Objective Assessment of Microperimetry Exam Using EEG Signals

Authors

Muhammad Najam Dar[1,2], Antonio Neto Severo de Castro[1], Kholood Janjua[3,4], Zoha Zahid Fazal[5], Mohammed Abdul Saqhlain Shaik[6], Tehmina Sheharyar[3,7]; Mohamed Ibrahim Ahmed[1,8], Yasir Jamal Sepah[9]

Affiliations

1. Research Affiliate, Byers Eye Institute, Stanford University, CA, USA
2. Senior Data Scientist, Ocular Imaging Research and Reading Center, CA, USA
3. Visiting Instructor, Byers Eye Institute, Stanford University, CA, USA
4. Department of Biomedical Engineering and Sciences, National University of Sciences and Technology, Islamabad, Pakistan
5. Postdoctoral Research Fellow, Ophthalmology, Stanford University, CA, USA
6. Data Scientist, Byers Eye Institute, Stanford University, CA, USA
7. Lecturer, Mirpur University of Science and Technology, Mirpur, AJK, Pakistan
8. Executive Director, Ocular Imaging Research and Reading Center, CA, USA
9. Assistant Professor, Byers Eye Institute, Stanford University, CA, USA

Correspondence

Yasir Jamal Sepah, MBBS
Assistant Professor, Byers Eye Institute
Stanford University School of Medicine
Palo Alto, California, United States 94303
Email: yjs@stanford.edu

Keywords

objective microperimetry, automated stimulus registration, EEG, visual evoked potential, deep learning.

Abstract

Purpose

To test the hypothesis that deep learning can decode single-trial cortical responses from electroencephalography (EEG) to individual, long-duration microperimetry (MP) stimuli, thereby providing a basis for an objective measure of visual pathway integrity. We also investigated whether occipital EEG signals could automatically register microperimetry stimuli, replacing the patient's manual button press to improve exam reliability.

Methods

We collected pilot EEG data from two healthy participants (12 trials total) during microperimetry (Nidek MP1) at low (3.2 cd/m²) and high (127 cd/m²) stimulus intensities. Occipital EEG signals (8-channel OpenBCI) were bandpass filtered (4–49 Hz), z-score normalized, and segmented into 600 ms epochs time-locked to stimuli. A bidirectional LSTM (Bi-LSTM) deep neural network then classified each epoch as stimulus or non-stimulus.

Results

In seven low-intensity trials (16 dB), the model achieved up to ~77.4% accuracy in distinguishing stimulus vs. non-stimulus EEG segments using the O1/O2 channels. In five high-intensity trials (0 dB), accuracy reached ~80%. Occipital electrodes (O1, O2) outperformed parietal or combined parieto-occipital montages for detecting the flash stimuli. Consistent with the neuroanatomy of the primary visual cortex, classification performance was maximal using the bilateral occipital electrodes (O1, O2).

Conclusion

Integrating occipital EEG with microperimetry provides a proof-of-concept objective framework that reduces reliance on subjective patient feedback and could enhance diagnostic precision. While our pilot study was limited by a small sample and lack of precise EEG–stimulus synchronization, the findings warrant further refinement for clinical translation.

Translational Relevance

Enabling automated stimulus registration can improve microperimetry's reliability in pediatric and challenging populations.

Introduction

Microperimetry (MP) is a visual field test that evaluates localized macular function by mapping retinal sensitivity, tracking fixation stability, and assessing structural aspects of visual function. It helps in the early diagnosis of age-related macular degeneration and diabetic macular edema by revealing subtle functional impairments. MP can detect early functional loss in para-atrophic regions of geographic atrophy (GA) and Stargardt Disease (SD), impacting over 5 million patients globally, enabling earlier intervention before central vision loss. It is also useful to track localized functional decline to model the progression of atrophic diseases such as GA and SD. MP helps bridge the gap between structural and functional assessments, offering early detection of macular dysfunction and guiding timely treatment decisions. This translational role is vital in clinical trials and patient care, as without MP, subtle functional improvements might be missed until structural changes become pronounced[1,2].

The MP test, despite its significance in ocular diagnostics, can be hampered by lengthy testing times and requires substantial patient cooperation[3], limiting its utility in clinical trial settings. The stimulus registration requires the patient to press the trigger when they can perceive the faintest flashlight stimulus to push the limits of vision by adjusting stimulus intensity. Age-related cognitive impairments[4], testing fatigue[3], inattentiveness or non-cooperation in patients[5], and extremely infrequent physiological scotoma assessment[6] are significant challenges in MP testing, resulting in unreliable assessments[7]. Microperimetry has shown promise in pediatric ophthalmology for diagnosing and treating various conditions, including hereditary retinal diseases, amblyopia, and nystagmus[8]. However, studies show that children exhibit poorer test-retest reliability, lower mean sensitivity, and reduced fixation stability compared to adults[9]. Additionally, children demonstrate significant practice effects, necessitating multiple examinations for accurate results[9]. Some patients may hesitate or over-respond to the stimulus[6, 7] and require multiple sessions to learn the manual stimulus registration protocol[10, 11]. All these issues stem from the subjective nature of the patient's response, underscoring the need for a more objective MP stimulus registration method.

Recent studies have explored various objective methods to assess retinal sensitivity and visual function, although each has limitations. For instance, deep learning applied to SD-OCT scans can predict enface retinal sensitivity maps with a point-wise error of ~2.73 dB (mean error ~1.66 dB), but this approach cannot yet replace microperimetry due to the small sample size and lack of test–retest validation[12]. Similarly, focal electroretinography (FERG) has been compared to microperimetry, and it was able to detect reduced macular sensitivity in intermediate AMD with higher accuracy than microperimetry, though differences in patient age and using only six stimuli in that study may have confounded the results[13]. Given the scarcity of research on objective microperimetry, other patient-independent assessments are being investigated: for example, dark adaptation recovery has been measured via custom EEG electrodes recording dark-adapted visual evoked potentials (DAVEP), providing an objective, faster, and portable test that reduces reliance on subjective patient responses[14]. Likewise, in multiple sclerosis patients, analyzing VEP signals with a deep neural network achieved about 88% accuracy in distinguishing stimulus versus no-stimulus and in 91% of cases corresponded with standard Humphrey visual

field and OCT findings[15]. Moreover, VEP signals from a head-mounted EEG device have outperformed standard automated perimetry in glaucoma detection[16], and abnormal VEP waveform latencies/amplitudes (N75, P100, N135) can reveal visual pathway disorders such as multiple sclerosis[17].

