Diagnosis Autism by Fisher Linear Discriminant Analysis FLDA via EEG

Mohammed J. Alhaddad, Mahmoud I. Kamel , Hussein M. Malibary, Ebtehal A. Alsaggaf, Khalid Thabit, Foud Dahlwi and Anas A. Hadi

King Abdulaziz University KAU
Jeddah, Saudi Arabia)
{ malhaddad, miali, hmalibary, eaalsaggaf, drthabit, fdehlawi, ahadi008}
@kau.edu.sa

Abstract

Diagnosis of autism is one of the difficult problems facing researchers. In this paper, Electroencephalogram (EEG) based Autism diagnosis using Fisher Linear Discriminat (FLD) Analysis is presented. Multivariate analyses of all the channels (via the concatenated signals) were used. Different preprocessing techniques, different ensemble averages, as well as, different feature extraction techniques are studied. The average correct rates are (90%). Raw data features and FFT features are used. Windsor Filtered Data gave the best mean and the lower standard deviation of both raw and FFT features. Over all, FFT features have a better correct rate of 88.14% and lower standard deviation 0.0404 than raw features.

Keywords: *Electroencephalogram; Autism; Automated diagnosis; Fisher's linear discriminant analysis*

1. Introduction

Autism is a disorder rather than an organic disease and diagnosis of autism is one of the difficult problems facing researchers and those interested in the field of signal processing and medicine. Therefore, there is a lot of research going on around the world today trying to use neuroscience such as EEG study to identify individuals with autism. Hence, a need for automatic detection of EEG signals has been sought by many researchers to diagnose autistic people. Furthermore, they report different findings regarding to discriminat patterns between normal and autism disorders [1, 2].

Many causes of autism have been proposed, but understanding of the theory of causation of autism and the other autism spectrum disorders is incomplete [19]. In this case, the phenomenological models are most appropriate to be applied than the mechanistic models. Mechanistic models typically involve physically interpretable parameters, allow deeper insights into system performance and better predictions, but they require a priori information on the system and often need more time and resources [20].

In recent years, there has been an increasing interest in applying machine learning methods to the automated detection of autism EEG signals [3, 4]. EEG signals analysis based on machine learning methods has three main steps: preprocessing, feature extraction, and classification.

The major goal of this paper is to utilize the Fisher's Linear Discriminat (FLD) analysis in detecting the autistic children based on EEG signal analysis. Thus, optimum preprocessing, as well as, optimum feature extraction techniques -which give the highest classification accuracy- are studied. The artifacts of the recorded EEG signals were removed by visual inspection. Then, different preprocessing techniques were applied such as Rereferencing, Filtering, Windsorizing, Scaling, Single epoch extraction and Feature vector construction. After preprocessing, the raw data and FFT were used as features. Dimensionality reduction using different decimation factors were applied for the raw data features extracted. Finally, the extracted features were classified using FLDA.

The layout of the paper is as follows. Section 2 focuses on the literature review, the experiments that were performed and the methods used for data preprocessing, feature extraction are described in Section 3. Classification is given in Section 4. Results are discussed in Section 5. This project sponsored by KACST1.

2. Literature Review

One of the earliest Literatures that used the EEG and was tested with disabled subjects was described by Oberman, L.M., et al., .In their work, their results support the hypothesis of a dysfunctional mirror neuron system in high-functioning individuals with ASD [5]. Parallel to the work of Oberman, L.M., et al, neurofeedback (NFB) training were developed that used changes in mu brain-activity correlated to analysis the data by signal statistic. The results showed decreases in amplitude but increases in phase coherence in mu rhythms [6].

An analysis of EEG background activity in Autism was applied in work [7]. They used Fourier methods to extract EEG features and used k nearest neighbors (KNN) to classify the two groups. In addition their findings have 82.4% discriminate between normal and autistic subjects. They also applied their work at beta band and had the same accuracy classification 82.4% [7].

