Journal of Applied Data Sciences
Vol. 5, No. 4, December 2024, pp. 1754-1764

Artificial Intelligence Techniques for Early Autism Detection in Toddlers: A Comparative Analysis

Qusai Y. Shambour[1,*], Nazem Qandeel[2], Yousef Alraba'nah[3], Anan Abumariam[4], Mohd Khaled Shambour[5]

[1,3]*Department of Software Engineering, Faculty of Information Technology, Al-Ahliyya Amman University, Amman 19111, Jordan*

[2,4]*Department of Special Education, Faculty of Educational Sciences, Al-Ahliyya Amman University, Amman 19111, Jordan*

[5]*The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research, Umm Al-Qura University, Makkah 24352, Saudi Arabia*

(Received: August 23, 2024; Revised: September 21, 2024; Accepted: September 29, 2024; Available online: October 15, 2024)

Abstract

Individuals with autism spectrum disorder are often characterized by complications in social interaction and communication, which can be attributed to disruptions in brain development affecting their perception and interaction with others. While ASD is not a treatable condition, early detection and diagnosis can significantly mitigate its effects. Recent advancements in artificial intelligence have enabled the development of novel diagnostic tools, which can detect ASD at an earlier stage than traditional methods. This study attempts to enhance and automate the diagnostic process by employing a variety of machine learning techniques to identify the most critical characteristics of ASD. To this end, we employed six state-of-the-art machine learning classification models, including Support Vector Machine, Random Forest, k-Nearest Neighbors, Logistic Regression, Decision Tree, and Naive Bayes classifiers, to analyze and predict ASD in toddlers. Our dataset comprises 1054 instances from a non-clinical ASD screening dataset for toddlers, sourced from the University of California-Irvine (UCI) Machine Learning repository. This data was collected using the ASDTests mobile application, which was developed based on Q-CHAT and AQ-10 screening tools. Our evaluation focused on a range of performance metrics, including accuracy, precision, recall, and F1-score, to assess the efficacy of each model. Notably, the Logistic Regression model demonstrated the highest accuracy in diagnosing ASD in toddlers, achieving a perfect score of 100% across all metrics.

Keywords: Autism Spectrum Disorder, Toddlers; Q-Chat-10, Artificial Intelligence, Machine Learning, Early Diagnosis, Classifier

1. Introduction

Autism spectrum disorder (ASD) is among the most prevalent neurodevelopmental disorders, which impair social interaction and communication and lead to restricted interests and repetitive behaviors. Individuals with ASD have difficulties understanding others' emotions and thoughts and often have difficulty communicating [1]. Traditional diagnostic methods for ASD, such as the Autism Diagnostic Interview (ADI) and the Autism Diagnostic Observation Schedule (ADOS), involve behavioral assessments conducted by healthcare providers and parents. These methods include several tests designed to evaluate specific aspects of day-to-day life, such as social interaction, attitude toward games and sports, and the use of everyday objects. However, these diagnostic processes can be lengthy and inefficient, often taking over three years in some cases. This extended timeline delays access to necessary interventions, such as speech therapy, behavioral treatment, and medications that improve daily functioning and social participation for individuals with ASD [2], [3].

The limitations of traditional diagnostic methods have significant consequences, including delayed intervention and reduced treatment efficacy. Children diagnosed with ASD between 18 and 36 months show a high level of diagnostic stability, which significantly enhances long-term outcomes and quality of life. Early intervention, initiated before 36 months, improves language development, cognition, social skills, and adaptive behaviors. It also aids in developing effective communication skills, reducing communication challenges, and promoting daily living skills. Additionally,

***Corresponding author: Qusai Y. Shambour (q.shambour@ammanu.edu.jo)**
DOI: https://doi.org/10.47738/jads.v5i4.353

early intervention is cost-effective, decreasing the need for intensive interventions and special education services in the future [4], [5], [6]. Considering these drawbacks in current diagnostic methods, researchers have turned to artificial intelligence (AI) and machine learning (ML) algorithms. The use of AI and ML algorithms to identify ASD has taken center-stage in recent times due to their ability to diagnose ASD more quickly and accurately, thereby minimizing the need for elaborate assessments, trained staff, and added resources for handling the growing number of cases associated with ASD [7], [8]. AI and ML have been used to develop models for detecting symptoms and treatment of several diseases such as diabetes, heart diseases, and cancer [9]. In several cases, researchers have discussed the potential of AI and ML in detecting ASD with better accuracy and effectiveness to mitigate the impact of ASD [10].

Recent studies have exploited ML for ASD detection, and their preliminary results are quite encouraging. For example, the study by Bousidrah et al. [11] focused on the use of an ML framework that takes into account the diagnosis of ASD in children. The findings of the proposed ML technique suggest that it is an efficient tool that would lead to ASD diagnosis at a very early stage. The predictive models of case detections would help healthcare professionals for diagnosing ASD in young children. The experimental results associated with ASD detection have provided important information for healthcare professionals to better understand the most significant features to explore while diagnosing an ASD case. In another example, Rasul et al. [2] proposed advanced ML techniques to detect ASD and extract relevant features. Through AI, the diagnosis of health conditions can be made more reasonable, and the time saved in the treatment process is of utmost value for those who are likely to be suffering from ASD. Besides, the study evaluated eight different well-known classification models based on some performance metrics like accuracy, precision, recall (sensitivity), specificity, F1-score, AUC, kappa, and log loss. Experiments showed that the support vector machine model works best for the children dataset and logistic regression for the adult dataset. The artificial neural network model achieves an impressive accuracy of 94.24% when the hyperparameters are finely tuned. The study also examines five very popular clustering algorithms, which would help the reader understand how the model behaves when true labels are absent.