The contributions of our research to further the field of objective perimetry are as follows:

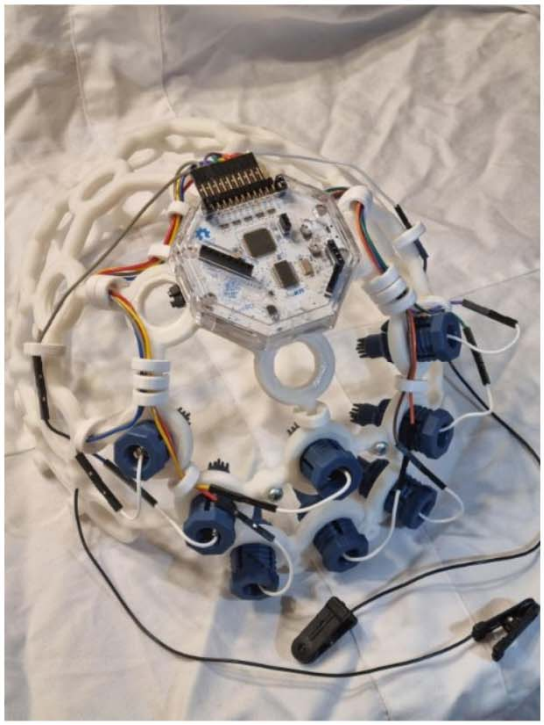
1. We introduce a paradigm that uses individual, long-duration visual stimuli of MP to evoke EEG responses, as opposed to traditional stimuli like flickers or geometric patterns.
2. We develop an LSTM-based deep learning model capable of decoding non-repetitive, variable intensity MP stimuli, with a comparison of various EEG electrodes

These contributions lay the groundwork for integrating EEG into clinical microperimetry, potentially improving test accessibility and reliability in patient populations where subjective responses are unreliable.

Methods

Participants and Data Acquisition

The dataset acquisition required EEG signals with timestamps corresponding to the presented MP stimuli to time-lock the effect of the flashing stimulus on the EEG signals. We used an OpenBCI complete ultracortex 8-channel EEG headset[18] with a cyton bio-amplifier board for signal acquisition. We adjusted EEG electrode locations to capture the response primarily from the visual cortex with P3, P4, PO3, PO4, POz, Oz, O1, and O2 channels[19] as shown in **Figure 1**. OpenBCI GUI was used for EEG signal recording in BDF format for offline processing. All the EEG signals were recorded with a 250 Hz sampling frequency.



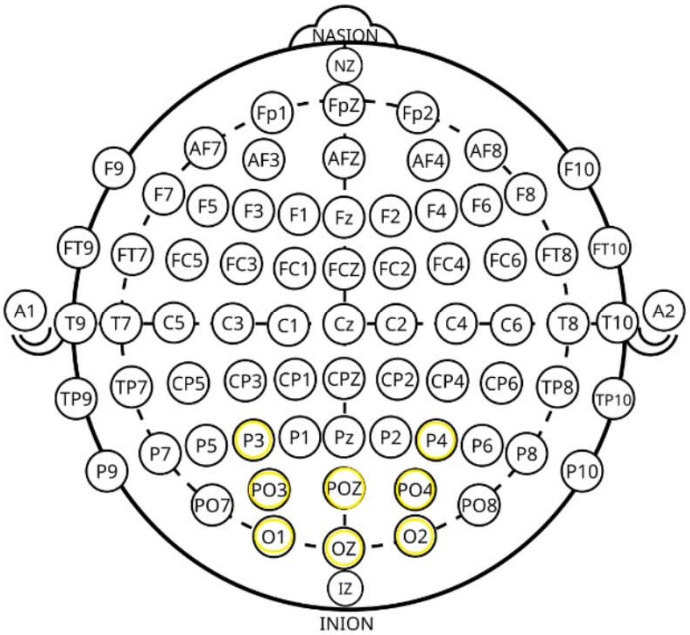


Figure 1: Customized channel configuration to capture the visual cortex response to flash stimulus of microperimetry.

We used the Nidek MP1 device for the microperimetry exam[20], with custom settings of Goldmann IV stimulus size, and enabled fixation monitoring. The trials were divided into two settings, one with high-intensity and the other with low-intensity stimulus. The highest possible luminance for Nidek MP1 is 0dB, while the lowest possible luminance is 20dB (1.27 cd/m2), which is equal to background luminance. During high-intensity settings, we used either 0dB (127 cd/m2) or 4dB (50.6 cd/m2) luminance of stimulus, while 16dB (3.2 cd/m2) was used for low-intensity stimulus settings. We chose 16 dB for low intensity stimulus, because it is slightly above the healthy macular sensitivity of 18.33 ± 2 dB for Nidek MP1[21], and historic microperimetry protocols shows that points missing 16dB stimulus threshold reliably indicate emerging scotomas before acuity loss[22]. Moreover, test–retest data in geographic atrophy show that a change of ± 4 dB in point-wise sensitivity reliably exceeds measurement variability, making 4 dB a meaningful near-floor benchmark for true progression versus noise[23]. All trials were performed using the right eye, while the left eye was covered during the exam. For synchronization, microperimetry exams and EEG recording using the OpenBCI GUI were started at the same time. For stimulus timestamp synchronization, the microperimetry screen video was recorded, showing the stimulus onset, stimulus offset, and the status of the trigger pressed by the patient after seeing the stimulus. The stimulus onset timestamp, stimulus offset timestamp, and patient response timestamp of pressing the trigger are used as ground truth for event-related potentials presented in **Figure 2**.

