Use of Artificial Intelligence in the Diagnostics of Autism Spectrum Disorder

Gabriele Mustika Kresnia, Arli Aditya Parikesit

Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta, Indonesia

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurologic development disorder; it is listed in the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V). Early diagnosis is critical in improving the life quality of individuals affected by ASD. Several studies show that Artificial Intelligence can be used in the diagnosis of ASD through biological means such as observing patient EEG data and surveying their genome. Articles were searched in the PUBMED database, ScienceDirect, and Springer Link between 2019 - 2020. Four papers were selected for review. The papers devised models that can accurately predict ASD in affected individuals, though some are based on old data and/or require testing on larger datasets to determine accuracy. As ASD diagnosis usually cannot be achieved before the individual shows symptoms, AI has the potential to improve ASD diagnosis in affected individuals. Further study to confirm the models and test on larger, more recent datasets would be required to develop more accurate models and achieve even better results.

Keywords: Artificial intelligence, autism spectrum disorder, EEG, social communication, genome

ABSTRAK

Autism spectrum disorder (ASD) merupakan salah satu gangguan perkembangan saraf yang tercantum pada Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V). Diagnosis dini sangat penting untuk meningkatkan kualitas hidup individu ASD. Beberapa penelitian menunjukkan bahwa kecerdasan buatan dapat digunakan untuk diagnosis ASD melalui metode berbasis biologis seperti mengamati data EEG pasien dan mensurvei genomnya. *Review* ini berbasis pencarian data antara tahun 2019 – 2020 di *database* PUBMED, ScienceDirect, dan Springer Link. Empat makalah kunci dipilih untuk ditinjau. Makalah-makalah tersebut mampu merancang model yang dapat memprediksi ASD secara akurat, meskipun beberapa aspek implementasinya didasarkan pada data usang dan/atau memerlukan pengujian pada kumpulan data yang lebih besar untuk menentukan akurasi. Mengingat diagnosis ASD biasanya tidak dapat dilakukan sebelum individu menunjukkan gejala, kecerdasan buatan berpotensi meningkatkan ketepatan diagnosis ASD. Masih diperlukan studi lanjutan untuk mengkonfirmasi model dan pengujian pada kumpulan data yang lebih besar dan lebih baru untuk mengembangkan model yang memiliki presisi lebih baik dan hasil lebih akurat.
Gabriele Mustika Kresnia, Arli Aditya Parikesit. Penggunaan *Artificial Intelligence* untuk Diagnosis *Autism Spectrum Disorder.*

Kata kunci: *Autism spectrum disorder*, EEG, *genome*, kecerdasan buatan, komunikasi sosial

INTRODUCTION

Autism Spectrum Disorder (ASD) is a disorder listed under the Diagnostic and Statistical Manual of Mental Disorders, 5th edition.[1,2] It is classified as a neurodevelopmental disorder, and according to the DSM-V, characterized notably by persistent deficits in communication skills and restrictive and repetitive patterns of behavior, interest, and activities. Autistic individuals also lack the social skills to develop, maintain and understand relationships, although symptoms may change or be invisible due to "masking"--a compensatory mechanism developed by the individual. There are also varying levels of ASD as defined by the DSM-V, and symptoms vary widely across autistic individuals. For this reason, diagnosing ASD is often difficult. There are also diagnostic biases; females are less likely than males to be diagnosed with ASD, and females are on average older at diagnosis than males, despite more "severe" symptoms.[1]

While certain members of the autism community are pushing for self-advocacy and against a cure as they view it to be a part of their life, many express a desire for intervention to help cope with some daily life challenges, such as sensory sensitivities or GI problems in certain autistic individuals.[2] An early diagnosis would benefit these individuals greatly, as it may help open the path for early intervention. Research has turned to investigating biomarkers for ASD instead of relying on the more behavioral symptoms, using models developed to predict the presence of ASD in individuals with specific features in their genome. Machine learning is an excellent tool in facilitating autism diagnosis and may be especially useful in navigating the varying apparent symptoms. The ultimate goal of this systematic review is to observe the use of artificial intelligence (AI) in diagnosing ASD, its limitations, and prospects of AI in the diagnostics of ASD.

