I.J. Image, Graphics and Signal Processing, **2012, 3, 35-41**
Published Online April 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2012.03.06

EEG based Autism Diagnosis Using Regularized Fisher Linear Discriminant Analysis

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

Faculty of Computing and Information Technology
King Abdulaziz University KAU, Jeddah, Saudi Arabia

(miali, hmalibary, Malhaddad, drthabit, fdehlawi, eaalsaggaf, ahadi008@kau.edu.sa)

Abstract— Diagnosis of autism is one of the difficult problems facing researchers. To reveal the discriminative pattern between autistic and normal children via electroencephalogram (EEG) analysis is a big challenge. The feature extraction is averaged Fast Fourier Transform (FFT) with the Regulated Fisher Linear Discriminant (RFLD) classifier.

Gaussinaty condition for the optimality of Regulated Fisher Linear Discriminant (RFLD) has been achieved by a well-conditioned appropriate preprocessing of the data, as well as optimal shrinkage technique for the Lambda parameter. Winsorised Filtered Data gave the best result.

Index Terms— Electroencephalogram, Automated diagnosis, Autism, Regularized Fisher's linear discriminant analysis, Fast Fourier Transform.

I. 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 [3]. 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 [4].

In recent years, there has been an increasing interest in applying machine learning methods to the automated detection of autism EEG signals [5, 6]. 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 Regularized Fisher's Linear Discriminat (RFLD)

analysis in detecting the autistic children based on EEG signal analysis. Thus, optimum preprocessing -which gives the highest classification accuracy- is studied. The artifacts of the recorded EEG signals were removed by visual inspection. Then, different preprocessing techniques were applied such as Re-referencing, Filtering, Winsorizing, Scaling, Single epoch extraction and Feature vector construction. After preprocessing, FFT was used as features. Dimensionality reduction using decimation factor 2 was applied. Finally, the extracted features were classified using RFLD.

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.

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.

II. 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 [7]. 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 [8].

An analysis of EEG background activity in Autism was applied in work [9]. 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% [9].

The significance of classification accuracy was measured by 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 [10]. 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 [11]. However simple and robust RFLD was not used before in autism diagnosis [12]

III. 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.

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[13], and BCI2000 [14].

  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[2,9]. 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 Re-referencing*: 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)[15].

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) *Winsorizing*: 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 Winsorised.

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 N_c denotes the number of channels, N_s denotes the number of temporal samples in one epoch and N_e denotes the number of epochs. Due to the epoch duration of 1s and the 256Hz, N_s always equals 256. Depending on the electrode configuration N_c equals 16.

Table1. Shown the different combined preprocessing techniques of the EEG signal which were used.

TABLE 1. THE DIFFERENT COMBINED PREPROCESSING TECHNIQUES OF THE EEG SIGNAL

	Re-referencing	Filter	Winsorizing	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
Winsorised Filtered Data	No	Yes	Yes	No
Norm Winsorised Data	No	No	Yes	Yes
Winsorised Filtered Ref Data	Yes	Yes	Yes	No
Norm Winsorised Filtered Ref Data	Yes	Yes	Yes	Yes

C. feature extraction

FFT feature extraction technique was used.

- *Data set*: Artifact free data of 1276 sec. were selected from each normal and autistic children

groups. A big concatenated matrix is constructed with dimension $N_e \times N_{cs}$, where N_e denotes the number of epochs of both Normal and Autism which equals 1276×2=2552, N_{cs} denotes the number of channels × the number of samples which equals 16×256=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[16].

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.

IV. REGULARIZED FISHER LINEAR DISCRIMINAT ANALYSIS

For known Gaussian distributions with the same covariance matrix for all classes, it can be shown that Linear Discriminant Analysis (LDA) is the optimal classifier in the sense that it minimizes the risk of misclassification for new samples drawn from the same distributions [17].

Over the last decade several more sophisticated non-linear classification methods, like support vector machines and random forests, have been proposed, but it is wise to try linear ones first (of course using shrinkage estimation), Fisher's method is still often used and performs well in many applications. Also, it is a linear combination of the measured variables, being easy to interpret [12]. The FLDA will choose w, which maximize:

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

In FLDA, The standard estimator for a covariance matrix is the empirical covariance. This estimator is unbiased and has under usual conditions good properties. But for extreme cases of high-dimensional data with only a few data points given- as our case - the estimation may become imprecise. This leads to a systematic error: Large eigenvalues of the original covariance matrix are estimated too large, and small eigenvalues are estimated too small; see Figure. 1. This error in the estimation degrades classification performance. Regularization is a common remedy for the systematic bias of the estimated covariance matrices [18].