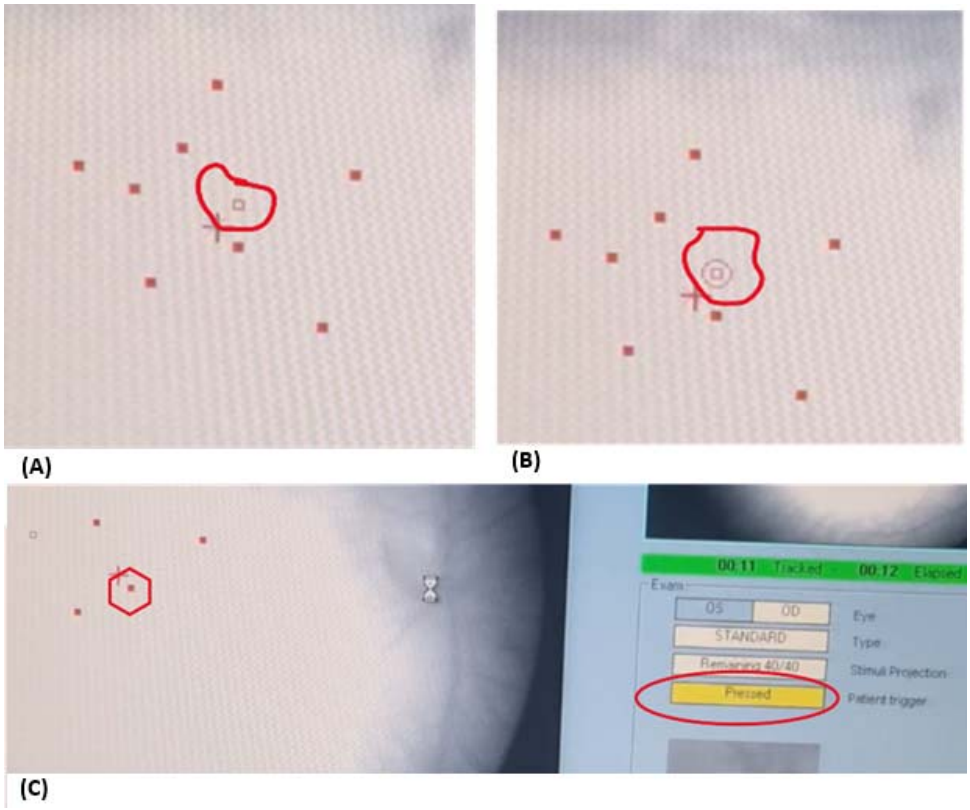


Figure 2: MP recorded video screenshots depicting timestamp extraction process using DataVyu software. (A) The hollow square shows target next location of stimulus to be presented to patient's retina (B) Stimulus onset timestamp when a circle appears enclosing the hollow square (C) Stimulus offset timestamp when outer circle disappears and square filled with red color, patient trigger will be shown pressed and separately logged as patient response timestamp.

The standard MP devices lack the ability to precisely time-lock MP stimulus with external devices' data streaming using hardware-based synchronization. Therefore, we performed offline synchronization of the MP stimulus with EEG signals. The precise stimulus timestamps were extracted by offline reviewing the MP screen video recording in slow motion (¼ of the original speed) using DataVyu software[24], a video annotation tool for timestamp extraction. The timestamps from DataVyu software were exported in Excel format for each trial. We minimized time-stamp misalignment between stimulus and EEG signals by recording timestamps with slow motion videos. The approximate timing error introduced by this manual approach of synchronization is 250 ms; therefore, we need to counter this misalignment in our methodology. The complete experiment protocol with EEG data acquisition during MP stimulus presentation and offline timestamp extraction using DataVyu is depicted in **Figure 3**.

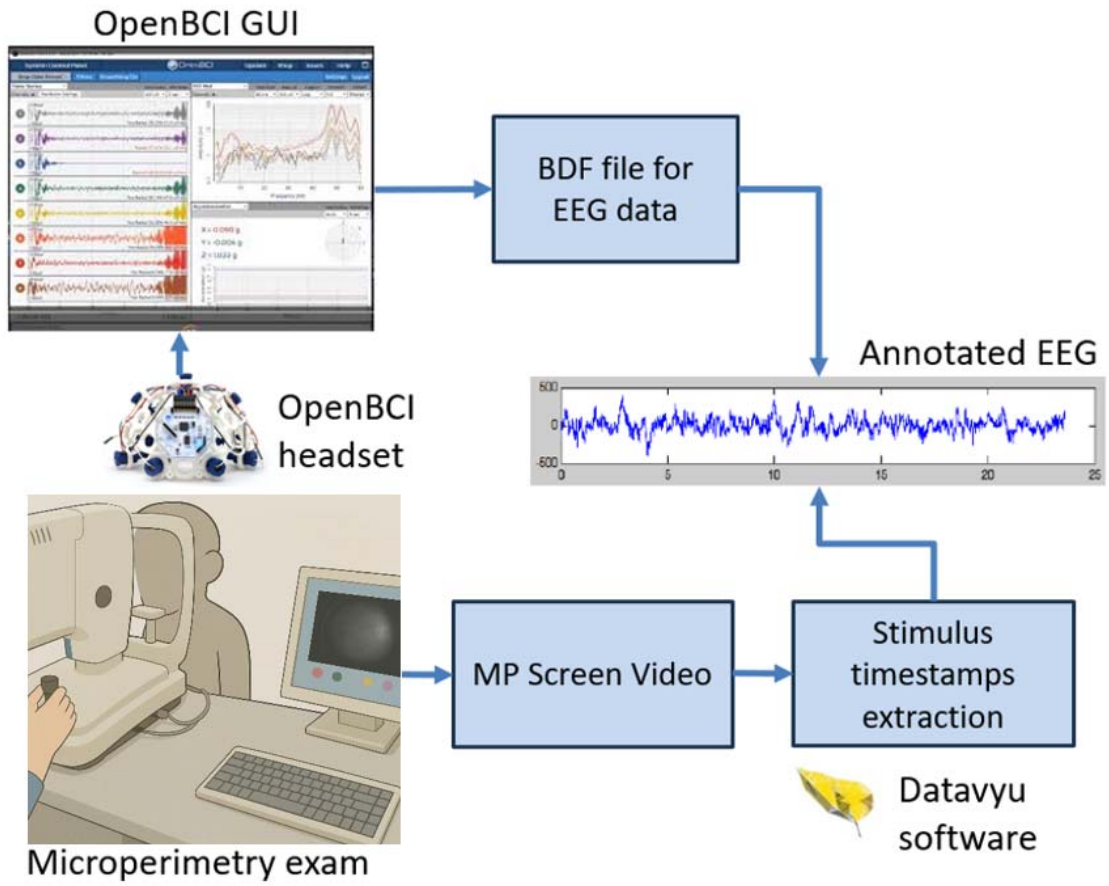


Figure 3: Flow chart for EEG data acquisition during microperimetry examination, and EEG data annotation for stimulus timestamps.

This pilot data for hypothesis testing contains 12 trials from two participants. Subject ID 1 is a Filipino male in his 30s, and subject ID 2 is a White female participant in her 20s, without any corrective lenses or medications used during trials. The male subject has a visual acuity of 20/20 while the female participant has a visual acuity of 20/15. Five trials were conducted with high-intensity and seven trials were conducted with low-intensity stimulus settings. The eye laterality, length of each trial, fixation stability rate, and number of stimuli are presented in **Table 1**. Informed consent was obtained by all the participants and guided about the use and expected outcomes of this study. This pilot study incorporates two subjects with normal visual functions to validate the hypothesis that advanced machine learning techniques can elicit neural responses of non-flickering and long-duration MP stimulus.

