Journal of
Computer Technology & Applications
https://journals.stmjournals.com/jocta

ISSN: 2229-6964 (Online)
ISSN: 2347-7229 (Print)
Volume 15, Issue 3, 2024
September–December
DOI (Journal): 10.37591/JoCTA

Research JoCTA

Autism Spectrum Disorder Prediction Using Classification Techniques: A Comparative Analysis

S. Srividhya[1,*], Lavanya S.R.[2]

Abstract

Autism spectrum disorder (ASD) is a multifaceted neurodevelopmental disorder marked by difficulties in social interaction, communication, and repetitive behaviors. Identifying and addressing ASD early is essential for enhancing the quality of life for those affected. Data mining techniques have emerged as powerful tools in analyzing large datasets to predict and diagnose ASD, aiding in early identification and intervention. This article presents a comprehensive comparative analysis of classification techniques employed in data mining for predicting ASD. ASD, characterized by diverse symptoms and complexities in diagnosis, necessitates advanced methodologies for early detection and intervention. The effectiveness of data mining techniques, such as decision trees, support vector machines, and k-nearest neighbors, is examined for predicting ASD. The comparative analysis focuses on accuracy, precision, and recall to evaluate the strengths and limitations of each technique. Findings from this study aim to provide insights into the applicability of classification methods in ASD prediction, guiding the development of robust models for early identification and intervention strategies. The article emphasizes future directions and challenges, aiming to improve the accuracy and practical implementation of ASD prediction models using data mining techniques.

Keywords: Autism spectrum disorder (ASD), classification techniques, decision tree, support vector machine, K-nearest neighbor

INTRODUCTION

Autism spectrum disorder (ASD) is a multifaceted neurodevelopmental disorder characterized by a variety of symptoms such as difficulties in social interaction, communication challenges, repetitive behaviors, and sensory sensitivities. The intricate nature of ASD necessitates early detection and intervention strategies to enhance the quality of life of the affected individuals. According to Pan [1], children with ASD who interacted with people less socially also showed less physical activity than those who interacted with adults more socially. Although it tends to affect people of all racial and cultural backgrounds, boys are more than four times more likely than girls to be identified as having autism. Although a child can be accurately diagnosed as early as two years old, most diagnoses are made after the age of four [2]. It is possible to intervene more quickly when it is discovered earlier. If autistic children receive the right care and sufficient training, it may be feasible to enhance their behavioral and communication skills [3, 4]. The diagnosis of autism is time- and money-intensive. Early identification of autism can greatly benefit patients by allowing the prescription of appropriate medications. It can prevent the patient's

*Author for Correspondence
S. Srividhya
E-mail: vidhyasai14@gmail.com

[1]Associate Professor, Department of Computer Science, KPR College of Arts Science and Research, Coimbatore, Tamil Nadu, India
[2]Associate Professor, Department of Computer Science, KPR College of Arts Science and Research, Coimbatore, Tamil Nadu, India

Received Date: July 20, 2024
Accepted Date: August 22, 2024
Published Date: September 12, 2024

Citation: S. Srividhya, Lavanya S.R. Autism Spectrum Disorder Prediction Using Classification Techniques: A Comparative Analysis. Journal of Computer Technology & Applications. 2024; 15(3): 66–71p.

illness from worsening and assist in reducing the long-term expenses related to a delayed diagnosis. Therefore, there is a great need for a quick, accurate, and simple screening test that can anticipate an individual's qualities related to autism and determine whether they need a full autism examination.

Data mining techniques have emerged as powerful tools for analyzing extensive datasets to predict and diagnose ASD, aiding in early identification and tailored intervention plans. This article presents a comparative analysis of various classification techniques employed in data mining for predicting ASD, aiming to elucidate their efficacy and suitability in ASD prediction models. According to [5], data-intensive machine learning techniques have great potential for examining the reproducibility of brain function patterns in more extensive and diverse datasets. This study aims to contribute to the existing knowledge base by offering a comprehensive comparative analysis of classification techniques within data mining for ASD prediction. The objective is to facilitate the development of more precise, efficient, and applicable models that aid in early identification and intervention strategies for individuals affected by ASD.