Even in light of those advances, however, comprehensive research on the effectiveness of several ML algorithms for ASD detection remains limited. This study aims to bridge that gap by identifying ASD in toddlers at an early age using a variety of ML classification models. The ML classification models that were thoroughly investigated include Support Vector Machine (SVM), Random Forest (RF), k-nearest neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), and Naive Bayes (NB). These classifiers were assessed in terms of accuracy, precision, recall, and F1 score. This study will be furthered in Section 2 with a literature review. Section 3 presents a demonstration of the dataset and classification methods used. Section 4 presents the experimental results, while a thorough discussion of those results takes place in Section 5. Finally, Section 6 concludes the findings.

2. Related Works

This section presents some studies that considered the effectiveness of several ML classification models applied for the early detection of ASD. Akter et al. [12] examined the use of ML techniques for the early detection of ASD. They investigated ASD datasets related to toddlers, adolescents, children, and adults using several feature transformation methods such as log, Z-score, and sine functions. Then they implemented several classification models on the transformed datasets and assessed the classification performance. Elaborating on the results, it was stated that SVMs performed the best for the toddler dataset, while Adaboost provided optimal results for the children dataset. For the adolescent dataset, Glmboost was effective, and Adaboost excelled once again on the adult datasets. Key feature transformations that improved classification included sine function on the toddler dataset and Z-score normalization on the children and adolescent datasets. This study emphasizes the significance of data pre-processing, feature engineering, and selection of appropriate ML classification models in the effective diagnosis of ASD.

A comprehensive approach was presented by Bala et al. [13] for utilizing ML for early diagnosis of ASD across different age groups. The authors collected ASD datasets for children, adolescents, toddlers, and adults. They attempted different feature selection techniques, ran several classifiers on these datasets, and evaluated performance based on several metrics including predictive kappa statistics, accuracy, F1-score, and AUROC. The study revealed SVM to be the most effective classifier, producing high accuracy: 97.82% for toddlers' data, 99.61% for children, 95.87% for adolescents, and 96.82% for adults. This study highlights the efficacy of ML models in the early detection of ASD, which would significantly reduce the costs and time involved in traditional methods of diagnosis.

Talukdar et al. [8] provided an exhaustive study on the most commonly used ML classifiers for diagnosing ASD in children and adolescents. Several algorithms, including NB, LR, SVM, and RF, were used to work with non-clinical children and adolescent's datasets. The study describes all of the specific signs and symptoms of ASD, namely, impediments in repetitive behavior, social communication, and restricted interests. It emphasizes the importance of early diagnosis and intervention, highlighting the two-stage diagnostic process for young children and the unique challenges faced in diagnosing older children and adults. The paper also discusses the genetic and environmental factors contributing to ASD and the various treatment and intervention strategies available.

Hasan et al. [14] presented a comprehensive evaluation of various ML models for early detection of ASD. The study leverages four standard ASD datasets encompassing adolescents, children, toddlers, and adults. The authors preprocess these datasets by handling missing values and encoding categorical features. They address class imbalance using a random-over-sampler strategy. Eight ML classifiers, including AdaBoost, RF, DT, K-NN, Gaussian Naive Bayes, LR, SVM, and Linear Discriminant Analysis, are employed to classify the feature-scaled datasets. The study identifies the best-performing classification models and feature scaling techniques for each dataset by comparing various statistical evaluation measures. The study concludes that proper tuning of ML models significantly enhances the prediction of ASD among different age groups. It emphasizes that the detailed feature importance analysis can aid healthcare practitioners in making informed decisions during ASD screening.

Prasad et al. [15] employed a design science approach, consisting of stages such as problem identification, objectives, design and development, demonstration, evaluation, and communication. Data was sourced from the National Hospital in Abuja, Nigeria, comprising 998 samples and 10 attributes, including preterm birth, gestational diabetes, and family history. Data analysis involved handling missing values, feature engineering, and correlation assessment. Principal Component Analysis (PCA) was utilized for feature selection. Subsequently, three classification algorithms were tested: DT, RF, and NB. The RF algorithm yielded the highest accuracy rate of 98.7% after k-fold cross-validation. Bayesian probability was also employed to assess the model's reliability. The researchers conclude that the model can assist medical professionals in early ASD diagnosis, enabling timely intervention and treatment.

Reghunathan et al. [16] focused on improving accuracy in detecting ASD affecting children, adolescents, and adults. They utilized various ML algorithms, especially emphasizing a LR model that performed well. The authors used a feature reduction technique called Cuckoo search to find highly discriminative features affecting the prediction of ASD in children, adolescents, and adults. Overall, this research strives towards improving ASD diagnosis through ML, reinforcing the importance of early detection and intervention. The findings indicated a path forward for the advancement of predictive models relying on medical data toward better accuracy concerning the identification of ASD.