Recently, the significance of classification accuracy was assessed empirically using different machine learning algorithms: the k-nearest neighbors (k-NN), SVM and naïve Bayesian classification (Bayes) algorithms with mMSE as a feature vector which described by William, B., T. Adrienne, and N. Charles [8]. They used Net Station software for acquisition data and Orange software for machine learning classification. Their accuracy classification is over 80% accuracy into control and high risk for autism HRA groups at age 9 months. Classification accuracy for boys was close to 100% at age 9 months and remains high (70% to 90%) at ages 12 and 18 months. For girls, classification accuracy was highest at age 6 months, but declines thereafter.

EEGLAB were used to extract evoked EEG features: raw EEG, CSD interpolated data, and back- projected IC features and also signal statistic was used to classify both groups. These data provide the first empirical demonstration of increased neural noise in those with ASD. Channel selection was based on an optimized electrode approach. Whereby the channel that showed the highest P1 amplitude [9]. However simple and robust FLD was not used before in autism diagnosis [14].

3. Materials and Methods

The whole process of methodologies used for automated diagnosis can be subdivided into a number of separated processing modules: Data Acquisition, pre-processing, feature extraction and classification.

[1] Sponsored by King Abdulaziz City for Science and Technology KACST, project 8-NAN106-3

A. *Experiment and Data Acquisition*

The model was conducted and tested with fifteen children from Saudi Arabia, Jeddah. It was done in the laboratory of King Abdulaziz University Hospital, where the EEG signals were recorded.

The procedure of experiment was follow:

- Subjects: The disorders consisted of eight children (5 boys and 3 girls, age 10–11 years). The control group consisted of four children (all of them are boys, age 10–11 years) without past or present neurological disorder.

  Recordings: The recordings were made with the subjects in a relaxed state in order to obtain as many artifact-free EEG data as possible. The recording system consists of the following components: g.tec EEGcap, 16 Ag/AgCl electrodes, g.tec GAMMAbox, g.tec USBamp[16], and BCI2000 [10].

  During the recording, the data were filtered using bandpass filter with frequency band (0.1-60) Hz and digitized at 256Hz. The notch filter was also used at 60Hz.

- Electrode selection: The ASD disorders have significantly values for discriminate between two subjects at electrodes FP1, F3, T5, F7, T3 and O1[2,7]. The electrodes which may give high accuracy were selected. The EEG were recorded using the international 10 – 20 system (channels FP1, FP2, F7, F3, Fz, F4, F8, T3, C4, Cz, C3, T5, Pz, O1, Oz and O2) with AFz as GND and right ear lobe as REF.

B. *Data Preprocessing*

1) Artifact Detection and removal: The artifacts of the recorded EEG signals were removed by visual inspection using BCI2000Viewer tool.
2) EEG Rereferencing: The selection of a suitable EEG reference can greatly influence the classification accuracy and sensitivity to artifacts. In this study we use common average referenced (CAR)[12].
3) Filters: A further software sixth order forward–backward Butterworth bandpass filter was used to filter the data with cut-off frequencies at 1.0 Hz and 30.0 Hz.
4) Windsorizing: Eye blinks; eye movement, muscle activity, or subject movement can cause large amplitude outliers in the EEG. To reduce the effects of such outliers, the data from each electrode were windsorized.
5) Normalization: The samples from each electrode were scaled to the interval $[-1, 1]$.
6) Feature vector construction: The samples from the selected electrodes were concatenated into feature vectors. The dimensionality of the feature vectors was $N_c \times N_s \times N_e$, where Nc denotes the number of channels, Ns denotes the number of temporal samples in one epoch and Ne denotes the number of epochs. Due to the epoch duration of 1s and the 256Hz, Ns always equals 256. Depending on the electrode configuration Nc equals 16.

As shown in Table 1. illustrates the different combined preprocessing techniques of the EEG signal which were used.