Section II presents a literature review. Section III describes the dataset used in this study. The proposed comparative technique is described in Section IV. The experimental results are presented in Section V. The suggested methodology based on experimental findings is presented in Section VI. Section VII presents conclusions and future enhancements.

BACKGROUND STUDY

Elements such as social, communication, and behavior-related concerns should be included in the therapy for children with autism [6]. Pisula provided an excellent intervention approach that aids in the development of suitable therapeutic strategies for individuals who have been diagnosed with autism. The key to successful intervention is to begin working with a child as soon as possible if there are reasonable indications that they may have autism. The ideal time to begin therapy with a child is prior to the 18th month of life. Speech therapy and therapeutic activities should be carefully planned and selected based on the unique needs of each child. Because of the attention issues these children exhibit, individual treatment sessions should begin with getting to know the child.

Bone et al. [7] created a model using the surface vector machine (SVM) classifier for training and cross-validation of the data. To diagnose autism, they conducted interviews and used two standardized exams, the social responsiveness scale, to gather data for this study. Yuan et al. [8] study on a machine learning model for diagnosing ASD is based on natural language processing (NLP). They initially digitally transformed all forms of unstructured and semi-structured data, pre-processed them, and then completed the categorization to create this prediction model. The procedure for detecting ASD was made simpler and more efficient by using this paradigm.

A prediction algorithm using past long-term patient records to identify all conceivable symptoms of ASD was established by Dutta et al. [9]. This approach can recognize both the common and uncommon forms of ASD. The system used to create the matrix was developed using both machine learning and colloquy theory, enabling any set of associations to produce rules. Additionally, it has been reported that their ASD detection technology operates faster because it requires only a single database query and uses a larger amount of RAM.

Altay et al. [10] employed K-nearest neighbors (KNN) and linear discriminant analysis (LDA). to categorize children aged 4 to 11 with ASD. They used a 70:30 split between testing and training. These algorithms decrease biased predictions and increase prediction accuracy.

A strategy for diagnosing ASD was developed by the investigator [11] to achieve greater classification accuracy using the fewest feature subsets possible. The University of California Irvine machine learning repository (UCI ML) repository provided the kid data set, which had 292 instances and 21 features. It was subsequently assessed using a binary firefly feature selection wrapper that leveraged swarm intelligence. To differentiate between ASD and no ASD class type, the author used the feature selection wrapper to choose 10 features from the ASD dataset's 21 features: "A1_Score," "A2_Score," "A3_Score," "A4_Score," "A5_Score," "A7_Score," "A8_Score," "A9_Score," and "A10_Score."The authors utilized ML classifier models, NB, J48 decision tree [12], SVM [13], KNN [14], and Multilayer Perceptron (MLP) [15], after dimensionality reduction, for categorizing ASD and no ASD class type. After selecting the best feature subset and training the classification models with the fewest possible behavior sets, the approach's findings demonstrated an average accuracy in the range of 92.12% to 97.95%, proving the effectiveness of the models.

Bekerom [16] examined the outcomes of using various machine learning algorithms, such as naive Bayes, SVM, and Random Forest algorithms, to identify ASD symptoms in children, including obesity, developmental delay, and decreased physical activity. Wall et al. [17] used a brief screening test and validation to classify autism, resulting in high sensitivity, specificity, and accuracy for both ADTree and functional tree. Using a sizable brain imaging dataset from the Autism Imaging Data Exchange (ABIDE I), Heinsfeld [18] used a deep learning algorithm and neural network to identify individuals with ASD. The results indicate an average classification accuracy of 70%, with values ranging from 66% to 71%. The Random Forest classifier produced a mean accuracy of 63% compared with the SVM classifier with 65% mean accuracy.

