

Article

Electroretinogram Analysis Using a Short-Time Fourier Transform and Machine Learning Techniques

Faisal Albasu ^{1,2,*} , Mikhail Kulyabin ^{3,*} , Aleksei Zhdanov ¹ , Anton Dolganov ¹ ,
Mikhail Ronkin ¹ , Vasilii Borisov ¹ , Leonid Dorosinsky ¹, Paul A. Constable ⁴ , Mohammed A. Al-masni ^{2,*} 
and Andreas Maier ³ 

- ¹ Engineering School of Information Technologies, Telecommunications and Control Systems, Ural Federal University Named after the First President of Russia B. N. Yeltsin, 620002 Yekaterinburg, Russia; a.e.zhdanov@urfu.ru (A.Z.); anton.dolganov@urfu.ru (A.D.); m.v.ronkin@urfu.ru (M.R.); v.i.borisov@urfu.ru (V.B.); l.g.dorosinsky@urfu.ru (L.D.)
 - ² Department of Artificial Intelligence and Data Science, College of Software & Convergence Technology, Daeyang AI Center, Sejong University, Seoul 05006, Republic of Korea
 - ³ Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany; andreas.maier@fau.de
 - ⁴ College of Nursing and Health Sciences, Caring Futures Institute, Flinders University, Adelaide, SA 5042, Australia; paul.constable@flinders.edu.au
- * Correspondence: falbasu@urfu.ru (F.A.); mikhail.kulyabin@fau.de (M.K.); m.almasani@sejong.ac.kr (M.A.A.-m.)
- † These authors contributed equally to this work.

Abstract: Electroretinography (ERG) is a non-invasive method of assessing retinal function by recording the retina's response to a brief flash of light. This study focused on optimizing the ERG waveform signal classification by utilizing Short-Time Fourier Transform (STFT) spectrogram preprocessing with a machine learning (ML) decision system. Several window functions of different sizes and window overlaps were compared to enhance feature extraction concerning specific ML algorithms. The obtained spectrograms were employed to train deep learning models alongside manual feature extraction for more classical ML models. Our findings demonstrated the superiority of utilizing the Visual Transformer architecture with a Hamming window function, showcasing its advantage in ERG signal classification. Also, as a result, we recommend the RF algorithm for scenarios necessitating manual feature extraction, particularly with the Boxcar (rectangular) or Bartlett window functions. By elucidating the optimal methodologies for feature extraction and classification, this study contributes to advancing the diagnostic capabilities of ERG analysis in clinical settings.

Keywords: electroretinography; biomedical signal processing algorithms; short-time Fourier transform; spectrogram; feature extraction; classification; machine learning; deep learning; neural network; retinal study



Citation: Albasu, F.; Kulyabin, M.; Zhdanov, A.; Dolganov, A.; Ronkin, M.; Borisov, V.; Dorosinsky, L.; Constable, P.A.; Al-masni, M.A.; Maier, A. Electroretinogram Analysis Using a Short-Time Fourier Transform and Machine Learning Techniques. *Bioengineering* **2024**, *11*, 866. <https://doi.org/10.3390/bioengineering11090866>

Academic Editors: Qifa Zhou and Xuejun Qian

Received: 4 July 2024

Revised: 18 August 2024

Accepted: 21 August 2024

Published: 26 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Electroretinography (ERG) is a non-invasive form of assessing the functional health of the retina through its response to light stimulation. The stimulation is presented as a series of interval-based light pulses, which trigger varying responses based on the state of retinal adaptation and the wavelength, duration, strength, and stimulating frequency of the light pulse [1]. Cone photoreceptors, which are responsible for photopic or 'daytime' vision, require more quanta for activation compared to the rod photoreceptors that function in scotopic or 'night-time' vision, requiring fewer quanta to activate [1,2]. The responses from the photoreceptors and post-receptoral neurons (bipolar, horizontal, amacrine, and ganglion cells) all contribute to the overall size and shape of the recorded one-dimensional (1D) ERG signal [1,3].

Several ERG recording methods, including full-field flash, pattern, and multifocal, can be utilized for the early detection and diagnosis of a wide variety of retinal-related diseases, including early diabetic retinopathy, glaucoma, retinal dystrophies, and age-related macular degeneration [1,2,4–6].

Typically, the full-field ERG (ffERG) signals last about 250 ms and have a frequency range of 0 to 300 Hz [7]. This test is crucial for assessing the functionality of the retina, which is essential for vision. As illustrated in Figure 1, the ERG signal primarily consists of two main components: the a-wave and the b-wave. The a-wave is the initial negative deflection in the ERG signal and is generated by the retina's photoreceptor cells (rods and cones). Following the a-wave, the b-wave is a positive deflection produced by the inner retinal cells, mainly the bipolar and Müller glial cells. These waves are crucial for understanding the retina's response to light stimuli. The main characteristics of these waves are their amplitudes and time to peak. The amplitude of the a-wave (V_a) and the b-wave (V_b) refers to the height of these waves, measured in microvolts (μV). These amplitudes reflect the strength of the response generated by the retinal cells. The time to peak of the a-wave (T_a) and the b-wave (T_b) represents the time it takes for these waves to reach their maximum height after the light stimulus, measured in milliseconds (ms) [1]. These time-domain features are essential for diagnosing various retinal conditions. In addition to the main a- and b-wave components, the ERG signal can also include other components such as the Oscillatory Potentials (OPs) and the Photopic Negative Response (PhNR). The OPs are high-frequency wavelets superimposed on the ascending limb of the b-wave. They are thought to originate from the inner retinal layers, particularly the amacrine cells, and are useful for evaluating inner retinal function. The PhNR is the negative wave following the b-wave peak and is shaped by the retinal ganglion cells. An additional test protocol is the Flicker ERG, recorded using a pulse presented at 30 Hz. The response primarily assesses the cone system's functionality; it is responsible for color vision and visual acuity under photopic conditions and is useful for diagnosing cone-related disorders [8].

A number of different ERG signals may be extracted based on the different electrophysiological protocols and clinical applications [9]. The scotopic 0.01 ERG response is obtained under dark-adapted conditions and is generated by rod photoreceptors with a dominant b-wave and minimal a-wave. The scotopic 2.0 ERG has a stronger flash strength when presented under dark-adapted (DA) conditions and has a mixed rod–cone response. Under light-adapted (LA) conditions, the photopic 2.0 ERG response is a cone-driven response of typically smaller amplitude owing to the fewer cones in the human retina [1].

Currently, the most widely used form of ERG analysis and feature extraction is time-domain analysis, which involves the identification of the *a*- and *b*-wave amplitudes and their corresponding time to peaks, usually by algorithms that find the peaks automatically that can then be checked by the clinician [1,2]. However, the time-domain features do not fully reveal the underlying energy contributions of the neural generators (photoreceptors, bipolar, amacrine, horizontal, and retinal ganglion cells). So, alternative methods using signal analysis have been explored to deconstruct the signal further [10]. These methods include Power Spectral Density (PSD) and Fourier Transform, as well as time–frequency-domain methods such as the Short-Time Fourier Transform (STFT), and Continuous and Discrete Wavelet Transforms [2]. Although these methods have yet to be explored as extensively as the time domain, they offer a more detailed analysis and additional features than those provided by pure time-domain analysis.

Regarding time–frequency analysis, the predominant research has been on Wavelet Transforms, with limited exploration of the Short-Time Fourier Transform (STFT) in analyzing ERG signals. Thus, this study uses STFT as an additional signal analytical approach to the ERG. As referenced in Section 2, the existing literature has predominantly employed STFT as a complementary technique to Wavelet Transform methods as a comparison. This highlights the opportunity to delve deeper into the potential benefits and insights that STFT could offer in the overall analysis of ERG signals.

