Towards Automated Analysis of Gaze Behavior from Consumer VR Devices for Neurological Diagnosis

Lio Schmitz[1,†], Markus Plack[1], Berkan Koyak[2], Muhammad Ehsan Ullah[2], Ahmad Aziz[2], Reinhard Klein[1], Zorah Lähner[1], Hannah Dröge[1]

[1] *Visual Computing Group, University of Bonn, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany*
[†] *E-mail: lschmitz@cs.uni-bonn.de*
www.uni-bonn.de

[2] *German Center for Neurodegenerative Diseases (DZNE), Center of Neurology, Department of Parkinson, Sleep and Movement Disorders, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany*
www.ukbonn.de

Recent studies have demonstrated that eye tracking is a valuable tool in the detection, classification and staging of neurodegenerative diseases such as Parkinson's Disease (PD). However, traditional methods for capturing gaze data often rely on expensive and non-engaging clinical equipment such as video-oculography, limiting their accessibility and scalability. In this work, we investigate the feasibility of using eye tracking data collected via consumer-grade virtual reality (VR) headsets to support neurological diagnostics in a more accessible and user-friendly manner.

This approach enables large-scale, low-cost, and remote assessments, which are particularly valuable in early detection and monitoring of neurodegenerative conditions. We show that relevant oculomotor features extracted from VR-based eye tracking can be used for predictive assessment. Despite the inherent noise and lower precision of consumer devices, careful preprocessing and robust feature engineering, including deep learning embeddings, mitigate these limitations. Our results demonstrate that both handcrafted and learned features from gaze behavior enable promising levels of classification performance. This research represents an important step towards scalable, automated, and accessible diagnostic tools for neurodegenerative diseases using ubiquitous VR technology.

Keywords: Machine Learning; Eye tracking; Virtual reality; Parkinson's Disease

1. Introduction

Eye tracking data has been shown to be a valuable tool in many medical applications, such as detecting concussions in athletes[1–4] or studying behavior to, *e.g.* , identify ADHD in children.[5–8] Many conditions that affect oculomotor function are marked by eye movement abnormalities such as slowed saccades, fixation instability, square wave jerks, impaired smooth pursuit, gaze-evoked nystagmus, vertical gaze palsy, reduced optokinetic nystagmus (OKN) and

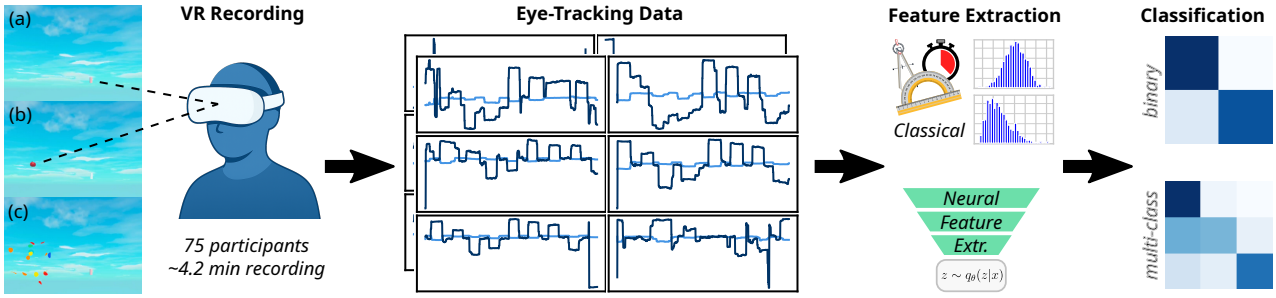


Fig. 1. Overview of our approach. We use raw eye tracking data (center) from a game-like VR experiment (left), where the user gaze starts in a resting position (a) and has to focus on dynamically appearing objects (b). Upon successful fixation, these objects playfully "explode" into confetti, providing immediate visual feedback (c). From multiple repetitions with different target locations we extract various types of features and perform classification (right).