3. Methodology

3.1. Data Collection

The study utilized the ASD Toddler dataset, which is publicly available from the UCI ML Repository [17]. The dataset was gathered by conducting ASD screenings on toddlers between the ages of 12 and 36 months. Data collection was conducted using a global online survey administered through the ASDTests mobile application, developed by Thabtah et al [17]. The dataset consists of 1,054 instances, with 17 independent variables (excluding the case number), and one dependent variable that indicates the ASD classification. The features consist of 10 behavioral aspects obtained from screening questions (Q-Chat-10) and 7 individual characteristics that offer demographic information, including age, gender, ethnicity, country of origin, and whether the child was born with jaundice (refer to table 1).

Although the dataset provides valuable insights, it is important to acknowledge certain limitations and potential biases. The small size of the dataset limits the development of a generalized model, as a larger dataset is crucial for constructing a robust and widely applicable classification model. Furthermore, relying exclusively on an online platform for data collection may restrict the participation of certain populations, which may result in the exclusion of lower-income groups or regions with limited online access. In addition, while the sample includes participants from various countries,

ISSN 2723-6471

there is a varying representation of geographic and cultural diversity. This could potentially impact the dataset's ability to accurately represent the broader global population of toddlers at risk for ASD. Nevertheless, the dataset is widely used in ASD research and has been validated in previous recent studies [8], [11], [12], [13], [15]. Despite these limitations, we believe that the dataset serves as a valuable resource for exploring the application of ML algorithms in ASD diagnosis.

Table 1. Description of the ASD toddler dataset

Instances	Attributes	Male/Female	Age	ASD/Normal
1054	17	735/319	12-36 (months)	735/319

3.2. Data Preprocessing

Initially, the preprocessing of the ASD toddler dataset comprises the encoding of categorical attributes and the scaling of the data using a standardization approach. It should be noted that the dataset does not have any missing or outlier values and thus remains reliable for further analysis. These preprocessing steps are crucial to allow subsequent ML models to achieve efficiency and effectiveness thus promoting reliable and accurate predictions. Before training the model, categorical variables are converted into numerical data. For nominal variables, one-hot encoding is applied, and ordinal variables are subjected to label encoding; for instance, the gender attribute takes numeric values where 0 is for females and 1 for males.

Furthermore, many features in the dataset have different scales. Therefore, we need feature scaling, whereby we carry out the transformation of the data to the same standardized scale. This will improve not only the model accuracy but also the speed of the learning process. In our work, we have used the min-max scaling normalization method to normalize each feature to the range between 0 to 1, which endows every feature with equal importance over the model, thus diminishing the risk of providing large ranges of feature values that could be overriding the learning process.

3.3. Apply Individual ML Classifiers

Six ML classifiers are employed, including SVM, RF, KNN, LR, DT, and NB. All these models are implemented using the Python programming language. The values for hyper-parameters tuning are displayed in table 2.

3.3.1. Support Vector Machine

SVM is a supervised ML model utilized for both regression and classification tasks. It segments the features by fitting the optimal hyperplane to split data points into different classes by maximizing the margin between classes and minimizing the classification errors. SVM works on both linear and non-linear data, hence comes in handy for high-dimensional datasets. The various kernel functions are used to map input data into higher-dimensional spaces where separation becomes feasible. With the ability to automatically set main parameters and minimize expected test errors, SVM is powerful in solving complex classification problems [18].

3.3.2. Random Forest

RF is an ensemble learning technique that merges several decision trees to form a robust model for classification and regression tasks. It constitutes a "forest" of trees trained on a bootstrap sample taken from the dataset and a subset of the features that are randomly chosen. This, termed bagging, increases model stability and reduces overfitting. When making their predictions, each tree in the forest casts a vote, with the final output determined through majority voting or averaging. RF is adept at handling complex datasets, providing rankings of feature importance, and producing reliable predictions in various fields. It is capable of generalization and hence tends to give better accuracy than a single decision tree, reducing some common issues like overfitting [19].

3.3.3. k-Nearest Neighbors

KNN is one of the simplest, yet effective ML models applicable to regression as well as classification. Its functionality is quite intuitive; for a given instance in the feature space, it finds K nearest data points and makes class membership or value predictions based on those points. Being a non-parametric and instance-based method, KNN is distribution-free and utilizes the whole training set to predict. The performance of KNN depends on selecting an appropriate K value and distance metric. Although KNN is versatile and intuitive, it is quite computationally arduous for huge datasets, especially those in high-dimensional spaces. Nevertheless, KNN has been decidedly usable in many domains, including text classification, image processing, pattern recognition, and recommendation systems [20], [21]. Simple and effective, KNN positioned itself as an excellent data mining algorithm [22].

3.3.4. Logistic Regression

LR serves as a common method in ML for binary classification tasks. It evaluates probabilities that a sample belongs to some specific class based on observations of its features. Generally, this process requires a fitting of a logistic function to the training set data, thereby transforming the input features into one probability score between 0 and 1. For LR to classify something, a labeled training dataset needs to fit the model. Then maximum likelihood estimation is used to fit the model parameters for the logistic function. With a new instance, the logistic regression then computes the probability of the instance being in a positive category by using the logistic function fitted in the previous step, assigning the class with the highest probability [23].