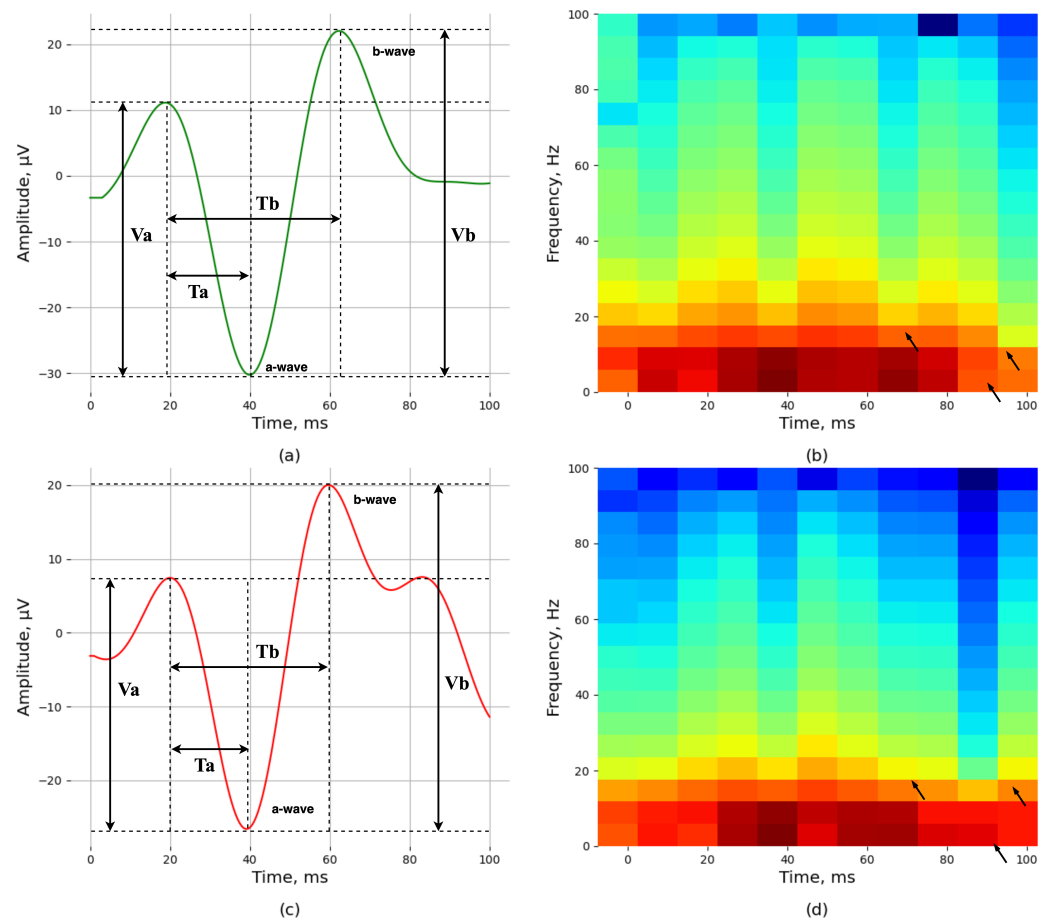


Figure 1. ERG representation of a healthy and an unhealthy Maximum (DA) 2.0 signal in time and time–frequency domains: (a) healthy signal in the time domain; (b) spectrogram representation of (a) in the time–frequency domain; (c) unhealthy signal in the time domain; (d) spectrogram representation of (c) in the time–frequency domain. The black arrows in (b,d) indicate that the spectrogram frequency distribution ranges from 0 to 80 ms and 0 to 15 Hz (maximum energy), 0 to 80 ms and 15 to 25 Hz (medium energy), and from 25 Hz (low energy). The key difference between the healthy signal (b) and the unhealthy signal (d) is evident in the energy distribution across the frequency bands. The healthy signal demonstrates a more diverse spread of energy, with maximum energy occurring at higher frequencies (0–20 Hz) compared to the unhealthy signal, where the maximum energy is concentrated at lower frequencies (0–15 Hz). This indicates that unhealthy signals tend to have more energy concentrated in the lower frequency bands, suggesting a potential marker for identifying signal health.

STFT could be selected as the most interpretable of the transformations mentioned above. The spectrogram is a 2D representation of the signal with the time on the horizontal axis and the frequency on the vertical axis, which can be given as follows:

$$STFT(\tau, f) = FFT(x(t) \cdot w(t - \tau)), \quad (1)$$

where FFT is the fast Fourier Transform of spectrum calculation; $STFT(\tau, f)$ corresponds to the representation of the input signal x with the window function w (with given length and form) for time position τ and frequency position f . Let us denote that in general the spectrum can be expressed as $\int_{-\infty}^{+\infty} x(t)w(t - \tau)e^{-j2\pi ft}dt$ [11].

Equation (1) provides a linear, unambiguous, and reversible relationship between the input (x) and output results ($STFT(\tau, f)$). The power spectrum density for the next processing can be given as $|STFT(\tau, f)|$ due to the complex origin of the equation [11].

STFT uses a sliding overlapping window function to convert the signals to a time–frequency domain using the fast Fourier Transform (FFT) algorithm. This produces a 2D spectrogram representation with the time on the horizontal axis, the frequency on the vertical axis, and the amplitude/power represented as a color map. Figure 1 depicts a healthy (top) and an unhealthy (bottom) signal with dystrophy in the time domain and their corresponding spectrogram representations calculated using STFT. We can see that the spectrogram gives us the signal frequency, which is in the range of 0–100 Hz, along with the time the frequency occurs and how much power that frequency contains, with red representing higher power frequencies and blue representing lower power frequencies.

The STFT spectrogram shows the energies within each frequency band from 0 to 100 Hz. The horizontal axis of the spectrogram denotes time bins (in milliseconds), and the vertical axis represents frequency bands in Hertz (Hz). It should be noted that the black arrows in Figure 1b show that the spectrogram frequency distribution is from 0 to 50 ms and 0 to 20 Hz (maximum energy), from 60 to 80 ms and 0 to 10 Hz (medium energy), and from 0 to 67 ms and 15 to 30 Hz (low energy). The black arrows in Figure 1d show that the spectrogram frequency distribution is from 0 to 80 ms and 0 to 15 Hz (maximum energy), from 0 to 80 ms and 15 to 25 Hz (medium energy), and from 25 Hz (low energy). The key difference between the healthy signal (Figure 1b) and the unhealthy signal (Figure 1d) is evident in the energy distribution across the frequency bands. The healthy signal shows a more diverse energy spread with the maximum energy occurring at higher frequencies (0–20 Hz) compared to the unhealthy signal, where the maximum energy is concentrated at lower frequencies (0–15 Hz). This indicates that unhealthy signals tend to have more energy concentrated in the lower frequency bands, suggesting a potential marker for identifying signal health.

This study compared various window functions for optimal feature extraction using STFT and spectrogram generation to classify the signals and determine which window yielded the best features for ERG signal classification. Several combinations of window function, window size, and window overlap were used to extract spectrogram images to train deep learning (DL) models, and manual feature extraction, which was used to train classical machine learning (ML) models. The results from both approaches were compared to determine which window yielded the best signal classification and whether DL had an advantage over the classical ML approaches. The main contributions of this study were the use of different window parameter combinations for feature extraction and the application of DL for classifying the extracted spectrogram images.

The paper is organized as follows: Section 2 reviews relevant studies and those employing STFT for feature extraction in similar fields. Section 3 presents the materials and methods used for the study, which include the ERG signal database and the pipeline for signal processing, feature extraction using the STFT, model building, and evaluation. Section 4 and Section 5 describe the results obtained from the analyses using multiple evaluation metrics and discuss the outcomes, and, finally, Section 6 concludes with the implications of the findings for the analysis of the ERG and future directions.

2. Related Works

ERG analysis methods can be divided broadly into three different approaches: time-domain analysis, which involves analyzing the amplitudes and time to peaks of the signal; frequency-domain analysis, which involves studying the frequencies of the signal; and time-frequency analysis, which involves studying the signal's frequencies at the time they occur along with their power and nonlinear methods [2]. Time-domain analysis, for the most part, is the most popular method used in the literature because it is fast and usually provides differences in amplitude or time to peak when there is a retinal disease. However, subtle or early functional changes may not be evident in time-domain analysis initially, such as in diabetes and glaucoma, thus the application of signal analysis may improve earlier diagnosis in both. In addition, signal analysis may also support classification between groups in early neurological disorders [10].

Several studies have used frequency-domain methods to analyze ERG signals. These methods provide a different perspective on signal analysis by providing spectral information unavailable in the time domain. Most studies in the frequency domain use the FT with the FFT algorithm [12] to convert the signal into the frequency domain before analyzing it. A few other methods, namely Power Spectral Density (PSD) and Linear Prediction (LP), have also been used. In [7,13], Gur et al. were able to find similarities between corneal and non-corneal ERG signals by using FFT and LP to identify specific frequencies in normal corneal ERGs under different conditions. After studying the Oscillatory Potentials (OPs) from the ERG signals of diabetic patients using the FFT, Vander Torren et al. [14] concluded that it was possible to express OPs quantitatively even in pathologies. Similarly, by studying photopic and scotopic ERGs in the Fourier spectrum and comparing them to the time domain, Li et al. [15] were also able to highlight differences in the dominant frequency and power between the scotopic and photopic ERGs. In a different study, Sieving et al. [16] used discrete Fourier Transform (DFT) to study Flicker ERGs cycle by cycle, extracting real-time harmonic components.

Using Welch's Power Spectral Density (PSD), Karimi et al. [17] were able to find significant differences in the frequency components in the scotopic and photopic ERGs of patients with and without retinitis pigmentosa. To search for signs of retinal pathologies in patients with stage I and II open-angle glaucoma, Zueva et al. [18] analyzed the frequency responses from Flicker and pattern ERGs by decomposing them into a Fourier Series.

While the frequency-domain methods mentioned above provide spectral information about ERG signals, they need to improve significantly regarding temporal information, which is crucial for ERG analysis. Time - frequency domain methods offer a way to obtain both spectral and temporal information from the signals and represent it in a 2D or 3D format. Unlike the classical FT, STFT allows us to visualize the signal's frequencies, the time window at which they occur, and how strong that frequency is at that point in time. This allows us to extract multi-dimensional features that are otherwise not accessible in the time domain or frequency domain alone.

To the best of our knowledge and findings, virtually all time-frequency ERG analysis studies are based on Continuous and Discrete Wavelet Transforms, except very few studies that included STFT as part of the analysis. In [19], STFT was one of the time-frequency methods used along with Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) to analyze the photopic ERG signals obtained from a healthy subject. In [20], STFT was applied along with CWT and DWT to analyze the effects of obesity on ERG signals. Three different responses (cone, rod, and maximal combined) were analyzed, and features were extracted using STFT, CWT, and DWT, after which the results from these methods were compared. In [21], STFT and DWT were used to determine the frequency components of the three photopic and Flicker 30Hz ERG signals of patients with Central Retinal Vein Occlusion (CRVO). More recently, ref. [8] used CWT to manually extract features from adult and pediatric signals, which were used to train a Decision Tree classifier by combining time-domain features (a- and b-wave amplitudes and implicit times) with the wavelet features. In a similar study to this paper, ref. [22] compared several mother wavelet combinations to determine which combination would better classify pediatric ERG signals.

It is worth noting that a significant drawback of STFT is its time-frequency resolution trade-off, which stems from the uncertainty principle (Gabor limit in signal processing) [23]. This means that it is impossible to achieve a high resolution for both the time and frequency components of the signal simultaneously; hence, a compromise has to be made between the two. Thus, the larger the window size, the better the frequency resolution and the lower the time resolution, and the smaller the window size, the better the time resolution, but the lower the frequency resolution.

