EEG-Based Computer-Aided Diagnosis to detect brain aging Autism Spectrum Disorder Using Wavelet Transform, Entropy, and ML algorithms

Bhupat Baraiya(202311021), Jenil Mansuriya(202311050)
Indian Institute of Information Technology Vadodara – International Campus Diu

***Abstract*—In this piece of work, the authors introduce an analytical system that uses machine learning to analysis of electroencephalogram (EEG) signals in order to detect the aging of the brain in ASD patients compared to normal persons. The objective is to investigate the possibility of distinguishing using EEG signals between the brains of individuals with ASD and Typically Developing (TD) ones. Applying a feature extraction and classification model on EEG. EEG was pre-processed with band-pass, notch and Independent Component Analysis (ICA) to eliminate artifacts,noises. Then the frequency decomposition was done using Daubechies-4 (db4) wavelet, then Shannon entropy and energy features were computed from EEG subbands and classified using a Random Forest model and SVM. This is proof of concept rather then full fledged model.**

***Index Terms*—EEG, Autism Spectrum Disorder, Discrete Wavelet Transform, Shannon Entropy, Computer-Aided Diagnosis.**

I. Introduction

Autism Spectrum Disorder (ASD) is a neuro developmental disorder. physical disorder that is a problem with social interactions. communication and behavioral reflexes. Conventional diagnosis relies heavily on behavioral evaluations such as DSM-5,SRS-22 etc, which require expert's supervision and is time-intensive. As an alternative, EEG-based computer-aided systems have gained attention due to their ability to capture neuro physiological activity and provide objective indicators of brain function and are less time consuming and EEG are less expensive than ECG and MEG
First segmentation of the EEG is performed, segmentation is important because we can't feed whole signal at once to the model, if we do segmentation we can extract more features which cover more detail and can help model to learn from more patterns available. After segmentation we also need to define some overlap which will help model see what it saw earlier and not completely forget the past patterns. After segmentation and overlap is done, Applying Fourier transform directly to such signals is less practically suitable because the nature of EEG signals is rather complex, nonlinear, and nonstationary. The Discrete Wavelet Transform (DWT) provides time-frequency representation, allowing extraction of dynamic brain activity information. DWT dismantles the frequency in various bands of frequency that we will explain further ahead in this paper. After DWT is applied we can apply any method to extract features here we have applied Shannon entropy to extract features from each subbands. Combined with entropy-based features and machine learning, EEG analysis can yield powerful tools for brain waves recognition.

II. Literature review

Sheikhani et al. [3], the datasets were recorded by 21 electrodes with both earlobes chosen as common referential electrodes and extracted from two groups: 10 (9 boys and 1 girl) ASD and 7 (4 boys and 3 girls) non-ASD children. A short time Fourier transform (STFT) technique was used to extract EEG signal features and then applied as an input to nearest neighbors (KNN) classifier to get classification accuracy up to 82.4%. Djemal et al. [4] proposed a Computer-Aided Diagnosis (CAD) system for ASD based on EEG signal analysis using DWT, Shannon entropy, and an ANN (artificial neural network) classifier and got an accuracy of 99.7% Dickinson et al.[5] found that the association between age and Peak Alpha Frequency (PAF) was stronger in the ASD group compared to controls. Dede et al.[6] analyzed 726 variables across a large cohort of 776 participants and cautioned against the use of diagnosis alone for categorizing neurobiological profiles. Shi et al. [7] introduced a Dynamic Residual Block (TDRB) to enhance time-domain feature extraction in their network architecture and the proposed time-frequency synergy network (TFSNet) achieved an accuracy of 98.68% on the University of Sheffield dataset

TABLE I
Distribution of Subjects by Gender and Group

Group	Female	Male
ASD	12	16
TD	15	13

III. Dataset Description

The dataset used in this project is the publicly available dataset called. "Electrophysiological Marks of Brain Aging in Autism. Spectrum Disorder" (Milne et al., 2021). The dataset includes EEG data of 28 ASD and 28 typically developed. The age ranges from 18-68 years. Data acquisition was recorded with the Biosemi ActiveTwo EEG system placed in 10-20 system as shown in the Fig-1, each recording is length 150-second (2.5 minutes) and were taken with eyes-closed in rest

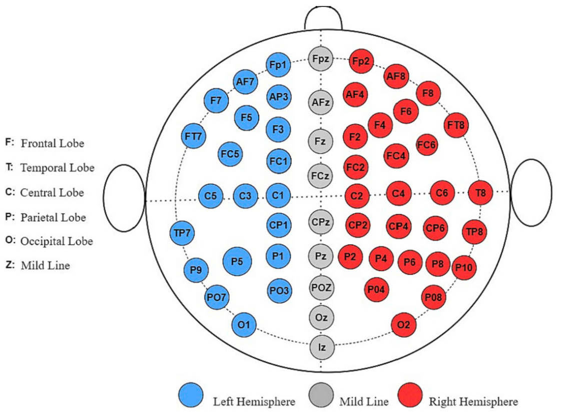

Fig. 1. Sample EEG dataset visualization showing electrode placement and signal pattern.

