Multiclass Classification of Retinal and Optic Nerve Disorders Using Time-Domain Features from Band-Separated PERG Signals

Sathees Kumar Nataraj
Department of Mechatronics Engineering, University of Technology Bahrain Salmabad, Bahrain
E-mail: sknataraj@utb.edu.bh

Keywords: pattern electroretinogram (PERG), time-domain features, frequency band analysis, multiclass classification, AdaBoost, machine learning, neural networks

Received: June 11, 2025

Pattern Electroretinogram (PERG) signals offer critical insights into the functional status of retinal ganglion cells and are widely used for diagnosing optic nerve and retinal pathologies. While existing research has predominantly emphasized binary classification or full-spectrum signal analysis, limited studies have addressed the use of frequency band-based feature extraction in multiclass classification scenarios. This study introduces a novel approach that segments PERG signals into defined frequency bands and extracts a comprehensive set of time-domain features to support five-class classification. The dataset, originally comprising 15 diagnostic classes, was consolidated into five clinically relevant categories to ensure balanced representation. From each frequency band, more than 30 time-domain features were derived per eye, capturing key waveform characteristics. A consolidated dataset of five clinically relevant diagnostic classes have been used. Multiple machine learning models were evaluated, with AdaBoost-ECOC achieving the best performance: 99.11% ± 1.26 accuracy, 82.92% ± 0.59 precision, and 83.03% ± 0.55 F1-score. These results demonstrate the effectiveness and efficiency of frequency-based time-domain feature extraction for PERG signal classification.

Povzetek: Študija prikazuje, da je mogoče PERG signale uspešno razvrščati z uporabo značilk iz različnih frekvenčnih pasov.

1 Introduction

Ocular electrophysiology provides an objective method to evaluate visual function by measuring electrical responses from the retina and visual pathways in response to light or pattern stimuli. Core diagnostic tools such as flash electroretinography (ERG) and visual evoked potentials (VEP) have been instrumental in clinical evaluations [1]. Flash ERG measures photoreceptor (a-wave) and bipolar cell (b-wave) activity, while VEP evaluates cortical responses, thereby assisting in optic nerve assessment [2]–[5]. While imaging techniques such as optical coherence tomography (OCT) and visual field analysis provide structural insights, they often detect pathology only in advanced stages. Electrophysiological techniques, including multifocal ERG (mfERG) and multifocal VEP, enable earlier detection by identifying functional impairments preceding structural changes [6].

1.1 Background

Electrophysiological signals represent the measurable electrical activity of biological tissues, including neurons and sensory organs [7]. These signals can be captured using non-invasive skin electrodes, invasive probes, or multielectrode arrays [8]–[10]. In ophthalmology, ERG and its variant pattern electroretinography (PERG) are widely used [11].

Electrophysiological signals are the measurable electrical activities generated by biological cells and tissues, such as neurons, muscle fibers, and sensory organs [7][8]. These signals can be captured using non-invasive skin electrodes, invasive probes, or multielectrode arrays. In ophthalmology, ERG and its derivative, pattern electroretinography (PERG), are used to record retinal responses [9] [10]. PERG is particularly effective in detecting early dysfunction of retinal ganglion cells (RGCs), which is critical in glaucoma diagnosis. It employs contrast-reversing checkerboard stimuli focused on the macula, generating P50 and N95 waveforms that reflect outer retinal and RGC activity, respectively. [11]. Pattern electroretinography (PERG) stands out for its sensitivity in detecting early dysfunction of retinal ganglion cells (RGCs), particularly in glaucoma [12]. PERG utilizes contrast-reversing checkerboard patterns focused on the macula to generate responses characterized by P50 and N95 waveforms, which reflect outer retinal and RGC activity, respectively [13], [14]. Recordings are obtained using corneal electrodes or skin-mounted sensors, with patients maintaining fixation on a central target [15]. These signals, however, are sensitive to noise from eye movements and fixation errors. To address this, standardized acquisition protocols set by the International Society for Clinical Electrophysiology of Vision (ISCEV) are combined with modern signal processing techniques such as signal averaging and noise reduction [16], [17].

Pattern electroretinography (PERG) has become a critical tool for assessing retinal and optic nerve function, with computational preprocessing techniques playing a pivotal role in enhancing signal quality. Recent studies have demonstrated the effectiveness of multi-wavelet

transformations and deep learning architectures in isolating PERG components from noise caused by eye movements, electrical interference, and poor fixation [1], [18]. For example, research using Ricker, Gaussian, and Morlet wavelets achieved up to 91% classification accuracy for specific ERG protocols by creating 3D wavelet stacks that capture complementary signal features [19]. These technical advancements have enabled more reliable detection of pathological patterns across diverse conditions from common disorders like glaucoma and retinitis pigmentosa to rare entities such as birdshot chorioretinopathy and paracentral acute middle maculopathy [1].

In addition to differentiating macular diseases from optic nerve disorders, PERG has gained traction for early detection in inherited maculopathies [20], [21]. Canedo et al. (2023) proposed novel wavelet-based modeling approaches to improve PERG signal analysis [22]. Studies by Asanad and Karanjia [20] and Holder et al. [21] also highlight its clinical value. Although ISCEV standards (e.g., Bach et al. [16]) ensure consistency, their implementation in automated, multi-disease classification systems remains limited [23].

Most existing machine learning (ML) studies apply binary classification (e.g., glaucoma vs. healthy) using ERG or OCT data, with few utilizing PERG for multi-pathology classification [24]. Although wavelet-based features like Daubechies transforms have shown promise in glaucoma staging [25]. their application to broader PERG-based disease classification is limited. Furthermore, the impact of varying band-pass filter ranges (e.g., 1–100 Hz vs. 5–50 Hz) remains underexplored [26].

1.2 Proposed methodology

While existing research has established machine learning to diagnose individual retinal diseases, most focus on binary classification tasks using limited datasets [27], This study utilizes the IOBA-PERG dataset, a comprehensive repository of 1,354 transient pattern electroretinography (PERG) responses [1]. The study address broader diagnostic challenges through two key approaches:

a. Frequency specific signal segmentation

Advanced signal processing techniques are applied to perform a detailed analysis of waveform morphology, focusing on features such as N35–P50–N95 latencies, oscillatory potentials, and spectral characteristics. Drawing from methodologies used in EEG signal processing, PERG signals are divided into five clinically relevant frequency bands to enhance diagnostic interpretability as shown in Figure 1.