MATERIAL and METHODS

Articles in the PUBMED database between 2019 - 2022 were searched. The search keyword is [("autism spectrum disorder" OR "autism") AND "artificial intelligence" AND "diagnosis"]. The flowchart depicting the details review method can be observed in **Figure.**

The included papers must be written in English, look into ASD diagnosis based on biological indicators, focus only on ASD diagnosis

Alamat Korespondensi *email: arli.parikesit@i3l.ac.id*

without looking into any other disorders and/or diseases, and must implement the use of artificial intelligence in the diagnostic process. In addition, the article must focus only on diagnostics and not intervention and/or cure for ASD. Based on the aforementioned inclusion criteria, four papers were chosen for review from the three databases.[3–6]

RESULTS

Based on the initial search in PUBMED, 79 papers were found in the database. Those papers were screened with the existing inclusion criteria (**Figure**), and four papers were eventually selected.

In the four papers selected for review, different algorithms used in the diagnostics of ASD were identified. The used datasets for the studies, the accuracy and sensitivities of the algorithms, and the application and short description of the subjects are summarized in **Table**.

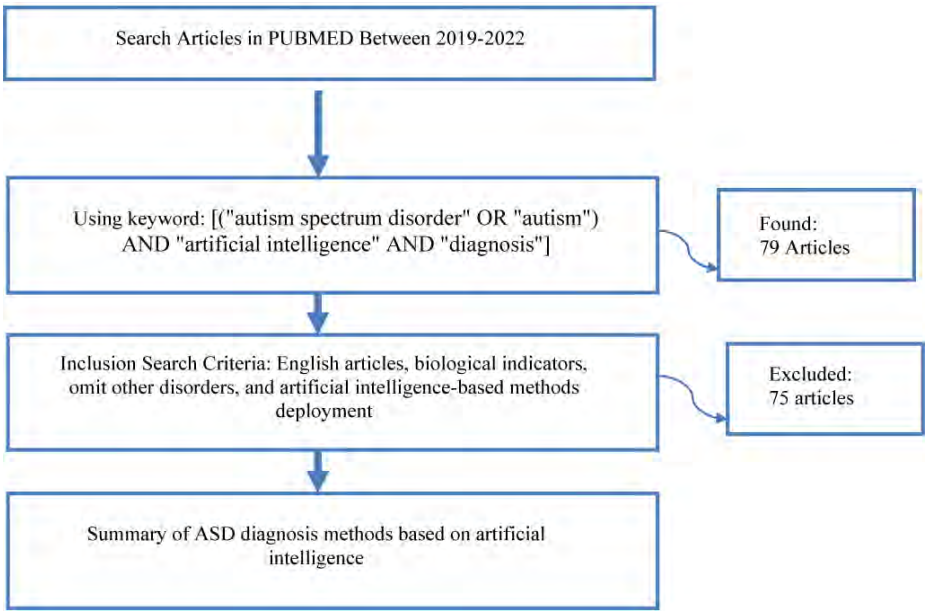

Figure. Flowchart of the review method

DISCUSSION

Machine Learning

Machine learning algorithms refer to a series of algorithms that allow computers to create a model that can improve analyzing patterns in data, learn, and adapt with the increase in available data added to the model.[7] There are various algorithms in this branch of artificial intelligence. Gök and his team used long non-coding RNA (lncRNA) gene data from The BrainSpan Atlas of the Developing Human Brain developmental transcriptome dataset to diagnose ASD, applying the Haar wavelet transform method to find the most contradistinguished features.[6] Class-balancing algorithms and then discretization is performed on the data, and then the Bayesian network classifier is used to recognize ASD risk genes versus normal genes. A Bayes network is a directed, acyclic graph whose vertices, V ={V1, V2,…, Vn}, analogous to random variables, and the directed edges, E, represent conditional probabilities between the random variables together with a set of local probability distributions P.[8] The model proposed by Gök achieved an excellent accuracy of 83.9% in classifying the lncRNA gene data of individuals who have ASD versus normal individuals.[6-7]