vestibuloocular reflex (VOR) gain as can be observed for Parkinson's disease (PD), atypical Parkinsonian disorders such as multiple system atrophy (MSA), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD), spinocerebellar ataxias (SCA), Huntington's disease (HD), frontotemporal dementia (FTD), Alzheimer's Disease and schizophrenia.[9,10] Of those, PD features decreased saccade amplitudes, quantified as task-specific hypometria, as a probable reflection of early nigrostriatal degeneration, whilst increased saccadic latency, and antisaccade errors, attributed to the non-dopaminergic degeneration of higher-level eye movement related circuits later in the disease course, thereby linking eye-tracking metrics to cognitive decline.[11–13] Oculomotor signatures – such as asymmetry of corrective saccades and VOR gain measured by video head-impulse tests (vHIT) to distinguish central causes (e.g., posterior circulation strokes or cerebellar ataxia, neuropathy and vestibular areflexia syndrome [CANVAS]) from peripheral vestibulopathies (e.g., vestibular neuritis), as well as slow or absent vertical saccades to differentiate early PSP from PD – highlight the broader diagnostic utility of eye tracking across diverse disorders, without the need for costly diagnostic measures such as alpha – synuclein seed amplification assays or PET-CT scans.[14–17]

At the same time, consumer-grade eye tracking hardware (webcams, phone cameras, VR headsets) has become more accessible and demonstrated potential for unintrusive, remote eye-tracking assessments, thus offering scalable tools to complement traditional methods for early detection and monitoring.[18,19] This enables remote assessment of problems and widely accessible diagnostic tools. However, it remains unclear whether these devices can measure data that is qualitatively sufficient for diagnostics.

In the following, we will focus on eye tracking data obtained from a virtual reality (VR) headset in the prediction of PD. VR headsets have several advantages over traditional camera recordings, *e.g.* eye tracking independent of head movement which allows for more reliable handling of movement artifacts, a built-in calibration for faster setup, and complete control over the visual environment which allows for repeatable, immersive experimental conditions with complete control over both foreground and background stimuli. By contrast, conventional camera-based systems typically require stricter environmental control and are more sensitive to external distractions. And, even though they are less common than phone cameras, they

can be bought at a reasonable price nowadays. By developing an automated analysis and classification of this data, we can simplify assessment of movement disorders such as PD. The experiment tracks the gaze of patients and healthy controls as they look at different objects in the form of fruits, according to simple instructions: The user's task is to look at pseudo-randomly appearing objects (see Fig. 1), starting from a reference resting position, in order to explode them and increase a score with each successfully performed action. This approach leverages gamification to enhance engagement, making the task more enjoyable and potentially more effective in sustaining user attention and compliance.

We analyze the applicability of three classes of features for predicting the diagnosis of PD using eye tracking data. *General features*, such as those established in prior work,[20] can easily be extracted but may underperform with low-quality or noisy recordings. *Task-evoked features* like reaction time, derived directly from the experimental setup, can capture meaningful responses but require careful computation and may still have limited predictive power (compare Fig. 4). *Learning-based features* obtained through neural network training offer a data-driven approach that may uncover latent patterns not accessible through traditional methods. We demonstrate that combining and selectively integrating these feature types yields improved predictive performance. This classification can potentially become a tool to help identify early oculomotor changes of movement disorders and monitor disease progression in a cost-effective, reproducible, and scalable fashion.

Contributions. This work takes a step towards accessible and automated analysis of eye tracking data from consumer-grade VR devices for the detection of oculomotor dysfunction. We explore the potential of such data for distinguishing PD from healthy controls and other movement disorders. In summary, our contributions are

- We design a gamified eye tracking experiment in virtual reality to capture controlled gaze behavior in a reproducible setting while remaining accessible and engaging through consumer VR hardware.
- We evaluate multiple feature classes for PD classification: (i) general statistical features, (ii) task-evoked metrics derived from the experimental design, and (iii) deep features learned via a supervised variational autoencoder.
- We demonstrate that combining these feature types improves classification performance across multiple metrics and extends to multiclass scenarios involving spinocerebellar ataxia patients.

2. Related Work

In this section, we discuss prior work from two intersecting areas: (1) the application of eye tracking in medical diagnostics, particularly for movement disorders, and (2) machine learning methods for oculographic data in both traditional and immersive environments.

2.1. *Ophthalmology for Diagnostics*

Eye movement behavior has long been studied as a non-invasive window into motor, coordinative, and cognitive function.[21,22] In clinical contexts, eye tracking has been used to assess

cognitive abilities, such as in dementia screening,[23] as well as to evaluate social behaviors,[24] and neurological conditions such as Parkinson's disease (PD).[11,25–27] Moreover, it has also been used to differentiate PD from related disorders like Progressive Supranuclear Palsy.[28] Eye tracking has found applications outside of neurodegenerative disease as well and has shown promise in concussion assessment, particularly in sports contexts.[1–4] Additionally, it has been explored as a tool for screening and diagnosing neurodevelopmental disorders, including attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder.[5–8] Beyond patient diagnosis, several studies investigate eye movements in decision making and clinical reasoning,[29–32] for instance by analyzing the connection between time-related costs in decision processes and eye movement[29] or exploring the relationship between diagnostic performance and visual search behaviors of the radiologists.[32]