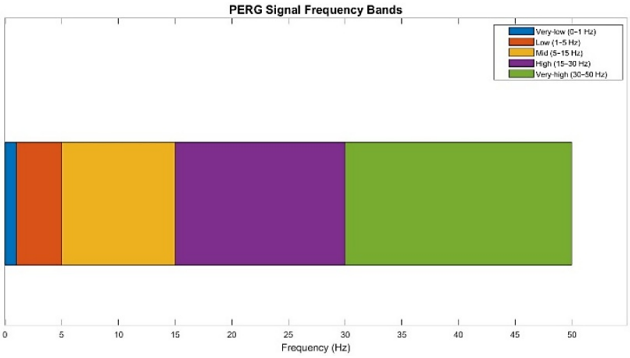

Figure 1: Block diagram of the proposed classification of retinal pathologies using machine learning

The signals are segmented into five clinically relevant frequency bands to enhance interpretability:

1. **Very-low (0–1 Hz):** Captures slow oscillatory components associated with metabolic retinal processes
2. **Low (1–5 Hz):** Isolates the N95 wave, indicative of inner retinal dysfunction
3. **Mid (5–15 Hz):** Emphasizes the P50 wave, representing macular function
4. **High (15–30 Hz):** Associated with transient glaucomatous responses
5. **Very-high (30–50 Hz):** May encode high-frequency oscillations in ischemic retinopathies

b. Multi-class classification framework

A frequency-optimized Machine Learning (ML) model [28] is developed to classify 15 retinal pathologies based on features from both time and frequency domains. The goal is to establish a machine learning-driven framework that incorporates band-optimized PERG features for the diagnosis of various retinal diseases, including early-stage glaucoma, diabetic macular edema, and retinitis pigmentosa. The study systematically evaluates the relevance of each frequency band, highlighting the advantages of frequency-domain over time-domain analysis.

1.3 Significance and contribution

Literature in this domain affirms the clinical relevance of PERG signals in assessing retinal function, yet significant gaps remain in analytical techniques used for classification and early detection. Most notably, while machine learning and artificial intelligence have been applied to ocular diagnostics, frequency-domain analysis of PERG signals remains underexplored. This research addresses that gap by investigating the diagnostic value of PERG frequency components through an ML framework trained on an open-access dataset from PhysioNet.

The proposed system, illustrated in Figure 2, employs machine learning algorithms to analyze the IOBA-PERG dataset and deliver multi-pathology classification outcomes. The implementation methodology and results are discussed in detail in the following sections of this article.

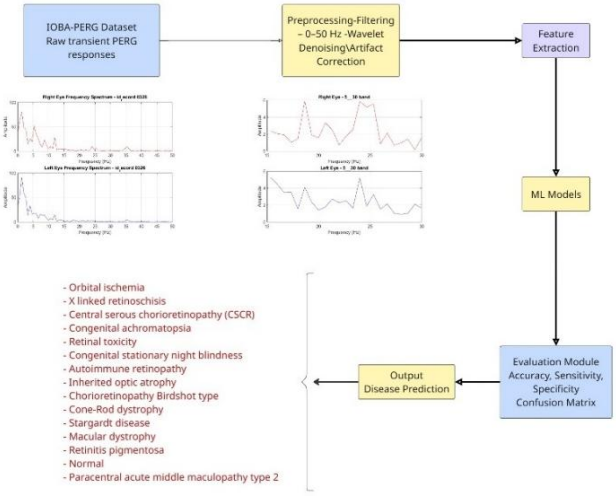

Figure 2: Block diagram of the proposed classification of retinal pathologies using machine learning

2 Methodology

This study aims to develop a robust multiclass classification framework for Pattern Electroretinogram (PERG) signals by leveraging frequency band-based feature extraction and supervised machine learning techniques. The research is driven by three core objectives:

1. to extract and evaluate meaningful time domain features from selected PERG frequency bands
2. to assess the performance of various machine learning classifiers, including neural networks and ensemble models, in accurately distinguishing five clinically relevant diagnostic categories, and
3. to explore the influence of class imbalance, feature dimensionality, and model interpretability in practical diagnostic scenarios.

These objectives provide a structured foundation for the study and guide the methodology described in the subsequent sections, including data collection, signal processing, classification, and performance evaluation.

2.1 Dataset description

In this research, the publicly available IOBA-PERG dataset from PhysioNet [1] have been used, which comprises 1,354 transient pattern electroretinogram (PERG) responses recorded from 304 individuals, encompassing a total of 336 independent session records. Each record contains bilateral retinal signals sampled at a rate of 170 Hz, alongside extensive clinical metadata including age, sex, and multiple diagnostic labels (diagnosis1, diagnosis2, diagnosis3). In this analysis, only the primary diagnostic field (diagnosis1) was used to define the ground truth labels for supervised multi-class classification system.

Although the complete dataset includes a broad range of retinal and optic nerve pathologies, a curated subset consisting of 15 clinically significant disease categories, including healthy controls, have been chosen. This selection was primarily motivated by the need to balance class representation and mitigate the adverse effects of sample size disparities. Categories with fewer than five recordings were excluded to prevent data imbalance, which can compromise classifier performance. Future work will aim to extend the analysis to include underrepresented classes once the present methodology has been validated.

The final dataset used for classification included 280 records, stratified across 15 diagnostic categories. Table 1 presents a summary of the most prevalent pathologies retained for analysis, along with brief clinical descriptions.

Table 1: Summary of Selected Diagnoses from the IOBA-PERG Dataset

Sl. No.	Name of the Disease	No. of Records	Explanation
1	Normal	106	No clinical signs of retinal dysfunction.
2	Retinitis Pigmentosa	47	Genetic retinal dystrophy affecting photoreceptor response.
3	Macular Dystrophy	33	Degenerative or inherited disorder of the macula.
4	Stargardt Disease	16	Juvenile macular degeneration caused by ABCA4 mutations.
5	Cone-Rod Dystrophy	14	Affects cone and rod photoreceptors, leading to progressive vision loss.
6	Birdshot Chorioretinopathy	12	Rare bilateral posterior uveitis characterized by creamy lesions.
7	Inherited Optic Atrophy	10	Genetically mediated degeneration of retinal ganglion cells.
8	Congenital Stationary Night Blindness	8	Non-progressive dysfunction of rod photoreceptors.
9	Retinal Toxicity	7	Damage induced by systemic or intravitreal medications.
10–15	Other Rare Disorders*	27	Autoimmune retinopathy, ischemia, achromatopsia, etc.
	Total	280	

*Includes: autoimmune retinopathy, orbital ischemia, X-linked retinoschisis, central serous chorioretinopathy, congenital achromatopsia, and paracentral acute middle maculopathy.

2.2 Preprocessing of PERG signals

All PERG recordings were initially subjected to preprocessing to enhance signal fidelity and minimize noise. The signals, acquired at 170 Hz, capture the dynamic activity of retinal cells within the relevant electrophysiological frequency range (0.5–50 Hz). Each session provided both left and right eye recordings. The raw signals were depicted in Figure 3 and Figure 4, represents a sample normal PERG signal and Macular Dystrophy signal accordingly.

