Support Vector Machine Classification of Autism and Typically Developing Children using Electroencephalograph and Recurrence Quantification Analysis Parameters

Thanga Aarthy M¹, Menaka R^{*1}

¹School of Electronics Engineering, Vellore Institute of Technology, Chennai 600127, India.

ABSTRACT

Introduction: Autism Spectrum Disorder (ASD) is a deficit in brain development. It is characterized by a wide spectrum of conditions, such as challenging social skills, repetitive behaviours and difficulty in speech. At present, diagnosing children with ASD at a younger age is challenging. Currently, visual analysis is practised at hospitals which may lead to misinterpretation hence a quantitative analysis is required for early detection and intervention.

Objectives: This research analyzes electroencephalography (EEG) signals of children with ASD and typically developing (TD) and to attain a better classification accuracy in identification.

Materials and Methods: 10 ASD and TD children were considered for the study. Children were made to sit in front of a visual screen and asked to watch a video for ten minutes. During this period EEG signals were acquired, to analyze the difference in characteristics between ASD and TD children.

Results: EEG signals were acquired from 19 channels. They are preprocessed and Recurrence Quantification Analysis (RQA) is applied to the resulting signal. The features extracted are then fed to different types of SVM classifiers. The responsive brain regions were identified and their contribution to RQA features was analyzed.

Conclusion: Responsive channels were identified as Fp1, Fp2, F3, F4, Fz, Cz, O1, O2, T3 and T5. The features extracted from these channels were fed to SVM classifiers out of which quadratic SVM presented an accuracy of 81.8%. In future, larger datasets must be considered for validation and different RQA parameters must be considered for better accuracy.

Key Words: ASD, EEG, RQA, SVM, TD

INTRODUCTION

Neurodevelopmental disorders in children are at an increasing rate every year. Autism Spectrum Disorder (ASD) is a disorder that pertains to the deficit in neuron interactions accompanied by social interaction impairment, repetitive behaviour, communication struggle, and constrained interests.¹The autism study commenced in the year 1960 by Victor Lotter in England. Autism's prevalence rate in 1990 lied among 10 and 16 per 10,000 inhabitants. Later in 2013, different studies predicted prevalence rate of autism was approximately 1% in some regions of the United States and in 2016 ASD was diagnosed as 1 in 36 children.^{2,3} Even in India, the ASD prevalence rate attained a hike. Juneja et al. presented the prevalence rate in India to be in a range of 0.15% to 1.01% (1 in 125 (3-6 years), 1 in

85 (6-9years)).^{4,5,6} In India still there is a lack of diagnosis and understanding of the disorder. Due to this lag, early detection lies in vain. Timely symptom identification and early diagnosis could pave the way to a better scenario for children with ASD. At present various tools are available such as M-CHAT (Modified Children Autism Test), Childhood Autism Rating Scale (CARS), Autism Diagnostic Observation Schedule (ADOS), etc.² Assessment is difficult as the symptoms correlate with other neurodevelopmental disorders and also cannot be identified promptly till a particular age. The quantitative analysis could provide useful findings compared to visual predictions. To recognize a biomarker in ASD children at an early age is a challenge to date. Many techniques such as Functional Magnetic Resonance Imaging (fMRI), Magnetic Resonance Imaging

Corresponding Author:

Menaka R, School of Electronics Engineering, Vellore Institute of Technology, Chennai 600127, India
Email: menaka.r@vit.ac.in

ISSN: 2231-2196 (Print)

ISSN: 0975-5241 (Online)