3.3.5. Decision Tree

DT can be considered a family of ML algorithms used primarily in classification and regression problems. These models are constructed very much like trees of repeated small partitions of smaller sets based on the values of the feature variables to predict an outcome. Each branch in the tree is associated with an outcome of a feature, and a leaf node generally signifies a class label. DTs are famous for their interpretability, feature importance evaluation, and ability to handle categorical and numerical data. These models are the building blocks of ensemble methods and find application in different domains, if one is to consider their interpretability in decision-making processes [24].

3.3.6. Naïve Bayes

NB is a probabilistic ML model used for data classification. It applies Bayes' theorem to compute the probability of an instance belonging to different classes. It is assumed that the features are independent, thereby simplifying the probability calculations. Hence, the model learns the probabilities with which an instance in the training data belongs to each of the different classes. When an unseen instance is fed into this model, it calculates the probabilities of each class based on the probabilistic knowledge it acquired during training [25].

Table 2. Hyperparameters setting of ML classifiers

Classifier	Parameters
SVM	C: 600.0, gamma: ‘scale’, kernel: ‘poly’
RF	n_estimators: 300, max_depth: 15, min_samples_leaf: 1, min_samples_split: 2
KNN	n_neighbors: 5, metric: ‘manhattan’, weights: ‘distance’
LR	C: 500, l1_ratio: 0.1, penalty: ‘l2’, solver: ‘saga’
DT	criterion: ‘gini’, max_depth: 10, min_samples_leaf: 1, min_samples_split: 2, splitter: ‘best’

3.4. Evaluating the Performance of M Classifiers

Several evaluation metrics including Accuracy, Precision, Recall, and F1-score were considered to assess and compare the performance of different classifiers. These metrics were chosen for their complementary nature and specific relevance to ASD diagnosis. The metrics were calculated using the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as presented in the confusion matrix (shown in table 3) [26], [27], [28]. After evaluation, we identified the classifiers that achieved the highest outcomes for the ASD toddler dataset. These metrics collectively offer a detailed understanding of classifier performance on ASD datasets, supporting the development of precise diagnostic tools and interventions for toddlers with ASD.

Table 3. Confusion Matrix

	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Accuracy: It provides a comprehensive assessment of a model's predictive capabilities, which is crucial in ASD diagnosis where both correct identification of ASD cases and correct exclusion of non-ASD cases are important. It represents the ratio of accurately classified instances to the total number of instances within a dataset. This metric shows the model's effectiveness in accurately identifying both true positives (correctly identified ASD cases) and true negatives (correctly identified non-ASD cases).

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

Precision: It measures how well a classifier can identify true positive cases. It calculates the number of predicted positive instances that match with actual positive cases, thus reducing false positives. Precision plays a key role with regard to evaluating the performance of a classifier, as it shows how well the classifier is at distinguishing individuals with ASD from those classified as such. A false positive can cause unnecessary stress for families, additional tests, and the potential for wasting resources in the clinical setting. High precision ensures that the model predicting ASD can be trusted, thus making it a very crucial element for reliable ASD screening.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (1)$$

Recall: It measures the extent to which a classifier can identify all true cases of ASD while reducing the number of missed diagnoses. It is calculated by taking the proportion of true ASD cases that were identified by the model against the total true ASD cases. Therefore, high recall is indispensable in preliminary scenarios to ensure those requiring further assessment or intervention do not pass through the net. Failure to make timely diagnoses or to implement intervention can have worse long-term outcomes for children with ASD.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (2)$$

F1-score: This metric provides a balance of how the classifier performs in bridging the gap between precision and recall and is very useful in the domain of ASD diagnosis. It helps in tackling the possible trade-off between the number of ASD detected (high recall) and the number of cases confirmed positive by the model (high precision). Such balancing is even important as, in both cases, ASD observation and over-diagnosis need to be reduced in clinical practice. This balanced approach is particularly valuable when working with an imbalanced dataset that has one class much larger than the other.

$$\text{F1} - \text{score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (3)$$

4. Experimental Results

This section presents and discusses the assessment findings of different ML classification models on the ASD toddler dataset. We used evaluation metrics such as accuracy, precision, recall, and F1-score to assess the performance of various classifiers. These metrics are crucial for evaluating classification models, especially in medical diagnosis where false positives and false negatives carry significant costs. We trained and tested the models using classification reports (figure 1, figure 2, figure 3, and figure 4) and confusion matrices (figure 5), allocating 75% of the dataset for training and 25% for testing.

Figure 1 illustrates the accuracy of the ML classification models, which measures the proportion of correctly classified instances out of the total instances. The LR model achieved a perfect accuracy of 100%, correctly classifying all instances. This high performance is due to its high TP rate, where 74 instances were correctly classified as ASD, and a high TN rate, where 190 instances were correctly identified as non-ASD. The SVM followed closely with an accuracy of 98.48%. RF and KNN achieved accuracy scores of 97.34% and 96.96%, respectively. The DT and NB models had lower accuracy scores of 95.07% and 89.98%, respectively. These results suggest that LR and SVM are highly effective in correctly identifying ASD and non-ASD cases in toddlers.