Thabtah and Peebles [19] proposed new rule-based machine learning (RML) criteria to examine the rationale behind the categorization. Except for toddlers, the study examined datasets from the UCI ML repository that included 1100 occurrences and 21 attributes in the child, adolescent, and adult categories. The RML result was subsequently tested against eight algorithms (RIPPER, RIDOR, Nnge, Bagging, Boosting, CART, C4.5, and PRISM) to assess the performance. RML outperformed each of the datasets in terms of the rate of accuracy, error, sensitivity, specificity, and harmonic mean.

PROPOSED COMPARATIVE ANALYSIS

The methodology proposed in this study consisted of three phases. The first phase involved handling missing values, the second phase involved the feature extraction process, and the final phase involved classification. The UCI ML Repository and Kaggle are the sources of the ASD datasets, which are freely accessible. Datasets from children, adolescents, and adults were used for this analysis. The child dataset had 21 attributes and 292 records. The adolescent dataset contained 21 attributes and 104 records. The adult dataset contained 21 attributes and 704 records. All three datasets had missing values, which were treated in the first phase of the analysis. The proposed methodology is illustrated in Figure 1.

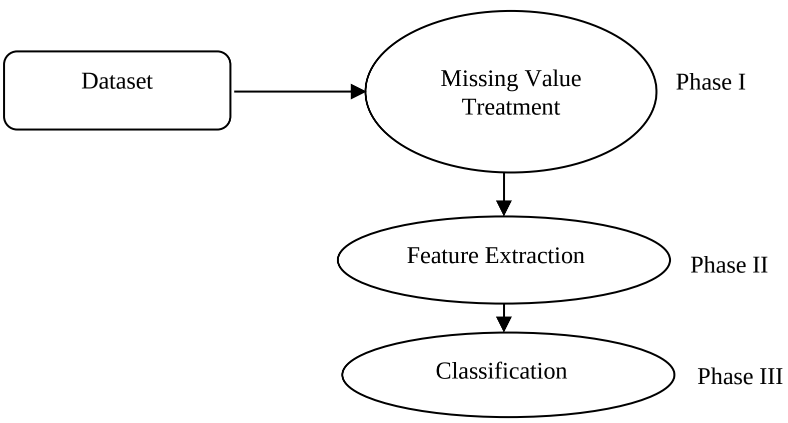

Figure 1. Proposed architecture.

The proposed methodology involves missing value treatment in the first phase. Missing values were handled by simply ignoring missing value instances. After ignoring the missing values, the dataset was completed for all values in all instances. The second phase deals with feature extraction. In the second phase, the technique involved factor analysis. Factor analysis is a statistical technique for reducing dimensionality and examining underlying structures in datasets by identifying patterns among observed variables. The main objective is to reveal the latent factors that account for the correlations between the variables.

Steps Involved in Factor Analysis

- *Step 1:* Gather the dataset.
- *Step 2:* Identify the initial factors and their loadings on each observed variable.
- *Step 3:* Include Kaiser's criterion (retaining factors with eigenvalues >1), screen plot examination, and theoretical implications.
- *Step 4:* Apply the rotation method Varimax to improve the interpretability of the factors. Rotation aims to achieve a simpler structure with higher factor loadings for a limited number of factors.
- *Step 5:* Interpret the factors based on the loading pattern Variables with high loadings on a particular factor were strongly related to this factor.
- *Step 6:* Utilize the identified factors as a reduced set of dimensions that captures the variance in the original dataset. Factor scores can be computed to represent individual observations of these dimensions.

With all the above steps performed in the factor analysis, the dataset had a limited number of features for the classification process.

The third phase of the proposed methodology involves the classification process. All the extracted features in Phase II are assigned to the classification algorithms for the prediction process. decision trees, support vector machines, and K-nearest neighbor algorithms were employed in the analysis process. The effectiveness of the proposed method was evaluated using performance metrics such as recall, precision, and accuracy (ACC).