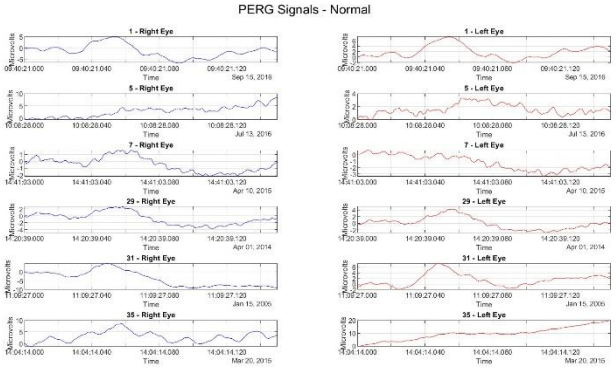

Figure 3: Raw PERG signal (Normal Subject)

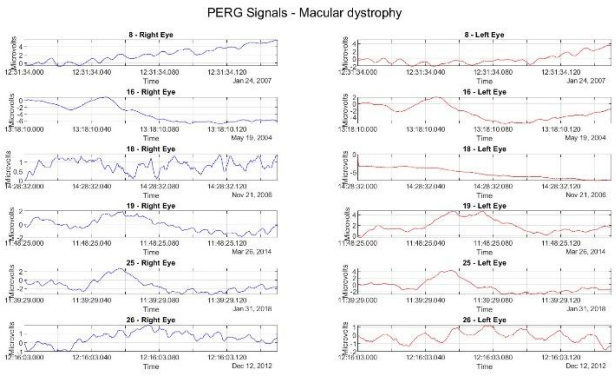

Figure 4: Raw PERG signal (Macular Dystrophy)

In addition, raw signals were filtered using a fourth-order zero-phase Butterworth bandpass filter (0.5–50 Hz). This choice was guided by standard electrophysiological analysis practices [2], effectively suppressing baseline drift and high-frequency electrical artifacts without distorting phase information. The zero-phase implementation (via MATLAB's filtfilt function) ensures that temporal relationships between signal components are preserved [29]. Post-filtering, spectral content was examined through Fast Fourier Transform (FFT) to assess signal quality and identify dominant frequency components [30]. This preprocessing step served both to denoise the data and to prepare it for subsequent band-specific decomposition and feature extraction. Figure 5 presents a sample raw PERG signal, filtered and with FFT applied, the frequency components in the 0–50 Hz range.

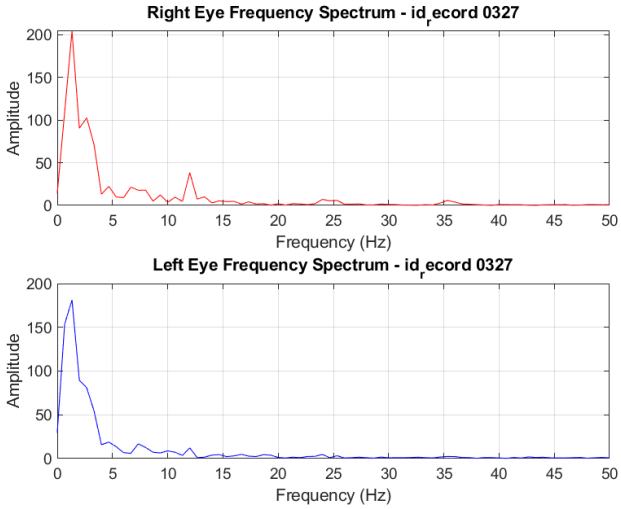

Figure 5: Frequency Spectrum of a raw PERG signal (0–50 Hz)

2.3 Frequency band decomposition

Subsequently, to limit and segregate physiologically meaningful signal components, the filtered PERG signals were decomposed into five distinct frequency bands, each reflecting different retinal or neurophysiological phenomena [25]. Table 2 list the proposed frequency band limits and its functional relevance.

Table 2: Frequency band range (Hz) - functional relevance

Band Label	Frequency Range (Hz)	Interpretation
Very low	0 - 1	Baseline fluctuations, non-neural drift
Low	1 - 5	Principal PERG components (e.g., N35, P50, N95)
Mid	5 - 15	Flicker responses, ganglion cell activity
High	15 - 30	Oscillatory potentials, secondary components
Very high	30 - 50	Noise-prone region, possible residual oscillations

Frequency spectra were analyzed for both eyes independently, and magnitude responses were retained for subsequent feature extraction. Figure 6–10: depicts the Band-specific frequency plots for both eyes.