Trial ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Duration (seconds)	92	122	94	103	92	171	186	188	191	185	191	213
Stimuli	39	36	40	45	44	75	75	75	84	65	71	57
Laterality	OD	OD	OD	OD	OD	OD	OD	OD	OD	OD	OD	OD
Subject ID	1	1	2	2	2	1	1	1	2	2	2	2
Fixation Stability rate (%)	92.4	63.9	94.7	85.4	87.0	93.5	91.9	91.0	91.6	85.4	79.6	72.8

Table 1: Details of data acquired including laterality, length of trial, fixation stability rate, and number of stimuli presented for each trial.

EEG Signal Processing

The EEG pre-processing was used to prepare the signal for a deep-learning model to classify stimulus and non-stimulus segments. The detailed framework is presented in **Figure 4**. EEG data were re-referenced using the common median reference (CMR) as a pre-processing step to improve signal quality.

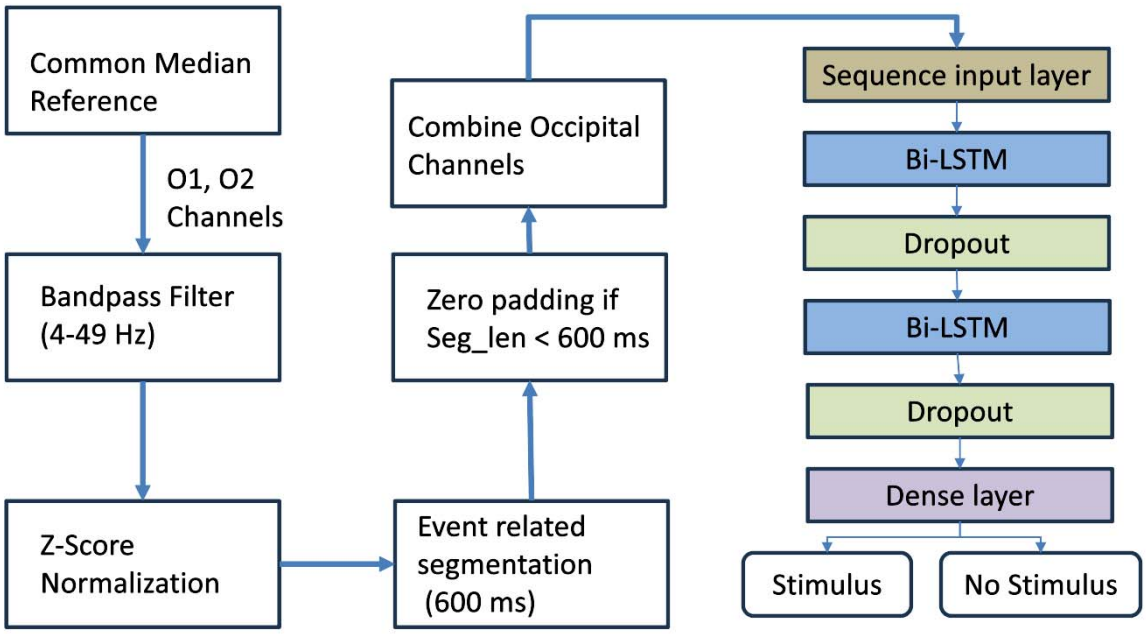


Figure 4: Flow chart for detection of stimulus and non-stimulus EEG segments using O1 and O2 channels.

The CMR re-referencing technique is less sensitive to outliers, highlights local neural activity by removing global fluctuations, environmental and muscle noise, and works well even when some channels are corrupted. The median of all eight channels was computed and subtracted from each channel. Suppose $S_i(t)$ is the EEG signal from channel i at time t, and $M(t)$ represents the median across all eight channels, and the CMR is represented by equation 1.

$$CMR_S_i(t) = S_i(t) - M(t) \qquad \textbf{(Equation 1)}$$

The artifact removal and standardization of the EEG signal is significant for uniform scaling of features to test our hypothesis of exploring deep learning architecture for subtle response elicitation of non-standard visual stimuli. To achieve this, we performed further artifact removal, and z-score based data normalization. After CMR, the slow drifts in EEG signals obscuring P1 and N1 peaks were eliminated by using the low cutoff frequency of 4 Hz. Similarly, the muscle and electrical interference noise was removed by using the high cutoff frequency of 49 Hz. To achieve this, a bandpass Butterworth filter of 4-49 Hz is designed with filter order 4 to remove both low and high-frequency noises. This bandpass filtering helps optimize the EEG signal for flash signal analysis on neuronal visual processing for parieto-occipital electrodes. After bandpass filtering, all channels were normalized using z-score normalization. For that purpose, the average of the signal is subtracted and then divided by the standard deviation to extract uniform scale features, required as input for a deep-learning architecture.

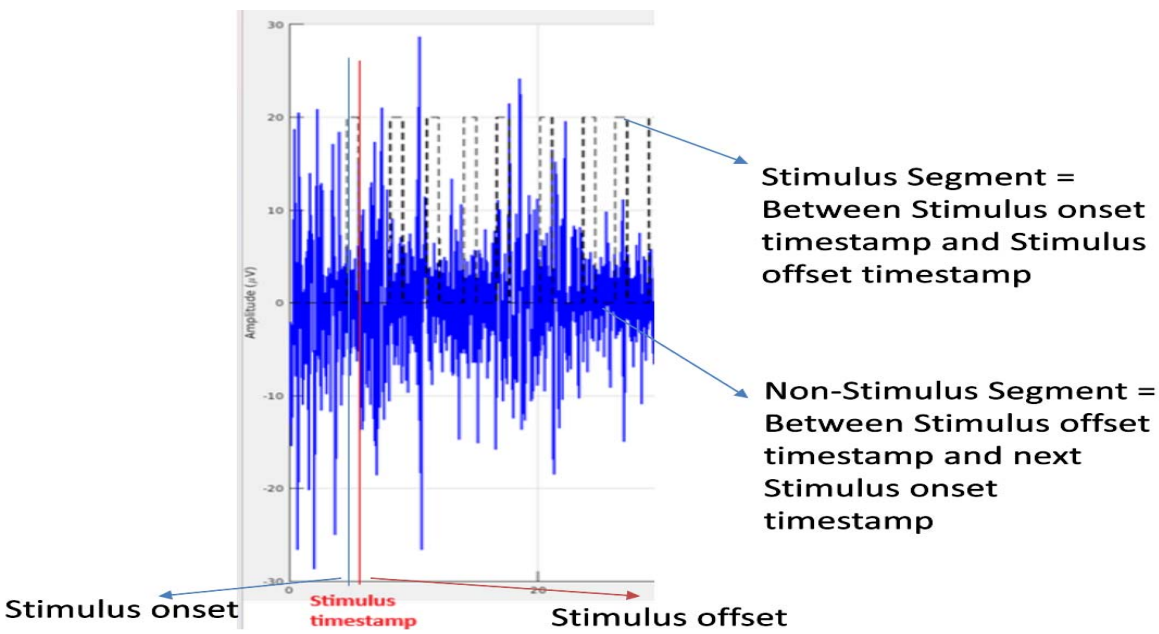


Figure 5: The O1 channel of EEG signal depicting stimulus timestamp, and segmentation of 600 ms chunk before the stimulus timestamp as stimulus segment and everywhere else as non-stimulus segment.