A variety of eye tracking tools have supported these investigations. High-fidelity systems such as the EyeLink 1000[20,29] and the iView X Hi-Speed[25] are commonly used in controlled lab settings. More recent efforts have also examined the feasibility of webcam-based eye tracking for remote cognitive assessments, despite current limitations in resolution and precision.[33] Immersive and extended reality platforms have opened new avenues for eye tracking in clinical research. In mixed reality, Daniol et al.[34] used the Microsoft HoloLens 2 to analyze eye tracking patterns in PD patients. In the virtual reality (VR) domain, Orlosky et al.[35] recorded eye movements of PD patients during VR tasks for subsequent clinical evaluation. Other VR-based efforts include tools for post-COVID symptom assessment[36] and dementia screening.[23] Adhanom et al.[37] provide a comprehensive review of eye tracking applications in VR. In our work, we also use immersive VR with the Meta Quest Pro headset to capture oculographic data from patients with PD.

2.2. *Machine Learning for Oculometrics*

Machine learning and deep learning techniques have been increasingly applied to eye tracking data across a range of medical contexts. Recent studies demonstrate their utility in detecting PD and monitoring its progression, highlighting the potential of eye tracking as a non-invasive digital biomarker.[38,39] These approaches typically operate on small-scale datasets and employ classifiers such as Naïve Bayes, Decision Trees, and Random Forests.[40,41] Other work extracts a feature set from video-based eye trackers, and trains an ensemble classifier composed of Support Vector Machines, Logistic Regression, and Random Forest models.[42] Additionally, tablet-based eye tracking has been used both to differentiate Parkinson's patients from healthy controls using Logistic Regression,[43] and to predict the severity of motor symptoms (e.g., MDS-UPDRS scores) using SVM classifiers.[44] There are specific eye movement abnormalities in PD patients, which leverage many of these approaches; an overview can be found in.[12,45]

Recently, deep learning models have been increasingly applied to PD detection and monitoring using eye tracking data. Mao *et al.*[46] proposed a classification pipeline that combines deep learning with ensemble methods to identify neurological disorders, using features such as pupil position extracted from 30 Hz infrared video eye tracking data. Their approach utilizes a long short-term memory (LSTM) network for initial temporal modeling, followed by a decision-tree ensemble (Random Forest) for final classification. Vodrahalli *et al.*[47] used a

500 Hz binocular infrared eye tracking system to detect eye movement disorders, including PD. Their method generates convolutional neural network (CNN) embeddings from gaze data recorded while patients read numbers on a page, which are then classified using a Random Forest model. Similarly, Uribarri et al.[48] applied state-of-the-art deep learning architectures to classify PD using 300 Hz eye tracking data. They used near-infrared illumination to track pupil movements, focusing their predictions on short fixation intervals. Reiner et al.[49] employed a Tobii Fusion Pro remote eye tracker to collect oculometric data, which was processed using a deep learning pipeline designed for visual feature extraction. In a related effort, Jiang et al.[50] developed a custom deep neural network for automated PD classification using eye tracking data acquired with an HTC Vive Pro Eye VR headset featuring integrated Tobii tracking with 120 Hz.[51] Deep learning-based eye tracking analysis has also been employed to detect other neurodegenerative and developmental conditions, including Alzheimer's disease,[52,53] dyslexia,[54–58] and autism spectrum disorder.[59–61]

In contrast to prior work, our approach targets PD (and spinocerebellar ataxia) classification using eye tracking data from consumer-grade VR hardware with lower sampling rates of 70 Hz. We employ a structured, gamified task and a robust feature extraction pipeline, showing that even under these constrained conditions, effective classification is possible. This performance is supported in part by higher-order statistical features (e.g., skewness, kurtosis), learned features, and task-evoked metrics that capture subtle but clinically relevant variations in gaze behavior.

3. Experimental Setup

In this work, we explore the potential of using eye tracking data from widely available VR headsets to enable more accessible approaches to support neurological diagnostics. To this end, we use a dataset collected in a virtual reality (VR) environment with the Meta Quest Pro, a consumer-grade headset featuring integrated infrared-based eye tracking. Despite its accessibility, the system introduces challenges related to signal noise and reduced accuracy in peripheral regions, which must be considered during data preprocessing and model design.

