Special Issue Published in International Journal of Trend in Research and Development (IJTRD),
ISSN: 2394-9333, www.ijtrd.com

Performance Analysis of Supervised Approaches for Autism Spectrum Disorder Detection

[1]Anurekha G and [2]Geetha P,
[1]Research Scholar, [2]Assistant Professor,
Department of Information Science and technology, CEG campus, Anna University, India

Abstract**:** Autism is an impairment of development in the central nervous system. It affects social coordination, emotions and motor activity of an individual. Many research works on autism detection have emerged in the recent times. The shortcomings of those works are, only limited amount of data is employed and an independent omics data is analyzed for prediction. Depth knowledge associated with the disorder has to be found for early detection. Multi-omics data analysis can provide more insight into the distinctiveness of the disorder. In the current existence of supervised machine learning algorithms have a greater notion in the field of medical sciences. In the proposed system, a novel technique which integrates genomics and proteomics data to classify the candidate gene from non-autistic gene using various machine learning algorithms Adaboost, Random Forest, J48 and KNN. The proposed work is tested on the standard dataset (NCBI and SFARI) and their performances are analyzed based on classification accuracy.

Keywords: *Autism, Impairment, Nervous System, Multi-Omics, Machine Learning.*

I. INTRODUCTION

Autism is an impairment of development in the central nervous system [12]. Children affected by autism will have less social coordination, swinging emotions, impaired motor activity and repeated actions. Cause of Autism is still a mystery. The various factor that causes autism is genetic, which is predominant and environmental to which parents exposed. Researchers have found that prenatal environment and stress also play a vital role in autism occurrence. Children with autism can make a remarkable improvement when the intervention is early. Parents do not have proper awareness of the disorder and ignore the early signs that their children exhibit. Hence it is very much necessary for the early prediction of the disorder. For the early detection of autism the various screening tool used are Autism Diagnosis Interview-Revised (ADI-R), Autism Diagnostic Observation Schedule-Generic (ADOS-G), Childhood Autism Rating Scale (CARS), Diagnostic Instrument for Social and Communication Disorders (DISCO), Gilliam Autism Rating Scale (GARS) and Modified Checklist Autism in Toddlers Revised (M-CHAT-R). These tools require the parent or the caretaker intervention and could analyze only the behavioural of the children but in-depth knowledge of study has to be done for the paediatricians to diagnose the disorder. Gene-based study and MRI image related studies are carried out for complete knowledge of the occurrence of the disorder. Various researchers have been indulged using the gene expression which provided better results. Considering multi-omics data may provide more insight into the complex disorder in the aspect of early detection and diagnosis. Multi-omics data are the collection of various omics like genomics, proteomics, transcriptomics, metabolomics, and epigenomics. Technology advances have driven to the effective and efficient inclusion of multi-omics data in the field of analysis of complex diseases. The various types of omics data have its own characteristics and they vary from the regular or normal pattern is associated with the disorder. However, analysis of single omics data will not have many dependencies indicating the reactive processes.

II. RELATED WORKS

Mary Randolph Gips [12] has proposed the biological aspects of Autism spectrum disorder. The biological factors of the disorder are surveyed in detail. The various causes of occurrences of the Autism disorder are discussed. DP Wall et.al., [11] proposed Alternating Decision Tree classifier to classify autism with non-autistic data based on questionaries. Though accuracy is 99 %, only limited amount of data was utilized for classification. Chin Lee Koo et.al., [10] projected various machine learning methods in the application of detecting gene-gene. and gene-environment interaction. Authors have incorporated neural network, support vector machine and random forest for their designated piece of research.

Valerio W.Hu and Yinglei Lai [9] proposed uncorrelated shrunken centroids and support vector machine to classify autistic gene from non-autistic gene based on microarray analysis. The algorithm could classify the severe, mild and moderate form of autism with an accuracy of 93% . Nurul Ridwah Mohd Isa et.al., [7] have proposed a fuzzy expert system that could classify the children with the severity associated. The children with autism are classified as mild, moderate and severe using various fuzzy rules.