Table 1. The Different Combined Preprocessing Techniques of the EEG Signal

	Rereferencing	Filter	Windsorizing	Normalization
Raw Data	No	No	No	No
Ref Data	Yes	No	No	No
Filtered Data	No	Yes	No	No
Filtered Ref Data	Yes	Yes	No	No
Norm Filtered Ref Data	Yes	Yes	No	Yes
Norm Filtered Data	No	Yes	No	Yes
Windsor Filtered Data	No	Yes	Yes	No
Norm Windsor Data	No	No	Yes	Yes
Windsor Filtered Ref Data	Yes	Yes	Yes	No
Norm Windsor Filtered Ref Data	Yes	Yes	Yes	Yes
Norm Windsor Filtered Ref Data	Yes	Yes	Yes	Yes

C. Feature Extraction

Two different feature extraction techniques are used: temporal and frequency domains i.e. raw data and FFT.

- Data set: Artifact free data of 1276 sec. were selected from each normal and autistic children group. A big concatenated matrix is constructed with dimension $Ne \times Ncs$, where Ne denotes the number of epochs of both Normal and Autism which equals $1276 \times 2 = 2552$, Ncs denotes the number of channels × the number of samples which equals $16 \times 256 = 4096$.
- Ensemble Averaging: Ensemble average is used to test the effect of removing white Gaussian noise on the accuracy.
- Frequency Features: the spectral analysis is an important method as the brain is known to generate task-dependent activity in relatively small frequency bands. It is a basic mathematical tool based on the Fourier transform allowing the study of the signal frequency spectrum. We applied Fast Fourier Transform FFT method on each epoch.

The Fourier Transform is defined by the following equation:

$$X(f) = F\{x(t)\} = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} dt \quad (1)$$

Where x(t) is the time domain signal, X(f) is the FFT, and f is the frequency to analyze[13].

D. Feature Selection

Due to the high dimension of raw EEG data, the data were downsampled from 256Hz to 128Hz. The Downsampling were done for raw EEG data only. In FFT frequencies from 1~50Hz were selected.

4. Fisher Linear Discriminant Analysis

Over the last decade several more sophisticated non-linear classification methods, like support vector machines and random forests, have been proposed, but Fisher's method is still often used and performs well in many applications. Also, the Fisher discriminant function is a linear combination of the measured variables, being easy to interpret [14]. The FLDA will choose w, which maximize:

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \tag{2}$$

5. Results and Discussion

All the models have been implemented using MATLAB software with BCI2000 software tools and results were compared from the classification accuracy point. FLD was applied without the use of ensemble average and using the ensemble average from 2 to30 ensembles and two different feature extraction techniques has been applied in the order stated below:

A. Original data (raw data) as features

10-fold cross-validation was used to estimate average classification accuracy of FLDA. The accuracy curves obtained using FLDA plotted against the ensemble average for all the 10 data types are presented in Figure 1. Windsored-filterd data as in Figure 2 gives the best accuracy compared with others

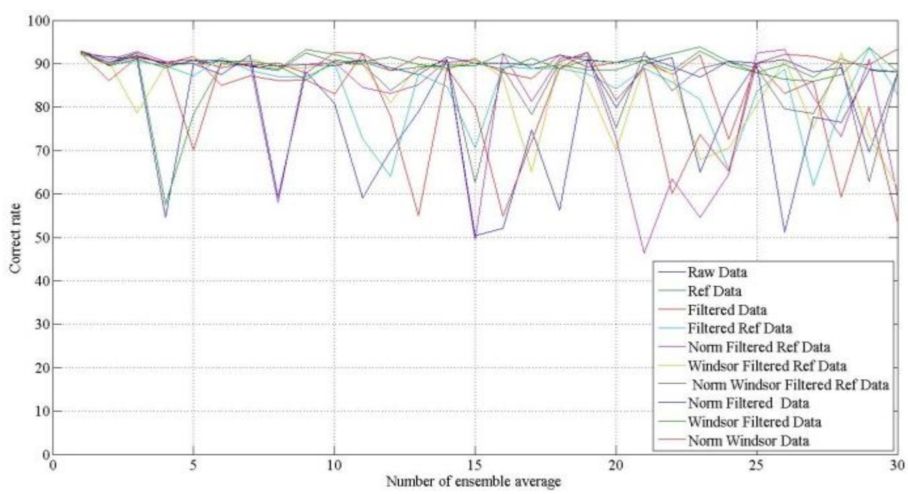

Figure 1. Correct Rates vs. Number of Ensemble Average obtained by Cross-validation with FLD using Raw Features for all Data Types

