows 11 Home OS. The method was executed using Java Development Kit (JDK) 1.8 and Apache NetBeans IDE 15, ensuring an efficient development and execution environment.

The dataset "Autism Prediction in Adults" employed in this work was sourced from Kaggle. It contains AQ scores (A1 to A10), age, gender, ethnicity, family history of autism, history of jaundice, country of residency, prior screening test status, the relationship of the test-taker, and a binary target label "Class/ASD" (1 indicating ASD, 0 indicating non-ASD). The dataset encompasses both demographic and clinical factors, providing a comprehensive array of features relevant to adult ASD prediction.

To evaluate the efficacy of the PAASD-HPT and benchmark models, many critical measures were employed: accuracy, precision, recall (sensitivity), F1-score, and specificity. These measures facilitate the assessment of the model's accurateness, its ability to discover genuine ASD cases, minimize false positives, and maintain balanced performance across all categories. The subsequent formulas were utilized:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

$$Specificity = \frac{TN}{TN+FP} \tag{7}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

The default and tuned parameter values for each algorithm is shown in Table 1. The algorithm selects the parameter value from between these ranges. For example, the value of the n_estimators parameter in the AdaBoost algorithm lies between 10 and 50. In KNN algorithm, the k value must be 1 to 5. In the experiments, the default parameter of n_estimators of Random Forest algorithm is 20 and the tuned parameter is 40.

**Table 1.** Default and Tuned Hyper-Parameter Value

| Algorithm | Hyper-Parameter | Value Ranges | Default Parameter | Tuned Parameter |
|---|---|---|---|---|
| **AdaBoost** | n_estimators | [10,50] | 10 | 50 |
| | weight_mass_percent | [40, 100] | 100 | 67 |
| | Seed | [1, 5] | 1 | 1 |
| **Decision Tree** | Confidence_factor | [0.1, 0.5] | 0.25 | 0.102 |
| | min_sample_leaf | [2, 5] | 2 | 4 |
| **KNN** | K | [1, 5] | 1 | 3 |
| **Random Forest** | n_estimators | [100, 500] | 20 | 155 |
| | max_depth | [10, 50] | 10 | 20 |
| **SVM** | logC | [1, 4] | 2.5 | 1.9 |
| | tolerance | [0.001, 0.005] | 0.001 | 0.005 |

The efficiency of the suggested approach is compared with the previous model Hajjej *et al.* [10], Rasul *et al.* [14] and default parameter. The AdaBoost algorithm metrics was shown in Table 2. The proposed model provides the highest values of 98.31%, 98.37%, 98.31%, 98.31% and 98.51% in accuracy, precision, recall, f1-score and specificity. The approach Hajjej *et al.* [10] has lowest accuracy (90%) compared to other approaches. The suggested method (Adaboost+SCA) increases 6.39%, 9.23%, and 2.20% accuracy compared to AdaBoost with default parameter, Hajjej *et al.* [10] and Rasul *et al.* [14] approach. The metrics precision, recall, f1-score and specificity increases 4%, 6%, 6%, and 5% compared to other approaches.

Table 3 shows the performance metrics for DT algorithm. The suggested method provides the highest values of 96.62%, 96.85%, 96.62%, 96.62%, and 97.02% in accuracy, precision, recall, f1-score and specificity. The approach Rasul *et al.* [14] has

**Table 2.** Performance metrics for AdaBoost Algorithm

| Metrics (%) | AdaBoost with Default Parameter | Hajjej et al. [10] | Rasul et al. [14] | AdaBoost+SCA |
|---|---|---|---|---|
| **Accuracy** | 92.40 | 90 | 96.19 | 98.31 |
| **Precision** | 92.80 | 93 | 95 | 98.37 |
| **Recall** | 92.40 | 93 | 91.94 | 98.31 |
| **F1-Score** | 92.36 | 93 | 93.44 | 98.31 |
| **Specificity** | 91.70 | 93 | 97.97 | 98.51 |