Figure 2 shows the results of the precision metric, which measures the proportion of true positives among all positive predictions made by the classification models. LR again achieves a perfect precision of 100%, indicating that all positive predictions made by the model are correct. SVM follows closely, with a precision of 97.76%, while RF and KNN achieve precision scores of 97.73% and 95.35%, respectively. The DT and NB models exhibit relatively lower precision scores of 93.74% and 83.33%, respectively. These results suggest that LR and SVM are highly effective in minimizing false positives and correctly identifying ASD cases in toddlers.

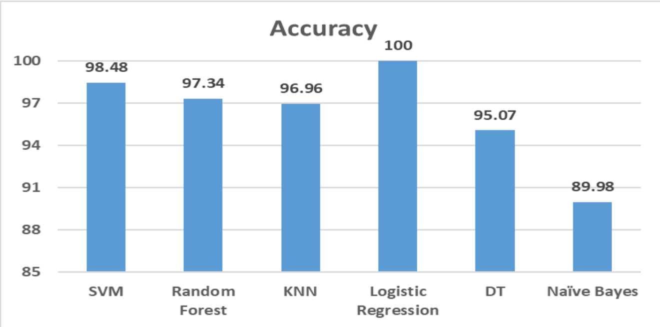

Figure 1. Accuracy comparison of various ML models used for ASD diagnosis in toddlers

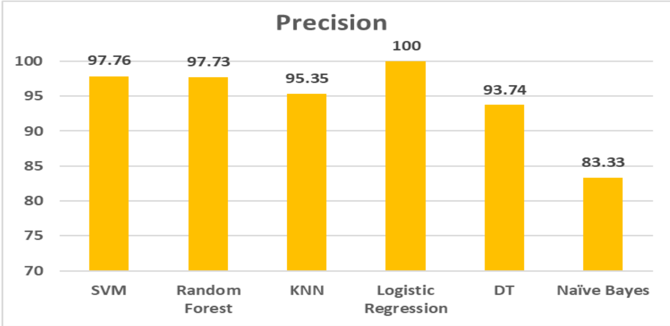

Figure 2. Precision comparison of various ML models, indicating the proportion of correct positive predictions for ASD cases

Figure 3 shows the recall of different models, and indicates the proportions of true positives among all actual positive instances; that is, they show how many of the relevant positives (in this case, exactly how many of the relevant ASD instances) the different models captured. The best performer was the LR model, which reached an impressive 100% recall; this means that the model was able to classify all of the ASD instances correctly. The next best was the SVM, with a very strong 98.53% recall. This means that the SVM was very effective at identifying almost all of the relevant true ASD instances. The KNN model performed slightly worse but still very respectably, with a 97.57% recall. The DT and NB models had lower recall scores of 94.1% and 90.26%, respectively. These results suggest that LR and SVM are highly effective in diminishing false negatives and identifying the real cases of ASD correctly.

Figure 4 presents the F1 score, which serves as a useful metric for delivering a balance between precision and recall. The F1 score accounts for false positives and false negatives and is interpreted such that the higher the score, the better the model performed in terms of its balance between precision and recall. The LR model achieved an F1 score of 100%. This indicates that the LR model is very reliable in making the correct classifications for toddlers with ASD. Following the LR model, the SVM model achieved an F1 score of 98.13%. Although the SVM model also performs quite well, the slight difference in the F1-score indicates that it is not quite as harmoniously balanced between precision and recall as the LR model. The SVM model, however, does enjoy a fairly sizeable margin of victory over the next model, the KNN model, which attained an F1 score of 96.33%. The DT and NB algorithms exhibit relatively lower F1 scores of 93.92% and 84.6%, respectively. These results suggest that LR and SVM are highly effective in achieving a balance between precision and recall, making them suitable for ASD diagnosis in toddlers.

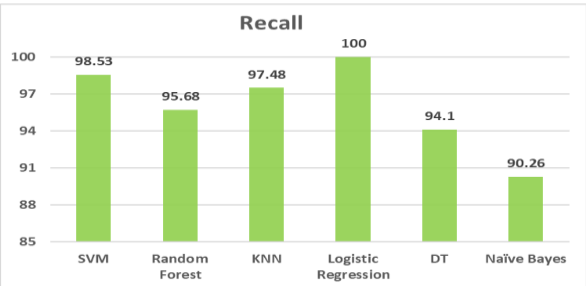

Figure 3. Recall comparison of various ML models, showing how effectively they identify true ASD cases from all positive instances

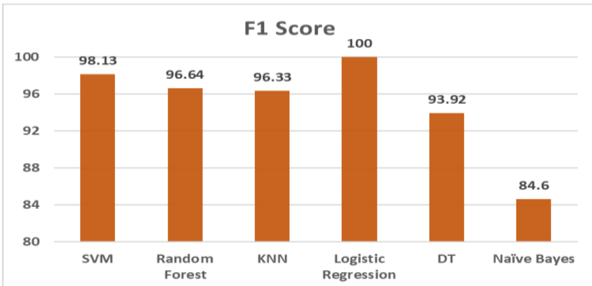

Figure 4. F1-score comparison of various ML models, illustrating the balance between precision and recall

The confusion matrices of all classification models illustrated in figure 5 offer a detailed insight into the performance of each classification model. They illustrate classes that were correctly and incorrectly predicted by each model, providing a breakdown of error types. Each matrix classified all the instances in the data set into four categories: TP, TN, FP, and FN to evaluate the model's predictive capability and its ability to obtain the desired outcomes.