In the absence of built-in hardware synchronization capabilities within the microperimetry device, we adopted a software-based preprocessing approach to approximate temporal

correspondence between visual stimuli and neural responses. To achieve this, we performed event-related segmentation on the normalized EEG data based on stimulus timestamps, as shown in Figure 5. The segmentation window of 600 ms before each stimulus offset timestamp is considered a stimulus segment. The rationale for choosing 600 ms pre-offset window is to consistently capture both the full stimulus duration, preceding early onset neural activity significant for assessing baseline dynamics, while avoiding inclusion of post-perceptual or motor-related processes. Secondly, the offset timestamp is captured using Datavyu software displaying recorded video of the MP screen; therefore, there is a time buffer between the actual stimulus offset, the patient triggering the push button, and the results displayed on the recorded video of the MP screen. The VEP response to a flash stimulus is normally 300 ms, but to cater to the potential lag, a 600 ms window is used in our methods. The other part of the signal outside the stimulus segment is considered a non-stimulus part, which is further divided into 600-ms segments to make all the segments equal in length. Zero padding was performed if any segment lengths were less than 600 ms. This preprocessing of windowing larger lengths was necessitated to ensure meaningful temporal analysis in the absence of hardware-level time-locking.

Deep Learning Model

The context of isolated single-flash, long-duration MP stimuli poses unique challenges for binary classification between the stimulus and the non-stimulus EEG segments, which necessitates the use of a more temporally expressive architecture. The single-flash and relatively long-duration of the MP device stimulus results in inherently low signal-to-noise (SNR) due to the absence of repetitive averaging and temporal dispersion of neural responses. The temporal dispersion is further complicated with the lack of native support of MP devices for hardware-level synchronization, which requires segmentation windowing of a length (600 ms) larger than the normal VEP pattern. This increased window size introduces the risk of overlapping noise and unrelated brain dynamics, making conventional feedforward or shallow recurrent models insufficient for discriminating subtle, temporally distributed signal patterns.

To address these MP-based methodological constraints, we employ a stacked bidirectional long short-term memory (BiLSTM) architecture. The model is configured with two sequential BiLSTM layers, each followed by a dropout layer. The dual BiLSTM layers capture temporal dependencies and contextual features from past and future states by processing input sequences in both forward and backward directions. The abstract-level temporal features were extracted by the first BiLSTM layer, while deeper-level patterns were extracted in the second BiLSTM layer. The dropout layer helps mitigate overfitting by randomly deactivating neurons during model training. The proposed architecture offers the ability to model long-range temporal dependencies within an extended EEG window length and context-aware sequence learning because of the bidirectional structure of LSTM. The stacking of multiple BiLSTM layers allows the network to extract higher-level spatiotemporal features sensitive to subtle stimulus-induced neural responses while being robust to background noise and timing variability. The mini-batch size of 8 was used with an ADAM optimizer during the training of this architecture.

Results

We first report overall model performance per trial for the two stimulus intensity conditions (Table 2), followed by a comparison of electrode montages (Table 3). We then describe how stimulus intensity and electrode selection affected classification accuracy and error patterns. The experimentation protocol was defined to test the hypothesis of the ability of deep-learning architecture to elicit subtle neural responses of visual stimulus in the context of the MP exam. For each of the 12 trials, the data was split into training, validation, and testing based on several stimuli presented during the trial. Validation and test sets were 15% and the train set was 70% for each trial. We have computed different performance metrics, including accuracy, precision, sensitivity, F1-score, specificity, and Cohen's Kappa. Accuracy measures the overall proportion of correct classification, while precision measures the proportion of stimulus predictions that are correct according to the ground truth. Specificity measures the correct classification of non-stimulus segment predictions, while the F-1 score balances precision and sensitivity. The detailed results of all the clinical trials are presented in **Table 2**. Because the sample is small, wide variability is expected, as shown in different trials for this pilot study of hypothesis testing.

Trial ID	Intensity	Accuracy	Precision	Sensitivity (TPR)	F1-Score	Specificity (TNR)	Cohen's Kappa
1	High (0/4 dB)	77.42	50.00	14.29	22.22	95.83	13.55
2	High (0/4 dB)	80.00	45.45	71.43	55.56	81.82	43.46
3	High (0/4 dB)	70.97	33.33	12.50	18.18	91.30	4.78
4	High (0/4 dB)	74.29	50.00	22.22	30.77	92.31	17.75
5	High (0/4 dB)	70.97	25.00	14.29	18.18	87.50	2.11
6	Low (16 dB)	75.44	50.00	28.57	36.36	90.70	22.52
7	Low (16 dB)	78.69	62.50	33.33	43.48	93.48	31.81
8	Low (16 dB)	58.06	31.03	60.00	40.91	57.45	13.24
9	Low (16 dB)	76.19	57.14	25.00	34.78	93.62	22.86
10	Low (16 dB)	77.42	46.15	46.15	46.15	85.71	31.87
11	Low (16 dB)	69.84	14.29	7.14	9.52	87.76	-6.21
12	Low (16 dB)	74.29	35.00	58.33	43.75	77.59	28.41

Table 2: Performance metrics in percentage for stimulus and non-stimulus EEG segment classification for all low intensity and high intensity stimuli trials.

High Intensity (0/4 dB) Results

For the high-intensity group, accuracy ranges from approximately 70.97% to 80.00%, depicting consistency in overall predictions. In one trial, a precision of 50.00% with a low sensitivity of 14.29% indicates that while half of the positive predictions are correct, many stimulus segments are misclassified. Another trial reflects a precision of 45.45% but a much higher sensitivity of 71.43%, depicting a model behaving aggressively to identify stimulus segments but at the cost of precision. Generally high (81.82% to 95.83%) values of specificity suggest that the model reliably identifies non-stimulus segments, mainly due to large proportions in the data. The large proportion of non-stimulus segments is because there is a gap of 2-3 sec between each stimulus presentation, while the stimulus only lasts for 200 ms. Cohen's Kappa values range from about 2.11 to 43.46, indicating that despite decent accuracy, the agreement may not be robust in some scenarios. This variability in performance metrics implies that the model behavior is sensitive to class imbalances of stimulus and non-stimulus segments.

