Combined EEG/ERG Features for Bipolar Disorders Diagnosis

Julie Muzzolon[1,4], Xiaoxi Ren[2], Steven Le Cam[1], Thomas Schwitzer[3], Valérie Louis Dorr[1]
1. Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France
2. Unité d'Imagerie Adaptative Diagnostique et Interventionnelle, Nancy, France
3. Pôle Hospitalo-Universitaire de Psychiatrie d'Adultes et d'Addictologie du Grand Nancy, Centre Psychothérapique de Nancy, Laxou, France
4. e-mail: julie.muzzolon@univ-lorraine.fr

***Abstract*—Bipolar Disorder (BD) is a disabling lifelong condition that remains misdiagnosed. Robust biomarkers are needed for a reliable and early diagnosis. Recent studies have demonstrated that electrophysiological ERG/EEG measurements hold relevant features for the diagnosis of BD. In this study, we propose a combined analysis of these modalities with promising performance for the detection of BD subjects with respect to controls.**

Keywords-ERG; EEG; DWT; bipolar disorders; SVM.

I. Introduction

Bipolar disorders (BD) are characterized by alternating manic and depressive episodes. Although these disorders are quite common, the diagnosis is often late [1] and subjective since it primarily relies on an interview guided by a clinician. Hence, there is a need for more robusts biomarkers independent of the subjective interpretations of patients and practitioners.

Previous studies have shown that psychiatric disorders in general affect the responses of retinal rod and cone cells [2]–[4], and that electroretinigram (ERG) responses to light stimuli can help in the differential diagnosis of mental disorders [5][6]. Electroencephalogram (EEG) alterations in responses recorded from primary visual cortex areas are also well-documented [7][8].The aim of this study is to assess the benefit of combining ERG with EEG measurements. To the best of our knowledge, no previous research work applied machine learning techniques to coupled ERG/EEG features for BD diagnosis.

Most studies focus on waveform amplitudes and latencies of a and b waves [5]. These temporal features are sensitive to noise and do not characterize the whole response waveforms. We then propose to extract time-frequency (TF) features from ERG and EEG responses using Discrete Wavelet Transform (DWT) [9]. The most significant coefficients according to the Wilcoxon rank sum test (alpha risk < 0.05) were selected.

Finally, we performed classification using Support Vector Machine (SVM) on 6 datasets : we studied the discriminating power of TF features against temporal features from EEG alone, ERG alone and combined ERG/EEG. Our database being rather modest in size, we performed stratified k-fold cross-validations to avoid overfitting. Averaged F1-score, Accuracy, Recall and Specificity scores are reported, as well as the standard deviation (SD) of these criteria over the tested folds.

In Section II, we introduce the data source and methods employed to collect the recordings, denoise the signals, extract the biomarkers and perform our predictions. In Section III, we describe the selected biomarkers and the prediction results. Finally, in Section IV, we conclude about the benefit of coupled ERG/EEG TF features in BD diagnosis.

II. Methods

A. Data source and protocol

ERG (right and left eyes averaged) and EEG (average of 4 electrodes over primary visual cortex of both hemispheres) responses to visual stimuli were recorded on euthymic bipolar patients (N = 30, Age (mean $\pm$ SD) = 47.5 ± 13.3, 67.7% women) and on healthy control subjects (N = 25, Age (mean $\pm$ SD) = 42.3 ± 14.8, 60.0% women) who were included in the BiMar study carried out by the CPN, Nancy, France. We used the Retinaute device (BioSerenity), a virtual reality headset fitted with electrodes that simultaneously records ERG and EEG responses. All stimuli were performed according to the International Society for Clinical Electrophysiology of Vision (ISCEV) standards [10][11].