Received: 22.01.2021

Revised: 19.02.2021

Accepted: 17.04.2021

Published: 20.09.2021

(MRI), Electroencephalography (EEG) and eye-tracking were present for identifying ASD biomarkers.⁷ Dvornek et al. presented research using fMRI functional connectivity measures and classified at an accuracy of 68.5% with Long Short Memory Task (LSTM).⁸ Shihab et al. proposed a method for the identification of a suitable biomarker for ASD using the eye tracking method. It is based on the ocular movement measurement.⁷ Jeste et al. suggests different ways in achieving EEG based biomarkers for neurodevelopmental disorders such as Attention Deficit Hyperactive Disorder (ADHD) and ASD.⁹ EEG is accepted to be the most suitable tool in analyzing stimuli based brain activity or at rest.^{9,10} Bosl et al. presented a method for identifying ASD using infants at the age of 3 months with the ASD diagnosed siblings and determined EEG as a potential tool.¹¹ Heunis et al. acquired scalp EEG signal from 7 ASD and TD (2-6 years) and concluded 92.9% accuracy using Support Vector Machine (SVM).¹² EEG is accepted as an efficient tool in other neurodevelopmental disorders as well. Various methods were used to analyze the EEG signal and its efficiency in discrimination. Oweis et al. extracted Hilbert Huang Transform features from EEG signals of seizure affected people and classified them at an accuracy of 94% with SVM.¹³ Saini et al. presented a method for lie detection using features extracted from EEG signals. Features extracted from EEG signals are time, time-frequency, frequency and statistical and classified using SVM.¹⁴ With the advancements in healthcare, EEG has become a highly potential tool and to attain prompt information proper features must be selected.^{15,16}

EEG signals exhibit nonlinearity. To analyze the nonlinear signal, phase space trajectory representation is important. This representation discloses the core information of any complex system. Recurrence plots (RP) presents valuable insights into the dynamic systems. Visual interpretation of RP could not be accurate and changes with each perception hence various complex measures were derived from RP known as Recurrence Quantification Analysis (RQA).^{17,18} In this research RQA is considered for the extraction of features from the acquired EEG signal. Tripathy et al. proposed a technique for sleep stage classification with the acquired sleep EEG signal. Extracted RQA and dispersion entropy features were classified and attained an accuracy of 85.51%.¹⁹ Ngamga et al. researched with invasive EEG recordings of 5 epilepsy subjects. RQA Features were extracted and a biomarker was identified.²⁰ Heunis et al. extracted RQA features from resting-state EEG signals acquired from children with ASD and TD and classified them at an accuracy of 92.9% using a support vector machine.²¹ RQA is concluded to be a suitable tool for chaotic nonlinear signals like EEG. The parameters which play a crucial role are embedding dimension, distance metric and neighbourhood selection. The proper selection of these pa-

rameters could result in good discrimination between ASD and TD.

The ASD and TD children were presented with the video stimuli. After extraction of features with RQA, the resulting features have to be fed to a classifier. Further to classify ASD and TD, different SVM classifiers were considered.

MATERIALS AND METHODS

EEG data were acquired from 10 (8M, 2F) typically developing children and 10 (7M, 3F) children with autism in the age group of 3 – 7 years. In this study, EEG records were collected from an acquisition system: Natus Nihon Ohden MEB9000 version 05-81, with a sensitivity of 7 microvolts. The data acquisition was performed using International 10-20 electrode systems. The EEG signals were recorded from 22 channels with a sampling frequency of 500Hz and filtered with a low pass filter and high pass filter at a frequency range of (0.53-70Hz). These electrodes are Fp1(channel 1), F7(channel 2), T3(channel 3), T5(channel 4), F3(channel 5), C3(channel 6), P3(channel 7), O1(channel 8), Fp2(channel 9), F8(channel 10), T4(channel 11), T6(channel 12), F4(channel 13), C4(channel 14), P4(channel 15), O2(channel 16), Fz(channel 17), Cz(channel 18), Pz(channel 19)..

A consent form was acquired from the participants. The subjects were made to sit in a quiet room to watch an animated cartoon video for a period of 10 minutes. The distance between the subject's eye and the 32" monitor was 55cm depending upon the height. The Ag/AgCl electrodes were then positioned on the scalp using a conductive gel and tapes. The workflow of the research is as shown in Figure 1.

Figure 1: Block Diagram of the proposed research.

