2024 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT) | 979-8-3503-6289-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/USBEREIT61901.2024.10583978

ERG Classification by Using ML Methods Based on Short-Time Fourier Transform

Faisal B. Albasu
Engineering School of Information Technologies, Telecommunications and Control Systems
Ural Federal University
Yekaterinburg, Russia
0000-0003-4584-1383

Mikhail Kulyabin
Pattern Recognition Lab
Friedrich-Alexander-Universität Erlangen-Numberg
Erlangen, Germany
0009-0007-0440-030X

Aleksei Zhdanov
Engineering School of Information Technologies, Telecommunications and Control Systems
Ural Federal University
Yekaterinburg, Russia
0000-0003-4725-3681

Anton Dolganov
Engineering School of Information Technologies, Telecommunications and Control Systems
Ural Federal University
Yekaterinburg, Russia
0000-0003-2318-9144

Vasilii Borisov
Engineering School of Information Technologies, Telecommunications and Control Systems
Ural Federal University
Yekaterinburg, Russia
0000-0003-0486-7552

Mikhail Ronkin
Engineering School of Information Technologies, Telecommunications and Control Systems
Ural Federal University
Yekaterinburg, Russia
0000-0001-5322-988X

***Abstract*—Electroretinography (ERG) is a noninvasive method for evaluating the functional status of the retina by using light stimulation. This method has demonstrated its efficacy in generating valuable features for ERG classification, as evidenced by existing research findings. This study uses the Short-Time Fourier Transform, a time-frequency domain method, to extract features from ERG signals, which are then used to train Decision Tree models for the classification of these signals. The main objective of the study is to compare different window functions, sizes and overlaps in order to determine which combination yields the most optimal features for the classification of ERG signal. The dataset used for this investigation is from a large database of electroretinogram signals consisting of 5 types of signals including Maximum 2.0 ERG response, Photopic 2.0 ERG response, Scotopic 2.0 ERG response, oscillatory Potentials, and Photopic 2.0 EGR Flicker response. The results from this investigation have indicated that this methodology has considerable potential, but also presents several challenges, such as imbalanced datasets and the resolution trade-off. This work has the potential to contribute to the development of more effective tools to assess the functional health of the retina, enhancing the diagnosis and treatment of various retinal diseases.**

***Index Terms*—Short-time Fourier Transform, Electroretinography, Feature Extraction, Spectrogram, Machine Learning Classification, Electroretinogram**

I. Introduction

Electroretinography (ERG) is a non-invasive method for evaluating the functional status of the retina by using light stimulation. This stimulation typically involves light pulses at intervals that elicit different responses depending on the type of cells being stimulated. Cone cells necessitate a high level of light for activation, whereas rod cells require minimal light and are primarily active in low light conditions [1], [2]. The responses of these cells are then captured as one-dimensional signals with a recording electrode [1], [3].

ERG signals have significant potential for detecting and diagnosing a wide range of early retinal-related diseases such as diabetic retinopathies, cone and rod dystrophies, and age-related macular degenerative disorders [1], [2], [4], [5]. In addition, it has been identified as a valuable tool in the preparation of cataract surgery and in the diagnosis of hereditary retinal disorders [3].

ERG signals are characterized by their brief duration, typically lasting up to 250ms, and have a frequency range of 0 to 1kHz, sampled at 2kHz. The key elements of an ERG signal include its amplitude (a, b) and latencies (la, lb), which are commonly referred to as its temporal components. Additionally, other components such as Oscillatory Potentials (OP), Photopic Negative Response, and Flicker ERG Response are often simultaneously recorded with the a and b components. These components are usually extracted manually and, with the assistance of a skilled clinician, can provide valuable insights into the signal's health status [6]. The a-wave amplitude signifies the initial negative response from the retina, followed by the b-wave amplitude representing a positive response. The latencies of these amplitudes are measured from the baseline of the a-wave amplitude to their respective troughs and peaks.