Table 1 provides several frequency and time-frequency domain methods used for analyzing ERG signals previously, as well as the types of ERG signals used for the studies.

The analyses presented in Table 1 indicate a prevalent preference among researchers for the FT in frequency-domain analysis and the Wavelet Transform in time-frequency-domain

analysis. This inclination towards the Wavelet Transform in time–frequency-domain analysis may stem from considering the time–frequency resolution trade-off. As different mother wavelets in Wavelet Analysis can impact signals uniquely, the STFT also exhibits variations in signal representation based on factors such as the chosen window function, window size, and overlaps between windows during signal processing. This warrants further investigation to ascertain the efficacy of STFT in ERG signal analysis, given its nuanced response to different signal characteristics.

Table 1. Related works.

Author(s)	Year	Method(s)	Signal(s)	No. of Subjects (Signals)
Gur et al. [13]	1979	FFT and LP	Corneal and Non-Corneal ERG	4 (N/A)
Gur et al. [7]	1980	FFT and LP	Corneal ERG	13 (N/A)
Van Der Torren et al. [14]	1988	FFT	Oscillatory Potentials	N/A (N/A)
Li et al. [15]	1990	Fourier Spectrum	Photopic and Scotopic ERGs	13 (23)
Sieving et al. [16]	1998	Discrete Fourier Transform	Flicker ERG	N/A (N/A)
Karimi et al. [17]	2012	Welch’s Power Spectral Density	Full-Field Photopic and Scotopic ERGs	N/A (54)
Alaql n.d. [19]	2016	Fourier Transform, STFT, CWT and DWT	Photopic ERG	N/A (N/A)
Zueva et al. [18]	2019	Fourier Series	Flicker and Pattern ERG	12 (N/A)
Erkyamaz et al. [20]	2021	STFT, CWT, DWT	Cone, Rod, Maximal ERG	40 (N/A)
Behbahani et al. [21]	2021	STFT, DWT	Photopic and Flicker ERG	20 (N/A)
Zhdanov et al. [8]	2022	CWT	Scotopic, Photopic, Maximum ERGs, Flicker and Oscillatory Potentials	N/A (425)
Kulyabin et al. [22]	2023	CWT	Photopic, Scotopic and Maximum	N/A (353)

3. Materials and Methods

3.1. Dataset

The database used for the study consisted of five ERG signal types: Maximum 2.0 ERG response, Photopic 2.0 ERG response, Scotopic 2.0 ERG response, Photopic 2.0 ERG Flicker response, and Scotopic 2.0 ERG Oscillatory Potentials. The database consisted of pediatric and adult patients with the signals recorded according to the ISCEV recording standard [1]. For detailed descriptions of the database and protocols employed in this study, please refer to [24]. All recordings in the dataset were made with the Tomey GmbH EP-1000 stimulator sampling at 2 kHz with a 0.1–300 Hz bandpass filter. The DTL fiber active electrode was utilized for ERG recordings. The DTL was composed of 7 cm long, low-mass spun nylon fibers impregnated with metallic silver. However, despite its comfort, DTL electrodes are commonly used in ERG due to their poor stability on the eye, leading to potential movement with blinks and subsequent variations in amplitude. Conversely, gold foil or contact lenses offer greater stability but are less comfortable. Thus, each type of electrode in ERG exhibits distinct strengths and weaknesses [25]. Two flash strengths were employed: DA 2.0 for scotopic responses and LA 2.0 for maximal responses. Flash stimuli consisted of a white light at $2 \text{ cd} \cdot \text{s} \cdot \text{m}^{-2}$ intensity on a $0 \text{ cd} \cdot \text{s} \cdot \text{m}^{-2}$ background, ensuring standardized conditions for eliciting retinal responses [26].

This study utilized only the scotopic maximum (DA2) ERG response signals from the dataset. DA2 has a duration of up to 250 ms. However, most samples in the dataset used for this study have a length of 100 ms, with some extending up to 250 ms. To maintain consistency and reduce noise, we truncated the longer signals to match the size of the majority of samples. The signals in the dataset provide information on diseases such as cone and rod retinal dystrophies. More details about the dataset are shown in [24].

Due to the highly unbalanced nature of the dataset, with the unhealthy samples being more than twice the number of the healthy samples, we used a balanced version of

the dataset to avoid having a biased and overfitted classifier. The dataset balancing was accomplished using the Imbalanced Learn Python library [22]. In this study, we utilized an under-sampling technique with the A11KNN function from the library, which uses the nearest neighbor algorithm to identify and contradict samples and their neighborhoods [27]. Table 2 demonstrates the distribution between the healthy and unhealthy signals in both the unbalanced and balanced datasets.

Table 2. Distribution of healthy and unhealthy ERG signals, including balanced and unbalanced databases.

Unbalanced Dataset		Balanced Dataset	
Healthy	Unhealthy	Healthy	Unhealthy
60	143	60	62

3.2. Spectrogram Conversion

Figure 2 shows the stages undertaken for the entire methodology. The major stages of the pipeline were the image and feature extraction stages, the classical ML stage, and the final DL stage.

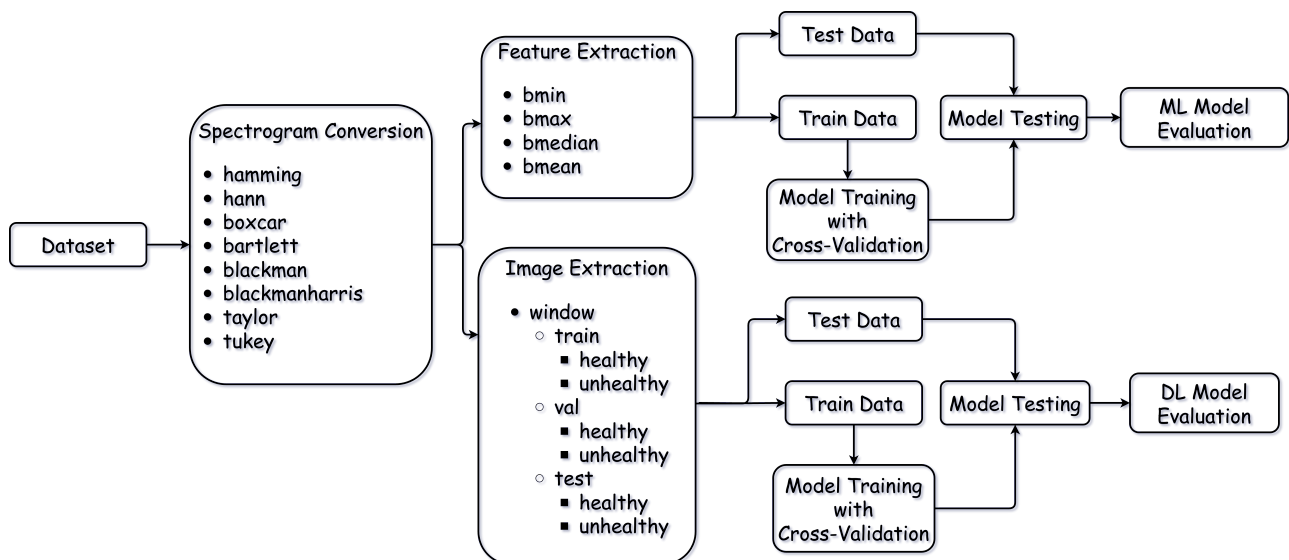


Figure 2. Complete study pipeline: from spectrogram conversion, feature extraction for ML, and image extraction for DL to data splitting and classifier evaluation.

The first stage involved data preparation before spectrogram image extraction for the DL models. The ERG signals were split into healthy and unhealthy classes. We then employed an 80:20 split for training and testing, with a training set further divided into 90:10 training and validation sets, resulting in a final 70:20:10 split of training, testing, and validation, respectively. This splitting occurred on the raw data rather than the extracted images to prevent data leakage, ensuring all images from the same signal remained together in either the training, testing, or validation sets. We then utilized various window functions to extract spectrograms, namely Boxcar, Hann, Hamming, Tukey, Bartlett, Blackman, Blackman–Harris, and Taylor windows [28]. The entire process, including data splitting and spectrogram extraction, was implemented using Python (version 3.11). Finally, the images were organized according to the Imagenet dataset structure [29].

3.2.1. Image Extraction

For the DL classifiers, spectrogram images were calculated using the STFT. Multiple spectrogram images were extracted for each signal using different combinations of the

window function, size, and overlap to obtain a large number of images for training the DL models. Given that each window has 21 size–overlap combinations and a total of 120 signals, this yielded 2520 spectrogram images for each window. Each window contained several sizes and overlaps used to extract the spectrogram images. The window sizes and overlaps were chosen as powers of 2, as recommended by the FFT algorithm [30]. This was implemented due to the time–frequency trade-off inherent in the conversion of signals, whereby signals with larger window sizes would yield better frequency-domain resolution (i.e., better feature representation), while signals with smaller window sizes would yield better time-domain resolution.

As previously described, the data was split prior to image extraction to avoid data leakage. All images from a single signal were saved in one location; hence, if one signal was in the validation set, then all images from that signal were in the same validation set.

3.2.2. Feature Extraction

The obtained spectrogram arrays were processed to extract specific features: maximum, minimum, median, and mean intensity values (bmax, bmin, bmedian, and bmean, respectively). This extraction process is essential for a thorough analysis of the ERG signals. Spectrograms are visual representations of the spectrum of frequencies in a signal as it varies with time. By converting ERG signals into spectrograms, we can analyze the frequency components and their changes over time, providing a deeper understanding of the retinal responses. The maximum intensity value (bmax) represents the highest energy point in the spectrogram, which can indicate the most robust response at a particular frequency. The minimum intensity value (bmin) shows the lowest energy point, helping to identify the baseline activity or noise level. The median intensity value (bmedian) provides the middle value of the intensity distribution, offering a measure of the central tendency that is less affected by outliers. The mean intensity value (bmean) gives the average energy level across the spectrogram, offering insight into the overall energy distribution of the signal. These features (bmax, bmin, bmedian, and bmean) are crucial for various reasons. First, they help quantify the ERG signal’s characteristics, making comparing signals from different patients or conditions easier. Second, these features can be used as input for ML algorithms to classify or predict retinal conditions based on ERG data. Finally, by analyzing these intensity values, researchers can identify patterns or abnormalities that might not be apparent in the time-domain analysis alone.

