SHORT-PAPER

Microsaccades as Indicators for Early-Stage Parkinson's Disease: An Explainable AI Approach

YITING WANG, Karolinska Institutet, Stockholm, Stockholm, Sweden

MATTIAS NILSSON, Karolinska Institutet, Stockholm, Stockholm, Sweden

Open Access Support provided by:

Karolinska Institutet

Published: 25 May 2025

ETRA '25: 2025 Symposium on Eye Tracking Research and Applications
May 26 - 29, 2025
Tokyo, Japan

Conference Sponsors:
SIGGRAPH

Microsaccades as Indicators for Early-Stage Parkinson's Disease: An Explainable AI Approach

Yiting Wang
Department of Clinical Neuroscience
Karolinska Institutet
Stockholm, Sweden
yiting.wang@ki.se

Mattias Nilsson
Department of Clinical Neuroscience
Karolinska Institutet
Stockholm, Sweden
mattias.nilsson@ki.se

Abstract

Abnormal fixational eye movements are frequently observed in individuals with Parkinson's disease (PD). While these involuntary movements have the potential to serve as objective indicators for early-stage PD screening, the specific role of microsaccades has yet to be validated using computational methods such as machine learning. In this study, we combine explainable AI (XAI) techniques with end-to-end classification to evaluate the contribution of microsaccades in differentiating early-stage PD from healthy controls using data from a simple fixation task. An attention mechanism is employed to analyze raw binocular eye position data, while integrated gradients are used to identify the most influential samples in the model's decision-making process. Experimental results show that it is feasible to distinguish early-stage PD based on raw eye tracking data. Importantly, visualizing integrated gradients across eye position over time indicates that microsaccades play a key role in classification, underscoring their potential as discriminative features for early-stage PD screening.

CCS Concepts

• **Applied computing** → **Health informatics**; • **Computing methodologies** → **Knowledge representation and reasoning**.

Keywords

Fixational Eye Movement, End-to-end Classification, Explainable AI, Early-stage Parkinson's Disease

ACM Reference Format:

Yiting Wang and Mattias Nilsson. 2025. Microsaccades as Indicators for Early-Stage Parkinson's Disease: An Explainable AI Approach. In *2025 Symposium on Eye Tracking Research and Applications (ETRA '25), May 26–29, 2025, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3715669.3723133

1 Introduction

Eye movements abnormalities are commonly observed symptoms in early-stage Parkinson's Disease (PD) [Turcano et al. 2019]. Among the various types of eye-tracking data, fixational eye movements are affected by PD because of the increased gaze instability during fixation attempts. Fixational eye movements are involuntary, small-scale movements that occur when the eyes are fixated on a target, playing a crucial role in maintaining visual acuity by preventing sensory adaptation, counteracting visual fading, and ensuring continuous perception [Martinez-Conde 2006]. Microsaccades, the largest subtype of fixational eye movements, occur approximately 1–2 times per second, with amplitudes typically around 1 degree [Martinez-Conde et al. 2009]. These small, involuntary movements, while subtle, reveal insights into underlying neurological activity and share a common oculomotor generator with large saccades. Research has shown that PD patients exhibit a higher frequency and increased amplitude of microsaccades [Alexander et al. 2018]. This pattern underscores the potential of microsaccade analysis as an objective and non-invasive tool for the early detection and monitoring of PD. However, limited research has been conducted on microsaccade analysis using computational methods such as machine learning, leaving a gap in fully understanding their role in diagnostic of PD.

End-to-end eye-tracking analysis utilizing deep learning networks, which directly process raw data instead of relying on human-crafted features, has gained growing attention in recent years [Hoppe and Bulling 2016; Lohr and Komogortsev 2022; Zemblys et al. 2019]. This methodology has been widely adopted and developed for event detection [Krishnamoorthy et al. 2021; Wang et al. 2025; Zemblys et al. 2019] and biometric identification [Lohr and Komogortsev 2022; Makowski et al. 2021; Raju et al. 2024], achieving impressive performance with good generalization ability on different datasets. Meanwhile, there has been a recent surge in its use within medical domains [Deng et al. 2022; Sriram et al. 2023; Uribarri et al. 2023]. However, the "black-box" nature of deep learning poses a significant barrier to its adoption in real-world clinical settings. To address this limitation, pioneers in the field have increasingly employed Explainable AI (XAI) techniques to provide interpretable explanations alongside model predictions [de Belen et al. 2021; Krakowczyk et al. 2023a]. XAI explains model decisions by linking them to input features and visualizing learned patterns. Visualization is relatively straightforward with image or video data, as the input can be directly displayed. However, with limited studies applying XAI to eye movement data, intuitive and clear visualizations remain rare. Moreover, to the best of our knowledge, no prior research has utilized XAI to illustrate the role of microsaccades in classifying early-stage PD.