We recorded ERG and EEG responses under dark-adapted (DA) and light-adapted (LA) conditions with a strength of 3.0 cd.s.m^{-2} (DA3.0, LA3.0). In total 16 and 32 flashes for DA3.0 and LA3.0 respectevily. A 30Hz flash LA3.0 (Flicker) was also repeated 16 times. Each stimulus triggers an electrical activity of a specific cell in the retina : the combined rod-cone activity can be studied with DA3.0 and cone activity only with LA3.0.

B. Signal denoising and preprocessing

50Hz powerline interference was removed with an infinite impulse response notch filter (center frequency = 50Hz, quality factor = 5). We did a 10-level DWT decomposition and set approximate coefficients and corresponding detail coefficients to zero to remove low frequencies (0-1 Hz) and high frequencies (above 62 Hz) [9]. The stimuli consisting of a repetition of flashes, we then segmented our signals into equal-size epochs starting 50 ms before each flash. Ouliers epochs were rejected and we worked on the averaged epoch.

C. Biomarkers selection

We selected the amplitude and latencie of a and b waves for DA3.0 and LA3.0 [11]. The retinal response to the Flicker stimulus is periodic, so we measure the amplitudes and latencies of the first trough and peak. The EEG responses result in a series of negative (N-waves) and positive waves (P-waves), but we focused on the P2-wave as it is the most robust [10].

In order to extract more relevant features, we computed a 6-level DWT analysis [9] that gives a synthetic and non

redundant representation of the ERG and EEG in both time and frequency domains. The sampling frequency of our signals being 1000 Hz, it allows us to analyze the energy content in the frequency ranges $[0, 8]$, $[16, 31]$, $[31, 62]$, $[62, 128]$, $[128, 256]$, and $[256, 512]$ Hz. We chose 'daubechies-4' wavelet since it gave the best reconstruction of our signals once the lowest energy coefficients were removed.

A nonparametric Wilcoxon rank sum test with an alpha risk of 0.05 was used to select coefficients significantly different between patients with BD and the healthy population.

D. Machine learning model and prediction evaluation

We conducted our classification on ERG, EEG and coupled ERG/EEG features. We analyzed wave time characteristics and TF coefficients separately. Classification was made using a linear SVM classifier that separates the two classes (1 = BD, 0 = controls)[12]. In order to evaluate the discriminating power of our model, we performed a stratified cross-validation, where our data set was randomly split into 5 folds within each the proportion of the classes is preserved : 4 folds constitute the training set (N = 44) and the 5^{th} fold is the test set (N = 11). We repeat this operation 10 times so we have 50 predictions for each dataset.

We recorded the accuracy, recall, specificity and F1-score at each step, then these scores are averaged. We also pay attention to variability in the predictions by computing the SD of the scores. A great recall (resp. specificity) means that only a few bipolar patients (resp. controls) will be misclassified.

III. Results

Temporal characteristics selection showed a significant greater a-wave amplitude for DA3.0 ($p < 0.05$) as long as a significant increase in LA3.0 a-wave latency ($p < 0.05$) in bipolar patients compared to controls. In contrast, the Flicker P2-wave amplitude is significantly higher ($p < 0.01$) in controls. We extracted 12 significant DWT coefficients, 7 in ERGs and 5 in EEGs while we had only 3 features in the time domain.

We obtained better classification results using TF features rather than temporal characteristics for any electrode, whether they are coupled or not, as shown in Table I.

TABLE I. SCORES (MEAN (SD)) FOR COUPLED AND NON COUPLED ERG AND EEG FEATURES

Electrode	Feature	$F1_{score}$	Accuracy	Recall	Specificity
EEG	Amp./Lat.	65.4 (12.8)	60.2 (11.3)	72.7 (21)	45.2 (20.5)
	DWT	75.5 (12.3)	73.1 (14.0)	76.7 (15.8)	68.8 (21.8)
ERG	Amp./Lat.	70.9 (10.1)	67.5 (11.3)	73.3 (13.9)	60.4 (20.4)
	DWT	76.5 (11.4)	74.4 (10.3)	79.7 (17.3)	68.0 (15.1)
EEG/ERG	Amp./Lat.	74.4 (9.6)	68.4 (11.5)	84.7 (14.2)	48.8 (20.7)
	DWT	82.8 (9.2)	80.4 (10.1)	87.3 (12.9)	72.0 (15.7)

Moreover, we show that combining EEG and ERG yields in greater scores with a decrease in the variability for most of the scores despite high standard deviations for EEG.