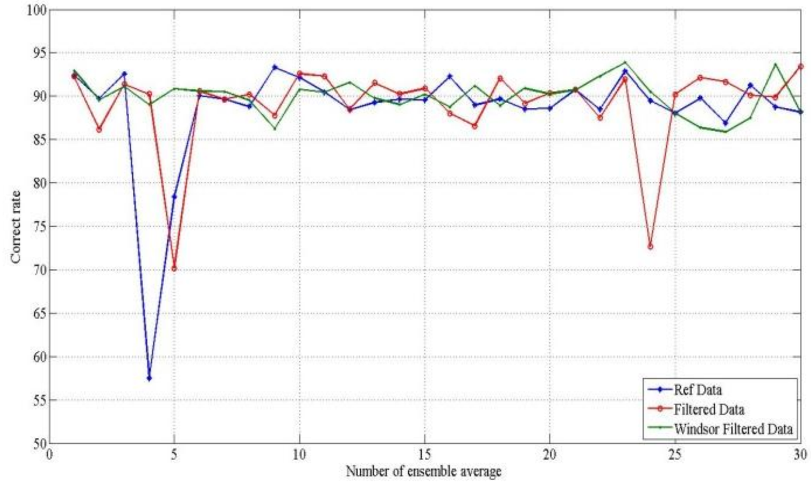

Figure 2. Correct Rates vs. Number of Ensemble Average obtained by Cross-validation with FLD using Raw Features for Best 3 Data Types

B, FFT features

FFT features were faster than raw features although there was no decimation here. Again, 10-fold cross-validation was used to estimate average classification accuracy of FLD. The accuracy curves obtained using FLDA plotted against the ensemble average for all the 10 data types are presented in Figure 3. Windsored-filterd data as in Fig. 4. gives the best accuracy compared with others.

The estimate of PSD or FFT of one EEG epoch has a chi-square distribution. In order to reduce the variance of FFT or PSD, it's necessary to average it over a number of segments [18]. All the programs which has been developed, as well as, the dataset which has been recorded and preprocessed, were located at www.mediafire.com/?m4uyv0l18cfcz3z.

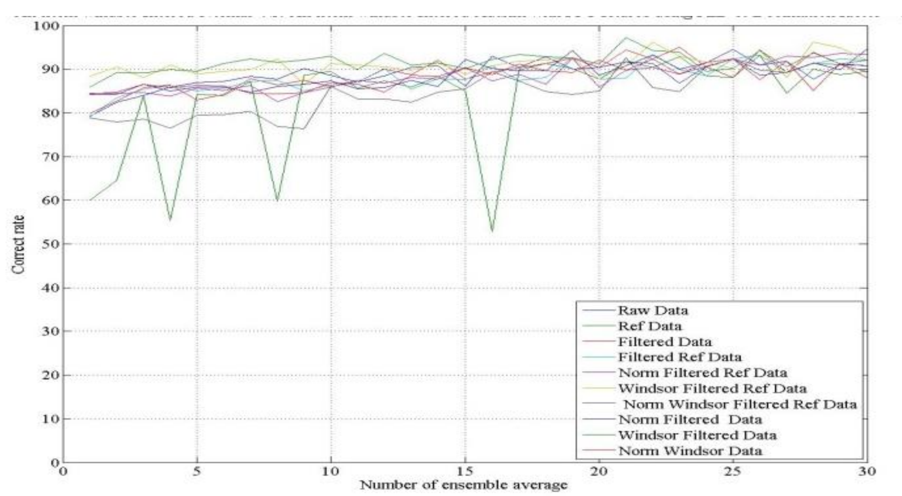

Figure 3. Correct Rates vs. Number of Ensemble Average obtained by Cross-validation with FLD using FFT Features for all Data Types