Various ERG signals can be extracted depending on the electrophysiological protocol and clinical application [7]. Under low light conditions, the Scotopic 2.0 ERG response is

generated primarily by rod photoreceptors that are stimulated by low light. The signal is mainly characterized by its low amplitude, high latency, and in most cases contains an insignificant a-wave of very low amplitude. When cone photoreceptors are stimulated by high intensity light, the maximum 2.0 ERG response is obtained. Its signals are characterized by high amplitudes and low latencies. The photopic 2.0 ERG response is mainly generated by the cone photoreceptors under medium light intensity conditions. However, photopic 2.0 ERG response signals typically have lower amplitudes than the Maximum 2.0 ERG response signals, except in certain cases where the signal is unhealthy. Furthermore, the amplitudes and latencies of Photopic 2.0 ERG responses are higher than those of Scotopic 2.0 ERG responses.

Time domain analysis stands as the predominant method for ERG analysis and feature extraction, widely utilized despite its susceptibility to distortions originating from various sources. Frequency domain analysis, on the other hand, has been employed in numerous studies to offer insights into frequency variations within ERG signals [2]. However, its drawback lies in the exclusion of the time element, limiting its ability to capture changes over time. In contrast, time-frequency domain analysis emerges as a comprehensive approach that allows for simultaneous examination of signals in both time and frequency domains. While this method may involve a slight compromise in resolution for each domain, it has demonstrated its efficacy in generating valuable features for ERG classification, as evidenced by existing research findings.

This study uses the Short-Time Fourier Transform which is a time-frequency domain method to extract features from the ERG signals which are then used to train Decision Tree models for the classification of ERG signals. The main objective of the study is to compare different window functions, sizes and overlaps in order to determine which combination yields the most optimal features for the classification of these signals.

II. Materials and Methods

Fig. 1 shows the data flow, feature extraction, and machine learning pipeline used for this study.

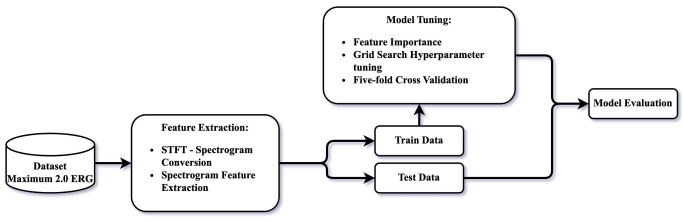

Fig. 1. Complete study pipeline: STFT feature extraction, Model training, testing and evaluation

A. Dataset

The dataset used for this study is from a large database of electroretinograms signals consisting for 5 types of signals including Maximum 2.0 ERG response, Photopic 2.0 ERG response, Scotopic 2.0 ERG response, Photopic 2.0 EGR Flicker response and Scotopic 2.0 ERG Oscillatory Potentials. The database consists of both paediatric and adult patients with the signals recorded according to the ISCEV recording standard [1]. A detailed description of the database and each of the protocols has been published in [8].

B. Data Preprocessing

For the sake of this study, only the maximum 2.0 ERG response signals were used. The Maximum 2.0 ERG response are the signals with the highest amplitude and tend to be the longest signals with a duration of up to 250 ms. However, the majority of the samples in the dataset used for this particular study have a length of 100ms with a few samples having a length of up to 250ms, hence in order to avoid confusion and noise, the long signals were cut to match the length of the short majority samples. The dataset contains 414 total signals with 340 unhealthy and 74 healthy signals.

C. Feature Extraction

To extract the features needed for this study, the signals were converted into the time-frequency domain and represented using the spectrogram. The spectrogram is obtained using the Short-Time Fourier Transform which uses a window function to localize the frequency of a signal in time by sliding the window over the signal and performing the Fourier transform on each segment in which the window is closed using the FFT algorithm [9], [10]. The STFT equation is denoted as:

$$STFT(\tau, f) = \int_{-\infty}^{+\infty} x(t)w(t-\tau)e^{-j2\pi ft}dt, \qquad (1)$$

where $STFT(\tau, f)$ represents the input signal x in the STFT with the window function w (with given length and form) for time position τ and frequency position f.