**Table 3.** Performance metrics for DT Algorithm

| Metrics (%) | DT with Default Parameter | Hajjej et al. [10] | Rasul et al. [14] | DT+SCA |
|---|---|---|---|---|
| **Accuracy** | 96.20 | 92 | 87.14 | 96.62 |
| **Precision** | 96.37 | 92 | 84.31 | 96.85 |
| **Recall** | 96.20 | 92 | 69.35 | 96.62 |
| **F1-Score** | 96.20 | 92 | 76.11 | 96.62 |
| **Specificity** | 96.54 | 92 | 94.59 | 97.02 |

lowest accuracy (87.14%) compared to other approaches. The proposed approach (DT+SCA) increases 5.021% and 10.87% accuracy compared to Hajjej *et al.* [10] and Rasul *et al.* [14] approach. The metrics precision, recall, f1-score and specificity increases 14.84%, 39.32%, 26.94% and 2.57% compared to Rasul *et al.* [14] approach.

Table 4 shows the performance metrics for KNN algorithm. The suggested method provides the highest values of 97.46%, 97.59%, 97.46%, 97.47%, and 97.76% in accuracy, precision, recall, f1-score and specificity. The approach Rasul *et al.* [14] has lowest accuracy (89.05%) compared to other approaches. The proposed approach (KNN+SCA) increases 4.79% and 9.44% accuracy compared to Hajjej *et al.* [10] and Rasul *et al.* [14] approach. The metrics precision, recall, f1-score and specificity increases 15.88%, 25.88%, 20.82%, and 4.08% compared to Rasul *et al.* [14] approach and 3%, 4%, 4%, and 4% for Hajjej *et al.* [10] approach.

**Table 4.** Performance metrics for KNN Algorithm

| Metrics (%) | KNN with Default Parameter | Hajjej et al. [10] | Rasul et al. [14] | KNN+SCA |
|---|---|---|---|---|
| **Accuracy** | 96.62 | 93 | 89.05 | 97.46 |
| **Precision** | 96.68 | 95 | 84.21 | 97.59 |
| **Recall** | 96.62 | 94 | 77.42 | 97.46 |
| **F1-Score** | 96.62 | 94 | 80.67 | 97.47 |
| **Specificity** | 96.81 | 94 | 93.92 | 97.76 |

Table 5 shows the performance metrics for RF algorithm. The suggested method provides the highest values of 98.73%, 98.76%, 98.73%, 98.73% and 98.88% in accuracy, precision, recall, f1-score and specificity. The approach RF with default parameter, Hajjej *et al.* [10] and Rasul *et al.* [14] has lowest accuracy (95.78%, 95% and 95.71%) compared to proposed approaches. The proposed approach (RF+SCA) increases more than 3% accuracy compared to other approach. The metrics precision, recall, f1-score and specificity increases 4%, 9%, 7%, and 3% compared to other approaches.

**Table 5.** Performance metrics for RF Algorithm

| Metrics (%) | RF with Default Parameter | Hajjej et al. [14] | Rasul et al. [18] | RF+SCA |
|---|---|---|---|---|
| **Accuracy** | 95.78 | 95 | 95.71 | 98.73 |
| **Precision** | 96.12 | 97 | 94.92 | 98.76 |
| **Recall** | 95.78 | 96 | 90.32 | 98.73 |
| **F1-Score** | 95.78 | 97 | 92.56 | 98.73 |
| **Specificity** | 96.28 | 97 | 97.97 | 98.88 |

Table 6 shows the performance metrics for SVM algorithm. The approach SVM with default parameter, Rasul *et al.* [14] and proposed approach has similar accuracy (96.20%, 96.19% and 96.20%) compared to Hajjej *et al.* [10] (90%). The proposed approach (SVM+SCA) increases 6.88% accuracy compared to Hajjej *et al.* [10] approach. The metrics precision, recall, f1-score and specificity increases 2.34%, 6.88%, 4.56%, and 3.92% compared to Hajjej *et al.* [10] approach.