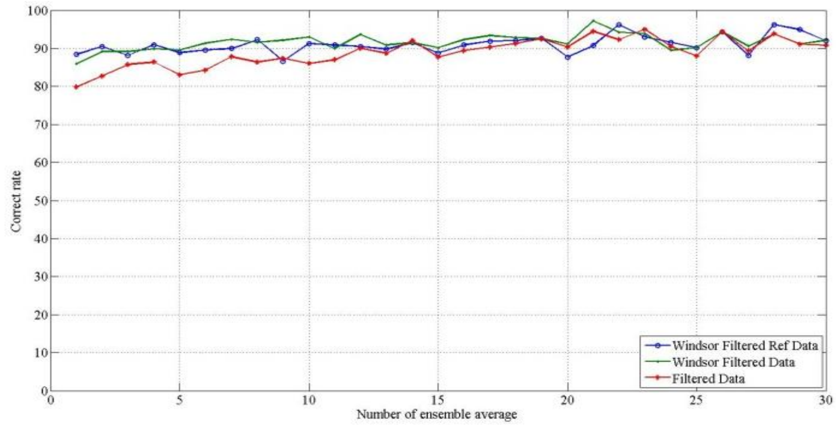

Figure 4. Correct Rates vs. Number of Ensemble Average obtained by Cross-validation with FLD using FFT Features for Best 3 Data Types

Table 2 shows the average of correct rate for raw and FFT features. The stared values are the highest. We can see that Windsor Filtered Data gives the best mean and the lower standard deviation for both raw and FFT features. For FFT, the second and the third best were Windsor Filtered Ref. Data and Filtered Data. On the other hand, Filtered Data and Ref. Data were the second and the third best results for raw features.

Over all, FFT features have a better correct rate of 88.14% and lower standard deviation 0.0404 than raw features.

Overtly-from EEG signal analysis viewpoint - there are discriminating patterns between normal and autistic children.

Improving the classification accuracy which had been given in [7], was due to the multivariate analysis of all the channels (i.e. via the concatenated signals), rather than studying the differences between of the corresponding channels of the normal and autistic children, as well as, the using of the Fisher Linear Discriminat Analysis. In order to give a concrete evidence of this discrimination, the small number of both the normal and autistic children (small dataset) should be increased.

Table 2. The Average of Correct Rate with Raw and FFT Features

	Average of Correct rate%	
Method of preprocessing	Raw data (mean±S.D)	FFT (mean±S.D)
Raw Data	88.16±0.075	88.90±0.033
Ref. Data	88.47±0.064*	83.28±0.117
Filtered Data	89.03±0.051*	88.95±0.037*
Filtered Ref. Data	84.19±0.085	88.20±0.032
Norm Filtered Ref. Data	80.60±0.148	88.15±0.032
Windsor Filtered Ref. Data	83.94±0.092	91.01±0.024*
Norm. Windsor Filtered Ref Data	86.40±0.079	84.58±0.05
Norm. Filtered Data	78.84±0.142	88.59±0.031
Windsor Filtered Data	89.97±0.02*	91.64±0.021*
Norm. Windsor Data	80.24±0.126	88.08±0.027
Mean	84.98±0.088	88.14±0.0404

6. Conclusion

In this paper, Electroencephalogram (EEG) based Autism diagnosis using Fisher Linear Discriminat (FLD) Analysis is presented. Different preprocessing techniques, different ensemble averages, as well as, different feature extraction techniques are studied. The average correct rates are (90%). Raw data features and FFT features are used. Windsor Filtered Data gave the best mean and the lower standard deviation of both raw and FFT features. Over all, FFT features have a better correct rate of 88.14% and lower standard deviation 0.0404 than raw features.

Acknowledgments

Many thanks go to all the subjects who volunteered to participate in the experiments described in this paper. We should not forget here to thank Dr. Ulrich Hoffmann et al [17]. His code helped us in developing many preprocessing algorithms. Finally, This research is considered as part of the main BCI project in the King Abdulaziz University that is funded by (King AbdulAziz City for Science and Technology) KACST, 8-NAN106-3.