To address the aforementioned challenges, this study aims to answer the following research questions: (1) Can end-to-end classification with eye-tracking position data effectively discriminate

ETRA '25, Tokyo, Japan

ACM ISBN 979-8-4007-1487-0/25/05
https://doi.org/10.1145/3715669.3723133

between early-stage PD and HC? (2) Can XAI techniques explain the role of microsaccades in the classification process? To this end, we employ a multi-head self-attention network to analyze raw eye-tracking position data and classify the learned features into early-stage PD or HC using a feed-forward network. Integrated Gradients, one of the most effective XAI methods, is utilized to compute attribution maps by integrating gradients from the input to a baseline. The visualization comparing the attribution maps with the raw eye-tracking position data provides an intuitive demonstration of the contribution of microsaccades to the decision-making process. The main contributions of this study are as follows:

- We investigate a novel approach using computational methods, specifically XAI, to validate the contribution of microsaccades in distinguishing early-stage PD from HC. This work advances the adoption of explainable deep learning models in clinical settings.
- An end-to-end classification framework, based on the multi-head self-attention mechanism, is applied directly to 4-dimensional eye-tracking position data, eliminating the biases associated with hand-crafted features.
- The visualization of attribution maps alongside raw data provides an intuitive understanding of how microsaccades influence the decision-making process, paving the way for broader applications of XAI in eye movement analysis.

2 Methodology

2.1 Method Overview

Figure 1 illustrates the overall structure of our proposed method. The input consists of 4-dimensional time-series eye-tracking position data, representing the horizontal and vertical positions of both eyes. A non-overlapping sliding window is applied to the input data to segment it into patches, which are then fed into the attention network. Each patch undergoes position encoding, and a classification token is appended before being processed by the multi-head self-attention mechanism. This mechanism enables the model to interpret the input data from multiple perspectives, capturing both global and local dependencies. The extracted features are passed through a feed-forward network, which serves as the classifier to determine whether the data belongs to the PD or HC group. After the model is trained, integrated gradients are computed for the corresponding classification target. Finally, the resulting attribution map is compared with the input data to assess whether the most significant contributions to the classification originate from microsaccades.

We define raw eye tracking position data as $X = (x^1, x^2, ..., x^N) \in \mathbb{R}^{N\times D}$, where N represents the sample size and D denotes the data dimension, which is 4, corresponding to the horizontal and vertical positions of both the left and right eyes. After applying the sliding window, the data can be represented as $X_p = (x_p^1, x_p^2, ..., x_p^B) \in \mathbb{R}^{B\times T\times D}$ where B is equal to N divided by T, and T denotes the time window.

2.2 End-to-End Classification

A common challenge in applying deep learning to medical datasets is their relatively small size. This limitation increases the risk of overfitting, making it crucial to explore simple yet efficient networks for medical applications. To address this, our study employs a single-layer multi-head attention mechanism to extract features directly from the raw eye-tracking data. Meanwhile, to further alleviate overfitting, we utilize a cosine annealing scheduler to dynamically adjust the learning rate during optimization. Additionally, weight decay is applied to regularize the model by penalizing large weights, helping to improve generalization. Early stopping is also implemented, terminating training if the testing loss does not decrease for 10 consecutive epochs.