Support Vector Machines

Abdolzadegan and his team employed the use of Support Vector Machines (SVM).[3] A support vector machine is a computer algorithm that can assign labels to objects through learning by example.[9] It is built based on statistical learning theory based on the principle of structural risk minimization.[10] EEG signals were obtained from the children; several features were extracted, clustered with artifacts removed, and then selected to regulate the most powerful features in early detection.[3] These features are then classified through SVM, chosen for its ability to affect the problem of nonlinear distribution of the training data. The paper also cites that SVM

Table. Review summary of the scientific articles

Ref.	Algorithm	Datasets	Application	Accuracy	Sensitivity	Description
(3)	SVM, KNN	EEG data from 34 children 3-12 years old with ASD and 11 control children in the same age range	Early ASD diagnosis using EEG data from children	SVM: 90.57% KNN: 72.77%	SVM: 99.91% KNN: 91.96%	Children above 5 were used to train the models, and they tested children below 5 to support diagnosis ASD
(4)	DL, ML (5 different algorithms)	Cytosine ('CpG') methylation was consistent genome-wide in 14 autism cases and 10 controls	Investigating the epigenetic footing of ASD and recognition of early biomarkers	DL : AUC 0.958–1.00 for detection of ASD ML : 4 out of 5 algorithms had an AUC of≧0.95	DL: 97.5% ML: >85% across all algorithms	Epigenetic dysregulation was recognized in several paramount suitor including previously linked to autism development
(5)	KNN	EEG data from 60 children ranging in ages 6-8, 30 normal cases and 30 ASD cases	Early ASD diagnosis using EEG data from children	81.67%	N/A	EEG signals of the brain's C3 and C4 channels and the topological features of its network used to distinguish ASD from normal cases at an early age
(6)	ML	366 annotated autism genes (positive data) stretching from 2128 disease genes with their 524 analogous advancing time points of 26 brain structures whose age range is between 8 weeks and 40 years.	Classification of autism risk gene	83.9%	90.2%	Prediction of autism risk gene using long-coding RNA (lncRNA) to see whether an lncRNA gene causes disease or not

KNN: K-Nearest Neighbor, SVM: Support Vector Machines, DL: Deep Learning, ML: Machine Learning, N/A: not mentioned by the authors

is chosen to reduce operational risks. Three datasets were used: the baseline EEG data (Baseline), taken when the children were sitting in a comfortable, relaxed position with their eyes open for 2 minutes, EEG data while the children were watching a 5-minute cartoon (with voice dataset), and EEG data while the children were watching the same cartoon, but with no voice (without voice dataset). SVM had a 90.57% mean accuracy rate across all three datasets in identifying normal children versus children with ASD; 15 most effective features of the EEG signal were selected; the general accuracy of the algorithm was increased with the addition of features.

K-Nearest Neighbor

Two papers in this review used the K-Nearest Neighbor (KNN) algorithm.[3,5] KNN was chosen for its simplicity and did not need a prior hypothesis to classify data. KNN is a decision rule widely used in pattern classification.[11] It classifies patterns based on a similarity measure. This classifying algorithm assigns observations previously not labeled to the class of examples that are most similar to the observation.[11] In this particular algorithm, there are two parameters: the Euclidean distance between the observation and the examples to be compared with, and the parameter k (hence the name K-Nearest Neighbor) that decides how many neighbors are chosen for the algorithm. The researcher chooses this parameter, and an appropriate value must be decided, as it has a convincing impact on the performance of the KNN algorithm. Some researchers suggest that k be set to the square root of the number of observations done, although the k value can be tuned to strike a good balance between overfitting and underfitting the pattern. Using a large k-value may reduce the variance caused by random error, however, it may also cause small but essential patterns to be overlooked. Also impacting the algorithm's accuracy is the classifier chosen to identify the observations.

