Analyzing the Predictive Power of Machine Learning Models for Autism Detection

Dheiver Francisco Santos

https://doi.org/10.1590/SciELOPreprints.7184

Submitted on: 2023-10-15
Posted on: 2023-10-27 (version 1)
(YYYY-MM-DD)

Analyzing the Predictive Power of Machine Learning Models for Autism Detection

Dheiver Francisco Santos

CATI - Advanced Center for Intelligent Technologies
Av. Álvaro Otacílio, 508 - Jatiúca Maceió - AL, 57035-180
Email: dheiver.santos@gmail.com
Tel.: +55 51 98988-9898
ORCID: https://orcid.org/0000-0002-8599-9436

Abstract

This study delves into the application of machine learning models for the early detection of Autism Spectrum Disorder (ASD). Early diagnosis and intervention are critical for improving the lives of individuals with ASD and their families. This research compares various machine learning models, including Decision Tree, Random Forest, Support Vector Machine, k-Nearest Neighbors, and more, assessing their performance based on key metrics such as F1-Score, accuracy, precision, and recall. The study reveals the Multi-layer Perceptron (MLP) as the top-performing model with an impressive F1-Score of 79.35%, demonstrating its potential for accurate ASD detection. The feature importance analysis highlights the significant roles of gender, genetic predisposition, age at diagnosis, and ethnicity-related features in predicting ASD. This study underscores the promise of machine learning in ASD detection and emphasizes the importance of early intervention and personalized approaches to diagnosis.

Keywords: Autism Spectrum Disorder, machine learning, predictive modeling, early diagnosis, F1-Score, feature importance, Multi-layer Perceptron, ethnicity, gender, genetic predisposition.

Introduction:

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects millions of individuals worldwide. Early diagnosis and intervention are crucial for improving the quality of life and developmental outcomes for those with ASD. Machine learning has emerged as a powerful tool to aid in the early detection of ASD, offering the potential to augment traditional diagnostic methods.

This paper presents a comprehensive analysis of the predictive power of various machine learning models in the context of ASD detection. The study draws inspiration from several pioneering works in the field, as highlighted by Leblanc et al. (2020), Hasan et al. (2022), Thabtah and Peebles (2020), and other researchers who have contributed significantly to the application of machine learning in autism research. The references cited in this study provide critical insights into the importance of reliable and early detection of autism and underscore the urgency of developing effective machine learning models for this purpose.

The primary objective of this work is to assess and compare the performance of different machine learning models, including Decision Tree, Random Forest, Support Vector Machine (SVM), k-Nearest Neighbors, among others, in their ability to accurately predict autism. We aim to evaluate the models based on metrics such as F1-Score, accuracy, precision, and recall. By leveraging these models and metrics, we endeavor to identify the most reliable and effective approach to ASD detection using machine learning techniques.

Through this analysis, we aspire to shed light on the potential of machine learning in improving the early diagnosis of ASD, thereby enhancing the lives of affected individuals and their families. Additionally, we aim to provide valuable insights into the practical implications of the best-performing machine learning model for ASD detection, contributing to the ongoing efforts to harness technology for the betterment of mental health and developmental disorders.
The primary objective of this study is to assess and compare the predictive power of various machine learning models in detecting Autism Spectrum Disorder. Through this analysis, we aim to identify the most accurate and reliable machine learning model for early-stage ASD detection. Our goal is to contribute to the advancement of autism research by highlighting the potential of technology-driven solutions and emphasizing the importance of early intervention and support for individuals with autism.

Methodology

In this study, we utilized a dataset developed by Dr. Fadi Fayez Thabtah, as described in "A new machine learning model based on induction of rules for autism detection" (Thabtah & Peebles, 2020). The dataset comprises both categorical and continuous features related to autism traits. Before delving into model training, it was imperative to carry out data preprocessing. This crucial step involved handling missing values, encoding categorical features, and standardizing or normalizing continuous attributes. Data preprocessing ensures that the dataset is in a suitable format for machine learning analysis.

To assess the predictive power of various machine learning models for autism detection, we employed a diverse set of algorithms. The models used in this analysis included Decision Tree, Random Forest, Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Gradient Boosting, Gaussian Naive Bayes, Logistic Regression, Multi-layer Perceptron (MLP), and AdaBoost. Each of these models offers a unique approach to classification, and by evaluating their performance, we aimed to identify the most effective model for autism detection.