The signals were converted using the Scipy Python scientific library [11], and the spectrogram visualization was done using the Matplotlib library [12].

The spectrogram gives us a 2D representation of the signals as shown in Fig. 2 with the horizontal axis representing the time, the vertical axis representing the frequency of the signal along with a colormap representing the amplitude of the frequency at a particular time stamp in the signal. The amplitude is usually denoted using brightness or color temperature with the brighter/warmer regions indicating high amplitude and darker/cooler regions indicating low amplitude [13].

The features extracted from the spectrogram include the min, max, median, and mean intensities of the spectrogram. These features represent the low, high, mid, and average amplitude regions of the signal, respectively.

For each iteration, a separate set of values was used for the spectrogram computing parameters, which are the window function, window size(NFFT/nperseg), and the overlap between segments when sliding the window(noverlap). The window sizes and overlaps were set as powers of 2, this was chosen because FFT works best when the number of data points is a power of 2.

979-8-3503-6289-3/24/$31.00 ©2024 IEEE

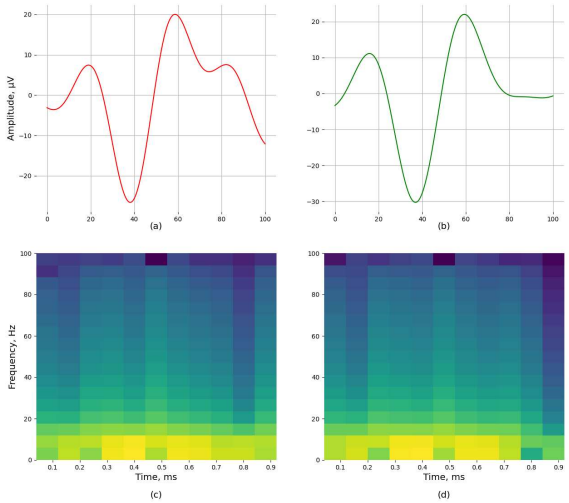

Fig. 2. Time domain and spectrogram representation of healthy and unhealthy signals in the time-frequency domain. Unealthy Signal (a) and its corresponding spectrogram representation in (c) and a healthy signal (b) with its corresponding spectrogram representation (d)

D. Machine Learning Pipeline

Before building the model, targets were encoded using label encoding, and the dataset was split into training and validation set.

The model used for the classification of signals based on the extracted features uses a decision tree algorithm implemented using the scikit-learn library to determine whether a signal is healthy or unhealthy [14]. The Decision Tree algorithm was chosen due to its interpretable nature, its ability to prioritize the most significant features based on feature importance and it's ability to work very well with unbalanced data.

Table I shows the initial hyperparameters for the decision tree model and the split size of the dataset.

TABLE I
TABLE OF INITIAL HYPERPARAMETERS USED FOR TRAINING

Hyperparameter	Value
max depth	7
min samples split	15
max leaf nodes	15
train test split	0.6:0.4
random state	42

For each model, hyperparameter tuning was performed using GridSearch with a 5-fold StratifiedKFold cross-validation. During the initial training the entire training set was fed to the model, after which for subsequent iterations non-zero features valued were selected via the feature importance attritube based on the Gini impurity reduction.

Given that the dataset used for the study is heavily unbalanced, multiple metrics were used for the evaluation of the models alongside the accuracy score, including precision, recall and F1.

TABLE II
VARIOUS CLASSIFICATION METRICS SHOWCASING THE RESULTS OF THE STUDY.