KNN had an accuracy rate of 72.77% from all three datasets used in the paper (Baseline, with voice, and without voice).[3] The paper did not mention the value of the k parameter used for their KNN algorithm; however, it did specify that the classifiers used for the KNN algorithm were similar to that of the SVM algorithm in the previous section. The reported sensitivity of 91.96% for the KNN classifier meant that this proposed method, according to the study, is advisable for the diagnosis of individuals. EEG data were also deployed, much like the previously mentioned study, using k = 4 and the Euclidean distance to classify the EEG features.[5] It has an accuracy level of 81.67%. This method employs the C3 channel of the brain, the topological features of its network (AD), and related features such as RADACC and RADMPL in the diagnosis of ASD. It was found that in the ASD graph samples, the topological AD feature is reduced.[5] They utilized the difference between the topological AD feature and its relationship with ACC and MPL values in the complex network of ASD and NC graphs as a criterion for the detection of ASD in the EEG samples.

Deep Learning

Deep learning is a sub-category of machine learning that contains a differing family of computational methods that consist of many data processing layers used for automated feature extraction and pattern recognition in large datasets.[11] Deep learning lets a computer build complex concepts out of more straightforward ideas; in essence, it uses a hierarchy of concepts formed by layering more difficult concepts on top of the simpler ones that define it as it accumulates knowledge from experiences.[12] It is also elaborated that gathering knowledge from experiences reduces the need for human operators to specify the parameters for the computer[11,12] formally. Through these hierarchical architectures, deep learning machines attempt to learn high-level abstractions in data. This branch of machine learning is applied across many artificial intelligence domains, such as natural language processing, computer vision, speech recognition, computer vision, and various other areas.[13]

The use of deep learning in prediction algorithms to diagnose autism via the DNA methylation status of newborn leukocyte DNA has already been deployed.[4] Based on cytosine ('CpG' or cg') loci (and related genes) that were epigenetically altered, the authors explored the potential epigenetic pathogenesis of autism spectrum disorder. They identified 230 CpG sites associated with 249 distinct genes that were significantly differently methylated between autism cases and the control, with a false discovery rate (FDR) of ≤ 0.05. These CpG sites include 160 that were differently hypomethylated and 70 that were hypermethylated. Among individual cytosine loci (and corresponding genes), the best predictive individual CpG markers (genes) i.e. AUC≧ 0.9 for autism detection were: cg20129082 (LOC100126784; NAV2), cg08590939 (OXCT1), cg11722376 (LOC389033), cg15371711 (MYL9), cg16678169 (ALS2CR4), cg15028160 (C19orf73) and cg01156550 (ASCL2). These criteria were used to distinguish between autism cases and normal cases. Using deep learning, they were able to achieve an AUC of ≧0.95. However, there were some limitations: first, the authors were incapable of performing expression analysis as they used dried blood spots for the method. However, they attempted to mitigate this problem by verifying whether any of the genes previously reported to be differentially expressed in the brains of individuals with autism were also differentially methylated in leukocyte DNA. The other constraint was that blood spots were obtained and archived in a period when an earlier classification system of autism was being used; that is, before the other subtypes of autism in the spectrum, such as Asperger's and Pervasive Developmental Disorder, was classified and brought together under one diagnosis of ASD per the DSM-V. Nevertheless, while the analysis was limited to only a small subset of ASD and a considerably small dataset was used, the authors maintained that they achieved a high accuracy of diagnosis. Thus their method is still viable, although further analyses of the larger number of cases to confirm their findings would be suitable.

Ethical Consideration and Outlook

Ethical considerations on deploying ASD diagnostics should be considered, and informed consent from the patient's guardian should be secured accordingly.[14] Including a massive amount of ASD patient's data for therapeutic and curative measures within the bioinformatics, frameworks would require a multidisciplinary approach among medical doctors, bioinformaticians, computer scientists, bioethicists, and other related fields.[15] As ASD diagnostics is considered a relatively new field, benchmarking for other established-study such as tuberculosis research are suggested.[16]

CONCLUSION

Four articles were reviewed on their use of artificial intelligence in diagnosing ASD based on biological features. Overall, the methods across all papers achieved good accuracy in diagnosing ASD, especially those focusing on early diagnosis in children. While artificial intelligence can be a powerful tool in the future of ASD diagnostics, understanding machine learning, psychology, and physiology is still required to create a sufficiently accurate model, as is typically expected from interdisciplinary research. Continuous technological advances and the availability of more data shows that machine learning has a potential future in aiding ASD diagnostics and therapeutical options.