Low Intensity (16 dB) Results

Performance in the seven low-intensity trials was more variable (accuracy range ~58% to 79%). One trial had an outlier low accuracy (~58%). Precision ranged from an extremely low 14% up to 62%, and sensitivity from 7% up to 60%. This indicates some low-intensity trials were very challenging, with the model missing most stimuli (as seen in one trial with only 7% sensitivity), whereas in other trials the model achieved moderate success. Notably, even when sensitivity was higher (~60%), precision tended to be moderate (~31%), again reflecting the precision-sensitivity trade-off. Most Cohen's kappa values were positive but low (several in the 20–30 range, with one trial yielding a slightly negative kappa of –6), underscoring the difficulty the model had consistently detecting faint stimuli.

In summary, the model could detect high-intensity flashes with moderate success and relatively consistent accuracy, but performance for low-intensity (dim) stimuli was more erratic and generally lower. This suggests that detecting very low luminance stimuli via EEG is feasible but prone to errors, likely due to weaker neural signals and lower signal-to-noise ratio.

Comparison of EEG Electrode Combinations:

During our experiments, we covered the parietal-occipital region using P3, P4, PO3, PO4, POz, O1, O2, and Oz EEG electrodes, key for visual processing. For the classification of subtle responses in EEG channels due to individual flash stimuli, we empirically tested different combinations of electrodes using the proposed methodology.

- **O1–O2 vs. others:** In general, a combination of O1 and O2 performs better than other combinations of electrodes as described in Table 3. This is because the occipital electrodes are positioned right over the primary visual cortex. O1 and O2 performed better than their combination with Oz, except in one of the twelve trials. This can be due to the ability of lateral occipital electrodes to capture more focal, lateralized visual responses compared to Oz.
- **Oz and noise:** The central location of the Oz electrode is also susceptible to more eye or muscle artifacts resulting in a low signal-to-noise ratio, reflecting a low-performance combination when added to O1 and O2.
- **Parietal channels' role:** These channels are primarily involved in spatial processing, attention, and integration of visual information. In some of the outliers, parietal channels performed better than occipital channels for stimulus detection. This can be due to the improvement in signal-to-noise ratio through broader coverage of visual and parietal cortices. These are occasional situations where parietal or broader parietal-occipital montages yield better flash stimulus detection than the standard occipital pair alone.
- **All channels vs. subsets:** Notably, using all 8 channels together did not produce the best results (Table 3), likely because the inclusion of non-visual channels added noise or irrelevant signals that confused the classifier, reinforcing the value of focusing on relevant channels.

Trial ID	Intensity	All Channels	All except O1, O2	O1, O2	O1, O2, Oz	P3, P4	PO3, PO4, POz
1	High (0/4 dB)	38.71	70.97	77.42	64.52	74.19	64.52
2	High (0/4 dB)	45.16	54.84	80.00	58.06	70.97	54.84
3	High (0/4 dB)	48.39	61.29	70.97	70.97	64.52	70.97
4	High (0/4 dB)	62.12	71.54	74.29	64.37	66.28	66.28

5	High (0/4 dB)	68.69	61.29	70.97	66.16	66.16	66.16
6	Low (16 dB)	66.67	70.18	75.44	50.88	71.93	49.12
7	Low (16 dB)	75.81	48.39	78.69	77.42	59.68	75.81
8	Low (16 dB)	61.90	61.90	58.06	55.56	55.56	55.56
9	Low (16 dB)	70.31	67.19	76.19	59.38	43.75	60.94
10	Low (16 dB)	72.12	70.48	77.42	66.75	66.75	72.12
11	Low (16 dB)	69.84	71.43	69.84	66.67	71.43	66.67
12	Low (16 dB)	73.24	57.75	74.29	76.06	69.01	76.06

Table 3: Accuracy comparison in percentage for analysis with different combinations of EEG electrodes.

Discussion and Implications

This pilot study demonstrates that single flash stimuli presented during a microperimetry exam can be detected via occipital EEG using deep learning. Our Bi-LSTM model achieved up to ~80% accuracy in classifying stimulus-present vs. stimulus-absent EEG segments under optimal conditions (high-intensity stimuli, occipital electrodes only), supporting our hypothesis that subtle visual cortex responses to individual 200 ms flashes can be captured without averaging. However, detection of lower-intensity stimuli was less reliable, and overall performance varied considerably between trials. The use of a default 200-ms stimulus duration of the MP exam is considerably longer than what is typically used to elicit standard flash VEPs (often 5–20 ms) or steady-state VEPs (which require flickering stimuli, e.g., 5–30 Hz). Because the MP1's flash duration (200 ms) is much longer than the brief flashes or flickering stimuli used in conventional VEP recordings, a classic time-locked VEP waveform (e.g., P100) was not evident in single trials. This compelled us to analyze longer 600 ms EEG segments with machine learning, rather than relying on clear peak identification. The trade-off of this approach is that it's less interpretable – we cannot easily point to a specific 'wave' corresponding to the stimulus – but it allows detection of stimuli in conditions where traditional VEP analysis fails.

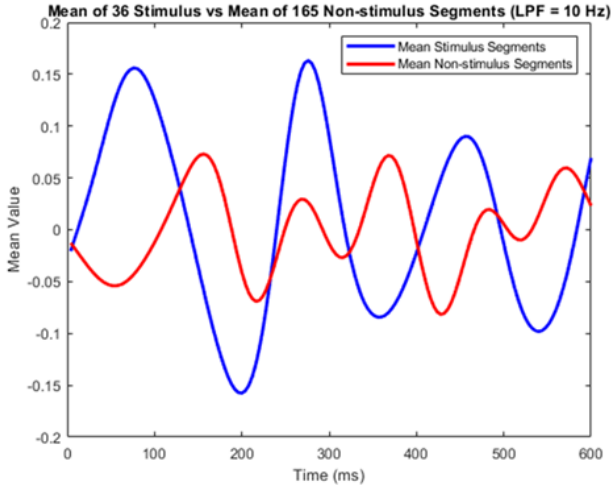
The results reveal a clear trade-off between precision and sensitivity. In scenarios where sensitivity is prioritized, precision tends to drop, which might lead to a higher number of false positives. The model's performance pattern – high specificity but varying balance between precision and sensitivity – suggests it was biased toward predicting the majority class (non-stimulus). In trials where the model 'played it safe' (high precision, low sensitivity), it avoided false positives but missed many actual stimuli, perhaps reflecting a stricter internal decision threshold or lower SNR for those trials. In trials with higher sensitivity, the model caught more stimuli but at the cost of more false alarms, indicating a different threshold balance. This variability could result from the small dataset and differing noise levels per trial. Future training strategies (e.g. adjusting decision thresholds, cost-sensitive loss functions, or class-balancing techniques) could improve this trade-off so that both false negatives and false positives are minimized.