The AVERT-MD (Avatar Evaluation in Real-Time for Movement Disorders) study has been granted ethical approval by the institutional review board of the University Hospital Bonn in compliance with the Helsinki Declaration (File number: 2024-367-BO). All participants are recruited either from the Department of Parkinson, Sleep and Movement Disorders at the University Hospital Bonn or from the observational study cohorts of the German Center for Neurodegenerative Diseases (DZNE): healthy individuals (DANCER), Parkinson's disease (DESCRIBE-PD), and spinocerebellar ataxia (SCA-Registry). Participants with congenital nystagmus, age-related macular degeneration, or head tremor were excluded from study participation. After providing written informed consent, demographic data were collected, and participants underwent clinical motor assessments consisting of the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and the Scale for the Rating and Assessment of Ataxia (SARA), which were administered by a movement disorders specialist certified in the aforementioned clinical assessment scales. At the beginning of the VR game, an instructional text will be shown to participants in VR, and the test will be run

without any interference from the examiner, only to be repeated in case of involuntary head movements.

The dataset is recorded through a controlled gaze-tracking task in which individual objects in the form of fruits appear sequentially at varying positions in the user's field of view. Participants are instructed to fixate on each new target upon its appearance, which is marked by the explosion of the object, followed by return to a centrally located reference object. The timing of each trial is participant-dependent and dynamically gated by real-time gaze detection, resulting in session durations in the order of several minutes. Each session yields continuous recordings of gaze vectors, the participant's positions, and 3D object coordinates for both central and target stimuli. The dataset includes recordings from 75 participants, comprising 33 individuals with Parkinson's disease (PD), mainly of Stage I or II of the Hoehn and Yahr staging system, which is well representative of the early stages of PD, 27 healthy controls (HC), and 15 individuals diagnosed with Ataxia, a separate neurological disorder characterized by motor impairments. The PD participants had an average UPDRS III score of 16.86 with a standard deviation of 10.27, and the PD group's male predominance reflects projected global prevalence ratios, rising from 1.46 in 2021 to 1.64 in 2050.[62] Recordings were performed at a 70 Hz sampling rate using the native Quest Pro SDK, without additional hardware modifications, resulting in 75 VR sequences with consistent task structure and spatial alignment.

By using the Meta Quest Pro, a widely available consumer VR headset with built-in eye tracking, the setup enables scalable data collection without the need for specialized hardware or lab environments. Note that the system has only been tested with this VR headset model, so its generalizability to other devices remains unclear.

4. Feature Extraction

To analyze gaze behavior, we extracted a range of features from the raw eye tracking data. These include handcrafted features based on task-evoked reaction metrics such as saccadic latency and velocity, spontaneous fixations and saccades, and higher-level representations learned via deep neural networks.

4.1. *Task-Evoked Saccadic Response Features*

To systematically analyze response dynamics in the eye movement behavior of the participants, we calculated two primary metrics from the gaze data: the *saccadic latency* (sl) and the *saccadic velocity* (sv). The saccadic latency refers to the time from appearance of the target to the start of the participants eye movement, as illustrated in Fig. 2. From the duration of the saccade and the distance to the target at movement start, we compute the saccadic velocity, the angular velocity of the eye movement in degrees per second.

To determine the initiation of eye movement toward the target, we analyze the angular distance between the gaze and the target object over time. As shown in Fig. 2, the angular distance remains approximately constant prior to movement initiation, reflecting gaze stability at a non-target location. Movement onset is identified as the point where the angular distance begins to decrease sharply. While methods based on fixed angular or velocity thresholds were