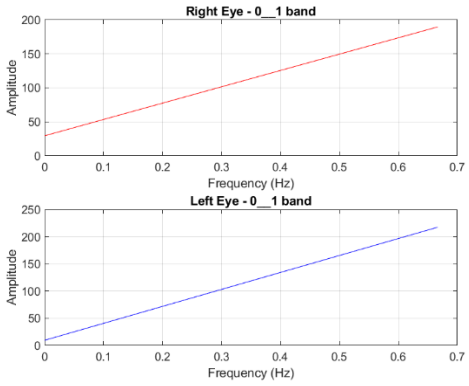

Figure 6: Frequency plot - (0 to 1 Hz) Right and Left eye

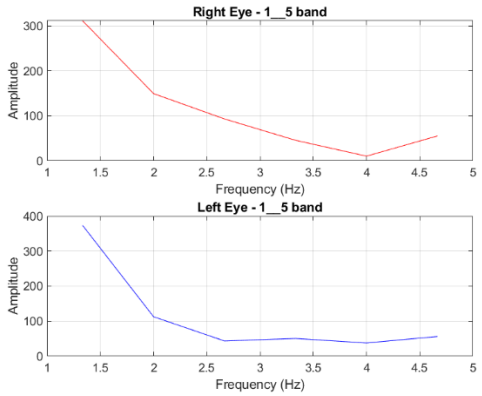

Figure 7: Frequency plot - (1 to 5 Hz) Right and Left eye

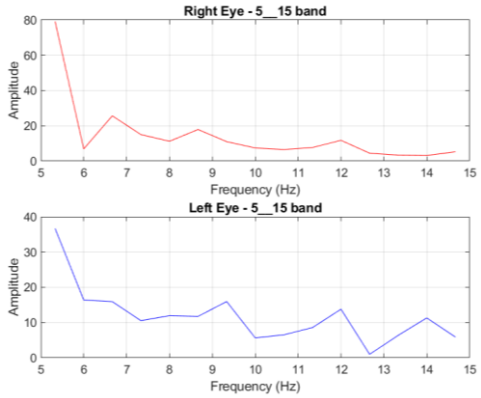

Figure 8: Frequency plot - (5 to 15 Hz) Right and Left eye

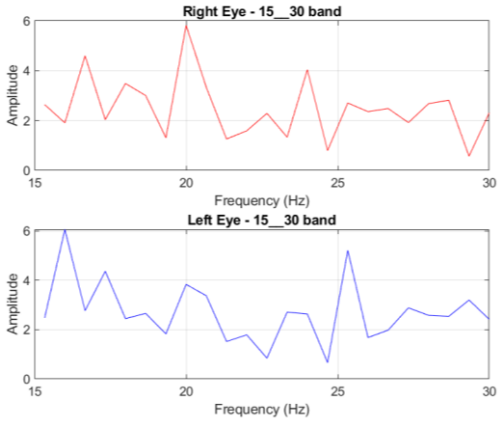

Figure 9: Frequency plot - (15 to 30 Hz) Right and Left eye

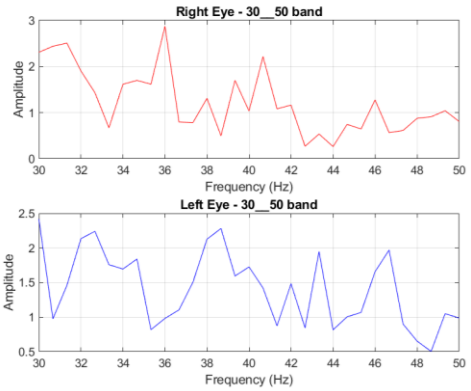

Figure 10: Frequency plot (15 to 30 Hz) Right and Left eye

This band-based separation enables detailed feature analysis across different physiological domains.

2.4 Time-domaing feature extraction

A comprehensive set of time-domain features was computed across each frequency band and for each eye separately. These features were selected to capture the statistical characteristics of the PERG waveforms. In total, over 11 scalar features were extracted per band per eye, including both traditional metrics and signal-processing-derived descriptors. Table 2 summarizes the key features used in this study.

Although the signals were filtered and segmented by frequency bands, the analysis in this work is limited to time-domain features. An initial comparative assessment considered features from time, frequency, and time-frequency domains. Among these, time-domain features demonstrated the most consistent and informative patterns, offering valuable insight into waveform behavior across different conditions. Based on these findings, the current analysis focuses exclusively on time-domain characteristics to provide a clear and interpretable foundation [31], [32].

Additional analysis involving frequency and time-frequency domain features will be presented in a subsequent study to expand the scope and diagnostic capabilities.

Table 3: Summary of extracted time-domain features

Feature	Description / Formula
Mean (μ)	Arithmetic average of signal amplitudes
Variance (σ^2)	Measure of amplitude dispersion
Standard deviation (σ)	variation of a amplitude dispersion
Frequency Variance (μ_f)	power-weighted variance of the frequency components
Root Mean Square (RMS)	Energy-based feature sensitive to amplitude
Peak-to-Peak	Difference between maximum and minimum values
Skewness	Degree of asymmetry in amplitude distribution
Kurtosis	Peakedness or flatness of amplitude distribution
Slope	Mean first-order derivative
Mean Absolute Value	Average of absolute signal values
Area Under Curve (AUC)	Integral of the waveform via trapezoidal approximation

The time-domain features used in this study were chosen to capture a broad spectrum of signal characteristics relevant to physiological interpretation. The mean and variance represent basic statistical measures, reflecting the average amplitude and the overall variability of the PERG signal. Root Mean Square (RMS) offers a measure of signal power, emphasizing higher amplitudes, while peak-to-peak captures the full dynamic range of the waveform by calculating the difference between its maximum and minimum values. To assess the shape and distribution of the signal, skewness is used to measure asymmetry, and kurtosis helps identify sharp peaks or flatness in the waveform. The slope feature captures the rate of change, which is important for detecting transitions or trends in signal morphology. Measures like mean absolute value (MAV) and area under the curve (AUC) summarize the magnitude and energy content of the signal.

Thus, Features were extracted separately for each eye across three frequency bands: B2 (1–5 Hz), B3 (5–15 Hz), and B4 (15–30 Hz), with 11 features per band. Bands B1 (0–1 Hz) and B5 (30–50 Hz) were excluded due to baseline drift and noise, respectively. This yielded 33 features per eye and 66 features per subject. The final dataset consisted of 280 subject recordings, each represented by a 66-dimensional feature vector.

2.5 Feature classification

The feature dataset used in this study was derived from Pattern Electroretinogram (PERG) signals recorded from both the left and right eyes of human subjects. Each subject's data included a variety of time-domain and frequency-domain features extracted from three predefined frequency bands: 1–5 Hz, 5–15 Hz, and 15–30 Hz.

For each subject, a total of 66 features were extracted. These included statistical metrics. An additional class label was assigned to each feature vector. The features were organized into a consolidated table, with dimensional alignment ensured by padding missing entries with NaN values. Each row represented a unique trial per subject, identified by a generated SubjectID.

The original dataset contained 15 distinct class labels corresponding to various clinical or experimental conditions. To address class imbalance and enhance classification robustness, these were regrouped into five broader categories:

Class 1: Normal signals (106 records)

Class 2: Retinitis Pigmentosa (47 records)

Class 3: Macular Dystrophy (33 records)

Class 4: Stargardt Disease (16 records)

Class 5: Combined group from original Classes 5–15, representing general or baseline cases

This relabeling strategy reduced the impact of underrepresented classes and improved classifier stability across trials. The final class distribution was verified to ensure balanced representation.

2.6 Data normalization

Prior to training, all feature values were normalized using Min-Max scaling (range normalization) to ensure uniformity in feature contribution and accelerate model convergence. Normalization was performed after class filtering but prior to data splitting, using MATLAB's normalize function with the 'range' method.

2.7 Classification models

This study evaluated the discriminative potential of the extracted features using five supervised machine learning classifiers under a consistent experimental protocol. The models included a Neural Network (PatternNet) with two hidden layers (15 and 10 neurons), trained using the Levenberg–Marquardt backpropagation algorithm and softmax output for multiclass prediction via one-hot encoding. Ensemble methods included Random Forest (100 trees with out-of-bag prediction) and Bagged Trees. To address class imbalance, the RUSBoost algorithm was employed with 100 decision tree learners and random undersampling of majority classes. The final model, AdaBoost [33], was implemented within an Error-Correcting Output Codes (ECOC) framework using decision trees (max split of 20) and a one-vs-all coding scheme.

All classifiers were trained and tested across 10 independent trials using an 80–20 train-test split. Performance was evaluated using standard metrics: accuracy, precision, recall, specificity, F1-score, training time, and model complexity (measured via trainable parameters, number of learners, or ECOC structure, depending on the model). Classifier interpretability was also assessed qualitatively. Confusion matrices were saved in both graphical and CSV formats to support class-wise error analysis. This structured setup enabled a rigorous comparison of neural and ensemble classifiers, offering insights into their relative accuracy, generalization, and applicability to multiclass PERG signal classification.