Overall, the LR emerged as the best performer, as all metrics showed a perfect 100% due to the complete identification of all ASD cases and no false positive cases. The SVM and RF also showed strong results, confirming their capacity to generate a highly precise diagnostic tool. The DT was reasonably effective, with a precision of 93.74%, a recall of 94.1%, and an F1-score of 93.92%. Finally, the NB was the weakest among those tested, but still potentially useful given its precision of 83.33%, recall of 90.26%, and F1 score of 84.6%, indicating its potential utility in less critical applications or as part of an ensemble method. These results highlight the importance of model selection in ASD classification tasks. While LR's flawless performance makes it a prime candidate, the strong showings of SVM and RF suggest their viability for enhancing early ASD diagnosis and intervention.

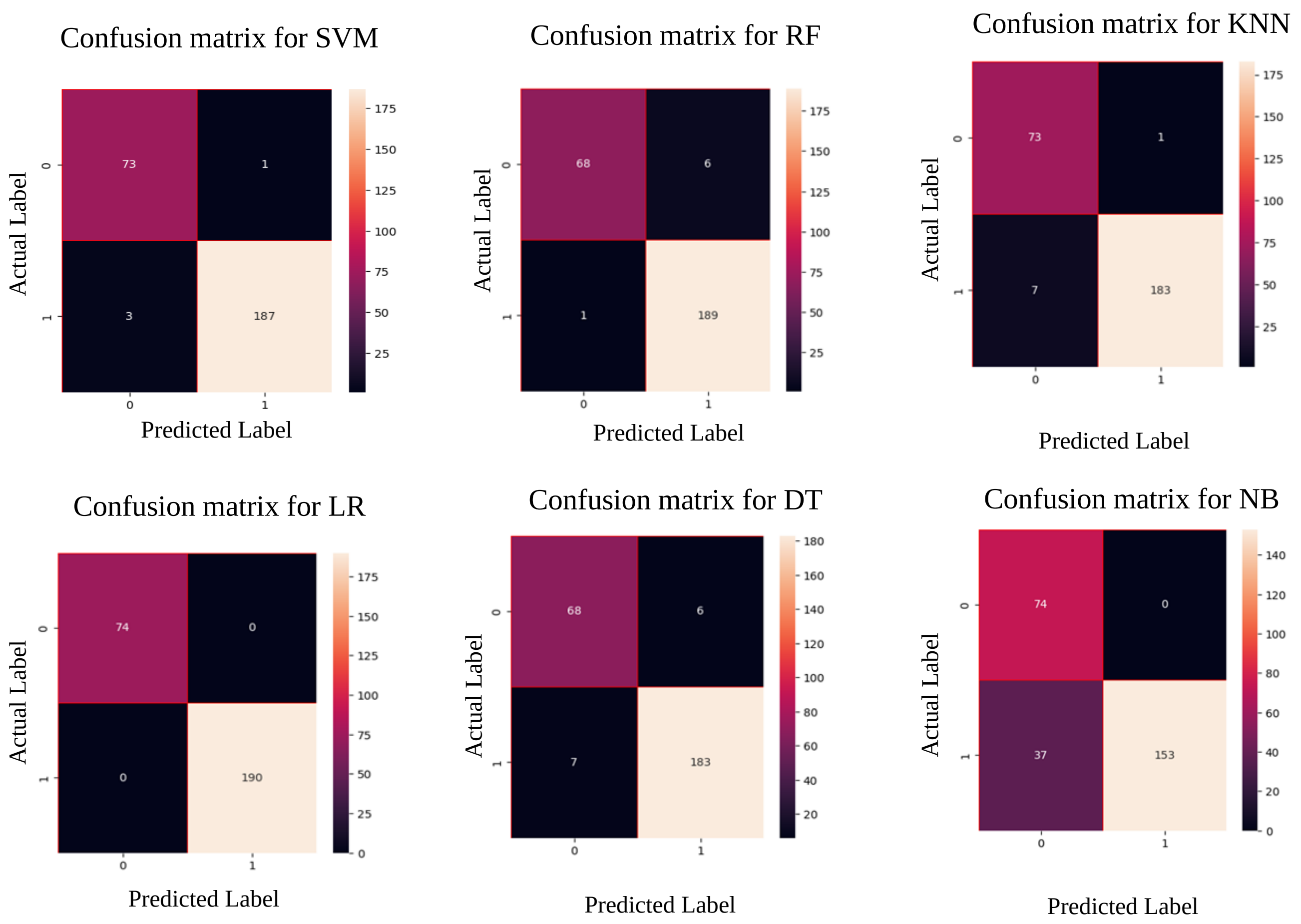

Figure 5. Confusion matrices of all classification models

5. Discussion

Based on the above results, we observed that the LR achieved the highest performance scores with 100% in all performance metrics. A perfect score meant that LR model set a linear boundary to separate the ASD and the non-ASD instances within the dataset. Although the performance metrics were perfect, additional validations in real-world scenarios are needed to ensure the generalization of LR. This study was based on clean data, which is free of unpredictable noises. In the application of LR, whether the model built on the clean data would still be able to show 100% in real clinical scenarios or in the case of collecting data from multiple hospitals or countries with different cultural backgrounds has yet to be detailed explored. SVM and RF also have impressive results, achieving accuracies of 98.48% and 97.34%, respectively. These results could be attributed to the effective handling of the high-dimensional spaces in SVM and the ensembled nature of RF in reducing the error ratio.