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

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

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

RESULTS AND DISCUSSIONS

To determine the performance measures, the suggested method is applied in stages to every category of the child, adolescent, and adult datasets. MATLAB 2016 was used to train the Deep neural networks (DNN) model on a PC equipped with an Intel Corei3 1.99 GHz processor and 12 GB RAM. Implementing this method in the three classifiers resulted in a good improvement in performance. SVM outperformed the other classifiers.

Table 1 shows that the proposed method outperformed and proved to be significant in terms of accuracy, precision, recall, and for the child, adolescent, and adult datasets. The support vector machine provided better results for the three datasets. Figure 2 shows a pictorial representation of a comparative analysis of the proposed and existing approaches.

From the above Figure 2, it is evident that the SVM shows better performance in terms of accuracy, precision, and recall for all three datasets used in the analysis. Accuracy, precision, and recall appear to be higher in the adult dataset for the SVM classifier than in the child and adolescent datasets.

Table 1. Performance measures of the proposed method in terms of accuracy, precision, recall, and F-measure.

Dataset	Metrics	Decision tree	SVM	KNN
Child	Accuracy	87.9	94.34	91.97
	Precision	87.99	92.23	91.95
	Recall	87.89	93.01	92.04
Adolescent	Accuracy	82.57	89.23	87.9
	Precision	81.61	92.34	87.93
	Recall	81.37	90.01	87.92
Adult	Accuracy	93.22	98.34	97.99
	Precision	93.29	98.01	97.95
	Recall	93.17	98.67	98.01

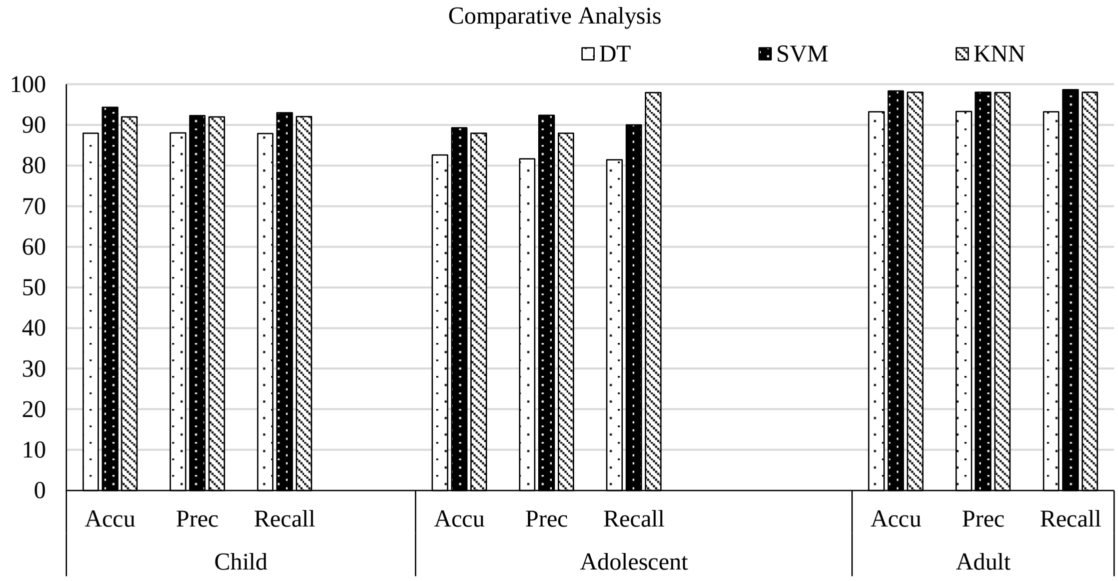

Figure 2. Comparative analysis of the proposed method in terms of accuracy, precision, recall, and F-measure.