Parameters	Acc.	F1	P	R
bartlett-32-4	**0.83**	**0.91**	0.81	**0.83**
bartlett-16-8	0.81	0.89	0.76	0.81
taylor-8-4	0.80	0.89	0.74	0.80
bartlett-64-16	0.79	0.88	0.71	0.79
blackman-16-8	**0.83**	**0.91**	**0.82**	**0.83**
bartlett-8-4	0.80	0.89	0.71	0.80
blackman-64-16	0.80	0.89	0.70	0.80
hamming-32-2	0.82	0.90	0.77	0.82
boxcar-32-2	0.81	0.90	0.73	0.81
blackman-8-4	0.82	0.90	0.76	0.82
boxcar-8-2	0.82	0.90	0.77	0.82

III. RESULTS

Table II shows the metric results for 11 iterations of which the parameters used for the feature extraction were chosen pseudo-randomly and fed into the model. The params column show the parameters used for the feature extraction with the window name being the first(e.g. bartlett, blackman), the first set of numbers being the window size(e.g. 32, 8, 16) and the last digit/set of digits being the overlap. It's worth pointing out that because the window size always has to be greater than the overlap the larger number is always the window size and the smaller number the overlap, so for example in the first row bartlett-32-4 denotes window function bartlett, window size of 32 and an overlap of 4. If the digits are 5 then the first 3 denote the window size and the last 2 the overlap.

For the precision and recall columns weighted averages were reported given that the model is learning from an unbalanced dataset and also apart from the initial iteration of the model training, all other iterations were based on feature importances.

IV. DISCUSSION

On the basis of the results in Table II, it can be seen that the models tend to favor larger samples, particularly in terms of accuracy. However, when considering other metrics such as precision, recall, and F1 scores, it can be concluded that while the models are not performing exceptionally, they are doing a relatively decent job given the unbalanced nature of the dataset. This implies that the unbalanced dataset is having a significant impact on the model's performance.

The results also indicate that the Bartlett window function with a size of 32 and an overlap of 4 produces the most optimal features of the model. However, this alone is not sufficient to draw a reasonable conclusion on the Bartlett window being the best window for the signals, as the limited number of features and the limited number of combinations in the current study do not provide a comprehensive evaluation.

It is also worth noting that, due to the trade-off between time and frequency resolution introduced by the short-time Fourier

transform, using a different combination of window size and overlap might produce a different or even potentially better set of features. This is because narrower windows tend to produce better time resolutions, while wider windows produce higher frequency resolutions [15].

Furthermore, the results suggest that the models do not perform particularly well in the minority samples, which could be due to the unbalanced nature of the dataset. This implies that the models may need to be adjusted to better handle imbalanced data or that the dataset itself may need to be balanced to improve the models' performance.

In summary, while the models are performing relatively well given the unbalanced nature of the dataset, there is still room for improvement, particularly in terms of balancing the dataset. The Bartlett window function with a size of 32 and an overlap of 4 produced the most optimal features, but further exploration of different window sizes and overlaps could potentially yield better results. The trade-off between time and frequency resolution introduced by the short-time Fourier transform also suggests that different combinations of window size and overlap might produce different or even better sets of features.

V. Conclusion

This study has provided an in-depth investigation into the application of the Short-Time Fourier Transform for feature extraction from electroretinography (ERG) signals, and the subsequent use of these features in Decision Tree models for the classification of these signals. Initial results from this investigation have indicated that this methodology has considerable potential, but also presents several challenges.

Among the most prominent challenges is dealing with an imbalanced dataset. As the dataset used in this study was heavily skewed toward unhealthy signals, it influenced the performance of the models. This suggests that the models perform poorly on the minority samples (healthy signals), while the high F1 score indicates a good performance on the majority samples (unhealthy signals).

Another challenge encountered during this study was the trade-off between time and frequency resolution introduced by the Short-Time Fourier Transform. This study found that different combinations of window size and overlap might produce different sets of features, which could potentially improve the performance of the models. Narrower windows were found to produce better time resolutions, while wider windows resulted in higher frequency resolutions.