KNN and DT showed satisfactory performance, achieving accuracies of 96.96% and 95.07% respectively. KNN showed remarkable results in terms of recall, suggesting its potential to accurately identify cases of ASD. Being prone to overfitting in small datasets, the DT may have less generalizability in this case. NB performed the worst of all models, achieving 89.98% accuracy. This can be explained by the assumption made by NB of conditional independence among features, which may not have held for the ASD dataset. Nevertheless, the relatively high recall of NB seems to indicate that it was able to identify most true cases of ASD, while it did generate many false positives in doing so.

Based on the comparative analysis, it is evident that LR outperformed SVM and RF for this particular dataset. However, with some adjustments, SVM and RF have the potential to be highly effective in diagnosing ASD. Models such as KNN and DT, although they may have lower accuracy, have shown promise for certain diagnostic requirements or as parts of ensemble methods. These findings highlight the significance of choosing the right model for ASD diagnosis. When the data does not allow for clear linear separation, more complex models like SVM and RF can be highly effective. They offer robust substitutes for datasets with non-linear relationships or complex feature interactions.

6. Conclusion

This study explored the application of artificial intelligence techniques to enhance early detection of ASD in toddlers. We employed six classification models: SVM, RF, KNN, LR, DT, and NB to analyze and predict ASD using a non-clinical ASD dataset. Our evaluation used metrics such as accuracy, precision, recall, and F1-score to assess the effectiveness of each model. The findings of this study indicate that in the context of ASD diagnosis, ML classification models can be employed by clinicians to detect cases of ASD in a more accurate and timely manner. Thus, the overall possibility to facilitate long-term results and diagnose relevant cases during the in-time periods to introduce the corresponding interventions. Not less importantly, the results have provided additional implications to explore and potentially adopt AI-driven techniques for the detection of ASD, thus, addressing the existing limitations regarding the use of traditional diagnostic methods.

LR outperformed all other models, achieving perfect scores across all metrics. SVM and RF also showed strong performance, indicating their robustness for ASD classification tasks. KNN demonstrated good accuracy and recall, making it a viable option for specific use cases. DT and NB, despite lower accuracy, demonstrated reasonable performance, indicating their suitability for some specific applications. These metrics would provide a comprehensive view of the strengths and weaknesses of each model, giving rise to the necessity of choosing an appropriate model per the requirements of the ASD classification task. The study results highlight the role of the ML algorithms in enhancing the precision and reliability of the early diagnosis of ASD, thus contributing to enhancing the prospects of good prognosis with early intervention.

Future research should focus on expanding the dataset to include more diverse populations to ensure that the findings can be applied to a broader population. It is also important to address any data preprocessing-related issues, as the imbalance of data concerning ML model performance can hinder its ability to classify cases accurately. Utilizing feature selection algorithms can enhance models' efficiency by identifying the most significant features for ASD classification. Optimization of hyper-parameters would be a step further to improve the performance of the models, which compiles to be accurate and resilient. Moreover, when integrating advanced ML and DL models with clinical expertise, ensemble and hybrid-based approaches may result in more advanced diagnostic tools. These tools should be reliable, accessible, and relatively easy to integrate into daily clinical settings to promote equity and transparency in usage.

7. Declarations

7.1. Author Contributions

Conceptualization: Q.Y.S., N.Q., Y.A., A.A., and M.K.S.; Methodology: N.Q. and M.K.S.; Software: Q.Y.S.; Validation: Q.Y.S., N.Q., and A.A.; Formal Analysis: Q.Y.S., N.Q., and A.A.; Investigation: Q.Y.S., Y.A., and A.A.; Resources: Y.A., A.A., and M.K.S.; Data Curation: Y.A. and A.A.; Writing Original Draft Preparation: Q.Y.S., N.Q.,

and Y.A.; Writing Review and Editing: N.Q., Y.A., and M.K.S.; Visualization: Q.Y.S., N.Q., and A.A.; All authors have read and agreed to the published version of the manuscript.

7.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.3. Funding

This work was funded by the Deanship of Scientific Research at Al-Ahliyya Amman University (Grant number: 6/21/2023-2024).

7.4. Institutional Review Board Statement

Not applicable.

7.5. Informed Consent Statement

Not applicable.

7.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. J. Landa, A. L. Gross, E. A. Stuart, and A. Faherty, "Developmental trajectories in children with and without autism spectrum disorders: the first 3 years," *Child development*, vol. 84, no. 2, pp. 429-442, 2013.