**Table 6.** Performance metrics for SVM Algorithm

| Metrics (%) | SVM with Default Parameter | Hajjej et al. [10] | Rasul et al. [14] | SVM+SCA |
|---|---|---|---|---|
| **Accuracy** | 96.20 | 90 | 96.19 | 96.20 |
| **Precision** | 96.48 | 94 | 93.55 | 96.48 |
| **Recall** | 96.20 | 90 | 93.55 | 96.20 |
| **F1-Score** | 96.20 | 92 | 93.55 | 96.20 |
| **Specificity** | 96.65 | 93 | 97.30 | 96.65 |

Hyperparameter tuning via the Sine Cosine Algorithm has proven to be an efficacious metaheuristic optimization method for enhancing the performance of ML approaches: AdaBoost, DT, RF, KNN and SVM. The SCA employs a population-based search approach that dynamically equilibrates exploration and exploitation by modifying candidate solutions using sine and cosine functions, facilitating effective navigation of intricate hyperparameter spaces. The proposed hyperparameter tuning approach achieves accuracy of 98.31%, 96.62%, 97.46%, 98.73% and 96.20% for Adaboost, DT, KNN, RF and SVM. Overall, the AdaBoost and RF achieve higher accuracy compared to DT, KNN and SVM.

## 4 Conclusion

This study introduces an effective predictive analytics framework for autism spectrum disorder prediction, termed PAASD_HPT, which employs hyper-parameter adjusting of machine learning models. The proposed approach uses SCA for tuning hyper parameter of ML algorithms. The suggested hyperparameter tuning method attains accuracies of 98.31% for AdaBoost, 96.62% for Decision Trees, 97.46% for K-Nearest Neighbors, 98.73% for Random Forest, and 96.20% for Support Vector Machines. Following hyperparameter optimization, Random Forest and AdaBoost outperform other models with a significant statistical difference. This model surpasses state-of-the-art methodologies; nonetheless, it may be evaluated using deep learning algorithms in the future to enhance accuracy for autism spectrum disease. The future research includes: Concentrate on advancing the amalgamation of machine learning and explainable AI methodologies to augment the precision and clarity of evaluation tools for ASD. It is essential to examine the incorporation of privacy-preserving methodologies, such as federated learning, to guarantee the confidentiality of patient data while enhancing diagnostic accuracy. Future research may evaluate the efficacy of various machine learning models across distinct age demographics to optimize diagnostic methodologies.