The eye blink signal was observed for 10 seconds with the eye open and eye close event. After thresholding, the resultant signal is further considered for analysis EEG is recorded simultaneously. The underlying neuronal activity can be analyzed with the recorded EEG signal.

FEATURE EXTRACTION

Recurrent Quantification Analysis

Eckmann et al. presented recurrence plots (RP) for finding out the dynamic characteristics of any system.²² RP is a representation of graphical analysis. The RP denotes specific state receptivity in phase space.¹⁷ Originally this graphical

analysis was practised to deliver better knowledge about the dynamic system with higher dimensions as it was difficult for visual interpretation and analysis in those phase space trajectories.^{22,23} The RP is represented below

$$R_{ij} = \Theta (\epsilon_i - ||x_i - x_j||) ; x_i \in R^m, i, j = 1, \dots, N \quad (2)$$

Which N denotes sum of considered states, Θ - Heaviside function, $|| \cdot ||$ - Norm, ϵ_i - Threshold distance, x_i - each state representation.

The figure 2 presents the RP of TD and ASD EEG signal.

Figure 2: (a) EEG plot of ASD signal acquired during video stimuli, (b) EEG plot of TD signal acquired during video stimuli, (c) & (d) represents recurrence plot of ASD and TD signal during audio/video stimuli

RP advantages are it can perform better on noisy data, nonlinear data and also short datasets. The condition $m > 2d + 1$ must be satisfied to prove the certainty, where d denotes the dimension of the attractor and m denotes the embedding dimension. Embedding dimension (m) and time delay (τ) must be selected precisely to achieve a significant interpretation.²³ In this study embedding dimension and time delay is fixed as 3 and 1 using the false nearest neighbour method (FNN).

After setting up the embedding dimension and time delay, a suitable neighbourhood method and distance metric must be selected. In this study, a fixed amount of nearest neighbours (FAN) Euclidean is selected for feature extraction. FAN is widely used in nonlinear data analysis and also for the comparison of two different systems. The Euclidean metric is used because it is less prone to noise. It is mathematically represented as

$$ED = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

where (x_1, y_1) and (x_2, y_2) are points in two dimensional space.

In this research, 10 RQA features were measured, are recurrence rate (RR), Entropy (ENT), Determinism (DET), Average Diagonal Length (L), Laminarity (LAM), Trapping Time (TT), Vertical entropy (ENTR_vert), Longest vertical line (V_{max}), longest diagonal line (L_{max}) and divergence which were mathematically represented in equations (4-13) below.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j} \quad (4)$$

$R_{i,j}$ - Recurrence plot, N - the size of the recurrence plot

$$DET = \frac{\sum_{l=d_{min}}^N l H_D(l)}{\sum_{i,j=1}^N R_{i,j}} \quad (5)$$

$H_D(l)$ - Histogram of the length of the diagonal line.

$$ENT = \sum_{l=d_{min}}^N p(l) \ln p(l) \text{ where,} \quad (6)$$

$$P(l) = \frac{H_D(l)}{\sum_{l=d_{min}}^N H_D(l)}$$

$P(l)$ - Probability of the diagonal lines

$$L = \frac{\sum_{l=l_{min}}^N l P(l)}{\sum_{l=l_{min}}^N P(l)} \quad (7)$$

$$LAM = \frac{\sum_{l=v_{min}}^N l H_v(l)}{\sum_{l=v_{min}}^N H_v(l)} \quad (8)$$

$H_v(l)$ - Histogram of vertical line structures

$$TT = \frac{\sum_{l=v_{min}}^N v P(v)}{\sum_{l=v_{min}}^N P(v)} \quad (9)$$

$$ENT_vert = \sum_{l=v_{min}}^N p(v) \ln p(v) \quad (10)$$

$$L_{max} = \max(\{l_i; i=1, \dots, N_l\}) \quad (11)$$

$$V_{max} = \max(\{V_i; i=1, \dots, N_v\}) \quad (12)$$

$$V_{max} = \max(\{V_i; i=1, \dots, N_v\}) \quad (13)$$

The parameter combination which showed higher accuracy were shown below. Each region contribution for the considered ten RQA features were shown in figure 3.