3 Results and discussion

The original dataset included 15 distinct class labels, which are listed in Table 4 along with their percentage distributions. These classes were later reorganized and grouped into five broader categories.

Table 4: Retinal pathology class distribution before merging

Value	Count	Percent (%)	Value	Count	Percent (%)	Value	Count	Percent (%)
1	5	1.79	6	7	2.50	11	14	5.00
2	16	5.71	7	8	2.86	12	33	11.79
3	5	1.79	8	8	2.86	13	47	16.79
4	5	1.79	9	10	3.57	14	106	37.86
5	3	1.07	10	12	4.29	15	1	0.36

Initial analysis was carried out using all 15 classes. Based on the results, the highest classification accuracy was 89.29%, achieved using the AdaBoost classifier. The associated performance metrics included a precision of 67.52 ± 6.18, recall of 68.83 ± 6.58, specificity of 99.23 ± 0.35, F1-score of 68.14 ± 6.20, and a training time of 0.13 ± 0.08 seconds. Other classifiers did not perform as well across these metrics. The analysis also revealed limitations in the dataset, particularly the low and uneven sample distribution across classes, one of which contained only a single sample. Detailed results from this 15-class classification are presented in Table 5.

Table 5: Performance comparison of the five classifiers used for the retinal pathology classification (before merging)

Trial	Accuracy			Precision			Recall			Specificity			F1			Training Time		
NN-LM	51.43	±	10.61	20.76	±	9.61	23.91	±	7.24	96	±	0.83	21.79	±	7.77	3.95	±	2.47
Random Forest	58.75	±	5.8	14.27	±	3.31	18.2	±	3.31	96.63	±	0.32	15.96	±	3.23	0.44	±	0.11
Bagged Trees	60	±	8.08	16.94	±	3.32	20.13	±	3.65	96.73	±	0.56	18.31	±	3.26	0.64	±	0.12
RUSBoost	17.5	±	21.31	16.89	±	20.68	18.25	±	19.72	94.11	±	1.31	16.86	±	20.56	1.02	±	0.4
AdaBoost	89.29	±	5.12	67.52	±	6.18	68.83	±	6.58	99.23	±	0.35	68.14	±	6.2	0.13	±	0.08

The revised five-class distribution, along with the number of samples and their respective percentages, is shown in Table 6.

Table 6: Retinal pathology class distribution after merging

Value	Count	Percent (%)
1	78	27.86
2	16	5.71
3	33	11.79
4	47	16.79
5	106	37.86

The performance of each classifier is presented from Table 7 to Table 11. The table 7 presents the results of 10 trials using the Multilayer Neural Network with Levenberg–Marquardt (MLNN-LM) algorithm for five-class classification. All metrics—accuracy, precision, recall, specificity, and F1-score—are reported in percentage terms. The model achieved high performance, with perfect scores in Trials 1, 7, and 9. An average accuracy of 96.43% ± 3.04 indicates strong predictive ability and reasonable stability. However, higher variability in precision (±9.56%) and recall (±8.79%), especially in Trial 6, points to challenges with class imbalance. The average F1-score of 93.12% ± 8.85 reflects a good balance between precision and recall. Training time ranged from 3.43 to 7.58 seconds, averaging 4.61 ± 1.19 seconds.

Table 7: Performance evaluation of retinal pathology classification using Multi layer neural

Trial	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 (%)	Training Time (seconds)
1	100	100	100	100	100	4.83
2	94.64	94.95	87.00	98.64	90.80	3.50
3	94.64	97.15	88.18	98.34	92.45	5.10
4	94.64	96.82	94.44	98.45	95.62	4.42
5	98.21	99.13	98.33	99.41	98.73	4.39
6	94.64	68.18	73.33	98.81	70.66	4.23
7	100	100	100	100	100	4.80
8	96.43	96.85	90.83	99.09	93.74	7.58
9	100	100	100	100	100	3.43
10	91.07	95.61	83.61	97.11	89.21	3.81
Average	96.43 ± 3.04	94.87 ± 9.56	91.57 ± 8.79	98.99 ± 0.92	93.12 ± 8.85	4.61 ± 1.19

Overall, the classifier is effective but sensitive to class distribution, suggesting the need for data balancing in future work.

Table 8: Performance evaluation of retinal pathology classification using Random Forest

Trial	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 (%)	Training Time (seconds)	Model Complexity
1	83.93	52.78	54.71	96.67	53.73	0.31	100
2	91.07	75.17	68.75	98.22	71.82	0.32	100
3	89.29	71.04	63.76	97.65	67.21	0.28	100
4	94.64	80.00	77.78	98.86	78.87	0.27	100
5	91.07	71.15	70.56	98.21	70.85	0.30	100
6	96.43	60.98	66.67	99.32	63.70	0.29	100
7	87.50	70.93	62.46	97.55	66.42	0.27	100
8	89.29	58.77	62.50	97.73	60.58	0.27	100
9	94.64	61.18	66.67	98.86	63.81	0.29	100
10	76.79	56.61	52.28	95.16	54.36	0.29	100
Average	89.46 ± 5.8	65.86 ± 8.94	64.61 ± 7.41	97.82 ± 1.21	65.13 ± 7.77	0.29 ± 0.02	100 ± 0

The table 8 summarizes the performance of the Random Forest (Tree Bagger) model across 10 trials for the five-class classification task. All evaluation metrics are expressed in percentages. The model achieved a maximum accuracy of 96.43% and a minimum of 76.79%, with an average of 89.46% ± 5.8, indicating moderate classification performance with some variability. Precision and recall showed noticeable fluctuations (65.86% ± 8.94 and 64.61% ± 7.41, respectively), highlighting occasional difficulty in consistently identifying positive cases. The F1-score averaged 65.13% ± 7.77, reflecting modest balance between precision and recall. Specificity remained high and consistent at 97.82% ± 1.21, confirming reliable identification of negative instances. Training time was efficient and consistent, averaging 0.29 ± 0.02 seconds, with a fixed model complexity. Overall, the classifier performed reasonably well but showed sensitivity to data imbalance and class variation.