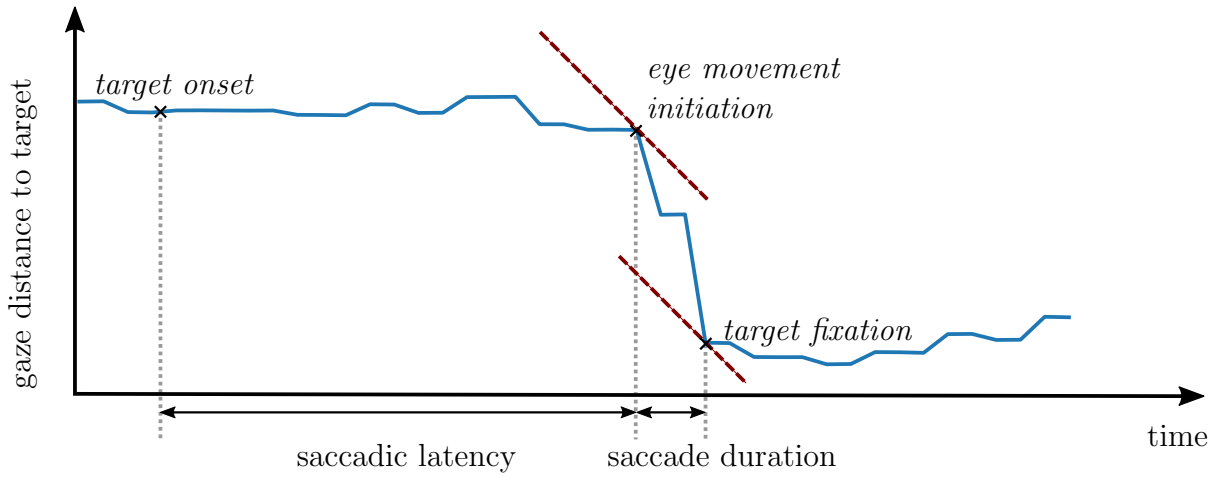


Fig. 2. Timeline illustrating the visual response to a stimulus. Eye movement initiation and target fixation are detected based on angular distance to the target over time. The blue curve represents the angular distance between the gaze and the target object. Eye movement initiation is defined as the point of closest approach to a reference line with a 45° negative slope. Target fixation is identified as the point where the gaze intersects a 45° positive slope reference line.

found to be sensitive to noise and inter-subject variability, we fit a reference line with a fixed slope of 45° to the angular distance profile and define the initiation point as the location of closest approach to this line. Fixation is defined similarly as the point, where the gaze intersects a 45° positive slope reference line. From the saccade duration, the saccadic velocity is calculated using the angular distance between the central fixation point and the target.

To prepare the data for classification, we extract statistical features from saccadic latency and velocity measurements recorded during the experiment. Each participant was exposed to $n = 127$ object appearances, with each trial $i \in \{1, \ldots, n\}$ yielding a pair of scalar values (ℓ_i, v_i) corresponding to their saccadic latency/sl and velocity/sv.

To generate meaningful participant-level features that characterize the empirical distributions of these sequences, we computed a set of descriptive statistics based on statistical moments over $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n)$ and $\boldsymbol{v} = (v_1, \ldots, v_n)$. The first and second central moments, representing the sample *mean* $\mu(x)$ and *variance* $\sigma(x)^2$,

$$\mu(x) = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \sigma(x)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu(x))^2, \tag{1}$$

quantify the average saccadic latency or velocity and the overall variability in oculomotor responses, respectively. The *standard deviation* $\sigma(x)$ is simply the square root of the variance. The third and fourth standardized moments, corresponding to *skewness* and *kurtosis*, respectively,

$$\text{skew}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu(x)}{\sigma(x)}\right)^3, \quad \text{kurt}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu(x)}{\sigma(x)}\right)^4, \tag{2}$$

whereby skewness quantifies the asymmetry of the distribution, with positive values indicating a longer right tail, while kurtosis measures the heaviness of the tails and sharpness of the peak,

relative to a normal distribution. To complement these moment-based features, we computed robust statistics that are less sensitive to outliers. The *median* offers a robust measure of central tendency, and represents the 50th percentile of the sequence, and the *interquartile range* $\mathrm{IQR} = Q_3 - Q_1$ is the difference between the 75th and 25th percentiles and captures the spread of the central 50% of observations. We also included the *minimum* and *maximum* values $\min(x) = \min_i x_i, \quad \max(x) = \max_i x_i$, which highlight the extremal responses and indicate anomalously fast or slow responses. The *range* $\mathrm{range}(x) = \max(x) - \min(x)$ measures the spread between the most extreme observations, reflecting the total span of variation.

Each statistic is computed independently for the saccadic latency vector $\boldsymbol{\ell}$ and the movement speed vector $\boldsymbol{v}$, resulting in a total of 18 features per participant.

4.2. *General Fixation and Saccade Features*

Saccadic eye movements and visual fixation provide important insights into motor and cognitive function, particularly in Parkinson's disease.[45,63] Therefore, we evaluated a variety of fixation- and saccade-based features from the gaze data. Fixations and saccades were identified using the Buscher method,[64] which classifies a fixation as a cluster of gaze points within a 1° visual angle for at least 100 ms. Saccades were inferred as the rapid transitions between consecutive fixations during free viewing. Similar to Jiang *et al.*[50] we extracted the total number of fixations (TNF), total duration of fixations (TDF), mean duration of fixations (MDF), and their number, as well as the scan path length (SL) and its duration (SD), and the rate of saccade/fixation. We compute them for both the interval of the target visibility and the complete recording. To extract additional insights into the feature distribution, we also compute statistical features as described in Sec. 4.