The features for each spectrogram output were extracted separately so that each window-size–overlap combination was distinct from the others. Table 3 shows the window sizes and overlaps used for the signal conversion. With 21 size–overlap combinations for each window, there were a total of 168 parameter combinations, from which the features of each resulting spectrogram were individually extracted. Figure 3 shows a spectrogram representation of signals based on windows used for the study with a window size of 32 and an overlap of 16. Each window represented the signal in a slightly different form based on the shape of the window used, namely (a) Hamming, (b) Hann, (c) Boxcar, (d) Bartlett, (e) Blackman, (f) Blackman–Harris, (g) Tukey, (h) Taylor.

Table 3. Table of window sizes and overlaps used for the spectrogram conversion. Each window was combined with each overlap for each iteration.

Window Sizes	Overlaps
128	64
64	32
32	16
16	8
8	4

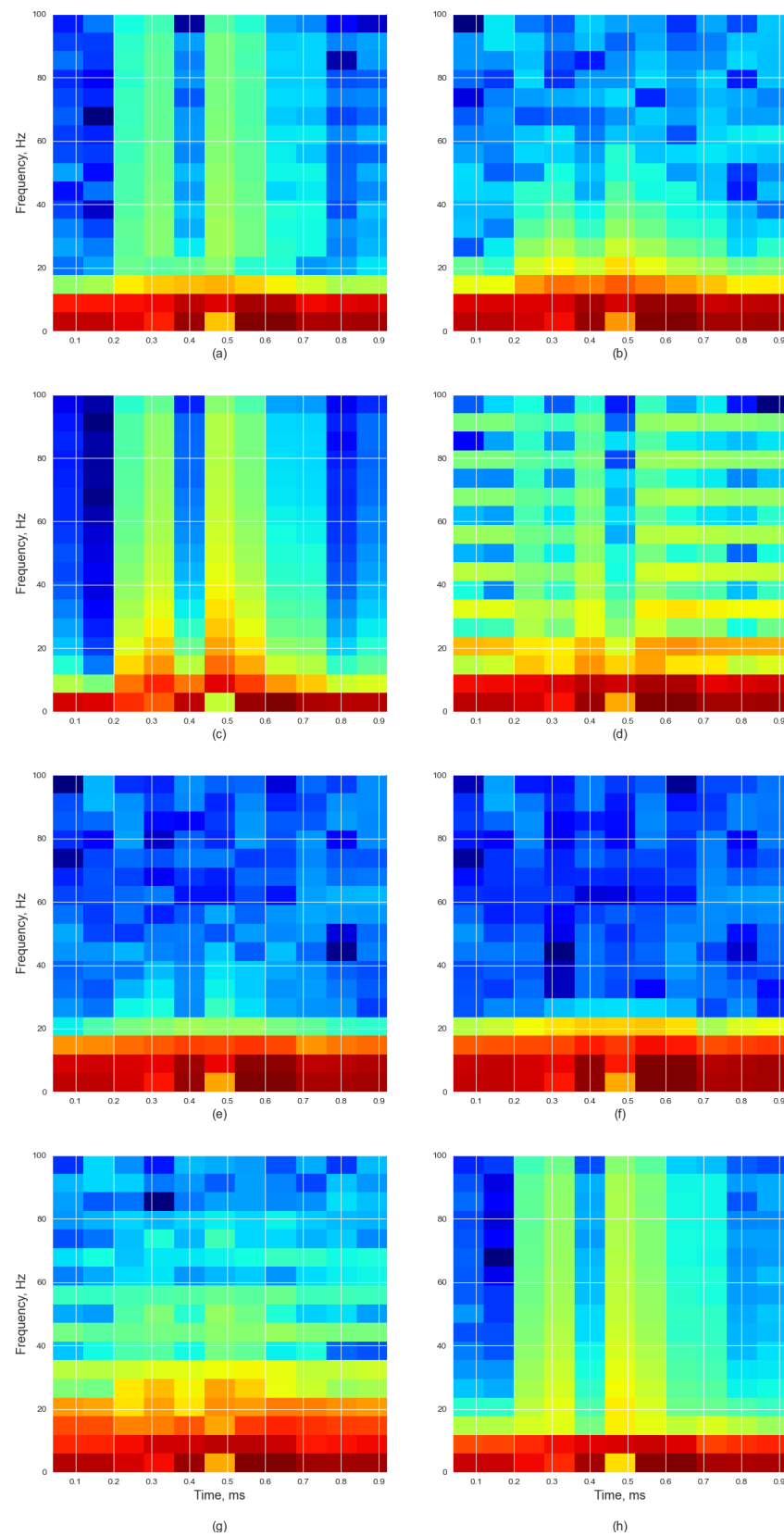


Figure 3. Spectrogram representation of signals based on windows used for the study with window size of 32 and an overlap of 16: (a) Hamming, (b) Hann, (c) Boxcar, (d) Bartlett, (e) Blackman, (f) Blackman–Harris, (g) Tukey, (h) Taylor.

3.3. Machine Learning Classifiers

Figure 2 demonstrates the training pipeline used in both the ML and DL approaches. As mentioned in Section 3.2.2, the ERG signal was first converted into a spectrogram from which the features bmin, bmax, bmedian, and bmean were extracted. This was obtained for each combination of window-size–overlap parameters. At each iteration, a different parameter combination was used to obtain the spectrogram before extracting the features for the classifiers.

Two classifiers were used for the standard ML approach: Decision Tree (DT) and Random Forest (RF) algorithms. These classifiers were implemented using the Scikit-Learn ML library [31]. These classifiers were chosen due to their ability to prioritize the most significant features in the feature set for the task. In addition to the DT classifier being able to prioritize the most significant features, the RF classifier, being an ensemble of multiple DT classifiers, can take advantage of various DTs, capitalizing on their advantages while minimizing their disadvantages. Table 4 shows the parameters used for the classifiers.

Table 4. Table of hyperparameters used.

Hyperparameters	DT	RF
Criterion	Gini	Gini
Max depth	10	10
No. of estimators	N/A	250
OOB score	N/A	True
Min samples split	2	2
Min samples leaf	1	1

We applied a 5-fold stratified k-fold cross-validation for the model training and tuned the hyperparameters to observe whether the classifier could yield better results.

3.4. Deep Learning Classification Models

ML techniques often struggle with the multifaceted nature of spectrogram-transformed data. ML methods require extensive feature engineering to identify relevant features, which is a time-consuming process and prone to human selection bias [32]. In contrast, DL offers a promising alternative, leveraging its ability to learn hierarchical representations automatically [33]. In this study, we used classical architectures: DenseNet121 [34], ResNet50 [35], VGG16, and VGG19 [36], as well as a new robust architecture, Visual Transformer (ViT) [37], that has been used for ERG classification in the time–frequency domain [22,38]. For this analysis, we use ViT Small (ViT_small_r26_s32_224) [39]. ViT is available at the HuggingFace Transformers repository [40]. The remaining models are available at the HuggingFace Timm repository [41].

We used an ADAM [42] optimizer with an initial learning rate of 0.001 for all the models. Each model was trained until convergence using early stopping criteria based on the validation loss. The cross-entropy loss function was utilized for the training. All experiments were performed on a single NVIDIA A100 graphics processing unit on a machine with two Intel Xeon Gold 6134 3.2 GHz and 96 GB RAM.

Random crop, translation, rotation, horizontal flip, and vertical flip were exclusively utilized on the images under consideration for the augmentation [43].

3.5. Metrics

To analyze the performance of the models, several metrics, including Accuracy, Precision, Recall, and F1-score, were calculated. These metrics provide a comprehensive view of the performance of each model:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (2)$$

$$Precision = \frac{TP}{TP + FP}, \quad (3)$$

$$Recall = \frac{TP}{TP + FN}, \quad (4)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (5)$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. We also evaluate the performance of each classification model using the receiver operating characteristic (ROC) with its area under the curve (AUC).

Combining these metrics in a binary diagnostic classification problem ensured a comprehensive model performance evaluation. Accuracy offers an overall success rate, where Precision and Recall become critical by focusing on the model's ability to correctly identify unhealthy patients without misclassifying healthy ones. The F1-score adjusts Precision and Recall, providing a metric that balances the importance of avoiding FP and FN. Together, these metrics address the multifaceted challenge of medical diagnosis, ensuring the model's performance is accurate, reliable, and clinically useful in distinguishing between healthy and disease classes.

4. Results

4.1. Performance of ML Classifiers

In this section, we demonstrate the performance of the ML classifiers (DT and RF) in distinguishing between healthy and unhealthy ERG signals. Several classification metrics were used for the evaluation to obtain a detailed understanding of the classifiers' behaviors, including Accuracy, F1-score, AUC, Precision, and Recall. Each window has 21 combinations of size and overlap; thus, the mean and standard deviation of the metrics were taken for each window. In addition, the feature importance distributions for the top classifiers have been included. These distributions reveal the features that were given priority and deemed significant during training. They were instrumental in classifying the healthy and unhealthy signals.