[2] R. A. Rasul, P. Saha, D. Bala, S. R. U. Karim, M. I. Abdullah, and B. Saha, "An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder," *Healthcare Analytics*, vol. 5, no. June, pp. 1-16, 2024.

[3] R. J. Landa, "Diagnosis of autism spectrum disorders in the first 3 years of life," *Nature Clinical Practice Neurology*, vol. 4, no. 3, pp. 138-147, 2008.

[4] G. Dawson, S. Rogers, J. Munson et al., "Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model," *Pediatrics*, vol. 125, no. 1, pp. 17-23, 2010.

[5] G. Vivanti, C. Dissanayake, and V. A. Team, "Outcome for children receiving the Early Start Denver Model before and after 48 months," *Journal of autism and developmental disorders*, vol. 46, no. 7, pp. 2441-2449, 2016.

[6] R. J. Landa, "Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders," *International Review of Psychiatry*, vol. 30, no. 1, pp. 25-39, 2018.

[7] D. Y. Song, S. Y. Kim, G. Bong, J. M. Kim, and H. J. Yoo, "The Use of Artificial Intelligence in Screening and Diagnosis of Autism Spectrum Disorder: A Literature Review," *Soa Chongsonyon Chongsin Uihak*, vol. 30, no. 4, pp. 145-152, 2019.

[8] 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, no. November, pp. 1-16, 2023.

[9] M. Shehab, L. Abualigah, Q. Shambour et al., "Machine learning in medical applications: A review of state-of-the-art methods," *Computers in Biology and Medicine*, vol. 145, no. June, pp. 105458, 2022.

[10] S. R. Shahamiri, and F. Thabtah, "Autism AI: a new autism screening system based on artificial intelligence," *Cognitive Computation*, vol. 12, no. 4, pp. 766-777, 2020.

[11] N. S. Bousidrah, Z. M. Khamees, and S. M. Ali, "Detection of Autism Spectrum Disorder by A Case Study Model Using Machine Learning Techniques An Experimental Analysis on Child, Adolescent and Datasets," *The Scientific Journal of University of Benghazi*, vol. 37, no. 1, pp. 87-100, 2024.

[12] T. Akter, M. S. Satu, M. I. Khan et al., "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," *IEEE Access*, vol. 7, no. Nov., pp. 166509-166527, 2019.

[13] M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, "Efficient machine learning models for early stage detection of autism spectrum disorder," *Algorithms*, vol. 15, no. 5, pp. 1-22, 2022.

ISSN 2723-6471

[14] S. M. M. 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,” *IEEE Access*, vol. 11, no. February, pp. 15038-15057, 2023.

[15] R. Prasad, F. Musa, and H. Muhammad, “Autism Spectrum Disorder Prediction Using Machine Learning and Design Science,” *International Journal of Experimental Research and Review*, vol. 39, no. May, pp. 213-228, 2024.

[16] R. K. Reghunathan, P. N. P. Venkidusamy, R. G. Kurup, B. George, and N. Thomas, “Machine Learning-Based Classification of Autism Spectrum Disorder across Age Groups,” *Engineering Proceedings*, vol. 62, no. 1, pp. 1-7, 2024.

[17] F. Thabtah, F. Kamalov, and K. Rajab, “A new computational intelligence approach to detect autistic features for autism screening,” *International Journal of Medical Informatics*, vol. 117, no. September, pp. 112-124, 2018.

[18] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing*, vol. 408, no. September, pp. 189-215, 2020.

[19] S. J. Rigatti, “Random Forest,” *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31-39, 2017.

[20] Q. Y. Shambour, A. H. Hussein, Q. M. Kharma, and M. M. Abualhaj, “Effective hybrid content-based collaborative filtering approach for requirements engineering,” *Computer Systems Science and Engineering*, vol. 40, no. 1, pp. 113-125, 2022.

[21] Q. Shambour, M. M. Abu-Alhaj, and M. M. Al-Tahrawi, “A hybrid collaborative filtering recommendation algorithm for requirements elicitation,” *International Journal of Computer Applications in Technology*, vol. 63, no. 1-2, pp. 135-146, 2020.

[22] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, “Learning k for kNN classification,” *ACM Transactions on Intelligent Systems and Technology,* vol. 8, no. 3, pp. 1-19, 2017.

[23] P. Schober, and T. R. Vetter, “Logistic Regression in Medical Research,” *Anesthesia & Analgesia*, vol. 132, no. 2, pp. 365-366, 2021.

[24] S. B. Kotsiantis, “Decision trees: a recent overview,” *Artificial Intelligence Review*, vol. 39, no. April, pp. 261-283, 2013.

[25] L. E. Sucar, "Bayesian Classifiers," Probabilistic Graphical Models: Principles and Applications, pp. 43-69, Cham: Springer International Publishing, 2021.

[26] D. Powers, “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.

[27] M. Kolhar, F. Al-Turjman, A. Alameen, and M. M. Abualhaj, “A Three Layered Decentralized IoT Biometric Architecture for City Lockdown During COVID-19 Outbreak,” *IEEE Access*, vol. 8, no. September, pp. 163608-163617, 2020.

[28] M. Sokolova, and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information processing & management*, vol. 45, no. 4, pp. 427-437, 2009.