To preserve the temporal information of time series data, position embeddings are added element-wise to the input patches. Additionally, a learnable classification token, shown as CLS in Figure 1, is concatenated to the input sequence and serves as the standard approach in classification [Devlin 2018]. Among various methods for computing attention weights, the scaled dot-product attention has become one of the most popular approaches, largely due to the widespread adoption of the Transformer model [Vaswani 2017]. The attention is computed as following:

$$Q = XW_q, W_q \in \mathbb{R}^{D\times d_k} \tag{1}$$

$$K = XW_k, W_q \in \mathbb{R}^{D\times d_k} \tag{2}$$

$$V = XW_v, W_q \in \mathbb{R}^{D\times d_v} \tag{3}$$

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V. \tag{4}$$

Multi-head self-attention with h heads is the linear transformation of concatenation of h scaled dot-product attention as follows:

$$\begin{aligned} MultiHead(Q, K, V) &= Concat(head_1, ..., head_h)W \\ head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V), \end{aligned} \tag{5}$$

where W is the linear transformation matrix after concatenation, W_i^Q, W_i^K, W_i^V transform Q, K, V to different head, correspondingly. The primary advantage of multi-head attention over single-head attention is its ability to capture different aspects of relationships between samples and enhances the model's generalization capabilities by reducing the risk of overfitting. Eye-tracking position data inherently contains information from both temporal and spatial domains. By employing multi-head attention, the model can effectively learn representations from these distinct domains, enabling a more comprehensive understanding of the data.

After extracting the features with multi-head self-attention mechanism, a fully connected feed-forward network serves as the classifier, discriminating between PD and HC. The cross entropy loss is used as the classifier loss:

$$\mathcal{L} = -\sum_i y_i log y_i', \tag{6}$$

where y_i is the ground truth for input x_p, while y_i' is the predicted label.

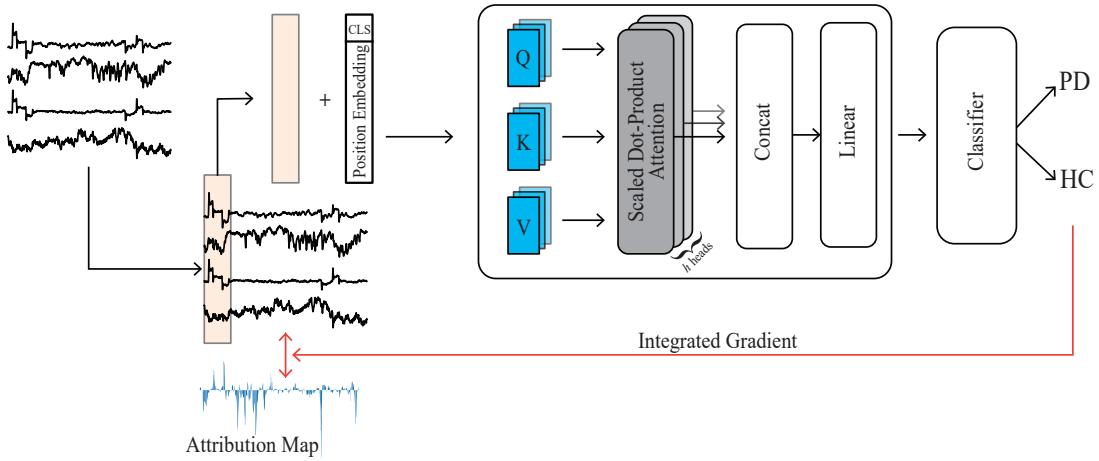

Figure 1: The overall structure of end-to-end classification and the comparison between raw data and attribution map

2.3 Explainable AI Method

In medical applications such as the early diagnosis of PD, interpretability of the model's learned features is crucial. Furthermore, no prior study has demonstrated the role of microsaccades using computational methods like XAI. To address this, we utilize Integrated Gradients (IG) [Sundararajan et al. 2017] for attributing the output of the attention-based classifier to its input features. By comparing the attribution map with the raw eye-tracking data, we aim to demonstrate that the samples contributing most significantly to the model's predictions originate from microsaccades. IG is chosen due to its superior performance compared to other gradient-based approaches [Qi et al. 2019], as well as its demonstrated effectiveness in analyzing eye-tracking data [Krakowczyk et al. 2023b,a].

IG computes the integral of the gradients along a path from a baseline input to the actual input. Formally, the attention-based classifier is represented as a function F, while $x' \in R^{B\times T\times D}$ denotes the baseline input. For the eye-tracking time-series data, we set the baseline to a zero embedding vector, ensuring a neutral reference point for feature attribution. Specifically, a straightline path from the baseline to the input is considered, and gradients are computed and cumulated at all points along the path. [Sundararajan et al. 2017] illustrated that IG satisfies a completeness axiom which the integrated gradients for all samples add up to the difference between the output of F at the input x and the baseline x', based on the fact that F is differentiable almost everywhere.