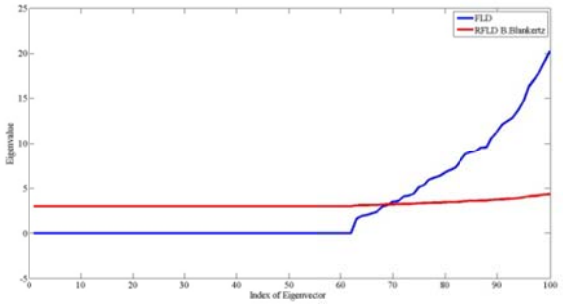

Figure 1. Eigenvalues of a given covariance matrix using FLD (blue line) and RFLD (red line)

For the RFLD we use:

$$\tilde{\Sigma}(\gamma) := (1-\gamma)\hat{\Sigma} + \gamma \nu \mathbf{I}$$

Where lambda was calculated using B. Blankertz et al. [18] method:

$$\gamma^* = \frac{n}{(n-1)^2} \frac{\sum_{i,j=1}^{d} \mathrm{var}_k(z_{ij}(k))}{\sum_{i \neq j} s_{ij}^2 + \sum_i (s_{ii} - \nu)^2} \quad (3)$$

V. 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. RFLD was applied without the use of ensemble average and using the ensemble average from 2 to30 ensembles using FFT feature extraction technique. 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 , its necessary to average it over a number of segments[19]. 10-fold cross-validation was used to estimate average classification accuracy of RFLD. The accuracy curves obtained using RFLDA plotted against the ensemble average for all the 10 data types are presented in Figure 2.

In figure 3. the best accuracy shown by Winsorised filtered data when compared it with others.

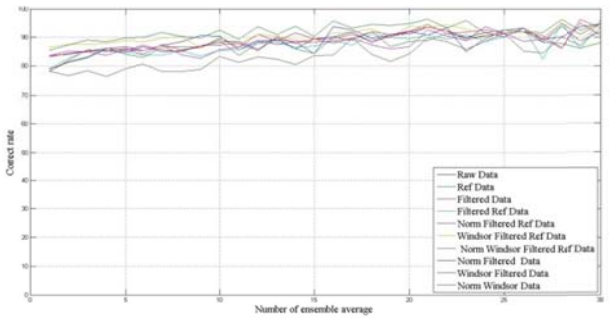

Figure 2. Correct rates vs. number of ensemble average obtained by cross-validation with FLD using FFT features for all data types

Table 2. Shows the average of correct rate for FFT features. The stared values are the higest. We can see that *Winsorised Filtered Data* gives the best mean and the lower standerd deviation. The second and the third best were *Winsorised Filtered Ref. Data* and *Filtered Data*.

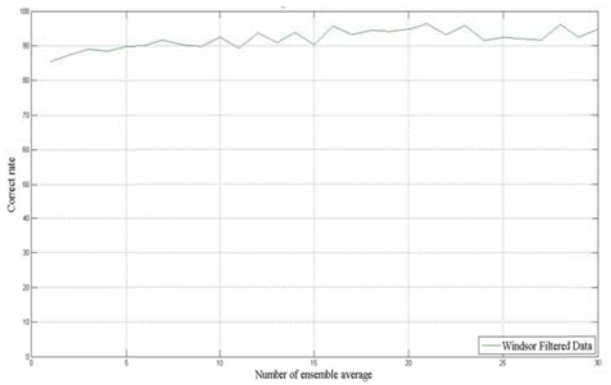

Figure 3. Correct rates vs. number of ensemble average obtained by cross-validation with FLD using FFT features for all data types

TABLE 2. THE AVERAGE OF CORRECT RATE WITH RAW AND FFT FEATURES

Method of preprocessing	Average of Correct rate % (mean ± S.D)
Raw Data	88.66±0.038
Ref. Data	87.25±0.033
Filtered Data	89.14±0.039*
Filtered Ref. Data	87.48±0.038
Norm Filtered Ref. Data	87.89±0.032
Winsorised Filtered Ref. Data	90.4±0.025*
Norm. Winsorised Filtered Ref Data	83.54±0.047
Norm. Filtered Data	88.87±0.032
Winsorised Filtered Data	92.06±0.027*
Norm. Winsorised Data	88.68±0.030
Mean	88.14±0.0404

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

Improving the classification accuracy which had been given in [9], 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 Regularized 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.

VI. CONCLUSION

In this paper, Electroencephalogram (EEG) based Autism diagnosis using Regularized Fisher Linear Discriminat (RFLD) Analysis is presented. Different preprocessing techniques, as well as, different ensemble averages are studied. The average correct rate is (92%). FFT features are used. Winsorised Filtered Data gave the best mean and the lower standard deviation for FFT features.

ACKNOWLEDGMENT

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 [20]. His code helped us in developing many preprocessing algorithms. Finally, we would like to thank our team for their efforts in the BCI project.