Figure 3: Brain Region-wise percentage contribution for RQA.

RR – Recurrence rate, DET – Determinism, ENTR – Entropy, ADL – Average Diagonal Line, LD – Longest Diagonal Line, DIV- Divergence, LAM – Laminarity, VENTR – Vertical Entropy, TT – Trapping Time, LVL – Longest Vertical Length

The channels which showed better response for the RQA features with the neighbourhood selection of FAN and distance metric Euclidean were shown in figure 3. The discriminating channels were Fp1, Fp2, F4, F8, O1, T3, T5, Fz and Cz. This parameter combination has shown differences in LD, ADL, DIV, VENTR and LVL in different regions. The recurrence features were fed to a classifier to classify children with ASD and TD.

CLASSIFICATION

Support Vector Machine:

Machine learning is an artificial intelligence application that enhances the computer able to access the data by itself and learn from the previously trained data. There are many machine learning algorithms such as K-means, Naïve Bayes, decision trees, SVM, etc. SVM classifier is known to work well with EEG signals. The main objective of SVM is to identify a suitable hyperplane in N-dimensional space. Hyperplanes act as boundaries that decide and classify the data points. Hyperplanes are formed with the help of support vectors. Support vectors are data points that are present nearer to the hyperplane. The position and orientation of the hyperplane depend on the support vectors. In this study, the extracted features were tested with different decision boundaries such as linear, cubic, quadratic and Gaussian. In

The input data fed to the classifier is $x_i = (x_1 \dots x_n)$. ϕ represent the map mapping to the N-Dimensional space y , to represent the label mathematically as $\{(x_1, y_1), \dots (x_m, y_m)\}$. The SVM identifies a hyperplane (w, b) shown in equation 14

$$\gamma = \min_i y_i \{ \langle w, \phi(x_i) \rangle - b \} \quad (14)$$

Some data points might not be linearly separable hence for such datasets special cases of hyperplanes are needed. A quadratic function is needed for such nonlinear datasets which are inseparable. As a thought, various shapes of hyperplanes could fit data better in certain instances. In this research, various hyperplanes have been tried for SVM classifiers for the acquired input dataset. The results of the different SVM classifiers were discussed in the next section.

RESULTS AND DISCUSSIONS

The SVM classifier for classification was developed in MATLAB environment ((MATLAB and Classification Learner R2019a, The MathWorks, Inc., Natick, MA, USA) for this study. The processing was performed on a laptop with a 1.90GHz quad-core CPU, 8GB of memory RAM and a 64-bit version of Windows. In this study, initially, the EEG signals were acquired at the selected nine electrodes in the presence of video stimuli. The acquired signals were pre-processed. The features from RQA were extracted from the preprocessed EEG signal using neighbourhood selection as FAN and Euclidean distance metric. The extracted features were fed to an SVM classifier with 5 fold cross-validation and principal component analysis. Their classification accuracy was presented in Table 3.

Table 3: Classification Accuracy for distinguishing ASD and TD using SVM classifier

S. No.	SVM Classifier Type	Classification Accuracy of distinguishing ASD and TD using SVM
1	Linear SVM	73.6%
2	Cubic SVM	79.2%
3	Quadratic SVM	81.8%
4	Fine Gaussian SVM	80.3%
5	Medium Gaussian SVM	73.6%
6	Coarse Gaussian SVM	59.7%

Figure 4: Classification accuracy using Quadratic SVM.

CONCLUSION

It is observed from table 3 that quadratic SVM achieved an accuracy of 81.8% compared to the other SVM classifier which is shown in figure 4. It is seen that frontal, temporal, occipital and prefrontal regions were responding well for the presented video stimuli RQA, higher classification accuracy can be achieved by using distance metric cosine, neighbourhood selection FAN with audio-video stimuli. Children with ASD show better discrimination from TD using audio-video stimuli. Each brain lobe is concerned with a specific purpose. Frontal and prefrontal lobe is responsible for its role in motor functions, problem-solving capability and memory. Temporal lobe function is for hearing, identifying emotion and long term memory storage. Occipital plays a role in visual processing. In this study, in the presence of video stimuli, RQA features with an SVM classifier provide a good result in distinguishing ASD and TD in the above-mentioned lobes.