Table 9: Performance evaluation of retinal pathology classification using Bagged Trees (Bootstrap Aggregation)

Trial	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 (%)	Training Time (seconds)
1	89.29	58.25	59.93	97.71	59.08	0.65
2	92.86	76.79	72.92	98.42	74.80	0.69
3	91.07	73.28	69.32	98.17	71.24	0.77
4	92.86	79.17	75.93	98.48	77.51	0.75
5	92.86	73.89	73.89	98.59	73.89	0.66
6	94.64	58.33	65.00	99.01	61.49	0.70
7	87.50	54.40	57.86	97.53	56.07	0.70
8	92.86	78.61	70.14	98.41	74.13	0.69
9	94.64	63.43	66.67	98.67	65.01	0.58
10	87.50	71.03	65.97	97.36	68.41	0.63
Average	91.61 ± 2.67	68.72 ± 9.28	67.76 ± 5.87	98.24 ± 0.54	68.16 ± 7.4	0.68 ± 0.06

Table 9 presents results from 10 trials using the Bagged Trees (Bootstrap Aggregation) model for five-class classification. Performance metrics are expressed as percentages. The model showed consistent accuracy, ranging from 87.50% to 94.64%, with an average of 91.61% ± 2.67, indicating good stability. Precision and recall averaged 68.72% ± 9.28 and 67.76% ± 5.87, respectively, suggesting moderate ability to detect positive instances with some variation across trials. The F1-score averaged 68.16% ± 7.4, reflecting reasonable balance between precision and recall. Specificity remained high at 98.24% ± 0.54, confirming consistent recognition of negative cases. Training times were efficient, averaging 0.68 ± 0.06 seconds, with constant model complexity. Overall, the classifier performed reliably, offering a balanced trade-off between accuracy and computational cost.

Table 10: Performance evaluation of retinal pathoalogy classification using RUS Boost

Trial	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 (%)	Training Time (seconds)
1	21.43	15.91	13.90	83.62	14.84	0.54
2	37.50	6.25	16.67	83.33	9.09	0.02
3	30.36	5.06	16.67	83.33	7.76	0.02
4	42.86	25.00	26.74	87.88	25.84	0.50
5	32.14	28.68	25.51	86.51	27.00	0.62
6	48.21	29.39	26.06	89.17	27.63	0.50
7	42.86	27.64	27.59	88.29	27.61	0.49
8	37.50	6.25	16.67	83.33	9.09	0.02
9	35.71	27.92	23.31	87.08	25.41	0.45
10	33.93	18.44	18.47	86.11	18.46	0.53
Average	36.25 ± 7.53	19.05 ± 10.12	21.16 ± 5.17	85.87 ± 2.29	19.27 ± 8.43	0.37 ± 0.24

Table 10 shows the results of 10 trials using the RUSBoost classifier for five-class classification. Compared to other models, RUSBoost delivered substantially lower performance, with an average accuracy of 36.25% ± 7.53. Precision and recall were also low, averaging 19.05% ± 10.12 and 21.16% ± 5.17, respectively, reflecting poor identification of true positives. The F1-score averaged 19.27% ± 8.43, indicating weak overall classification quality. Although specificity remained relatively stable at 85.87% ± 2.29, this reflects the model's tendency to favor the majority (negative) classes. Training time averaged 0.37 ± 0.24 seconds, with model complexity fluctuating (average 70 ± 48.3). These results suggest that RUSBoost struggled with the dataset, likely due to severe class imbalance and insufficient representative samples for minority classes.

Table 11 summarizes the performance of the AdaBoost classifier implemented via the ECOC (Error-Correcting Output Codes) framework. The model showed exceptionally high and consistent results, achieving an average accuracy of 99.11% ± 1.26 across 10 trials. Precision, recall, and F1-score were all highly stable, averaging 82.92% ± 0.59, 83.15% ± 0.59, and 83.03% ± 0.55, respectively. Specificity was nearly perfect at 99.77% ± 0.32, indicating excellent identification of negative cases. Training time remained minimal (0.05 ± 0.01 seconds) and model complexity low (5.7 ± 0.48), highlighting the efficiency of this approach. Overall,

AdaBoost-ECOC delivered the best balance of performance, consistency, and computational efficiency among all classifiers tested.

Table 11: Performance evaluation of retinal pathoalogy classification using AdaBoost (via ECOC Framework)

Trial	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 (%)	Training Time (seconds)
1	100	83.33	83.33	100	83.33	0.05
2	98.21	82.58	83.33	99.52	82.95	0.04
3	98.21	82.41	83.33	99.57	82.87	0.05
4	100	83.33	83.33	100	83.33	0.05
5	100	83.33	83.33	100	83.33	0.06
6	100	83.33	83.33	100	83.33	0.05
7	96.43	81.67	81.48	99.12	81.57	0.04
8	98.21	82.58	83.33	99.52	82.95	0.03
9	100	83.33	83.33	100	83.33	0.05
10	100	83.33	83.33	100	83.33	0.05
Average	99.11 ± 1.26	82.92 ± 0.59	83.15 ± 0.59	99.77 ± 0.32	83.03 ± 0.55	0.05 ± 0.01

Table 12: Performance comparison of the five classifiers used for the retinal pathology classifications

Classifier	Accuracy			Precision			Recall			Specificity			F1			Training Time		
NN-LM	96.43	±	3	94.87	±	9.56	91.57	±	8.8	98.99	±	0.9	93.12	±	8.9	4.6	±	1.2
Random Forest	89.46	±	5.80	65.86	±	8.94	64.61	±	7.41	97.82	±	1.21	65.13	±	7.77	0.29	±	0.02
Bagged Trees	91.61	±	2.67	68.72	±	9.28	67.76	±	5.87	98.24	±	0.54	68.16	±	7.40	0.68	±	0.06
RUSBoost	36.25	±	7.53	19.05	±	10.12	21.16	±	5.17	85.87	±	2.29	19.27	±	8.43	0.37	±	0.24
AdaBoost	99.11	±	1.26	82.92	±	0.59	83.15	±	0.59	99.77	±	0.32	83.03	±	0.55	0.05	±	0.01

Table 12 presents a comparative analysis of five classification models evaluated over 10 independent trials using key performance metrics: accuracy, precision, recall, specificity, F1-score, and training time.

The AdaBoost (ECOC) model outperformed all other classifiers, achieving the highest average accuracy (99.11% ± 1.26) and exceptional consistency across all metrics. Its precision, recall, and F1-score remained tightly clustered around 83%, with minimal variability, while requiring the least training time (0.05 ± 0.01 seconds) and low model complexity, making it both highly effective and computationally efficient.

The Neural Network with Levenberg–Marquardt (NN-LM) also delivered strong performance, with an average accuracy of 96.43% ± 3.04 and the highest precision (94.87% ± 9.56) among all models. However, its training time was significantly higher (4.61 ± 1.19 seconds), indicating greater computational cost.

Bagged Trees provided a good balance, achieving 91.61% ± 2.67 accuracy with moderate precision and recall values around 68%, and reasonable training efficiency. In contrast, Random Forest (Tree Bagger) showed slightly lower accuracy (89.46% ± 5.80) and higher variability across precision and recall, reflecting less stable behavior across trials.

The RUSBoost classifier, which incorporates random under sampling to address class imbalance, underperformed significantly in all metrics, with accuracy at 36.25% ± 7.53 and an F1-score of just 19.27% ± 8.43. While the intention of addressing imbalance was valid, the aggressive under sampling likely led to loss of information, reducing classification quality.