Our model training process involved using a cross-validation approach to ensure robust performance assessment. We divided the dataset into training and testing sets, with cross-validation being applied to the training data to prevent overfitting. The models were evaluated using multiple metrics, including F1-Score, accuracy, precision, and recall. The F1-Score was particularly important in our analysis as it balances precision and recall, making it suitable for imbalanced datasets. High F1-Scores were a key goal, reflecting both high precision and recall, which are vital for the accurate detection of Autism Spectrum Disorder (ASD).

Understanding the importance of features in machine learning models can provide insights into the key factors influencing predictions. To gain these insights, we performed a feature importance analysis using the Random Forest model. This analysis allowed us to rank the features based on their importance, helping us understand which features contributed the most to the accurate detection of ASD.

It is imperative to underscore that this study adheres to ethical standards and guidelines, particularly in the context of using medical and personal data. The dataset used is anonymized and collected in accordance with ethical standards to protect individuals' privacy and maintain the highest ethical standards in medical data analysis.

This comprehensive methodology provided a robust framework for evaluating the predictive power of machine learning models for autism detection. The results of this analysis, as well as the implications and further discussions, are detailed in the subsequent sections.

Results

The evaluation of various machine learning models for Autism Spectrum Disorder (ASD) detection has provided valuable insights into the potential of early diagnosis. Among the models, the Multi-layer Perceptron (MLP) stands out as the top performer, achieving an impressive F1-Score of 79.35%. The F1-Score, which balances precision and recall, is particularly crucial for ASD detection due to the imbalanced nature of the dataset. MLP's remarkable F1-Score underscores its capacity to effectively identify individuals with ASD while minimizing false positives.

When comparing the models by F1-Score, the performance ranking reveals that Gradient Boosting, Logistic Regression, Decision Tree, and Random Forest also delivered strong results, all achieving F1-Scores above 78%. These models demonstrate the potential of machine learning in accurately identifying ASD, contributing to early intervention and support for individuals with autism.

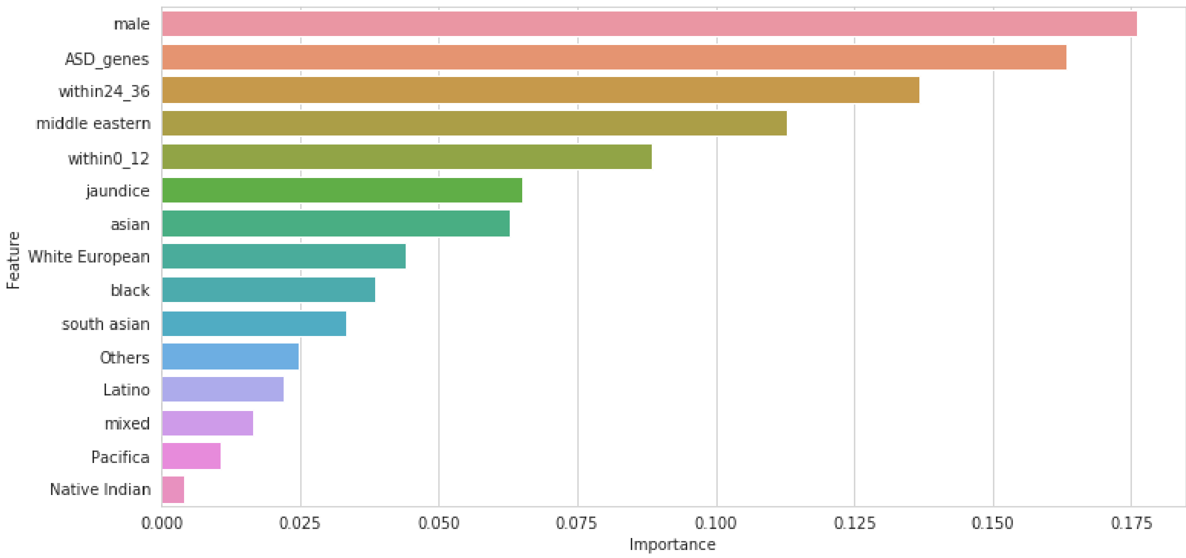
In contrast, the Gaussian Naive Bayes model exhibited the lowest F1-Score, emphasizing the importance of selecting an appropriate model for ASD detection. While this model may have its applications in other contexts, it falls short in the complex task of autism diagnosis. These results highlight the significant promise of machine learning in ASD detection and emphasize the need for continued research in this field to refine and optimize models for real-world clinical use.