Acknowledgement

The authors would like to thank the Department of Research and Community Service (LPPM) of Indonesia International Institute for Life Science for supporting this study. The authors declare that there is no competing interest.

REFERENCES

1. Murray AL, Allison C, Smith PL, Baron-Cohen S, Booth T, Auyeung B. Investigating diagnostic bias in autism spectrum conditions: An item response theory analysis of sex bias in the AQ-10. Autism Res [Internet]. 2017 [cited 2022 Jan 20];10(5):790–800. Available from: https://pubmed.ncbi.nlm.nih.gov/27891820/
2. Bagatell N. From cure to community: Transforming notions of autism. Ethos [Internet]. 2010 [cited 2022 Jan 20];38(1):33–55. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1548-1352.2009.01080.x
3. Abdolzadegan D, Moattar MH, Ghoshuni M. A robust method for early diagnosis of autism spectrum disorder from EEG signals based on feature selection and DBSCAN method. Biocybern Biomed Eng. 2020;40(1):482–93.
4. Bahado-Singh RO, Vishweswaraiah S, Aydas B, Mishra NK, Yilmaz A, Guda C, et al. Artificial intelligence analysis of newborn leucocyte epigenomic markers for the prediction of autism. Brain Res. 2019;1724:146457.
5. Bajestani GS, Behrooz M, Khani AG, Nouri-Baygi M, Mollaei A. Diagnosis of autism spectrum disorder based on complex network features. Comput Methods Programs Biomed. 2019;177:277–83.
6. Gök M. A novel machine learning model to predict autism spectrum disorders risk gene. Neural Comput Appl 2018 3110 [Internet]. 2018 [cited 2022 Jan 20];31(10):6711–7. Available from: https://link.springer.com/article/10.1007/s00521-018-3502-5
7. Alpaydin E. Introduction to machine learning. [cited 2022 Jan 20];682. Available from: https://mitpress.mit.edu/books/introduction-machine-learning-fourth-edition
8. Kaderali L, Radde N. Inferring gene regulatory networks from expression data. Stud Comput Intell. 2008;94:33–74.
9. Noble WS. What is a support vector machine? Nat Biotechnol 2006 2412 [Internet]. 2006 [cited 2022 Jan 20];24(12):1565–7. Available from: https://www.nature.com/articles/nbt1206-1565
10. Zhang Y. Support vector machine classification algorithm and its application. Commun Comput Inf Sci [Internet]. 2012 [cited 2022 Jan 20];308 CCIS(PART 2):179–86. Available from: https://link.springer.com/chapter/10.1007/978-3-642-34041-3_27
11. Wu Y, Ianakiev K, Govindaraju V. Improved k-nearest neighbor classification. Pattern Recognit. 2002;35(10):2311–8.
12. Bengio Y, Goodfellow I, Courville A. MIT press. Deep learning. 2017;1:1-7
13. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing 2016;187:27–48.
14. Anurogo D, Parikesit AA. Troubled helix – Tinjauan multiperspektif genetika dalam bioetika. Cermin Dunia Kedokt [Internet]. 2021 [cited 2021 Mar 9];48(3):147–53. Available from: http://103.13.36.125/index.php/CDK/article/viewt/1331
15. Parikesit AA. Introductory chapter: The contribution of bioinformatics as blueprint lead for drug design. In: Glavic I, editor. Molecular insight of drug design [Internet]. London: InTech; 2018 [cited 2018 Aug 29]:7. Available from: http://www.intechopen.com/books/molecular-insight-of-drug-design/introductory-chapter-the-contribution-of-bioinformatics-as-blueprint-lead-for-drug-design
16. Satya PGAN, Parikesit AA. Revolution in detecting tuberculosis using radiology with application of deep learning algorithm. Cermin Dunia Kedokt [Internet]. 2021;48(4):261–3. Available from: http://www.cdkjournal.com/index.php/CDK/issue/view/98