3 Experiments

3.1 Dataset

In this study, we utilize the dataset collected by Tsitsi *et al.* [Tsitsi et al. 2021], comprising 50 patients with PD and 43 age-matched healthy controls (HC). The PD cohort primarily consists of individuals in the early stages of the disease, characterized by unilateral or mild-to-moderate symptoms, with a median disease duration of 2 years. Eye movements were recorded using the Tobii Pro Spectrum system without a chinrest, operating at a 1200 Hz sampling rate. During data collection, participants were instructed to fixate on a 6 mm diameter black dot positioned at the center of a bright white screen, performing a simple fixation task. The fixation target remained visible for 15 seconds, followed by a 5-second blank interval in each trial. Participants completed a total of eight successive trials within 3 minutes, making the task brief and minimally demanding. In total, the dataset includes 400 trials for the PD group (6000 seconds) and 344 trials for the HC group (5160 seconds).

3.2 Data Preprocessing

To address missing values, outliers, and noise introduced by the eye tracker, and to enhance the quality and efficiency of subsequent end-to-end classification, we implement a four-step preprocessing pipeline on data from all 93 participants after converting the eye-tracking position data to degrees.

- Invalid trials were excluded based on two criteria: (1) the trial duration was less than 1000 ms, or (2) more than 70% of the samples in the trial were missing. After this, 44 trials from PD group and 35 trials from HC group have been removed.
- Blink detection and removal were performed by identifying successive missing values in the eye-tracking data. Blinks typically result in an average data loss of approximately 200 ms, while eyelid closure and reopening significantly impact eye movement signals [Collewijn et al. 1985]. To account for these effects, we removed consecutive missing data segments of 200 ms, along with an additional 300 ms of data before and after each blink, to ensure the elimination of any residual signal distortion associated with the blink event.
- After removing blinks, the remaining missing samples were filled using linear interpolation to prevent errors during smoothing and classification.
- The data were smoothed using a low-pass filter with a Bartlett window of 20 ms. This preprocessing step reduced noise, resulting in a cleaner and smoother dataset compared to the original raw data.

3.3 Implementation Details

Considering the size of our dataset, separating training, validating and testing datasets would lead to limited data samples in all

datasets. Therefore, we use five-fold cross-validation where each fold grantees to have almost equal numbers of PD and HC subjects.

Tuned hyper-parameters include the choice of optimizer, the learning rate, the weight decay of optimizer, non-overlapping time-window, number of heads in the multi-head self-attention mechanism. The selected tuning ranges and final values for the model parameters are presented in Table 1. We implement our proposed method using Python 3.11.5 and PyTorch 2.5.1 with CUDA 12.4, IG is implemented with Captum library [Kokhlikyan et al. 2020]. All code is running on one NVIDIA GeForce RTX 4080 SUPER GPU.

3.4 Classification Results

Classifying early-stage PD and age-matched HC using raw eye-tracking position data collected during a simple fixation task is challenging. Our primary goal is to demonstrate that raw positional eye-tracking data possess discriminative power for distinguishing between early-stage PD and HC. To evaluate the classification results, we calculate accuracy, F1 score, and receiver operating characteristic (ROC). In addition to sample-level performance obtained directly from the model, we apply hard voting to aggregate predictions for all samples belonging to a single subject, yielding subject-level predictions and corresponding performance metrics. Table 2 shows that the attention-based classifier achieves above-chance accuracy, demonstrating its ability to discriminate between early-stage PD and HC. Compared to the chance level (50% accuracy), the accuracy improvement is statistically significant at the sample level (($p < 0.001$)), but not at the subject level. Fold three achieves the highest subject-level accuracy and F1 score, both exceeding 70% and statistically significant($p < 0.001$ on the sample level, $p = 0.039$ on the subject level). On average, subject-level performance surpasses sample-level results, achieving approximately 65% accuracy and 62% F1 score. A permutation test is conducted, in which labels are randomly assigned to each sample, yielding an average accuracy of 52.33% over 100 iterations. Although our classification performance is only moderate in absolute terms, the accuracy remains statistically significant compared to the permutation test ($p < 0.001$).