References

[1] Fabricius, "The Savant Hypothesis: Is autism a signal-processing problem?", Medical Hypotheses,ScienceDirect, **(2010)**.

[2] Behnam H, Sheikhani A, Mohammadi MR, Noroozian M and Golabi P, "Analyses of EEG background activity in Autism disorders with fast Fourier transform and short time Fourier measure," in International Conference on Intelligent and Advanced Systems 2007,IEEE paper 10368672, **(2007)**, pp. 1240 – 1244.

[3] Schipul SASE and Just MA, "Applying Machine Learning Techniques to Brain Imaging Characteristics to Distinguish Between Individuals with Autism and Neurotypical Controls ", **(2010)**.

[4] Bosl CAN, "Using EEGs to Diagnose Autism Spectrum Disorders in Infants: Machine-Learning System Finds Differences in Brain Connectivity", **(2011)**.

[5] Oberman LM, Hubbard EM, McCleery JP, Altschuler EL, Ramachandran VS and Pineda JA, "EEG evidence for mirror neuron dysfunction in autism spectrum disorders", Cognitive Brain Research,ScienceDirect, vol. 24, **(2005)**, pp. 190-198.

[6] Pineda JA, Brang D, Hecht E, Edwards L, Carey S, Bacon M, Futagaki S, Suk D, Tom J and Birnbaum C, "Positive behavioral and electrophysiological changes following neurofeedback training in children with autism", Research in Autism Spectrum Disorders,ScienceDirect, vol. 2, **(2008)**, pp. 557-581.

[7] Sheikhani A, Behnam H, Mohammadi MR, Noroozian M and Golabi P, "Connectivity analysis of quantitative Electroencephalogram background activity in Autism disorders with short time Fourier transform and Coherence values", **(2008)**, pp. 207-212.

[8] William B, Adrienne T and Charles N, "EEG complexity as a biomarker for autism spectrum disorder risk", BMC Medicine, vol. 9, **(2011)**.

[9] Milne E, "Increased Intra-Participant Variability in Children with Autistic Spectrum Disorders: Evidence from Single-Trial Analysis of Evoked EEG", Frontiers in Psychology, vol. 2, **(2011)**.

[10] Schalk G and Mellinger J, A Practical Guide to Brain-Computer Interfacing with BCI2000: Springer **(2010)**.

[11] Von V, Diplom-Mathematiker and Krauledat M, "Analysis of Nonstationarities in EEG Signals for Improving Brain-Computer Interface Performance", **(2008)**.

[12] Kamel MI, Alhaddad M, Malibary H, Hadi AA "Improving P300 Speller by Common Average Reference (CAR)", To be published.

[13] Monson HH, "Statistical digital signal processing and modeling", John Wiley & Sons, **(1996)**.

[14] Croux C, Filzmoser P and Joossens K, "Classification efficiencies for robust linear discriminant analysis", Statistica Sinica, vol. 18, **(2008)**, pp. 581-599.

[15] Duda RO, Hart PE and Stork DG, "Pattern classification", vol. 2: wiley New York, **(2001)**.

[16] http://www.gtec.at

[17] Hoffmann U, Vesin JM, Ebrahimi T, Diserens K, "An efficient P300-based brain–computer interface for disabled subjects", Journal of Neuroscience Methods, vol. 167, **(2008)**, pp. 115–125

[18] Vos JE, "Representation in the frequency domain of non-stationary EEGs", G Dolce, H Künkel, Editors , Computerized EEG analysis, Gustav Fischer Verlag, Stuttgart, **(1975)**, pp. 41–50

[19] Trottier G, Srivastava L, Walker CD, "Etiology of infantile autism: a review of recent advances in genetic and neurobiological research", J Psychiatry Neurosci., vol. 24, no. 2, **(1999)**, pp. 103–115.

[20] Velten K, "Mathematical Modeling and Simulation Introduction for Scientists and Engineers", WILEY-VCH Verlag GmbH & Co.KGaA, Weinheim, **(2009)**.