Despite these challenges, the study has yielded valuable insights into the use of the Short-Time Fourier Transform for ERG signal analysis and feature extraction. In particular, the Bartlett window function with a size of 32 and an overlap of 4 was found to produce the most optimal features for the model. However, further investigation is needed to confirm this finding, as the number of features and combinations explored in this study was limited.

Going forward, future iterations of this study will aim to further refine the feature extraction process and investigate more comprehensive strategies for dealing with issues such as imbalanced datasets and the resolution trade-off. The goal will be to extract features that will improve the performance of the models and provide more accurate classifications of ERG signals. Ultimately, this work has the potential to contribute to the development of more effective tools for assessing the functional health of the retina, enhancing the diagnosis and treatment of various retinal diseases.

Appendix

Aleksei Zhdanov, Anton Dolganov, Vasilii Borisov, Mikhail Ronkin, Vyacheslav Ponomarev, Dario Zanca, September 16, 2022, "OculusGraphy: Ophthalmic Electrophysiological Signals Database", IEEE Dataport, doi: https://dx.doi.org/10.21227/r1wb-pg25.

References

[1] A. G. Robson, L. J. Frishman, J. Grigg, R. Hamilton, B. G. Jeffrey, M. Kondo, S. Li, and D. L. McCulloch, "ISCEV Standard for full-field clinical electroretinography (2022 update)," *Documenta Ophthalmologica*, vol. 144, no. 3, pp. 165–177, Jun. 2022.

[2] S. Behbahani, H. Ahmadieh, and S. Rajan, "Feature Extraction Methods for Electroretinogram Signal Analysis: A Review," *IEEE Access*, vol. 9, pp. 116 879–116 897, 2021.

[3] A. Balicka, A. Trbolová, and T. Vrbovská, "Electroretinography (A Review)," *Folia Veterinaria*, vol. 60, no. 1, pp. 53–58, Mar. 2016.

[4] A. Wood, T. Margrain, and A. M. Binns, "Detection of Early Age-Related Macular Degeneration Using Novel Functional Parameters of the Focal Cone Electroretinogram," *PLOS ONE*, vol. 9, no. 5, p. e96742, May 2014.

[5] M. Nebbioso, R. Grenga, and P. Karavitis, "Early Detection of Macular Changes With Multifocal ERG in Patients on Antimalarial Drug Therapy," *Journal of Ocular Pharmacology and Therapeutics*, vol. 25, no. 3, pp. 249–258, Jun. 2009.

[6] A. Zhdanov, A. Dolganov, D. Zanca, V. Borisov, and M. Ronkin, "Advanced Analysis of Electroretinograms Based on Wavelet Scalogram Processing," *Applied Sciences*, vol. 12, no. 23, p. 12365, Jan. 2022.

[7] A. E. Zhdanov, V. I. Borisov, A. Y. Dolganov, E. Lucian, X. Bao, and V. N. Kazaijkin, "OculusGraphy: Filtering of Electroretinography Response in Adults," in *2021 IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM)*, Jun. 2021, pp. 395–398.

[8] F. B. Albasu, S. Dey, A. Y. Dolganov, O. E. Hamzaoui, W. M. Mustafa, and A. E. Zhdanov, "OculusGraphy: Description and Time Domain Analysis of Full-Field Electroretinograms Database," in *2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*. Yekaterinburg, Russian Federation: IEEE, May 2023, pp. 64–67.

[9] S. Nisar, O. U. Khan, and M. Tariq, "An Efficient Adaptive Window Size Selection Method for Improving Spectrogram Visualization," *Computational Intelligence and Neuroscience*, vol. 2016, p. e6172453, Aug. 2016.

[10] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series."

[11] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[12] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[13] A. M. Alaql, "Analysis and Processing of Human Electroretinogram."

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[15] W. Heisenberg, "The actual content of quantum theoretical kinematics and mechanics," Dec. 1983.