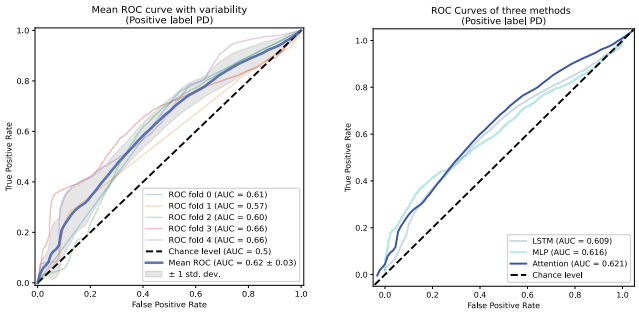

Figure 2: Mean Receiver Operating Characteristic (ROC) curve for early-stage Parkinson's disease classification. The left panel illustrates the five-fold cross-validation performance of the attention network, while the right panel compares different classification methods.

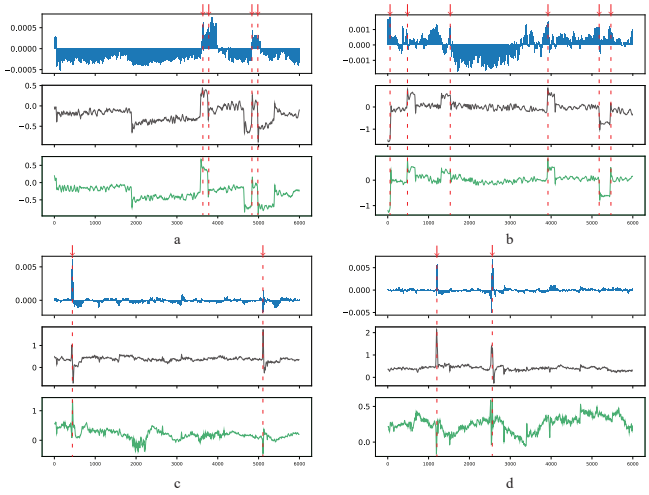

Figure 3: Comparison between attribution map and raw positional eye-tracking data over 5s. The red dashed lines indicate the alignment between microsaccades and the samples that contribute most to the classification results. (a) and (b) represent two PD subjects, while (c) and (d) represent two HC subjects.

Figure 2 presents the ROC curves for each fold of the cross-validation process of our proposed method (left) and a comparison among different classification approaches (right). The mean ROC curve of the attention network achieves an AUC of 0.62 ± 0.03, indicating moderate classification performance. Although the primary objective of this work is to introduce the novel application of XAI in highlighting the role of microsaccades in PD classification, we also compare the performance of our attention network with a standard LSTM model and Multilayer Perceptron (MLP) trained on traditional feature-based data. The AUC comparison, presented in the right panel of Figure 2, indicates comparable performance across different methods. Future work will focus on enhancing the performance of the end-to-end algorithm on our dataset.

3.5 Contribution of Microsaccades

Figure 3 visualizes the comparison between the attribution map and raw positional eye-tracking data. In each sub-figure, the three plots (from top to bottom) respectively represent the attribution map, the left eye's horizontal position, and the right eye's horizontal position. The X-axis represents the number of samples, with 1200 samples corresponding to 1 second due to the 1200 Hz sampling rate. The Y-axis represents the actual values without normalization. The red dashed lines highlight the alignment between microsaccades and the samples contributing most to the classification results. For the attribution map, higher values indicate greater contributions to the classification, with positive values representing contributions toward the target class and negative values representing contributions against it. In general, the highest positive values in the attribution map correspond to samples associated with microsaccades, illustrating that microsaccades play a crucial role in the decision-making process. Quantitatively, the Pearson correlation

Table 1: Tuning ranges and final values of hyperparameters

	Optimizer	Learning Rate	Time Window	Num Heads	Weight Decay
Options	AdamW, Adam, RMSprop	1e-6 to 1e-1	120,240,360,480,600	1 to 6	1e-12 to 1e-2
Final Value	RMSprop	8e-3	360	4	1e-4

Table 2: Accuracy and F1 score (%) for each fold and their overall average across all folds, reported at both sample-Level and subject-Level

	Fold 0		Fold 1		Fold 2		Fold 3		Fold 4		Avg.	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc	F1
Sample Level	60.80	65.13	56.11	40.38	61.11	63.70	62.69	62.04	60.04	52.52	60.15	58.48
Subject Level	57.14	60.87	61.90	42.86	66.67	69.57	71.43	72.73	66.67	63.16	64.76	61.84

coefficient (r) between the average IG scores, calculated from the start to the end of a microsaccade, and peak velocity ($r = 0.49$), as well as the correlation between IG scores and amplitude ($r = 0.55$), demonstrate that the subject shown in Figure 3(a) exhibits a correlation between the average IG score and the microsaccade features of peak velocity and amplitude. However, given the moderate classification performance, not all microsaccades contribute positively to feature extraction. For instance, some microsaccades around sample 1800 in Figure 3(a) and sample 1000 in Figure 3(b) do not exhibit a positive role. Despite these limitations, this study demonstrates, from a computational perspective using XAI, that microsaccades are important in distinguishing between early-stage PD and HC during a fixation task.