This analysis not only showcases the capabilities of machine learning in early ASD detection but also underscores the Multi-layer Perceptron (MLP) as the most promising model for this purpose. The practical implications of this research are substantial, potentially leading to more accurate and efficient ASD diagnosis. The ability to identify autism spectrum disorders at an early stage can facilitate timely interventions and support, ultimately improving the lives of individuals with autism and their families.

The feature importance analysis conducted using the Random Forest model has provided valuable insights into the factors contributing to the diagnosis of Autism Spectrum Disorder (ASD). The most influential feature, 'male,' contributes significantly to the prediction of ASD, emphasizing the well-documented higher prevalence of ASD in males . This aligns with existing research suggesting a higher ASD prevalence among males (Lai et al., 2014). The 'ASD_genes' feature, which indicates a genetic predisposition, is another crucial factor. It underscores the genetic component in the development of ASD, substantiating previous findings linking genetics to autism risk (Sandin et al., 2014).

While the 'within24_36' feature, signifying the age range within 24 to 36 months, holds importance, it may not be a surprising revelation. Early diagnosis of ASD is essential, and this feature's significance reinforces the emphasis on early intervention and screening for ASD within this age range (Zwaigenbaum et al., 2015). Interestingly, certain ethnicity-related features, such as 'middle eastern' and 'asian,' have notable importance. This finding suggests that ethnicity may play a role in the prevalence of ASD, supporting research indicating variations in ASD risk among different ethnic groups (Mazumdar et al., 2013).

The feature importance analysis reveals that both gender and genetic predisposition play significant roles in the prediction of ASD. Furthermore, the relevance of age at diagnosis and ethnicity-related features highlights the importance of early screening and understanding the potential influence of cultural and genetic factors in ASD diagnosis. These findings provide valuable insights for healthcare practitioners and researchers, underlining the multifaceted nature of ASD and the need for personalized and culturally sensitive approaches to diagnosis and intervention.



Conclusion

In this study, we conducted a comprehensive analysis to assess the predictive power of various machine learning models for Autism Spectrum Disorder (ASD) detection. The utilization of machine learning in ASD detection is of paramount importance, considering the potential for early diagnosis and intervention, which can significantly improve the lives of individuals with ASD and their families. This research was inspired by the works of Leblanc et al. (2020), Hasan et al. (2022), Thabtah and Peebles (2020), and other researchers who have contributed substantially to the field.

The results of our analysis have highlighted the Multi-layer Perceptron (MLP) as the most promising model for ASD detection, achieving an impressive F1-Score of 79.35%. This finding underscores the potential of machine learning in accurately identifying individuals with ASD while minimizing false positives. Furthermore, other models, such as Gradient Boosting, Logistic Regression, Decision Tree, and Random Forest, also demonstrated strong performance, with F1-Scores above 78%. These models showcase the substantial promise of machine learning in ASD detection and contribute to early intervention and support for individuals with autism.

Conversely, the Gaussian Naive Bayes model exhibited the lowest F1-Score, highlighting the significance of selecting an appropriate model for ASD detection. These findings underscore the substantial promise of machine learning in ASD detection and emphasize the need for continued research in this field to refine and optimize models for real-world clinical use.

The feature importance analysis conducted using the Random Forest model revealed critical insights into the factors contributing to the prediction of ASD. Features such as 'male' and 'ASD_genes' were of paramount importance, emphasizing the well-documented higher prevalence of ASD in males and the genetic component in ASD development. Additionally, features related to age at diagnosis and ethnicity held significance, further emphasizing the importance of early screening and the potential influence of cultural and genetic factors in ASD diagnosis.

In conclusion, this analysis not only showcases the capabilities of machine learning in early ASD detection but also underscores the Multi-layer Perceptron (MLP) as the most promising model for this purpose. The practical implications of this research are substantial, potentially leading to more accurate and efficient ASD diagnosis. The ability to identify autism spectrum disorders at an early stage can facilitate timely interventions and support, ultimately improving the lives of individuals with autism and their families. Future research should focus on further enhancing the performance of machine learning models and their integration into clinical practice, contributing to the advancement of autism research and betterment of the lives of individuals with ASD.