Overall, AdaBoost via ECOC emerged as the most robust and efficient model, followed by NN-LM for accuracy-focused scenarios and Bagged Trees for balance between performance and training time,. RUSBoost was

not suitable in this context due to its consistently poor results.

The confusion matrices of the AdaBoost and MLNN-LM classifiers corresponding to their lowest classification accuracy trials were analyzed to identify specific patterns of misclassification. These are presented in Figure 11 and Figure 12, respectively. From the AdaBoost confusion matrix, it is evident that the class formed by merging original classes 5–15 exhibited a misclassification rate of 11.1%, corresponding to 2 incorrectly predicted samples. In the case of the MLNN-LM classifier, Class 2 (Retinitis Pigmentosa) showed a misclassification rate of 33.3% (1 out of 3 samples), while Class 3 (Macular Dystrophy) had a misclassification rate of 11.1% (1 out of 9 samples). These results highlight specific weaknesses in class-level prediction, particularly for classes with fewer training samples.

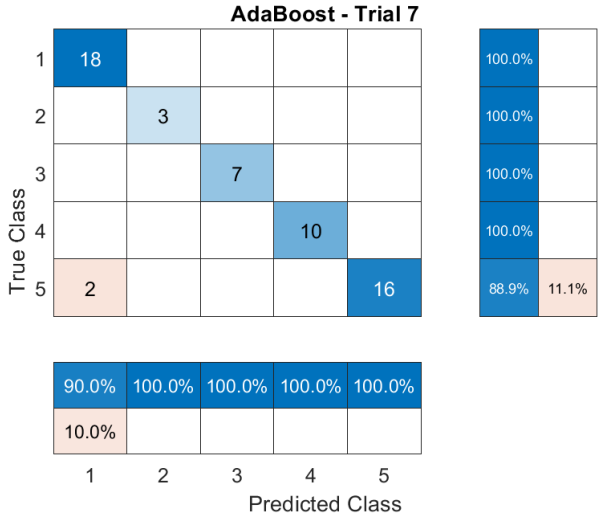

Figure 11: Confusion Matrix of AdaBoost via ECOC with Minimum Accuracy of 96.43%

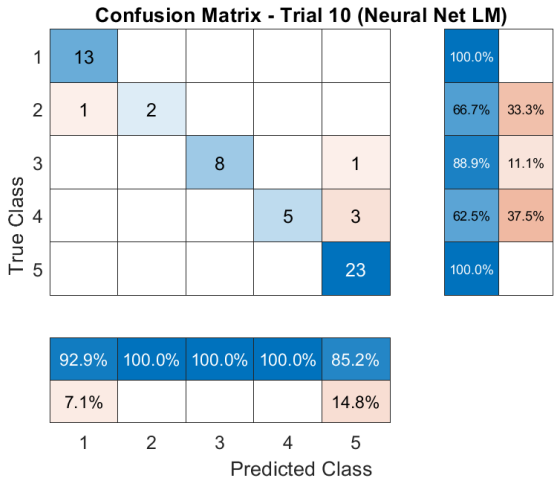

Figure 12: Confusion Matrix of MLNN-LM with Minimum Accuracy of 91.07%

4 Conclusion

This study presented a five-class classification framework for distinguishing retinal and optic nerve pathologies using time-domain features extracted from band-separated PERG signals. Among the various models evaluated, AdaBoost via the ECOC framework demonstrated the highest and most consistent classification performance, while the neural network using the Levenberg–Marquardt algorithm also achieved high accuracy with strong precision. The comparative evaluation revealed the strengths and limitations of each classifier, particularly in handling class imbalance and small sample sizes.

Despite challenges posed by uneven class distributions and limited data for certain categories, the results highlight the feasibility and promise of the proposed approach. The time-domain analysis alone has produced positive outcomes, establishing a strong foundation for further exploration.

As this work represents an initial phase, future research will build upon these findings by incorporating more advanced and focused analyses. Specifically, feature extraction will be refined by using selected frequency bands rather than the full spectrum, enabling targeted comparisons through the use of intelligent techniques and optimization algorithms. In addition to time-domain analysis, future studies will introduce frequency-domain and time-frequency domain features to improve model discrimination. Another planned direction involves the segmentation and analysis of specific waveform components—namely the N35, P50, and N95 waves—on a per-record basis, with the goal of enhancing physiological interpretability and improving classification performance.

References

[1] I. Fernández, R. Cuadrado-Asensio, Y. Larriba, C. Rueda, and R. M. Coco-Martín, ‘A comprehensive dataset of pattern electroretinograms for ocular electrophysiology research’, *Sci. Data*, vol. 11, no. 1, p. 1013, 2024. https://doi.org/10.1038/s41597-024-03857-1

[2] J. V. Odom *et al.*, ‘ISCEV standard for clinical visual evoked potentials (2009 update)’, *Doc. Ophthalmol.*, vol. 120, pp. 111–119, 2010. https://doi.org/10.1007/s10633-009-9195-4

[3] D. L. McCulloch *et al.*, ‘ISCEV Standard for full-field clinical electroretinography (2015 update)’, *Doc. Ophthalmol.*, vol. 130, no. 1, pp. 1–12, 2015. https://doi.org/10.1007/s10633-015-9504-z

[4] S. K. S. K. Nataraj *et al.*, ‘Statistical cross-correlation band features based thought controlled communication system’, *AI Commun.*, vol. 8, no. 4, pp. 20–28, Aug. 2016, doi: 10.3233/AIC-160703. https://doi.org/10.3233/aic-160703

[5] R. Hamilton *et al.*, ‘VEP estimation of visual

acuity: a systematic review', *Doc. Ophthalmol.*, vol. 142, pp. 25–74, 2021. https://doi.org/10.1007/s10633-020-09770-3

[6] D. C. Hood *et al.*, 'ISCEV standard for clinical multifocal electroretinography (mfERG)(2011 edition)', *Doc. Ophthalmol.*, vol. 124, pp. 1–13, 2012. https://doi.org/10.1007/s10633-011-9296-8

[7] J. Bures, M. Petran, and J. Zachar, *Electrophysiological methods in biological research*. Academic Press, 2013. https://doi.org/10.1016/c2013-0-10442-4

[8] R. N. Weinreb and P. T. Khaw, 'Primary open-angle glaucoma', *Lancet*, vol. 363, no. 9422, pp. 1711–1720, 2004. https://doi.org/10.3109/9780203007204-15

[9] N. V Thakor, 'Biopotentials and electrophysiology measurement', *Meas. Instrumentation, Sensors Handb.*, vol. 74, 1999. https://doi.org/10.1201/9780415876179.ch74

[10] J. Rettinger, S. Schwarz, and W. Schwarz, *Electrophysiology*. Springer, 2022. https://doi.org/10.1007/978-3-030-86482-8_3

[11] D. A. Thompson, M. Bach, J. J. McAnany, M. Šuštar Habjan, S. Viswanathan, and A. G. Robson, 'ISCEV standard for clinical pattern electroretinography (2024 update)', *Doc. Ophthalmol.*, vol. 148, no. 2, pp. 75–85, 2024.