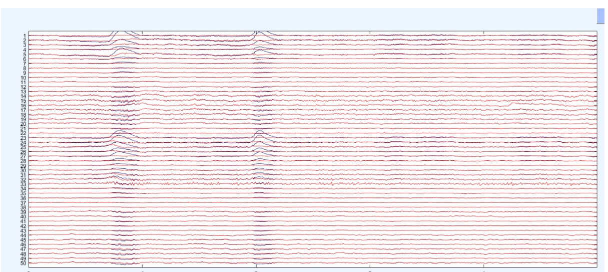

Fig. 2. EEG preprocessing and artifact removal process using ICA and filtering.

paradigm. The dataset does not have consistent number of electrodes for each patient. However the most frequent 56 electrodes which were used to collect data as shown below topography.

IV. Methodology

A. Preprocessing

EEG signals were preprocessed in order to remove artifacts and to denoisify and to enhance signal quality. First it was passed through a band-pass filter(0.5-60Hz), with which we can rebuff the slow. noise of electrodes in scalp which is usually less than. 0.5 Hz and we eliminate the rest of signal of frequency. above 60Hz since most useful features are less than 60Hz and above them are merely useful. The Second filter we used was notch filter at 60Hz to filter the electrical line noise. which exists in all electrical appliances in india at 60Hz and Europe and at 50Hz in USA. After getting the clean data we are now required to eliminate such artifacts as eye blink, muscles to eliminate these artifacts, movement, etc. is used to signal the heart beat we start with ICA decomposition of the signals followed by classification and remove that signal of artifacts to the primary EEG data. This was done in EEGLAB, a MATLAB based toolbox, the ICLabel component automatic component using the ICLabel plugin classification. once this has been done we obtain our clean brain signals that are primarily signals of the brain and no other noise or artifacts,Fig-2 indicates the scaled image of how the main EEG the blue is the removal of the artifacts followed by transformation of signals the lines are those we were taking out.

B. Feature Extraction

Wavelet Decomposition: Here we will use DWT to decompose the signal to analyses both domain of frequency and time, We could have used STFT(Short-Time Fourier Transform) or CWT However, one obvious drawback of SFTF is that it only works on frequency domain, but on complex topic like EEG both time and frequency are important and for the CWT technique is that it requires excessive calculations. Therefore, we used the DWT in the proposed work Here we have used Daubechies-four (db4) mother wavelet, it will help us dismantles the signals into four detail coefficients (D1–D4), and a approximate coefficient (A4). This DWT process is a recursive process as shown in the Fig-3, the wavelet coefficients are corresponding to several EEG subbands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–15 Hz), beta (15–30 Hz), and gamma (30–60 Hz) as shown in Fig-4.

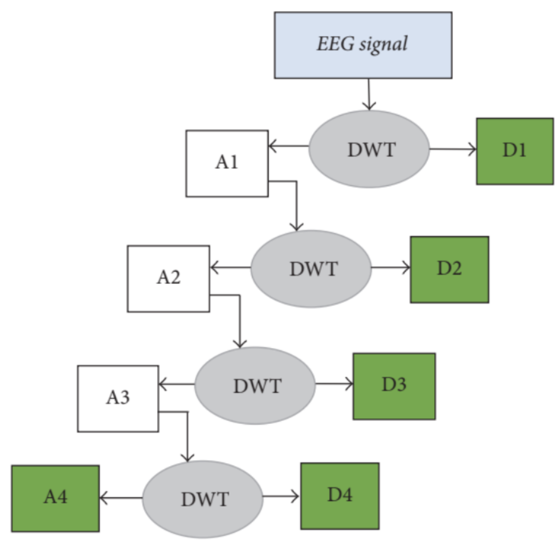

Fig. 3. EEG signal decomposition using Discrete Wavelet Transform (DWT).

Wavelet coefficients	EEG subbands	Frequency (Hz)
D1	—	128–256
D2	—	64–128
D3	Gamma	32–64
D4	Beta	16–32
A4	Alpha, theta, delta	0–16

Fig. 4. Detail and approximation coefficients from DWT analysis.

After we get this we can apply Shannon entropy on this sub bands to extract the features and energy of all the bands.

$$H = -\sum_{i} p_i \log_2(p_i) \tag{1}$$

where p_i represents the normalized amplitude probabilities.

C. Model training

There are various segmentation and overlap combinations and so for each of that combination we would yield different amount of segments. We then apply Shannon entropy to this different datasets and extract features for all of them, the more the segmentation the more the features and hence more parameters to learn from, these features like entropy and energy of the signals are then stored in the .csv files which we would use to train and test two of our models one is Random Forest and other is SVM in the classic split of 80/20, where 80% files were used to train the model and rest to test the model. The reason to not use an ANN here can be justified by the size of dataset and the availability of the data for the variance that is present here.