Figure 4 shows the mean accuracies of the DT and RF ML models using various windows. Detailed values of the results are given in Appendix A. All the windows have the same upper bounds, with an Accuracy of 70.83% and AUC of 72.22% for RF and an Accuracy of 66.67% and AUC of 67.14% for DT. Tables A1–A3 show this in detail. The results showed minimal variation between the windows; some even had identical mean scores and standard deviations. It is worth noting that the RF-based classifiers still performed slightly better than their DT-based counterparts, as reported by the higher mean scores.

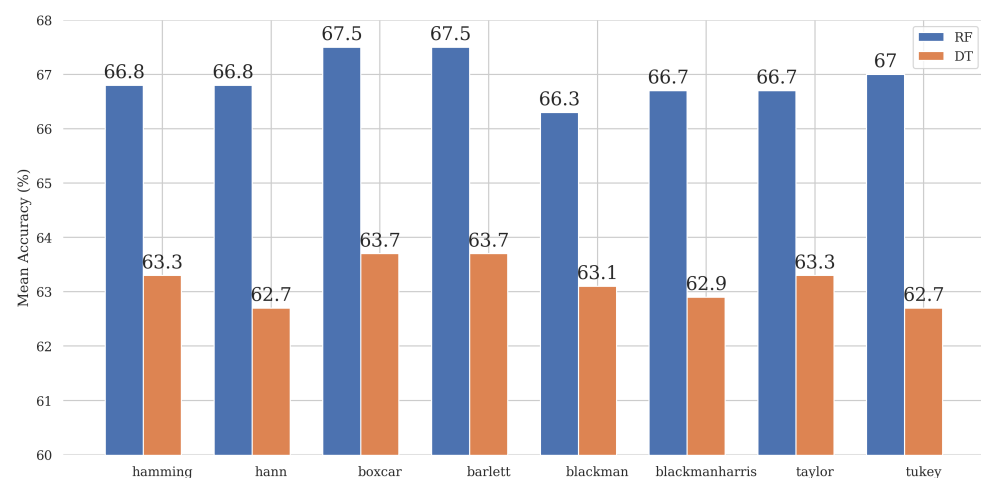


Figure 4. Mean accuracies of the analyzed DT and RF ML classifiers for all test windows.

The feature importance distributions in Figure 5 indicate that the most important features are the maximum, average, and median intensities. The significance of these features alternates between the maximum and mean intensities, while the minimum intensity had little to no significance. This can be seen in the feature importance distribution of the other top classifiers in Appendix B. Figure A1 shows that bmean holds the highest significance for the model, with its median value approaching 0.4. This observation was consistent across all cases presented in Appendix B as can be seen in Figures A3 and A4, except for Figure A2, where bmax exhibited a significantly larger interquartile range despite having a higher median value than bmean. In the remaining cases, the ranking of parameters by significance followed the sequence of bmax, bmean, bmedian, and bmin. Additionally, it is noteworthy that the variability in bmin was consistently the lowest across the figures, reinforcing its lesser impact on the model's performance.

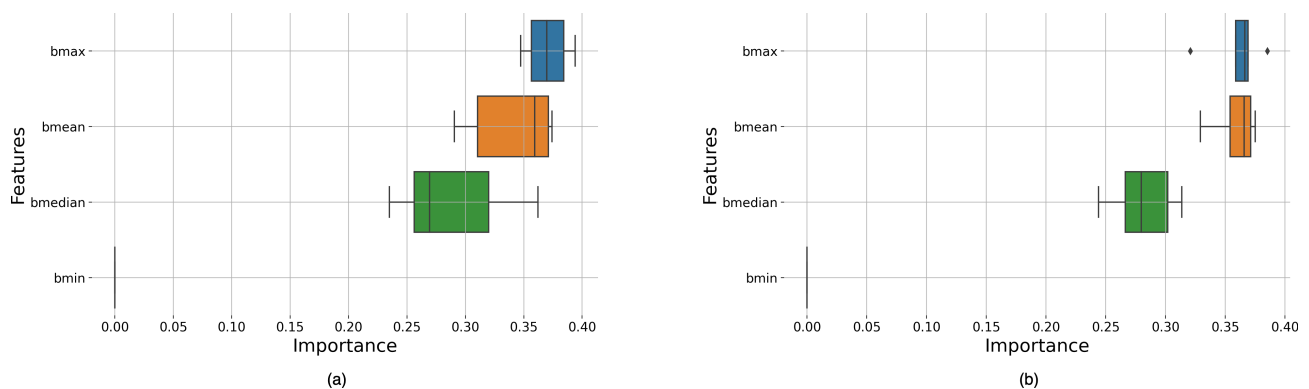


Figure 5. Feature importance distribution for classifiers with best performance. (a) RF classifier with Boxcar window, size 128, and an overlap of 64, (b) RF classifier with Bartlett window, size 128, and an overlap of 64.

Figure 1 illustrates that both healthy and unhealthy signal spectrograms have low energy distribution within a similar range, whereas the medium and maximum energy distribution have a more comprehensive range and higher variation in intensities. As a result, the average intensities also exhibited a more pronounced difference between the healthy and unhealthy signals.

4.2. Performance of DL Classifiers

Figure 6 illustrates the results obtained from the different DL models (DenseNet, ResNet, VGG16, VGG19, and ViT) used in the analysis with various windows. These results showed higher variation between the windows, with the highest performance in terms of metrics being the Hamming window (Accuracy of 81.2% with ViT Small), regardless of the architecture. This window is well known for its ability to suppress the side lobes as sharply as possible while keeping the main one narrow enough [44]. The mentioned property allows one to extract most features from each window position. Among the architectures, ViT Small showed the best results. Figure 7a shows the receiver operating characteristic (ROC) curves for the ViT model with different windows. Figure 7b shows the ROC curves of different DL models with the Hamming window. Detailed results are provided in Table A4.

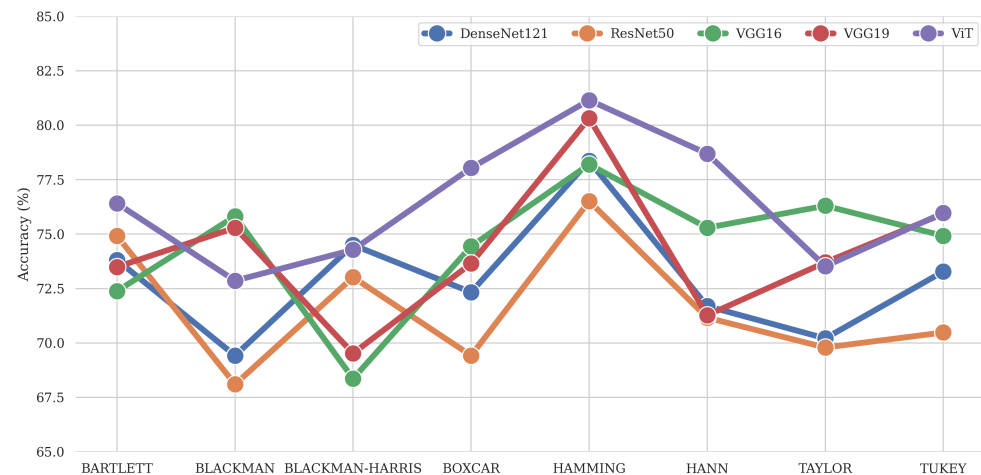


Figure 6. Classification accuracies of the analyzed DL architectures for all test windows.

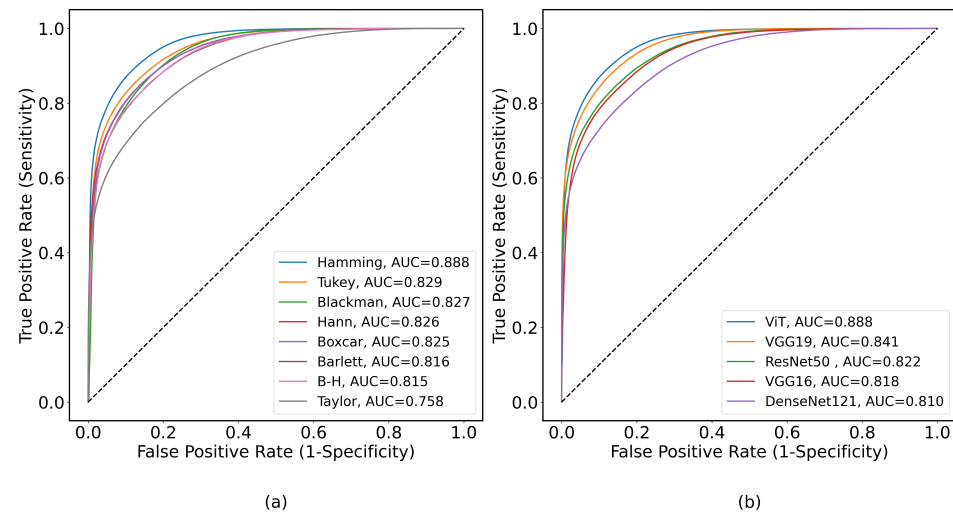


Figure 7. ROC curves for binary classification of ERG signals with corresponding AUCs for all tested windows using ViT model in (a) and for all tested models with Hamming window in (b).

5. Discussion

STFT is a method based on the FT proposed as a solution to the lack of temporal information from the classical FT method. The use of windows that slide across the signal helps extract the spectral and temporal information of the signal and present it in the form of a spectrogram. However, since it is impossible to obtain high temporal and spectral resolution simultaneously, it is necessary to determine which window, window size, and overlap between the sliding window would yield the most optimal features of the ERG signal for classification. As shown in Appendix A, there was very little difference in the results for the window functions studied. However, we can see significant differences in performance depending on window size and overlap.

The classifiers with the best results were those with larger window sizes; given that larger window sizes provided better frequency resolutions, this could indicate that signals with higher frequency resolutions produce the most optimal features.