REFERENCES

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

[2] H. BehnamA, A. SheikhaniB, M. R. MohammadiC, M. NoroozianD, and P. GolabiE, "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 p1240 - 1244

[3] Trottier G, Srivastava L, Walker CD. Etiology of infantile autism: a review of recent advances in genetic and neurobiological research. J Psychiatry Neurosci. 1999;24(2):103–115

[4] Kai Velten "Mathematical Modeling and Simulation Introduction for Scientists and Engineers" 2009 WILEY-VCH Verlag GmbH & Co KGaA, Weinheim

[5] S. A. S. E. Schipul , M. A. Just "Applying Machine Learning Techniques to Brain Imaging Characteristics to Distinguish Between Individuals with Autism and Neurotypical Controls " 2010.

[6] C. A. N. Bosl, "Using EEGs to Diagnose Autism Spectrum Disorders in Infants: Machine-Learning System Finds Differences in Brain Connectivity," 2011.

[7] L. M. Oberman, E. M. Hubbard, J. P. McCleery, E. L. Altschuler, V. S. Ramachandran, and J. A. Pineda, "EEG evidence for mirror neuron dysfunction in autism spectrum disorders," Cognitive Brain Research,ScienceDirect, vol. 24, pp. 190-198, 2005.

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

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

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

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

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

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

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

[15] Mahmoud I. Kamel, Mohammed Alhaddad, Hussein Malibary, Anas A. Hadi. "Improving P300 Speller by Common Average Reference (CAR)". To be published.

[16] H. H. Monson, Statistical digital signal processing and modeling: John Wiley & Sons, 1996.

[17] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern classification, 2nd ed.Wiley, New York, (2001).

[18] B. Blankertz et al. (eds.), Brain–Computer Interfaces, The Frontiers Collection, Springer-Verlag Berlin Heidelberg 2010

[19] J.E. Vos, 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

[20] Ulrich Hoffmann , Jean-Marc Vesin, Touradj Ebrahimi, Karin Diserens, "An efficient P300-based brain–computer interface for disabled subjects", Journal of Neuroscience Methods 167 (2008), pp. 115–125

AUTHORS

Dr. Mohammed J. Alhaddad received his Master from Essex University in 2001, and he obtained PhD from the school of Computer Science and Electronic Engineering, University of Essex in 2006, UK. He became Chairman of Information Technology Department at King Abdul-Aziz University. His research interests are: network Security, Artificial Intelligence, Robots, Brain Computer Interface BCI, and Radio Frequency Identification RFID, Data Mining, Semantic Query Optimization, co-operative query answering, distributed databases and Deductive databases.

Dr. M.I. Kamel Ali. Born in 1955, Cairo, Egypt, Bsc. from Electronic and communication department (1978) Cairo university. Ph.D.Systems and Computer Engineering, 1991, Al-Azhar University. Visiting Professor, University of Al Ain - United Arab Emirates, Al Ain, United Arab Emirates(1993). 1993 - 2002: Consultant, Research Center, Cairo University, Cairo, Egypt. 2002-2012 King Abdulaziz University (Computer Science Department). Research Interests:- Industrial Automatic Control, Modeling and

Simulation, Artificial Intelligence, Pattern Recognition, Brain Computer Interface

Dr. Hussein Malibary, MD, PhD .Saudi National, born in Makkah ,Saudi Arabia .Graduated from Louis Pasteuar University France in 1976 then training in neurology,certified in neurology and neurophysiology .Joined the Faculty of Medicine at King Abdulaziz University as assisrtant professor in 1981.Co-founder of Saudi Neurolgy Society and of Saudi neurology training program.Member of the American Academy of Neuroilogy.former Dean of the Faculty of Medicine Publshed 20 scientific papers in National and international journals and presented 34 abstracts in Neurology national and international conferences.

Dr. Khalid O. Thabit received the Ph.D. degree in Computer Science from the University of Rice, USA in 1981. He received B.S. degree in Computer Science from Massachusetts Institute of Technology, USA and M.S. from University of Southern California, USA. His Ph.D. research thesis was published as a book and the thesis was referenced and cited in more than 30 publications and five books. His thesis was selected as one of the top in the computer science major among 1000 Universities. His research interests span a broad range of areas from compilers to Arabic text and speech recognition and Knowledge-based systems. Dr. Thabit is an Assistant Professor at the Department of Computer Science in the Faculty of Computing and Information Technology, King Abdul Aziz University (KAU), Jeddah, Saudi Arabia. He served as the Chairman of the Department of Mathematics for 2 years (1982-1984) at KAU, and as the Chairman of the Department of Computer Science for 8 years (1985-1991) at KAU. Dr. Thabit is serving as a member in various administrative committees at KAU. He also served as a supervisor for many Master students, and one PhD.His main research interests are development of Arabic text to speech system and the computer generation of more than 28 million Arabic compound words. His current research interest is in Brain Computer interface.