Arjun Krishnan et.al., [4] have anticipated a machine learning approach based on a human brain-specific functional gene interaction network to represent a genome-wide prediction of autism-associated genes. Network-based classifier, Adaboost, Shared K nearest neighbor and the Louvain community finding algorithm is used to cluster the network into distinct modules of tightly coupled genes with an accuracy of 83.3% . Yehudit Hasin et.al., [3] have proposed a framework for integration Multi-omics data to address disease prediction. Multi-omics can provide researchers with a greater understanding of the flow of information from the original cause of the disorder. Network model-based approach and simple correlation are used for integration. Micheal Turewicz et.al., [1] have proposed both statistical and machine learning approach in analyzing the proteomics data. The statistical method involves a high throughput t-test whereas in machine learning approach ensemble based feature selection and support vector machine is used for classifying the group of patients as disordered or healthy.

The observations from the literature survey are that various machine learning algorithms is used in autism prediction and classification based on single omic perspective. Hence this can be enhanced by using multi-omics features for improving the performance. The accuracy and absolute insightful knowledge about the disease has to be identified by applying supervised machine learning approaches.

III. SYSTEM FRAMEWORK FOR ASD CLASSIFICATION

The proposed work have Muti-omics integration, Data preprocessing, Multi-omics clasiification and performance analysis process as specified in Fig 1.

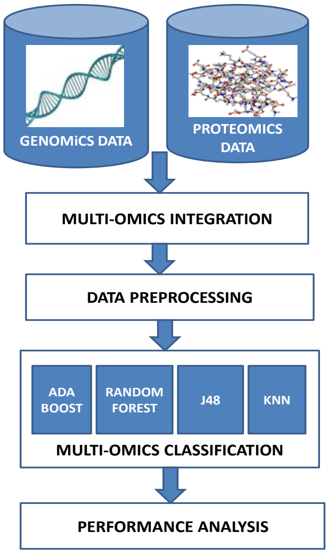

Fig. 1 Architecture Diagram for ASD Classification using Supervised Machine Learning Approaches.

A. Multi-omics Data Integration

Multi-omics based prediction of the complex disorder is a revolutionary approach on comparing with single omics. The drawback in single omics approach restricts to consequential change rather than causative. Autism spectrum disorder is influenced by several genetic and environmental factors hence multi-omics featured data can provide more insight. The intricacy of multi-omics approach is the complexity of data and integration of heterogeneous data. In this work genomic and proteomic data from NCBI database and SFARI database are integrated. The genes are mapped to its corresponding protein and an independent data file is formed for further classification of data

B. Data Preprocessing

The data integration method may be loosely controlled so that the resultant data may have values that are deviated from the generalized ones. So it is necessary to identify and remove the noise i.e. irrelevant data. Extreme value analysis an outlier detection technique is applied to eliminate the extreme values as the gene score and the protein interaction number will not contain extreme values. The outcome of data preprocessing module is noise exempted data.

C. Multi-omics data classification

Supervised learning algorithms AdaBoost, J48, KNN and Random Forest are employed for classification of scattered multi-omics data.

D. Implementation

In this work ncbi gene dataset and gene sfari data set are used, which in turn contain features Gene name, Gene symbol, chromosome, ASD interactions, gene category, gene score, syndromic, gene expressions and no of reports. WEKA 3.9.0 tool is used for implementation.

E. Performance Analysis

The performance of machine learning algorithms Adaboost, Random Forest, J48 and KNN are tabulated in Table 1 with precision, recall and classification accuracy. From the table, it is evidently proved that tree-based J48 algorithm have high accuracy for classifying scattered multi-omics featured proved that tree-based J48 algorithm have high accuracy for classifying scattered multi-omics featured data on comparing with AdaBoost, KNN and Random Forest.

Table 1: Performance Analysis for ASD classification using supervised machine learning approaches.

Algorithms	**Precision %**	**Recall %**	**Accuracy %**
ADABOOST	91.8	92.0	91.6
J48	92.3	92.4	92.4
KNN	90.9	91.4	91.4
RANDOM FOREST	91.3	91.8	91.8

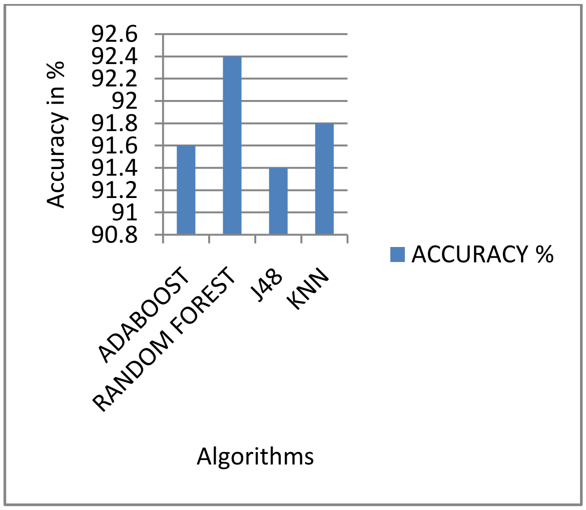

Fig. 2 Bar Chart for ASD Classification Accuracy Using Supervised Machine Learning Approaches.