In future, validation must be carried out with the larger datasets. Results have to be compared with various parameter combinations of RQA since the best parameter combination could provide better accuracy. Autism is a wide range of spectrum disorders so in the future lobe, the response must be analyzed in-depth which can lead to individual training and lead them to a better life.

ACKNOWLEDGEMENT

Funding: EEG datasets were acquired from Sri Ramachandra Medical College and Research Institute, India. Department of Science and Technology, India (Ref no: SEED/TIDE/092/2016) support this study and the authors thank them for their financial assistance.

Ethical approval: Human participants who participated in the study were in accord with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments, or comparable ethical standards.

Informed consent: Consent forms were acquired from all participants involved in the study.

Conflict of interest: The authors declare that they have no conflict of interest concerning the contents of this article.

Author's Contribution:

Thanga Aarthy M: This author contributed coding and results from observation for the research

Menaka R: This author contributed proof of concept for the research.

REFERENCES

1. Whitney G, Lauren BS, Charly N, Amy MW. Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation. *J Child Psychol Psychiatry*. 2013 May;54(5):582-90. Available from: <https://pubmed.ncbi.nlm.nih.gov/23078094> DOI: 10.1111/jcpp.12008.
2. Helena B, Cristiane SP, Daniela B, Deborah R, Fabio S, Joana P, et al. Autism spectrum disorders: an overview on diagnosis and treatment. *Braz J Psychiatry*. 2013;35(1):62-72. Available from: <https://pubmed.ncbi.nlm.nih.gov/24142129> DOI: 10.1590/1516-4446-2013-S104.
3. Zablotsky B, Black, LI, Blumberg SJ. The estimated prevalence of children with diagnosed developmental disabilities in the United States, 2014–2016. *NCHS Data Brief*. 2017; 291(1):1-8. Available from: <https://pubmed.ncbi.nlm.nih.gov/29235982/>.
4. Monica J, Smitha S. Autism Spectrum Disorder - An Indian Perspective. *Recent Advances in Autism*. Italy: Smjournals; 2018.
5. Rudra A, Belmonte MK, Soni PK, Banerjee S, Mukerji S, et al. Prevalence of autism spectrum disorder and autistic symptoms in a school-based cohort of children in Kolkata, India. *Autism Res*. 2017 Oct;10(10):1597-1605. Available from: <https://pubmed.ncbi.nlm.nih.gov/28544637/> DOI: 10.1002/aur.1812.
6. Raina SK, Vishav C, Ashok KB, Dinesh K, Seema S, et al. Prevalence of Autism Spectrum Disorder among Rural, Urban, and Tribal Children (1-10 Years of Age). *J Neurosci Rural Pract*. 2017 Jul-Sep; 8(3): 368–374. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5488556/> DOI: 10.4103/jnrp.jnrp_329_16
7. Ammar IS, Faten AD, Ali HK. Data Analysis and Classification of Autism Spectrum Disorder Using Principal Component Analysis. *Adv Bioinformatics*. 2020;2020:1-8. Available from: <https://www.hindawi.com/journals/abi/2020/3407907/> DOI: 10.1155/2020/3407907
8. Nicha CD, Pamela V, Kevin AP, James SD. Identifying Autism from Resting-State fMRI Using Long Short Term Memory Networks. *Mach Learn Med Imaging*. 2017 Sep;10541:362-370. Available from: <https://pubmed.ncbi.nlm.nih.gov/29104967/> DOI: 10.1007/978-3-319-67389-9_42.
9. Jeste SS, Frohlich J, Loo SK. Electrophysiological biomarkers of diagnosis and outcome in neurodevelopmental disorders. *Curr Opin Neurol*. 2015 Apr;28(2):110-6. Available from: <https://pubmed.ncbi.nlm.nih.gov/25710286/> DOI: 10.1097/WCO.0000000000000181.
10. Lonnie Z, Melanie P. Autism spectrum disorder: advances in diagnosis and evaluation. *BMJ*. 2018 May 21;361:k1674. Available from: <https://pubmed.ncbi.nlm.nih.gov/29784657/> DOI: 10.1136/bmj.k1674.
11. Bosl WJ, Helen T-F, Charles AN. EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach. *Sci Rep*. 2018; 8:1-20. Available from: <https://www.nature.com/articles/s41598-018-24318-x.pdf> DOI:10.1038/s41598-018-24318-x
12. Heunis T, Aldrich C, Peters JM, Jeste SS, Sahin M, et al. (2018). Recurrence quantification analysis of resting-state EEG signals in autism spectrum disorder – a systematic methodological exploration of technical and demographic confounders in the search for biomarkers. *BMC Med*. 2018; 16(1):1-12. <https://bmcmecine.biomedcentral.com/articles/10.1186/s12916-018-1086-7> DOI: 10.1186/s12916-018-1086-7
13. Oweis Rj, Abdulhay EW. Seizure classification in EEG signals utilizing Hilbert-Huang transform. *BioMed Eng OnLine*. 2011; 10(38):1-17. Available from: <https://doi.org/10.1186/1475-925X-10-38>