4 Conclusion and Discussion

In this paper, we propose a computational method to visualize the contribution of microsaccades in classifying early-stage PD from raw positional eye-tracking data. A multi-head self-attention network is employed to extract features from positional eye-tracking data, followed by a fully connected network to classify these features into PD and HC. Additionally, the Integrated Gradients method is utilized to compute the attribution map of the trained model, highlighting the importance of learned features in the decision-making process. The visualization comparing the attribution map with the raw data reveals that microsaccades play a crucial role in the classification, as the most positively contributing samples are predominantly associated with microsaccades. Our proposed method provides the first computational evidence of microsaccades' significance in classifying early-stage PD and has the potential to be extended to other medical applications, promoting more transparent and explainable deep learning approaches.

Since this study involves clinical data, it was conducted under the approval of an ethics committee and in compliance with established ethical guidelines. To ensure participant privacy, all data were anonymized and associated only with subject numbers, rather than names or any personally identifiable information of the patients or healthy controls. Furthermore, no personal information was accessible to the researchers at any stage of the study. This strict anonymization process safeguards the privacy and confidentiality of all participants, aligning with ethical standards for handling sensitive clinical data.

5 Limitation and Future Work

Future work will focus on addressing several limitations to enhance model performance. Data scarcity, a prevalent challenge in medical applications, likely contributed to the moderate classification performance observed. We propose exploring data augmentation techniques to expand our dataset. Regarding data preprocessing, the current approach of removing 300ms before and after blinks presents a theoretical risk of eliminating microsaccades. We intend to conduct validation experiments to ensure that our interpolation methods preserve microsaccade dynamics. Model architecture represents another significant avenue for improvement, as our current model was not specifically optimized for eye-tracking data and likely suffered from subject variance. Additionally, we plan to validate our methodology across multiple datasets to mitigate variability and reduce potential sample bias inherent to individual datasets.

Acknowledgments

The authors wish to thank the Promobilia Foundation and the Marianne Bernadotte Research Foundation for financial support.

References

Robert G Alexander, Stephen L Macknik, and Susana Martinez-Conde. 2018. Microsaccade characteristics in neurological and ophthalmic disease. *Frontiers in neurology* 9 (2018), 144.

Han Collewijn, Johannes Van Der Steen, and Robert M Steinman. 1985. Human eye movements associated with blinks and prolonged eyelid closure. *Journal of neurophysiology* 54, 1 (1985), 11–27.

Ryan Anthony Jalova de Belen, Tomasz Bednarz, and Arcot Sowmya. 2021. Eyexplain autism: Interactive system for eye tracking data analysis and deep neural network interpretation for autism spectrum disorder diagnosis. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

Shuwen Deng, Paul Prasse, David R. Reich, Sabine Dziemian, Maja Stegenwallner-Schütz, Daniel Krakowczyk, Silvia Makowski, Nicolas Langer, Tobias Scheffer, and Lena A. Jäger. 2022. Detection of ADHD based on Eye Movements during Natural Viewing. https://doi.org/10.48550/arXiv.2207.01377 arXiv:2207.01377 [cs].

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

Sabrina Hoppe and Andreas Bulling. 2016. End-to-End Eye Movement Detection Using Convolutional Neural Networks. https://doi.org/10.48550/arXiv.1609.02452 arXiv:1609.02452 [cs].

Narine Kokhlikyan, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov, Natalia Kliushkina, Carlos Araya, Siqi Yan, et al. 2020. Captum: A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv:2009.07896* (2020).

Daniel Krakowczyk, David Robert Reich, Paul Prasse, Sebastian Lapuschkin, Lena Ann Jäger, and Tobias Scheffer. 2023b. Selection of XAI Methods Matters: Evaluation of Feature Attribution Methods for Oculomotoric Biometric Identification. In *Gaze Meets Machine Learning Workshop*. PMLR, 66–97.