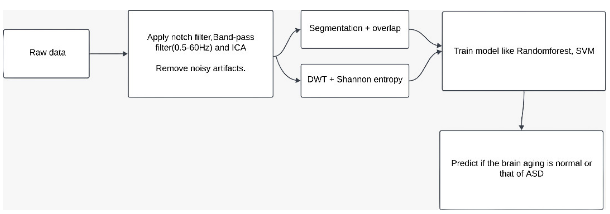

Fig. 5. System design for the methodology proposed.

D. Block diagram of proposed methodology

The Fig-5 shows the system design using block diagram for the methodology we used first we take the data, then pre-process it, then comes the segmentation and DWT part and after that feature extraction and training out model. A noteworthy point to consider here is that we also considered doing DWT on the whole signal by doing this step we are essentially increasing our features on which we can train our model along with the segmented parts.

V. Results and Discussion

We got various accuracy for various combinations of segmentation and overlapping of the EEG signals. The first segmentation and overlap were of 10 sec and 5 sec respectively,This achieved an accuracy of 65.31% on Random forest and 63.17% on SVM. The biggest segment we tried was of 25 sec with an overlap of 20 sec. This resulted into an accuracy of 72.67% on Random forest and around 73.87% on the SVM model. We have tried various segmentation and overlapping strategies, as shown in Table-2. We got highest accuracy at 5 sec segmentation and 4 sec overlap. The accuracy we achieved was 79.17% no Random forest and 77.15% on SVM. The reason behind achieving high accuracy at higher segmentation is that we get more features to learn from, although we tried two 5 second segmentation variant one with 3 seconds overlap and other with 4 sec overlap, the later one achieved higher accuracy because we provided more history to see during the training, so the model will learn from a considerable amount of past state of signals along with present state of the signals .

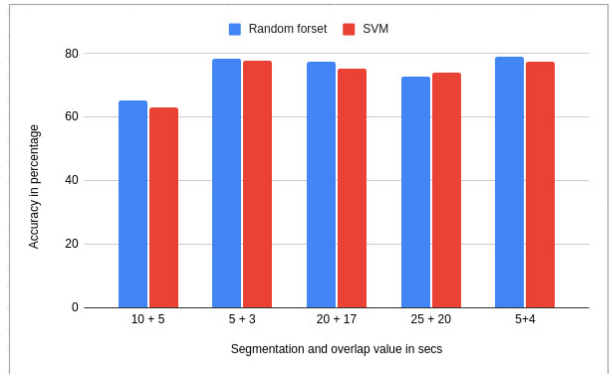

Fig. 6. Results of the experiments and accuracy.

TABLE II
CLASSIFIER ACCURACY WITH DIFFERENT SEGMENTATION AND OVERLAP SETTINGS.

Segmentation + overlap	Random Forest (%)	SVM (%)
10 + 5	65.31	63.14
5 + 3	78.47	77.72
20 + 17	77.50	75.36
25 + 20	72.67	73.87
5 + 4	79.17	77.45

VI. Conclusion

The results serve as a robust proof-of-concept that this CAD framework can identify objective biomarkers to assist in the complex ASD diagnostic process. The accuracy can be further increased by using a large dataset or combination of various datasets. Most of the paper we reviewed had implemented the later one, along with the dataset of Sheffield University other authors have combined dataset from KAU university which is also an open-source dataset. So for now this work can be considered as a proof of concept and the future work can focus on working on large dataset or by combining the datasets.

References

[1] A. Sheikhani et al., “Connectivity analysis of quantitative EEG background activity in Autism disorders,” *Proc. Int. Congress on Image and Signal Processing*, 2008.

[2] M. Ahmadlou, H. Adeli, and A. Adeli, “Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder,” *J. Clin. Neurophysiol.*, vol. 27, no. 5, pp. 328–333, 2010.

[3] M. Alhaddad et al., “Diagnosis of autism by Fisher linear discriminant analysis via EEG,” *Int. J. Bio-Science and Bio-Technology*, vol. 4, no. 2, pp. 45–54, 2012.

[4] R. Djemal et al., “EEG-based computer aided diagnosis of autism spectrum disorder using wavelet, entropy, and ANN,” *BioMed Research Int.*, 2017.

[5] A.Dickinson et al., “Electrophysiological signatures of brain aging in autism spectrum disorder,” CORTEX, vol. 148, pp. 139–151, Jan. 2022, doi: 10.1016/j.cortex.2021.09.022.

[6] A.J.Dede, W.Xiao, N.Vaci, M.X.Cohen, and E.Milne, “Exploring EEG resting state differences in autism: sparse findings from a large cohort,” Mol. Autism, vol. 16, no. 13, Feb. 2025, doi: 10.1186/s13229-025-00647-3.

[7] L. Shi et.al., “TFSNet: A Time–Frequency Synergy Network Based on EEG Signals for Autism Spectrum Disorder Classification,” Brain Sci., vol. 15, no. 7, Art. no. 684, Jun. 2025, doi: 10.3390/brainsci15070684.