[12] A. B. Suttle *et al.*, 'Assessment of the drug interaction potential and single-and repeat-dose pharmacokinetics of the BRAF inhibitor dabrafenib', *J. Clin. Pharmacol.*, vol. 55, no. 4, pp. 392–400, 2015. https://doi.org/10.1007/s10633-024-09970-1

[13] G. A. Omura, 'Progress in gynecologic cancer research: the Gynecologic Oncology Group experience', in *Seminars in oncology*, 2008, vol. 35, no. 5, pp. 507–521. https://doi.org/10.1053/j.seminoncol.2008.07.007

[14] V. Parisi, 'Electrophysiological assessment of glaucomatous visual dysfunction during treatment with cytidine-5′-diphosphocholine (citicoline): a study of 8 years of follow-up', *Doc. Ophthalmol.*, vol. 110, pp. 91–102, 2005. https://doi.org/10.1007/s10633-005-7348-7

[15] G. E. Holder, 'Pattern electroretinography (PERG) and an integrated approach to visual pathway diagnosis', *Prog. Retin. Eye Res.*, vol. 20, no. 4, pp. 531–561, 2001. https://doi.org/10.1016/s1350-9462(00)00030-6

[16] M. Bach and M. B. Hoffmann, 'Update on the pattern electroretinogram in glaucoma', *Optom. Vis. Sci.*, vol. 85, no. 6, pp. 386–395, 2008. https://doi.org/10.1097/opx.0b013e318177ebf3

[17] A. G. Robson *et al.*, 'ISCEV Standard for full-field clinical electroretinography (2022 update)', *Doc. Ophthalmol.*, vol. 144, no. 3, pp. 165–177, 2022. https://doi.org/10.1007/s10633-022-09872-0

[18] A. L. Goldberger *et al.*, 'PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals', *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000. https://doi.org/10.1161/01.cir.101.23.e215

[19] M. Kulyabin, A. Zhdanov, A. Dolganov, M. Ronkin, V. Borisov, and A. Maier, 'Enhancing electroretinogram classification with multi-wavelet analysis and visual transformer', *Sensors*, vol. 23, no. 21, p. 8727, 2023. https://doi.org/10.3390/s23218727

[20] S. Asanad and R. Karanjia, 'Multifocal electroretinogram', in *StatPearls [Internet]*, StatPearls Publishing, 2022. https://doi.org/10.5005/jp/books/10550_15

[21] G. E. Holder, A. G. Robson, C. R. Hogg, M. Kurz-Levin, N. Lois, and A. C. Bird, 'Pattern ERG: clinical overview, and some observations on associated fundus autofluorescence imaging in inherited maculopathy', *Doc. Ophthalmol.*, vol. 106, no. 1, p. 17, 2003. https://doi.org/10.1023/a:1022471623467

[22] C. Canedo, I. Fernández, R. M. Coco, R. Cuadrado, and C. Rueda, 'Novel modeling proposals for the analysis of pattern electroretinogram signals', in *Statistical Methods at the Forefront of Biomedical Advances*, Springer, 2023, pp. 255–273. https://doi.org/10.1007/978-3-031-32729-2_11

[23] P. A. Constable, M. Bach, L. J. Frishman, B. G. Jeffrey, A. G. Robson, and I. S. for C. E. of V. http://www. iscev. org, 'ISCEV Standard for clinical electro-oculography (2017 update)', *Doc. Ophthalmol.*, vol. 134, pp. 1–9, 2017. https://doi.org/10.1007/s10633-017-9573-2

[24] E. J. Ahn, Y. I. Shin, Y. K. Kim, J. W. Jeoung, and K. H. Park, 'Hemifield-based analysis of pattern electroretinography in normal subjects and patients with preperimetric glaucoma', *Sci. Rep.*, vol. 14, no. 1, p. 5116, 2024. https://doi.org/10.1038/s41598-024-55601-9

[25] M. Bach *et al.*, 'ISCEV standard for clinical pattern electroretinography (PERG): 2012 update', *Doc. Ophthalmol.*, vol. 126, pp. 1–7, 2013. https://doi.org/10.1007/s10633-012-9353-y

[26] A. Aykut, B. Akgün, A. S. Sezenöz, M. O. Sevik, and Ö. Şahin, 'Diagnosing retinal disorders with artificial intelligence: the role of large language models in interpreting pattern electroretinography data', *J. Heal. Sci. Med.*, vol. 7, no. 5, pp. 538–542, 2024. https://doi.org/10.32322/jhsm.1506378

[27] M. Kulyabin *et al.*, 'Synthetic Electroretinogram Signal Generation Using Conditional Generative Adversarial Network for Enhancing Classification of Autism Spectrum Disorder', *arXiv Prepr. arXiv2407.08166*, 2024. https://doi.org/10.1007/s10633-025-10019-0

[28] Q. Li and J. Zhou, ‘A Comparative Analysis of Extreme Gradient Boosting, Decision Tree, Support Vector Machines, and Random Forest Algorithm in Data Analysis of College Students’ Psychological Health’, *Informatica*, vol. 49, no. 15, 2025. https://doi.org/10.31449/inf.v49i15.7004

[29] O. Garćıa-Alarćon, S. Puga-Guzḿan, and J. Moreno-Valenzuela, ‘On parameter identification of the Furuta pendulum’, *Procedia Eng.*, vol. 35, pp. 77–84, 2012. https://doi.org/10.1016/j.proeng.2012.04.167

[30] T. Harčarik, J. Bocko, and K. Masláková, ‘Frequency analysis of acoustic signal using the Fast Fourier Transformation in MATLAB’, *Procedia Eng.*, vol. 48, pp. 199–204, 2012. https://doi.org/10.1016/j.proeng.2012.09.505

[31] P. Monsalve *et al.*, ‘Next generation PERG method: expanding the response dynamic range and capturing response adaptation’, *Transl. Vis. Sci. Technol.*, vol. 6, no. 3, p. 5, 2017. https://doi.org/10.1167/tvst.6.3.5

[32] E. D. Übeyli, ‘Statistics over features: EEG signals analysis’, *Comput. Biol. Med.*, vol. 39, no. 8, pp. 733–741, 2009. https://doi.org/10.1016/j.compbiomed.2009.06.001

[33] R.-Y. Rong *et al.*, ‘Machine-learning model of eye metastasis in nasopharyngeal carcinoma based on the AdaBoost method’, 2023. https://doi.org/10.21203/rs.3.rs-3300766/v1