We can also observe that the RF-based models outperformed the DT-based models. This was expected given that RF uses an ensemble of DTs, hence being able to capitalize on multiple trees rather than a single individual tree to make its predictions. Table A3 shows that Boxcar and Bartlett have the highest mean scores and the most significant variance, AUC 69.3% and 64.4%, due to these windows having multiple high-value score classifiers. As reported in Tables A1 and A2, it can be seen that all windows have nearly identical scores, with Boxcar and Bartlett having better classification Accuracy with higher

scores than the others at 70.8%, suggesting that these windows might have slightly better effects on the extracted features than the rest. Thus, the models' results and performance regarding the window functions were similar. However, there were differences in the metrics regarding the window size and overlaps. One plausible explanation for this was that the window function itself did not affect the signal significantly as much as the size and overlap of the windows do because the latter two determine the signal's resolution. This effect is described in Appendix A where all windows have the same maximum value for each metric; however, the Boxcar window, which is a square window and does not change anything in the signal, has the highest mean and variance because it has multiple window sizes with the maximum value metrics. On the other hand, the Bartlett and Boxcar window functions have the best performance among the analyzed window types. The Bartlett window, with an almost triangular shape, is known for being used to prevent the generation of too many oscillations in the frequency domain [28]. The results of using the window are the same as for the basic Boxcar window. This was likely to be due to the relatively small and straightforward (interpreted) feature space. The STD analysis in Table A3 also shows that the smallest values are obtained for the Hamming and Hann window function cases for the RF decision algorithm and the Hann and Tukey window function for the DT algorithm.

This is different from DL methods. We assume that it is about automatic feature extraction. Figure 6 shows that the difference between the windowed features is more significant than the manual feature extraction approach. DL methods extract more features from the signal than can be extracted manually. Due to this, the average metrics values are higher than for the manual feature extraction cases. Comparing both approaches, we can conclude the perceptiveness of the modern DL approaches. However, we must also note that manual feature extraction with STFT can be considered the most explainable approach. Given that DL architectures do a better job at learning and extracting features at a wider scale, they can still be used as feature extractors alongside a classical model for the final classification. This will be explored in future research as it provides the potential to expand the feature space without the need for manual feature extraction.

6. Conclusions

This study investigated various window functions for STFT calculation (and spectrogram generation) to classify ERG signals. The spectrogram images were extracted using several combinations of well-known window functions, window sizes, and window overlap values, and the manual features were extracted to train the classical ML model using the same methods. Based on the comparison of the results of the two approaches, DL can be recommended. In terms of Accuracy, the ViT Small architecture with the Hamming window showed the best performance among the combinations of DL models with window types (81%). However, if manual feature extraction is required, a RF with a rectangular Boxcar window or Bartlett window can be recommended as an alternative to the DL approach. In the study, the mean Accuracy in these cases was 67.5%.

The results of the analysis of ERG using Short-Time Fourier Transform and ML techniques are, of course, dependent on the size of the dataset used for training, thus necessitating a large original sample. To address this limitation, expanding dataset volumes and promoting open data sharing within the electrophysiology community could enhance the diversity and representation of synthetic waveforms. Although these preliminary results have been generated with a relatively small sample set, it is one of the largest in the world by data quantity [38]. Moreover, we are actively developing larger synthetic datasets to support clinical studies [45].

Another limitation of this study was the feature space used in the ML approach; the analysis only used four features: the minimum, maximum, median, and mean brightness of the spectrogram; this could be a reason why there was little to no difference between the windows, this limitation was not encountered in the DL approach as the automatic feature extraction of DL architectures gave it access to a more prominent feature space.

Hence, in future studies, we will look at expanding the feature space for the ML classifiers, as expanding the feature space for manual feature extraction approaches contributes to improved Accuracy and other metrics while maintaining the overall explainability of the system. This solution is essential for the developed algorithm to be easily understandable in medical applications.

Author Contributions: Conceptualization, F.A., A.Z. and M.K.; methodology, F.A., A.Z. and M.K.; software, F.A., M.K. and A.D.; validation, F.A., M.K. and A.Z.; formal analysis, F.A. and M.K.; investigation, F.A.; writing—original draft preparation, F.A. and M.K.; writing—review and editing, M.A.A.-m., V.B., P.A.C. and M.R.; visualization, L.D., F.A. and M.K.; supervision, M.A.A.-m. and A.M.; project administration, A.Z.; funding acquisition, A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Zhdanov, A.E.; Dolganov, A.Y.; Borisov, V.I.; Lucian, E.; Bao, X.; Kazajkin, V.N.; Ponomarev, V.O.; Lizunov, A.V.; Ivliev, S.A. 355 OculusGraphy: Pediatric and Adults Electroretinograms Database, 2020. <https://doi.org/10.21227/y0fh-5v04>, accessed on 1 April 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Experiment Results in Details

Tables A1 and A2 show the experiment results in details for different combinations of signal window type and decision algorithm (DT or RF) along with statistical analysis of the results in Table A3. Detailed results for the DL experiment are shown in Table A4.

Table A1. Decision Tree results.

Size–Overlap	Accuracy	F1	AUC	Precision	Recall
Hamming DT					
128-64	0.667	0.664	0.671	0.667	0.671
128-32-64-16	0.625	0.619	0.633	0.625	0.633
64-8	0.667	0.664	0.671	0.667	0.671
64-4	0.625	0.619	0.633	0.625	0.633
64-2	0.667	0.664	0.671	0.667	0.671
32-16-8-2	0.625	0.619	0.633	0.625	0.633
4-2	0.625	0.619	0.633	0.625	0.633
Hann DT					
128-64-4-2	0.667	0.664	0.671	0.667	0.671
Boxcar DT					
128-64	0.667	0.664	0.671	0.667	0.671
128-32	0.625	0.619	0.633	0.625	0.633
128-16	0.667	0.664	0.671	0.667	0.671
128-8-64-16	0.625	0.619	0.633	0.625	0.633
64-8	0.667	0.664	0.671	0.667	0.671
64-4	0.667	0.664	0.671	0.667	0.671
64-2-4-2	0.625	0.619	0.633	0.625	0.633
Bartlett DT					
128-64-128-8	0.667	0.664	0.671	0.667	0.671
128-4-4-2	0.625	0.619	0.633	0.625	0.633

Table A1. *Cont.*

Size–Overlap	Accuracy	F1	AUC	Precision	Recall
Blackman DT					
128-64	0.667	0.664	0.671	0.667	0.671
128-32-4-2	0.625	0.619	0.633	0.625	0.633
64-4	0.667	0.664	0.671	0.667	0.671
64-2	0.667	0.664	0.671	0.667	0.671
Blackman–Harris DT					
128-64	0.667	0.664	0.671	0.667	0.671
128-32-4-2	0.625	0.619	0.633	0.625	0.633
64-32	0.667	0.664	0.671	0.667	0.671
Taylor DT					
128-64	0.667	0.664	0.671	0.667	0.671
128-32-4-2	0.625	0.619	0.633	0.625	0.633
64-8	0.667	0.664	0.671	0.667	0.671
64-4	0.625	0.619	0.633	0.625	0.633
64-2	0.667	0.664	0.671	0.667	0.671
Tukey DT					
128-64-4-2	0.667	0.664	0.671	0.667	0.671

Table A2. Random Forest results.

Size–Overlap	Accuracy	F1	AUC	Precision	Recall
Hamming RF					
128-64-32-4	0.708	0.704	0.722	0.708	0.722
32-2	0.708	0.704	0.722	0.708	0.722
16-8-4-2	0.667	0.657	0.688	0.667	0.688
Hann RF					
128-64	0.708	0.704	0.722	0.708	0.722
128-32-4-2	0.667	0.657	0.688	0.667	0.688
Boxcar RF					
128-64	0.708	0.704	0.722	0.708	0.722
128-32	0.667	0.657	0.688	0.667	0.688
128-16	0.708	0.704	0.722	0.708	0.722
128-8	0.708	0.704	0.722	0.708	0.722
128-4	0.667	0.657	0.688	0.667	0.688
128-2	0.708	0.704	0.722	0.708	0.722
64-32-64-4	0.667	0.657	0.688	0.667	0.688
64-2	0.708	0.704	0.722	0.708	0.722
32-16	0.667	0.657	0.688	0.667	0.688
32-8	0.667	0.657	0.688	0.667	0.688
32-4	0.667	0.657	0.688	0.667	0.688
32-2	0.708	0.704	0.722	0.708	0.722
16-8-4-2	0.667	0.657	0.688	0.667	0.688
Bartlett RF					
128-64	0.708	0.704	0.722	0.708	0.722
128-32	0.667	0.657	0.688	0.667	0.688
128-16	0.708	0.704	0.722	0.708	0.722
128-8	0.708	0.704	0.722	0.708	0.722
128-4	0.667	0.657	0.688	0.667	0.688
128-2	0.708	0.704	0.722	0.708	0.722
64-32-64-4	0.667	0.657	0.688	0.667	0.688

Table A2. Cont.

Size–Overlap	Accuracy	F1	AUC	Precision	Recall
Bartlett RF					
64-2	0.708	0.704	0.722	0.708	0.722
32-16–32-4	0.667	0.657	0.688	0.667	0.688
32-2	0.708	0.704	0.722	0.708	0.722
16-8–4-2	0.667	0.657	0.688	0.667	0.688
Blackman RF					
128-64	0.708	0.704	0.722	0.708	0.722
128-32–4-2	0.667	0.657	0.688	0.667	0.688
Blackman–Harris RF					
128-64	0.708	0.704	0.722	0.708	0.722
128-32–4-2	0.667	0.657	0.688	0.667	0.688
Taylor RF					
128-64	0.708	0.704	0.722	0.708	0.722
128-32–4-2	0.667	0.657	0.688	0.667	0.688
Tukey RF					
128-64	0.667	0.664	0.671	0.667	0.671
128-32–32-4	0.667	0.657	0.688	0.667	0.688
32-2	0.708	0.704	0.722	0.708	0.722
16-8	0.667	0.657	0.688	0.667	0.688
16-4	0.625	0.608	0.651	0.625	0.651
16-2–4-2	0.667	0.657	0.688	0.667	0.688

Table A3. Random Forest and Decision Tree averaged results.