CONCLUSION

The proposed work emphasizes the early detection of ASDs. Research on ASD prediction using classification techniques has provided valuable insights into the effectiveness of various methods for diagnosing and predicting ASD. Through a comprehensive comparative analysis, this study highlighted the strengths and limitations of different classification algorithms in accurately identifying individuals at risk for ASD. The comparative analysis presented in this study offers valuable guidance for practitioners and researchers working towards more accurate and efficient ASD prediction models. As advancements in technology and data collection methodologies continue, integrating these findings into clinical practice holds immense promise in aiding early identification and intervention for individuals on the autism spectrum. Continued research and collaboration across disciplines is essential to harness the full potential of machine learning techniques to improve ASD prediction and ultimately enhance the lives of individuals affected by this condition. In this study, a factor analysis was used to reduce the number of attributes. Based on the contribution of the minimal benefit, the attributes were reduced in the dataset. The different evaluation parameters, such as accuracy, precision, and recall, yielded clinically acceptable results using SVM compared to decision tree and K-nearest neighbor algorithms. The proposed method can be further refined by advancing techniques to address imbalance issues in the datasets.

REFERENCES

1. Pan CY. Objectively measured physical activity between children with autism spectrum disorders and children without disabilities during inclusive recess settings in Taiwan. J Autism Dev Disord. 2008;38:1292–301. DOI: 10.1007/s10803-007-0518-6.
2. Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. Arch Gen Psychiatry. 2006;63:694–701. DOI: 10.1001/archpsyc.63.6.694.
3. Corsello CM. Early intervention in autism. Infants Young Child. 2005;18:74–85. DOI: 10.1097/00001163-200504000-00002.
4. Handleman JS, Harris SL, editors. Preschool Education Programs for Children with Autism. Austin, TX: Pro-Ed; 2001.
5. Varoquaux G, Thirion B. How machine learning is shaping cognitive neuroimaging. GigaScience. 2014;3:28. DOI: 10.1186/2047-217X-3-28.
6. Gumińska N, Zając M, Piórkowski P. People with Autism in Society – Challenge of 21st Century. Case of Poland. Procedia Soc Behav Sci. 2015;174:576–83. DOI: 10.1016/j.sbspro.2015.01.586.
7. Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS. Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion. J Child Psychol Psychiatry Allied Discip. 2016;57:927–37. DOI: 10.1111/jcpp.12559.
8. Yuan J, Holtz C, Smith T, Luo J. Autism spectrum disorder detection from semi-structured and unstructured medical data. EURASIP J Bioinformatics Syst Biol. 2017;2017:3. DOI: 10.1186/s13637-017-0057-1.
9. Dutta SR, Datta S, Roy M. Using cogency and machine learning for autism detection from a preliminary symptom. In: Confluence 9th Int Conf Cloud Comput Data Sci Eng. 2019:331–6. DOI: 10.1109/CONFLUENCE.2019.8776993.
10. Altay O, Ulas M. Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children. In: 6th Int Symp Digit Forensic Secur. (ISDFS). 2018;1–4. DOI: 10.1109/ISDFS.2018.8355354.
11. Vaishali R, Sasikala R. A machine learning based approach to classify autism with optimum behaviour sets. Int J Eng Technol. 2018;7:18.
12. Quinlan JR. C4.5: Programs for Machine Learning. Morgan Kaufmann Series in Machine Learning. San Francisco (CA): Elsevier; 2014.
13. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. Improvements to Platt's SMO algorithm for SVM classifier design. Neural Comput. 2001;13:637–49. DOI: 10.1162/089976601300014493.
14. Brighton H, Mellish C. Advances in instance selection for instance-based learning algorithms. Data Min Knowl Discov. 2002;6:153–72. DOI: 10.1023/A:1014043630878.
15. Abe S. Pattern Classification: Neuro-fuzzy Methods and Their Comparison. London: Springer-Verlag; 2001.
16. Vakadkar K, Purkayastha D, Krishnan D. Detection of autism spectrum disorder in children using machine learning techniques. SN computer science. 2021 Sep;2:1–9.
17. Wall D, Kosmicki J, DeLuca T, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. Transl Psychiatry. 2012;2:e100. DOI: 10.1038/tp.2012.10.
18. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. Neuroimage Clin. 2018;17:16–23. DOI: 10.1016/j.nicl.2017.08.017.
19. Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. Health Inform J. 2020;26:264–86. DOI: 10.1177/1460458218824711.