Existing studies for objective microperimetry lack spatial specificity and require extensive averaging, which is not feasible with non-standard stimuli of MP for VEP pattern detection. Can deep learning measure subtle cortical responses to sparse photic stimuli of Microperimetry? We investigated this question by analyzing parieto-occipital responses to a single-flash stimulus during microperimetry exam using deep learning architecture without relying on signal averaging

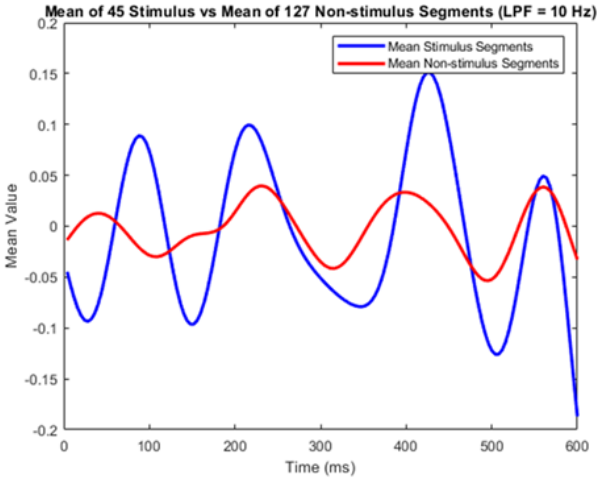
for VEP detection. Our study explores a novel deep learning-based method to automatically register the MP stimulus without the conscious effort of the patient using their electroencephalogram (EEG) signals. In this study, we discovered that the individual flashlight stimulus of 200 ms used in MP tests generates subtle responses in the O1 and O2 electrodes of EEG, which can be measured with a deep learning architecture. The proposed method has the potential to improve the reliability of MP with more frequent flashing to the patient's blind spot. Fixation losses can be reduced as the focus of the patient will not shift from the fixation point to manual stimulus registration.

Given the exploratory nature of the hypothesis and encouraging findings, our study also has certain limitations. One of the limitations is to explore the longer EEG segments after classification by a stacked BiLSTM model to be further validated for VEP pattern detection. Given the scope of our study, we visualized the average segments of all the trials for the detection of VEP for hypothesis validation purposes. For each trial, we have more non-stimulus segments than stimulus segments, but the average of those two types of segments is visually distinguishable even in the absence of the VEP pattern, as depicted in Figure 6. The amplitude of the mean of stimulus segments is generally higher than non-stimulus segments with low frequency. The plots in Figure 5 are passed to a low pass filter of 10 Hz for visual distinction. These visualizations show that even in the absence of a VEP pattern for individual flash stimuli with given limitations, how deep learning model can distinguish the difference between stimulus and non-stimulus segments.

Another limitation is that the model was trained and tested on data segments derived from the same trial. This within-trial splitting approach may not accurately reflect the model's performance on entirely new, unseen data. The future work will incorporate the cross-trial and cross-subject models' performance. One more limitation of our study is the small sample size of two participants with 12 trials, as this is the pilot feasibility study for a deep-learning-based objective MP exam using neural signals. A more diverse participant pool with different ages, ocular pathologies, and media opacities is essential to establish generalizability for the feasibility of using objective microperimetry in clinical practice.



(A) Average Trial ID 2 segments



(B) Average of Trial ID 4 segments

Figure 6: Averaging of all the segments to determine the existence of distinguishable VEP patterns from stimulus and non-stimulus segments (A) segments within trail ID 2 (B) segments within trail ID 4.

Improvement in our results underscores the need for hardware-based synchronization between EEG and Microperimetry devices, exploring ways to modulate the MP stimulus according to the standard for capturing the VEP signal. This external trigger-based synchronization is expected to minimize the misalignment between the EEG signal and actual stimulus presentation from 250 ms to around 2 ms, potentially increasing the efficiency and reliability of stimulus detection. The results also depict the need for model training combining CNN and LSTM for extracting tempo-spectral features and repetitive patterns, especially to address the inconsistencies observed across different intensity conditions. By focusing on optimizing the balance between key experimentation setups and enhancing model calibration, future iterations of the study could achieve more reliable and robust performance required for clinical usage. Our approach could be extended to pediatric or cognitively impaired patients to objectively measure their retinal sensitivity using the Microperimetry exam. With hard-level synchronization, we will be able to do real-time classification of stimulus and non-stimulus EEG segments in a future study to objectively measure Microperimetry outcome during the examination.

In summary, our study provides an initial validation that EEG signals can reflect single flash stimuli in a microperimetry context, though significant technical improvements are needed to reach clinical viability. We next focus on addressing the current limitations (synchronization, stimuli, generalizability) to move this concept from proof-of-principle toward a practical translational tool.

Conclusion

We developed and tested a novel method to objectively register microperimetry stimuli using EEG. Our results show that even a dim (3.2 cd/m²) flash stimulus can be detected via occipital EEG with reasonable accuracy (~75% or higher in multiple trials), especially in distinguishing when no stimulus is present. Lateral occipital electrodes (O1, O2) provided the most robust signals for this task. This proof-of-concept in two healthy subjects suggests that EEG signals can augment microperimetry by removing reliance on the patient's button press. However, translating this into practice will require addressing synchronization issues and validating the approach in larger, diverse populations. With further refinement – including hardware trigger integration for precise timing and more advanced modeling – EEG-based stimulus detection could enhance the speed, reliability, and accessibility of visual field testing. We plan to implement real-time EEG stimulus detection in future studies, bringing objective microperimetry closer to clinical reality.