Model	Window	Metric	Accuracy	F1	AUC	Precision	Recall
RF	Hamming	Mean	0.668	0.659	0.689	0.671	0.691
		STD	0.009	0.010	0.008	0.013	0.010
	Hann	Mean	0.668	0.659	0.689	0.667	0.687
		STD	0.009	0.010	0.008	0.013	0.014
	Boxcar	Mean	0.675	0.666	0.693	0.679	0.697
		STD	0.017	0.019	0.015	0.019	0.016
	Bartlett	Mean	0.675	0.666	0.693	0.679	0.697
		STD	0.017	0.019	0.015	0.019	0.016
	Blackman	Mean	0.663	0.654	0.681	0.669	0.689
		STD	0.018	0.018	0.021	0.009	0.008
	Blackman–Harris	Mean	0.667	0.658	0.687	0.669	0.689
		STD	0.013	0.013	0.014	0.009	0.008
	Taylor	Mean	0.667	0.657	0.687	0.667	0.687
		STD	0.013	0.015	0.011	0.019	0.018
	Tukey	Mean	0.670	0.662	0.691	0.667	0.687
		STD	0.013	0.014	0.010	0.013	0.012
DT	Hamming	Mean	0.633	0.628	0.641	0.633	0.641
		STD	0.017	0.018	0.015	0.017	0.015
	Hann	Mean	0.627	0.621	0.635	0.627	0.635
		STD	0.009	0.010	0.008	0.009	0.008
	Boxcar	Mean	0.637	0.632	0.644	0.635	0.642
		STD	0.019	0.021	0.018	0.018	0.017
	Bartlett	Mean	0.637	0.632	0.644	0.631	0.638
		STD	0.019	0.021	0.018	0.020	0.019
	Blackman	Mean	0.631	0.626	0.639	0.633	0.641
		STD	0.015	0.016	0.014	0.017	0.015
	Blackman–Harris	Mean	0.629	0.623	0.637	0.629	0.637
		STD	0.013	0.014	0.011	0.013	0.011
	Taylor	Mean	0.633	0.628	0.641	0.633	0.641
		STD	0.017	0.018	0.015	0.017	0.015
	Tukey	Mean	0.627	0.621	0.635	0.629	0.637
		STD	0.009	0.010	0.008	0.013	0.011

Table A4. Results of DL models.

Model	Window	Accuracy	F1	AUC	Precision	Recall
DenseNet121	Bartlett	0.738	0.735	0.828	0.732	0.738
DenseNet121	Blackman	0.694	0.685	0.760	0.684	0.694
DenseNet121	Blackman–Harris	0.745	0.734	0.790	0.732	0.745
DenseNet121	Boxcar	0.723	0.719	0.779	0.717	0.723
DenseNet121	Hamming	0.784	0.764	0.810	0.766	0.784
DenseNet121	Hann	0.717	0.705	0.798	0.704	0.717
DenseNet121	Taylor	0.702	0.692	0.745	0.691	0.702
DenseNet121	Tukey	0.733	0.728	0.814	0.725	0.733
ResNet50	Bartlett	0.749	0.738	0.813	0.736	0.749
ResNet50	Blackman	0.681	0.664	0.784	0.670	0.681
ResNet50	Blackman–Harris	0.730	0.721	0.817	0.719	0.730
ResNet50	Boxcar	0.694	0.689	0.766	0.687	0.694
ResNet50	Hamming	0.765	0.751	0.822	0.750	0.765
ResNet50	Hann	0.712	0.692	0.803	0.698	0.712
ResNet50	Taylor	0.698	0.688	0.775	0.687	0.698
ResNet50	Tukey	0.705	0.698	0.787	0.696	0.705
VGG16	Bartlett	0.724	0.713	0.760	0.712	0.724
VGG16	Blackman	0.758	0.753	0.823	0.750	0.758
VGG16	Blackman–Harris	0.684	0.669	0.737	0.672	0.684
VGG16	Boxcar	0.744	0.737	0.779	0.734	0.744
VGG16	Hamming	0.782	0.773	0.818	0.770	0.782
VGG16	Hann	0.753	0.748	0.821	0.745	0.753
VGG16	Taylor	0.763	0.753	0.781	0.750	0.763
VGG16	Tukey	0.749	0.748	0.793	0.746	0.749
VGG19	Bartlett	0.735	0.724	0.815	0.722	0.735
VGG19	Blackman	0.753	0.746	0.788	0.743	0.753
VGG19	Blackman–Harris	0.695	0.685	0.728	0.685	0.695
VGG19	Boxcar	0.737	0.731	0.795	0.728	0.737
VGG19	Hamming	0.803	0.791	0.841	0.788	0.803
VGG19	Hann	0.713	0.703	0.756	0.702	0.713
VGG19	Taylor	0.737	0.725	0.772	0.724	0.737
VGG19	Tukey	0.759	0.745	0.817	0.744	0.759
ViT Small	Bartlett	0.764	0.752	0.816	0.750	0.764
ViT Small	Blackman	0.729	0.722	0.827	0.719	0.729
ViT Small	Blackman–Harris	0.743	0.740	0.815	0.737	0.743
ViT Small	Boxcar	0.780	0.776	0.825	0.774	0.780
ViT Small	Hamming	0.812	0.801	0.888	0.798	0.812
ViT Small	Hann	0.787	0.771	0.826	0.771	0.787
ViT Small	Taylor	0.735	0.731	0.758	0.729	0.735
ViT Small	Tukey	0.760	0.748	0.829	0.746	0.760

Appendix B. Feature Importance Distributions of Other Top Classifiers

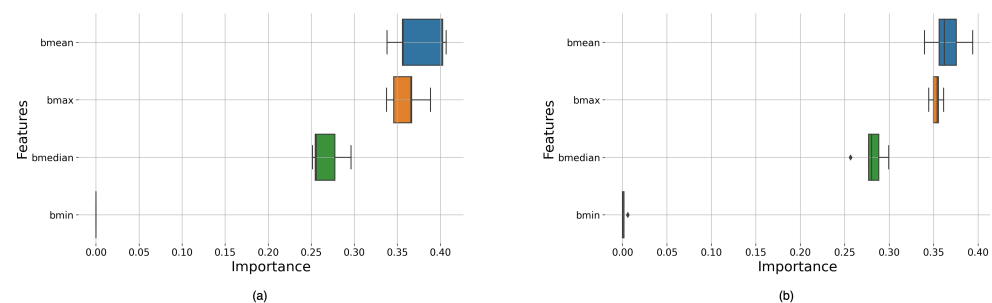


Figure A1. (a) RF classifier with Boxcar window, size 128, and an overlap of 32, (b) RF classifier with Bartlett window, size 128, and an overlap of 32.

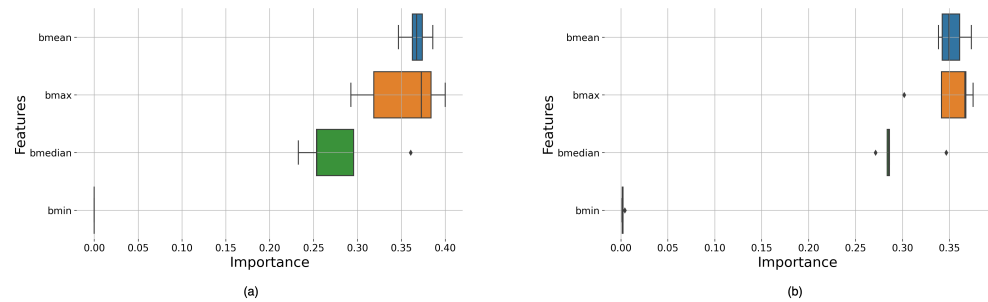


Figure A2. (a) RF classifier with Boxcar window, size 128, and an overlap of 16, (b) RF classifier with Bartlett window, size 128, and an overlap of 16.

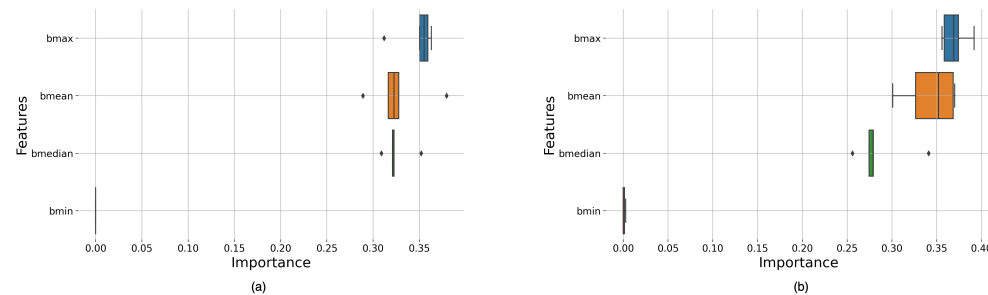


Figure A3. (a) RF classifier with Boxcar window, size 64, and an overlap of 2, (b) RF classifier with Bartlett window, size 64, and an overlap of 2.