CONCLUSION AND FUTURE WORK

The classification of Autism spectrum disorder gene from a non-autistic gene can be done effectively by the J48 algorithm. In this work, comparison of functioning and efficiency of the various machine learning algorithm like AdaBoost, Random forest, J48 and K nearest neighbor is carried out and the performance analysis clearly demonstrates the variance in performance of the different algorithm with respect to classification towards appropriate class. In future, the study may be extended by considering other omics data like transcriptomics and metabolomics.

References

[1] Turewicz, Michael, Michael Kohl, Maike Ahrens, Gerhard Mayer, Julian Uszkoreit, Wael Naboulsi, Thilo Bracht et al. "BioInfra. Prot: A Comprehensive Proteomics Workflow Including Data Standardization, Protein Inference, Expression Analysis and Data Publication" *Journal of Biotechnology*, vol. 261, pp. 116-125, Nov 2017.

[2] James Cusack, "Participation and the gradual path to a better life for autistic people", *Sage Publication, vol*. 21, pp.131-132, 2017

[3] Hasin, Yehudit, Marcus Seldin, and Aldons Lusis, "Multi-omics approaches to disease", *Genome biology*,18.83, May 2017

[4] Krishnan, Arjun, et al., "Genome-wide characterization of genetic and functional dysregulation in autism spectrum disorder" ,bioRxiv, June 2016

[5] Watanabe, Takamitsu, Takeshi Otowa, Osamu Abe, Hitoshi Kuwabara, Yuta Aoki, Tatsunobu Natsubori, Hidemasa Takao et al. "Oxytocin receptor gene variations predict neural and behavioral response to oxytocin in autism.", *Social cognitive and affective neuroscience* 12, vol. 3, pp. 496-506, 2016

[6] Isa, Nurul Ridwah Mohd, Marina Yusoff, Noor Elaiza Khalid, Nooritawati Tahir, and Azlina Wati binti Nikmat, "Autism severity level detection using fuzzy expert system" *In Robotics and Manufacturing Automation (ROMA),IEEE International Symposium* on, pp. 218-223,Oct 2015.

[7] Prahbhjot Malhi and Pratibha Singhi, "A retrospective study of toddlers with autism spectrum disorder: Clinical and developmental profile", Annals *of Indian Academy of Neurology*, Nov 2014

[8] Hu, Valerie W., and Yinglei Lai, "Developing a predictive gene classifier for autism spectrum disorders based upon differential gene expression profiles of phenotypic subgroups," *North American journal of medicine & science* 6.3, July 2013

[9] Koo, Ching Lee, et al. "A review for detecting gene-gene interactions using machine learning methods in genetic epidemiology", *Biomed research international*,vol 2013, pp.13, August 2013

[10] D.P.Wall, Kosmicki, J., Deluca, T. F., Harstad, E., & Fusaro, V. A.,"Use of machine learning to shorten observation-based screening and diagnosis of autism", *Translational psychiatry*, 2(4), April 2012

[11] M. Randolph-GIps," Autism: a systems biology disease", In Healthcare Informatics, *Imaging and Systems Biology (HISB)*, First IEEE International Conference, pp. 359-366, IEEE ,July 2011

[12] A. Kannappan A. Tamilarasi E. I. Papageorgiou, "Analyzing the performance of fuzzy cognitive maps with non-linear hebbian learning algorithm in predicting autistic disorder", *Expert Systems with Applications*, vol. 38 no. 3 pp. 1282-1, 2011.

[13] N. Andrea Witwer and Luc Lecavalier," Autism screening tools: An evaluation of the Social Communication Questionnaire and the Developmental Behaviour Checklist–Autism Screening Algorithm" , *Journal of Intellectual & Developmental Disability*, vol.32, pp.179-187, jul 2009

[14] Web Medical. (n.d). Autism Symptoms Causes Treatment and More. Retrieve from http://www.webmd.com/brain/autism/mental-health-autism.