References

1. Yang, Y. and Dunbar, H., 2021. Clinical perspectives and trends: microperimetry as a trial endpoint in retinal disease. *Ophthalmologica*, *244*(5), pp.418-450.
2. Wu, Z., Hadoux, X., Jannaud, M., Martin, K.R., van Wijngaarden, P. and Guymer, R.H., 2023. Systematic underestimation of visual sensitivity loss on microperimetry: implications for testing protocols in clinical trials. *Translational Vision Science & Technology*, *12*(7), pp.11-11.
3. Csaky, K.G., Patel, P.J., Sepah, Y.J., Birch, D.G., Do, D.V., Ip, M.S., Guymer, R.H., Luu, C.D., Gune, S., Lin, H. and Ferrara, D., 2019. Microperimetry for geographic atrophy secondary to age-related macular degeneration. Survey of ophthalmology, 64(3), pp.353-364.
4. Coulibaly, L.M., Mohamed, H., Fuchs, P., Schmidt-Erfurth, U. and Reiter, G.S., 2024. Inter and intradevice assessment of microperimetry testing in aging eyes. Scientific Reports, 14(1), p.1049.
5. Markowitz, S.N. and Reyes, S.V., 2013. Microperimetry and clinical practice: an evidence-based review. Canadian journal of ophthalmology, 48(5), pp.350-357.
6. Josan, A.S., Farrance, I., Taylor, L.J., Adeyoju, D., Buckley, T.M., Jolly, J.K. and MacLaren, R.E., 2023. Microperimetry Reliability Assessed From Fixation Performance. *Translational Vision Science & Technology*, *12*(5), pp.21-21.
7. Pfau, M., Jolly, J.K., Wu, Z., Denniss, J., Lad, E.M., Guymer, R.H., Fleckenstein, M., Holz, F.G. and Schmitz-Valckenberg, S., 2021. Fundus-controlled perimetry (microperimetry): application as outcome measure in clinical trials. *Progress in Retinal and Eye Research*, *82*, p.100907.
8. Khubieva, R.R., 2021. Microperimetry in Children's Practice. Ophthalmology in Russia, 18(4), pp.807-814.
9. Jones, P.R., Yasoubi, N., Nardini, M. and Rubin, G.S., 2016. Feasibility of macular integrity assessment (MAIA) microperimetry in children: sensitivity, reliability, and fixation stability in healthy observers. Investigative ophthalmology & visual science, 57(14), pp.6349-6359.
10. Wu, Z., Ayton, L.N., Guymer, R.H. and Luu, C.D., 2013. Intrasession test–retest variability of microperimetry in age-related macular degeneration. Investigative ophthalmology & visual science, 54(12), pp.7378-7385.
11. Wong, E.N., De Soyza, J.D., Mackey, D.A., Constable, I.J. and Chen, F.K., 2017. Intersession test–retest variability of microperimetry in type 2 macular telangiectasia. Translational Vision Science & Technology, 6(6), pp.7-7.
12. Seeböck, P., Vogl, W.D., Waldstein, S.M., Orlando, J.I., Baratsits, M., Alten, T., Arikan, M., Mylonas, G., Bogunović, H. and Schmidt-Erfurth, U., 2022. Linking function and structure with ReSensNet: predicting retinal sensitivity from OCT using deep learning. *Ophthalmology Retina*, *6*(6), pp.501-511.
13. Messenio, D., Babbi, A., Guglielmi, A. and Airaldi, M., 2022. Focal electroretinogram and microperimetry testing of photoreceptor-retinal pigment epithelium function in intermediate age-related macular degeneration. *Acta Ophthalmologica*, *100*(3), pp.277-284.
14. Versek, C., Banijamali, S.M.A., Bex, P., Lashkari, K., Kamarthi, S. and Sridhar, S., 2021. Portable diagnostic system for age-related macular degeneration screening using visual evoked potentials. *Eye and brain*, pp.111-127.
15. Banijamali, S.M.A., Versek, C., Babinski, K., Kamarthi, S., Green-LaRoche, D. and Sridhar, S., 2024. Portable multi-focal visual evoked potential diagnostics for multiple sclerosis/optic neuritis patients. *Documenta Ophthalmologica*, *149*(1), pp.23-45.

16. Schmetterer, L., Scholl, H., Garhöfer, G., Janeschitz-Kriegl, L., Corvi, F., Sadda, S.R. and Medeiros, F.A., 2023. Endpoints for clinical trials in ophthalmology. *Progress in Retinal and Eye Research*, p.101160.
17. Vidal-Jordana, A., Rovira, A., Arrambide, G., Otero-Romero, S., Río, J., Comabella, M., Nos, C., Castilló, J., Galan, I., Cabello, S. and Moncho, D., 2021. Optic nerve topography in multiple sclerosis diagnosis: the utility of visual evoked potentials. *Neurology*, *96*(4), pp.e482-e490.
18. Aldridge, A., Barnes, E., Bethel, C.L., Carruth, D.W., Kocturova, M., Pleva, M. and Juhar, J., 2019, April. Accessible electroencephalograms (EEGs): A comparative review with openbci's ultracortex mark IV headset. In *2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)* (pp. 1-6). IEEE.
19. Xie, J., Du, G., Xu, G., Zhao, X., Fang, P., Li, M., Cao, G., Li, G., Xue, T. and Zhang, Y., 2019. Performance Evaluation of Visual Noise Imposed Stochastic Resonance Effect on Brain-Computer Interface Application: A Comparison Between Motion-Reversing Simple Ring and Complex Checkerboard Patterns. *Frontiers in Neuroscience*, *13*, p.1192.
20. Chen, F.K., Patel, P.J., Xing, W., Bunce, C., Egan, C., Tufail, A.T., Coffey, P.J., Rubin, G.S. and Da Cruz, L., 2009. Test–retest variability of microperimetry using the Nidek MP1 in patients with macular disease. *Investigative ophthalmology & visual science*, *50*(7), pp.3464-3472.
21. Shah, V.A. and Chalam, K.V., 2008. Values for macular perimetry using the MP-1 microperimeter in normal subjects. Ophthalmic research, 41(1), pp.9-13.
22. Jarc-Vidmar, M., Popovič, P. and Hawlina, M., 2006. Mapping of central visual function by microperimetry and autofluorescence in patients with Best's vitelliform dystrophy. Eye, 20(6), pp.688-696.
23. Alibhai, A.Y., Mehta, N., Hickson-Curran, S., Moreira-Neto, C., Levine, E.S., Reichel, E., Duker, J.S. and Waheed, N.K., 2020. Test–retest variability of microperimetry in geographic atrophy. International Journal of Retina and Vitreous, 6, pp.1-6.
24. Datavyu Team (2014). Datavyu: A Video Coding Tool. Databrary Project, New York University. URL http://datavyu.org.

Disclosures

Authors' Contributions

YJS and MIA conceptualized the study. MASS, ASDN, KJ, TS and ZZF collected the data. MND analyzed the data. MND drafted the paper. All authors reviewed and approved the final draft.

Conflict of Interest

None.

Funding

This project was funded by the Byers Eye Institute at Stanford Medicine.

Ethical Approval

The Stanford Institutional Review Board approved this study.

Acknowledgments

None.