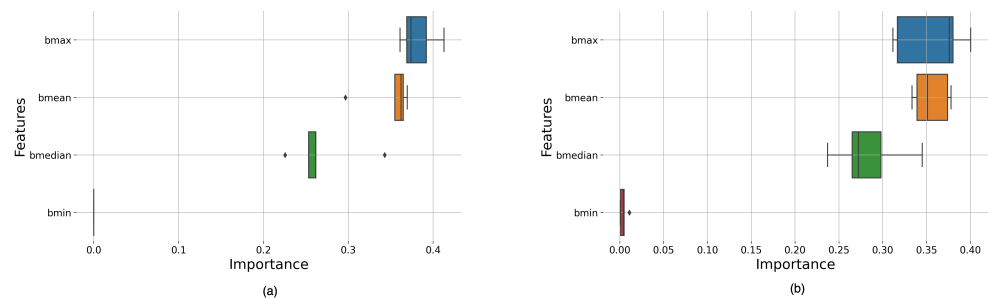


Figure A4. (a) RF classifier with Boxcar window, size 32, and an overlap of 2, (b) RF classifier with Bartlett window, size 32, and an overlap of 2.

References

1. Robson, A.G.; Frishman, L.J.; Grigg, J.; Hamilton, R.; Jeffrey, B.G.; Kondo, M.; Li, S.; McCulloch, D.L. ISCEV Standard for Full-Field Clinical Electroretinography (2022 Update). *Doc. Ophthalmol.* **2022**, *144*, 165–177. [[CrossRef](#)] [[PubMed](#)]
2. Behbahani, S.; Ahmadi, H.; Rajan, S. Feature Extraction Methods for Electroretinogram Signal Analysis: A Review. *IEEE Access* **2021**, *9*, 116879–116897. [[CrossRef](#)]
3. Balicka, A.; Trbolová, A.; Vrbová, T. Electroretinography (A Review). *Folia Vet.* **2016**, *60*, 53–58. [[CrossRef](#)]
4. Wood, A.; Margrain, T.; Binns, A.M. Detection of Early Age-Related Macular Degeneration Using Novel Functional Parameters of the Focal Cone Electroretinogram. *PLoS ONE* **2014**, *9*, e96742. [[CrossRef](#)] [[PubMed](#)]
5. Nebbioso, M.; Grenga, R.; Karavitis, P. Early Detection of Macular Changes With Multifocal ERG in Patients on Antimalarial Drug Therapy. *J. Ocul. Pharmacol. Ther.* **2009**, *25*, 249–258. [[CrossRef](#)]
6. Maa, A.Y.; Feuer, W.J.; Davis, C.Q.; Pillow, E.K.; Brown, T.D.; Caywood, R.M.; Chasan, J.E.; Fransen, S.R. A novel device for accurate and efficient testing for vision-threatening diabetic retinopathy. *J. Diabetes Complicat.* **2016**, *30*, 524–532. [[CrossRef](#)]
7. Gur, M.; Zeevi, Y. Frequency-Domain Analysis of the Human Electroretinogram. *J. Opt. Soc. Am.* **1980**, *70*, 53. [[CrossRef](#)]
8. Zhdanov, A.; Dolganov, A.; Zanca, D.; Borisov, V.; Ronkin, M. Advanced Analysis of Electroretinograms Based on Wavelet Scalogram Processing. *Appl. Sci.* **2022**, *12*, 12365. [[CrossRef](#)]
9. Zhdanov, A.E.; Borisov, V.I.; Dolganov, A.Y.; Lucian, E.; Bao, X.; Kazajkin, V.N. OculusGraphy: Filtering of Electroretinography Response in Adults. In Proceedings of the 2021 IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM), Souza, The Altai Republic, 30 June–4 July 2021; pp. 395–398. [[CrossRef](#)]
10. Constable, P.A.; Lim, J.K.; Thompson, D.A. Retinal electrophysiology in central nervous system disorders. A review of human and mouse studies. *Front. Neurosci.* **2023**, *17*, 1215097. [[CrossRef](#)]
11. Gröchenig, K. *Foundations of Time-Frequency Analysis*; Springer Science & Business Media: New York, NY, USA, 2013.

12. Cooley, J.W.; Tukey, J.W. An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. Comput.* **1965**, *19*, 297–301. [CrossRef]
13. Gur, M.; Gath, I. Time and Frequency Analysis of Simultaneously Recorded Corneal and Non-Corneal Electroretinogram. *J. Biomed. Eng.* **1979**, *1*, 172–174. [CrossRef]
14. Van Der Torren, K.; Groeneweg, G.; Van Lith, G. Measuring Oscillatory Potentials: Fourier Analysis. *Doc. Ophthalmol.* **1988**, *69*, 153–159. [CrossRef]
15. Li, X.X.; Yuan, N. Measurement of the Oscillatory Potentials of the Electroretinogram in the Domains of Frequency and Time. *Doc. Ophthalmol.* **1990**, *76*, 65–71. [CrossRef]
16. Sieving, P.A.; Arnold, E.B.; Jamison, J.; Liepa, A.; Coats, C. Submicrovolt Flicker Electroretinogram: Cycle-by-Cycle Recording of Multiple Harmonics with Statistical Estimation of Measurement Uncertainty. *Investig. Ophthalmol. Vis. Sci.* **1998**, *39*, 1462–1469.
17. Hassan-Karimi, H.; Jafarzadehpur, E.; Blouri, B.; Hashemi, H.; Sadeghi, A.Z.; Mirzajani, A. Frequency Domain Electroretinography in Retinitis Pigmentosa versus Normal Eyes. *J. Ophthalmic Vis. Res.* **2012**, *7*, 34–38.
18. Vladimirova, Z.M. Assessment of the Amplitude-Frequency Characteristics of the Retina with Its Stimulation by Flicker and Chess Pattern-Reversed Incentives and Their Use to Obtain New Formalized Signs of Retinal Pathologies. *Biomed. J. Sci. Tech. Res.* **2019**, *19*, 14575–14583. [CrossRef]
19. Alaql, A.M. Analysis and Processing of Human Electroretinogram. Master's Thesis, University of South Florida, Tampa, FL, USA, 2016.
20. Erkamaz, O.; Senyer Yapici, İ.; Uzun Arslan, R. Effects of Obesity on Time-Frequency Components of Electroretinogram Signal Using Continuous Wavelet Transform. *Biomed. Signal Process. Control* **2021**, *66*, 102398. [CrossRef]
21. Behbahani, S.; Moridani, M.K.; Ramezani, A.; Sabbaghi, H. Investigating the frequency characteristics of the electroretinogram signal in patients with central retinal vein occlusion. *Med. Sci. J.* **2021**, *31*, 205–217. [CrossRef]
22. Kulyabin, M.; Zhdanov, A.; Dolganov, A.; Maier, A. Optimal Combination of Mother Wavelet and AI Model for Precise Classification of Pediatric Electroretinogram Signals. *Sensors* **2023**, *23*, 5813. [CrossRef]
23. Heisenberg, W. The Actual Content of Quantum Theoretical Kinematics and Mechanics. 1983. Available online: <https://ntrs.nasa.gov/citations/19840008978> (accessed on 1 January 2020).
24. Albasu, F.B.; Dey, S.; Dolganov, A.Y.; Hamzaoui, O.E.; Mustafa, W.M.; Zhdanov, A.E. OculusGraphy: Description and Time Domain Analysis of Full-Field Electroretinograms Database. In Proceedings of the 2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT), Yekaterinburg, Russia, 15–17 May 2023; pp. 64–67. [CrossRef]
25. Kuze, M.; Uji, Y. Comparison between Dawson, Trick, and Litzkow electrode and contact lens electrodes used in clinical electroretinography. *Jpn. J. Ophthalmol.* **2000**, *44*, 374–380. [CrossRef]
26. Yip, Y.W.Y.; Man, T.C.; Pang, C.P.; Brelén, M.E. Improving the quality of electroretinogram recordings using active electrodes. *Exp. Eye Res.* **2018**, *176*, 46–52. [CrossRef]
27. Lemaitre, G.; Nogueira, F. Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
28. Prabhu, K.M.M. *Window Functions and Their Applications in Signal Processing*; Taylor & Francis: Abingdon, UK, 2014. [CrossRef]
29. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *arXiv* **2015**. [CrossRef]
30. Pfister, H. Discrete-Time Signal Processing. Lecture Note. 2017. Available online: <http://pfister.ee.duke.edu/courses/ece485/dtsp.pdf> (accessed on 1 January 2020).
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
32. Zheng, A.; Casari, A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*, 1st ed.; O'Reilly: Beijing, China; Boston, MA, USA, 2018.
33. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
34. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
37. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**. [CrossRef]
38. Kulyabin, M.; Zhdanov, A.; Dolganov, A.; Ronkin, M.; Borisov, V.; Maier, A. Enhancing Electroretinogram Classification with Multi-Wavelet Analysis and Visual Transformer. *Sensors* **2023**, *23*, 8727. [CrossRef]
39. Wu, K.; Zhang, J.; Peng, H.; Liu, M.; Xiao, B.; Fu, J.; Yuan, L. TinyViT: Fast Pretraining Distillation for Small Vision Transformers. In Proceedings of the Computer Vision—ECCV 2022; Lecture Notes in Computer Science, Tel Aviv, Israel, 23–27 October 2022; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer: Cham, Switzerland, 2022; pp. 68–85. [CrossRef]
40. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv* **2020**. [CrossRef]

41. Wightman, R. PyTorch Image Models. 2019. Available online: <https://github.com/rwightman/pytorch-image-models> (accessed on 1 January 2020). [CrossRef]
42. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
43. Mumuni, A.; Mumuni, F. Data augmentation: A comprehensive survey of modern approaches. *Array* **2022**, *16*, 100258. [CrossRef]
44. Harris, F.J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **1978**, *66*, 51–83. [CrossRef]
45. Kulyabin, M.; Zhdanov, A.; Maier, A.; Loh, L.; Estevez, J.J.; Constable, P.A. Generating Synthetic Light-Adapted Electroretinogram Waveforms Using Artificial Intelligence to Improve Classification of Retinal Conditions in Under-Represented Populations. *J. Ophthalmol.* **2024**, *2024*, 1990419. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.