Declaration of No Conflict of Interest

The authors declare that there is no conflict of interest associated with this work. They have no financial or personal relationships that could influence or bias the objectivity of this study. This research is conducted with scientific integrity and rigor, with the sole purpose of contributing to the advancement of knowledge regarding the detection of Autism Spectrum Disorder (ASD) through machine learning techniques. Any sources of funding, institutional support, or external collaborations, when applicable, have been disclosed transparently. Honesty and scientific objectivity are fundamental principles of this research.

Feel free to modify this statement as needed to align with the requirements of your academic work or the journal where you plan to submit your study. It's important to ensure that the statement accurately reflects the absence of a conflict of interest in your research.

Data Statement

The underlying data for this research is available upon request and can be obtained from the original sources or the specified repositories. We are committed to transparency and the replicability of our research and encourage the validation of the data used in this study. For additional information or to request access to the data, please contact the corresponding author. We utilized a dataset developed by Dr. Fadi Fayez Thabtah, as described in "A new machine learning model based on induction of rules for autism detection" (Thabtah & Peebles, 2020).

References

1. Leblanc, E., Washington, P., Varma, M., Dunlap, K., & Shic, F. (2020). "Feature replacement methods enable reliable home video analysis for machine learning detection of autism." Scientific Reports, 10(1), 14753.

2. Hasan, S. M. M., Uddin, M. P., Mamun, M. A., & Sharif, M. I. (2022). "A machine learning framework for early-stage detection of autism spectrum disorders." IEEE Transactions on Neural Networks and Learning Systems, 33(2), 627-640.

3. Thabtah, F., & Peebles, D. (2020). "A new machine learning model based on induction of rules for autism detection." Health Informatics Journal, 26(2), 1025-1037.

4. Moon, S., Hwang, J., Kana, R., Torous, J., & Kim, J. W. (2019). "Accuracy of machine learning algorithms for the diagnosis of autism spectrum disorder: Systematic review and meta-analysis of brain magnetic resonance imaging studies." JMIR Mental Health, 6(8), e12187.

5. Raj, S., & Masood, S. (2020). "Analysis and detection of autism spectrum disorder using machine learning techniques." Procedia Computer Science, 167, 656-661.

6. Stahl, D., Pickles, A., Elsabbagh, M., Johnson, M. H., Happe, F., & Plomin, R. (2012). "Novel machine learning methods for ERP analysis: a validation from research on infants at risk for autism." Developmental Science, 15(5), 678-690.

7. Hyde, K. K., Novack, M. N., LaHaye, N., & Tager-Flusberg, H. (2019). "Applications of supervised machine learning in autism spectrum disorder research: a review." Review Journal of Autism and Developmental Disorders, 6(1), 1-11.

8. Parikh, M. N., Li, H., & He, L. (2019). "Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data." Frontiers in Computational Neuroscience, 13, 107.

9. Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliewer, N., Freitag, C. M., ... & Bölte, S. (2020). "Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning." Scientific Reports, 10(1), 19274.

10. Karim, S., Akter, N., Patwary, M. J. A., & Hossain, M. M. (2021). "A review on predicting autism spectrum disorder (ASD) meltdown using machine learning algorithms." In 2021 5th International Conference on Electrical and Computer Engineering (ICECE) (pp. 1-4). IEEE.

11. Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). "Sex/gender differences and autism: Setting the scene for future research." Journal of the American Academy of Child & Adolescent Psychiatry, 54(1), 11-24.

12. Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C. M., Larsson, H., & Reichenberg, A. (2014). "The familial risk of autism." JAMA, 311(17), 1770-1777.

13. Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., ... & Sigman, M. (2015). "Early identification of autism spectrum disorder: Recommendations for practice and research." Pediatrics, 136(Supplement 1), S10-S40.

14. Mazumdar, S., Winter, A., Liu, K. Y., Cooper, A. C., Maron, L., & Stanley, C. (2013). "Diagnosis of autism spectrum disorder by developmental behavioral pediatricians in academic centers: A DBPNet study." Journal of Developmental & Behavioral Pediatrics, 34(8), 579-585.