14. Saini N, Bhardwaj S, Agarwal R. Classification of EEG signals using a hybrid combination of features for lie detection. *Neural Comput & Applic*. 2020;32(1):3777–3787. Available from: <https://doi.org/10.1007/s00521-019-04078-z>
15. Miranda IM, Aranha C, Ladeira M. Classification of EEG Signals using Genetic Programming for Feature Construction. *GECCO '19: Proceedings of the Genetic and Evolutionary Computation Conference*. 2019:1275-1283. Available from: <https://doi.org/10.1145/3321707.3321737>.
16. Kannathal N, Acharya UR, Limb CM, Sadasivana PK. (2005) Characterization of EEG—A comparative study. *Comput Methods Programs Biomed*. 2005 Oct;80(1):17-23. Available from: <https://pubmed.ncbi.nlm.nih.gov/16099533/> DOI: 10.1016/j.cmpb.2005.06.005
17. Webber CL, Marwan N. *Recurrence Quantification Analysis Theory and Best Practices*. Switzerland: Springer; 2015.
18. Acharya UR, Vinitha Sree S, Chattopadhyay S, Yu W, Alvin APC. Application of recurrence quantification analysis for the automated identification of epileptic eeg signals. *Int J Neural Syst*. 2011 Jun;21(3):199-211. Available from: <https://pubmed.ncbi.nlm.nih.gov/21656923/> DOI: 10.1142/S0129065711002808.
19. Tripathy RK, Acharya UR. Use of features from RR-time series and EEG signals for automated classification of sleep stages in a deep neural network framework. *BBE*. 2018;38(4):890-902. Available from: <https://doi.org/10.1016/j.bbe.2018.05.005>.
20. Ngamgaa EJ, Bialonskib S, Marwana N, Kurths J, et al. Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic EEG data. *Phys. Lett. A*. 2016;380(16):1419-1425. Available from: <https://doi.org/10.1016/j.physleta.2016.02.024>.
21. Courchesne E, Pierce K, Schumann CM, et al. Mapping early brain development in autism. *Neuron*. 2007 Oct;56(2):399-413. Available from: <https://doi.org/10.1016/j.neuron.2007.10.016>.
22. Eckmann JP, kamphorst SO, Ruelle D. Recurrence Plots of Dynamical Systems. *EPL*. 1987;4(9):973-977.
23. Marwan N. A historical review of recurrence plots. *EPJ*. 2017 Sep;164(1): 3–12. Available from: <https://arxiv.org/pdf/1709.09971.pdf>.