Daniel G. Krakowczyk, Paul Prasse, David R. Reich, Sebastian Lapuschkin, Tobias Scheffer, and Lena A. Jäger. 2023a. Bridging the Gap: Gaze Events as Interpretable Concepts to Explain Deep Neural Sequence Models. In *2023 Symposium on Eye Tracking Research and Applications*. ACM, Tubingen Germany, 1–8. https://doi.org/10.1145/3588015.3588412

Amrutha Krishnamoorthy, Vijayasimha Reddy Sindhura, Devarakonda Gowtham, C Jyotsna, and Joseph Amudha. 2021. StimulEye: An intelligent tool for feature extraction and event detection from raw eye gaze data. *Journal of Intelligent & Fuzzy Systems* 41, 5 (2021), 5737–5745.

Dillon Lohr and Oleg V Komogortsev. 2022. Eye know you too: Toward viable end-to-end eye movement biometrics for user authentication. *IEEE Transactions on Information Forensics and Security* 17 (2022), 3151–3164.

Silvia Makowski, Paul Prasse, David R Reich, Daniel Krakowczyk, Lena A Jäger, and Tobias Scheffer. 2021. DeepEyedentificationLive: Oculomotoric biometric identification and presentation-attack detection using deep neural networks. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3, 4 (2021), 506–518.

Susana Martinez-Conde. 2006. Fixational eye movements in normal and pathological vision. *Progress in brain research* 154 (2006), 151–176.

Susana Martinez-Conde, Stephen L. Macknik, Xoana G. Troncoso, and David H. Hubel. 2009. Microsaccades: a neurophysiological analysis. *Trends in Neurosciences* 32, 9 (Sept. 2009), 463–475. https://doi.org/10.1016/j.tins.2009.05.006

Zhongang Qi, Saeed Khorram, and Fuxin Li. 2019. Visualizing Deep Networks by Optimizing with Integrated Gradients.. In *CVPR workshops*, Vol. 2. 1–4.

Mehedi Hasan Raju, Lee Friedman, Dillon J Lohr, and Oleg V Komogortsev. 2024. Analysis of Embeddings Learned by End-to-End Machine Learning Eye Movement-driven Biometrics Pipeline. *arXiv preprint arXiv:2402.16399* (2024).

Harshinee Sriram, Cristina Conati, and Thalia Field. 2023. Classification of Alzheimers Disease with Deep Learning on Eye-tracking Data. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*. 104–113. https://doi.org/10.1145/3577190.3614149 arXiv:2309.12574 [cs].

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning*. PMLR, 3319–3328.

Panagiota Tsitsi, Mattias Nilsson Benfatto, Gustaf Öqvist Seimyr, Olof Larsson, Per Svenningsson, and Ioanna Markaki. 2021. Fixation duration and pupil size as diagnostic tools in Parkinson's disease. *Journal of Parkinson's Disease* 11, 2 (2021), 865–875.

Pierpaolo Turcano, John J Chen, Britta L Bureau, and Rodolfo Savica. 2019. Early ophthalmologic features of Parkinson's disease: a review of preceding clinical and diagnostic markers. *Journal of neurology* 266 (2019), 2103–2111.

Gonzalo Uribarri, Simon Ekman von Huth, Josefine Waldthaler, Per Svenningsson, and Erik Fransén. 2023. Deep Learning for Time Series Classification of Parkinson's Disease Eye Tracking Data. *arXiv preprint arXiv:2311.16381* (2023).

A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).

Xin Wang, Lizhou Fan, Haiyun Li, Xiaochan Bi, Wenjing Jiang, and Xin Ma. 2025. Skip-AttSeqNet: Leveraging skip connection and attention-driven Seq2seq model to enhance eye movement event detection in Parkinson's disease. *Biomedical Signal Processing and Control* 99 (Jan. 2025), 106862. https://doi.org/10.1016/j.bspc.2024.106862

Raimondas Zemblys, Diederick C. Niehorster, and Kenneth Holmqvist. 2019. gazeNet: End-to-end eye-movement event detection with deep neural networks. *Behavior Research Methods* 51, 2 (April 2019), 840–864. https://doi.org/10.3758/s13428-018-1133-5