## References


1) Issac A, Halemani K, Shetty A, Thimmappa L, Vijay VR, Koni K, et al. The global prevalence of autism spectrum disorder in children: a systematic review and meta-analysis. *Osong Public Health and Research Perspectives*. 2025;16(1). https://doi.org/10.24171/j.phrp.2024.0286.
2) Zhao W, Xu S, Zhang Y, Li D, Zhu C, Wang K. The application of extended reality in treating children with autism spectrum disorder. *Neuroscience Bulletin*. 2024;40(8):1189–1204. https://doi.org/10.1007/s12264-024-01190-6.
3) Bousidrah NS, Khamees ZM, Ali SM. Detection of Autism Spectrum Disorder by A Case Study Model Using Machine Learning Techniques An Experimental Analysis on Child, Adolescent and Datasets. *Scientific Journal of University of Benghazi*. 2024;37(1). https://doi:10.37376/sjuob.v37i1.5943.
4) Thalor M, Patil S. Psychological Assessment of Autism Spectrum Disorders using Machine Learning. *International Journal of Educational and Psychological Sciences*. 2024;2(1):65–70. https://doi:10.59890/ijeps.v2i1.1297.
5) Elbattah M, Ibrahim OAS, Dequen G. Improving autism spectrum disorder diagnosis using machine learning techniques. *Frontiers in Neuroinformatics*. 2024;18. https://doi.org/10.3389/fninf.2024.1529839.
6) Subhash AR, Motagi AKU. An Efficient Autism Spectrum Disorder Classification in Different Age Groups using Machine Learning Models. *International Journal of Online & Biomedical Engineering*. 2024;20(9):17–38. https://doi.org/10.3991/ijoe.v20i09.48831.
7) Abdelwahab MM, Al-Karawi KA, Hasanin EM, Semary HE. Autism spectrum disorder prediction in children using machine learning. *Journal of Disability Research*. 2024;3(1). https://doi.org/10.57197/JDR-2023-0064.
8) Ayub R, Ansarullah SI, Samdani F, Akhtar A, Afzal M, Ahmed SM. Enhancing early detection of autistic spectrum disorder in children using machine learning approaches. *Journal of King Saud University-Science*. 2024;36(10). https://doi.org/10.1016/j.jksus.2024.103468.
9) Farooq MS, Tehseen R, Sabir M, Atal Z. Detection of autism spectrum disorder (ASD) in children and adults using machine learning. *Scientific Reports*. 2023;13(1). https://doi:10.1038/s41598-023-35910-1.
10) Hajjej F, Ayouni S, Alohali MA, Maddeh M. Novel Framework for Autism Spectrum Disorder Identification and Tailored Education With Effective Data Mining and Ensemble Learning Techniques. *IEEE Access*. 2024;12:35448–35461. http://dx.doi.org/10.1109/ACCESS.2024.3349988.
11) Kavadi DP, Chirra VRR, Kumar PR, Veesam SB, Yeruva S, Pappala LK. A Hybrid Machine Learning Model for Accurate Autism Diagnosis. *IEEE Access*. 2024;12:194911–194921. https://doi.org/10.1109/ACCESS.2024.3520009.

12) Reghunathan RK, Venkidusamy PNP, Kurup RG, George B, Thomas N. Machine Learning-Based Classification of Autism Spectrum Disorder across Age Groups. *Engineering Proceedings*. 2024;62(1). http://dx.doi.org/10.3390/engproc2024062012.
13) Jeon I, Kim M, So D, Kim EY, Nam Y, Kim S, et al. Reliable Autism Spectrum Disorder Diagnosis for Pediatrics Using Machine Learning and Explainable AI. *Diagnostics*. 2024;14(22). https://doi.org/10.3390/diagnostics14222504.
14) Rasul RA, Saha P, Bala D, Karim SRU, Abdullah MI, Saha B. An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder. *Healthcare Analytics*. 2024;5. https://doi.org/10.1016/j.health.2023.100293.
15) Ben-Sasson A, Guedalia J, Nativ L, Ilan K, Shaham M, Gabis LV. A Prediction Model of Autism Spectrum Diagnosis from Well-Baby Electronic Data Using Machine Learning. *Children*. 2024;11(4). https://doi.org/10.3390/children11040534.
16) Haque N, Islam T, Erfan M. An exploration of machine learning approaches for early Autism Spectrum Disorder detection. *Healthcare Analytics*. 2025;7. https://doi.org/10.1016/j.jnha.2024.100379.
17) Autism Prediction. Available from: https://www.kaggle.com/competitions/autismdiagnosis/data.
18) Rajam FJ, Swakkin BRK. MAPLE: A Novel Processing Technique for Adult Autism Prediction. *Engineering, Technology & Applied Science Research*. 2025;15(3):23901–23906. https://doi.org/10.48084/etasr.10864.
19) Rajam FJ, Swakkin BRK. ASD Pipeline An Ensemble Machine Learning Framework Integrating Feature Selection, Behavioural Clustering, and Class Rebalancing for Accurate Autism Spectrum Disorder Prediction. *International Journal of Novel Research and Development*. 2025;10(5):b217–b230. https://dx.doi.org/10.56975/ijnrd.v10i5.306593.
20) Mirjalili S. SCA: a sine cosine algorithm for solving optimization problems. *Knowledge-based systems*. 2016;96:120–133. https://doi.org/10.1016/j.knosys.2015.12.022.