Finally, coupled EEG-ERG TF showed the best results with a high recall ($> 87\%$) meaning that a few bipolar patients will remain undiagnosed, whereas the specificity is lower (72%).

IV. Conclusion and Future Work

Our first results suggest that the TF features give a more precise representation of the ERG and EEG signals compared to the amplitudes and latencies of the waves. They also suggest that coupled ERG/EEG provides greater discrimination and more reliable predictions, making it highly beneficial for BD diagnosis. However, the relatively small data set might limit the generalizability of the obtained results. Our future work will focus on improving these results by including more flash stimuli and testing other machine learning classifiers.

References

[1] M. Berk *et al.*, "Setting the stage: From prodrome to treatment resistance in bipolar disorder", *Bipolar Disorders*, vol. 9, no. 7, pp. 671–678, 2007.

[2] S. M. Silverstein, D. L. Demmin, J. B. Schallek, and S. I. Fradkin, "Measures of retinal structure and function as biomarkers in neurology and psychiatry", *Biomarkers in Neuropsychiatry*, vol. 2, p. 100 018, 2020.

[3] A. Tan, T. Schwitzer, J. Conart, and K. Angioi-Duprez, "Study of retinal structure and function in patients with major depressive disorder, bipolar disorder or schizophrenia: A review of the literature", *Journal Français d'Ophtalmologie*, vol. 43, no. 5, e157–e166, 2020.

[4] T. Schwitzer, J. Lavoie, A. Giersch, R. Schwan, and V. Laprevote, "The emerging field of retinal electrophysiological measurements in psychiatric research: A review of the findings and the perspectives in major depressive disorder", *Journal of Psychiatric Research*, vol. 70, pp. 113–120, 2015.

[5] M. Hébert *et al.*, "The electroretinogram may differentiate schizophrenia from bipolar disorder", *Biological Psychiatry*, vol. 87, no. 3, pp. 263–270, 2020.

[6] T. Schwitzer *et al.*, "Retinal electroretinogram features can detect depression state and treatment response in adults: A machine learning approach", *Journal of Affective Disorders*, vol. 306, pp. 208–214, 2022.

[7] E. Bubl *et al.*, "Retinal dysfunction of contrast processing in major depression also apparent in cortical activity", *European Archives of Psychiatry and Clinical Neuroscience*, vol. 265, no. 4, pp. 343–350, Jun. 2015.

[8] K. Tursini *et al.*, "Subsequent and simultaneous electrophysiological investigation of the retina and the visual cortex in neurodegenerative and psychiatric diseases: What are the forecasts for the medicine of tomorrow?", *Frontiers in Psychiatry*, vol. 14, p. 1 167 654, 2023.

[9] P. B. Patil and M. S. Chavan, "A wavelet based method for denoising of biomedical signal", in *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, 2012, pp. 278–283.

[10] J. V. Odom *et al.*, "ISCEV standard for clinical visual evoked potentials: (2016 update)", *Documenta Ophthalmologica*, vol. 133, no. 1, pp. 1–9, 2016.

[11] A. G. Robson *et al.*, "ISCEV standard for full-field clinical electroretinography (2022 update)", *Documenta Ophthalmologica*, vol. 144, no. 3, pp. 165–177, 2022.

[12] D. A. Pisner and D. M. Schnyer, "Chapter 6 - support vector machine", in *Machine Learning*, A. Mechelli and S. Vieira, Eds., Academic Press, 2020, pp. 101–121.