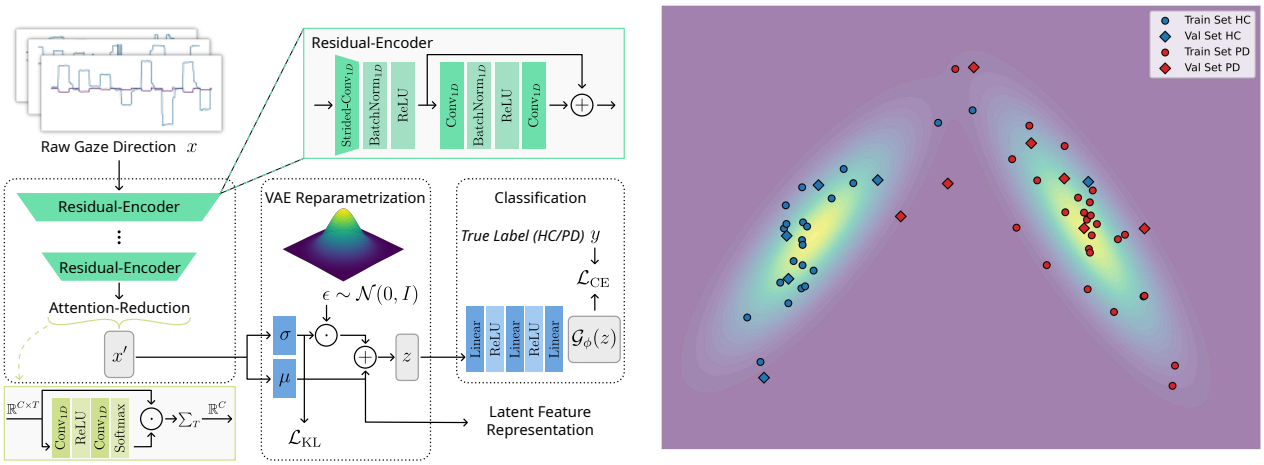
4.3. *Deep Feature Extraction*

To capture higher-level representations of gaze behavior beyond handcrafted features, we employ a deep neural network on the gaze data as shown in Fig. 3. We use a 1D convolutional neural network encoder with residual connections to extract features from the input. To summarize the temporal dynamics of the extracted features, we apply attention inspired average pooling. These pooled features are passed through a reparameterization layer, yielding latent representations which are subsequently processed by a classification module. The model outputs a 4-dimensional latent embedding $z \sim q_\theta(z|x)$ that captures the underlying structure of gaze behavior and is used for subsequent classification.

Model parameters θ and ϕ are optimized by minimizing a composite loss function that combines a cross-entropy loss with a regularization term based on the Kullback–Leibler (KL) divergence,

$$\min_{\theta,\phi} \; \mathcal{L}_{\mathrm{CE}}\left(\mathcal{G}_\phi(z), y\right) + \lambda \cdot \mathcal{L}_{\mathrm{KL}}\left(q_\theta(z|x)|p(z)\right), \tag{3}$$

where x is the input, y is the ground-truth label, and $\mathcal{G}_\phi(z)$ denotes the classification head applied to the sampled latent representation. The KL divergence term encourages the approximate posterior $q_\theta(z|x)$ to remain close to the prior $p(z) = \mathcal{N}(0, I)$.



(a) Neural Network Architecture **(b)** Feature Space Visualization

Fig. 3. (a) Architecture of the 1D convolutional network for gaze representation learning. (b) PCA projection of the learned feature space showing task-specific clustering.

5. Experiments

We now assess the ability of the features introduced in Sec. 4 to discriminate between healthy individuals (HC) and those with Parkinson's disease (PD) by evaluating their performance in a binary classification task using standard machine learning classifiers. We then extend this analysis to a more challenging three-class classification problem by additionally including spinocerebellar ataxia participants.

5.1. *Prediction of Parkinson's Disease*

To identify the most discriminative features, we begin with a statistical analysis based on the independent two-sample t-test, comparing feature distributions between PD and HC groups. Features yielding p-values less than 0.05 are considered statistically significant and are hypothesized to contribute meaningfully to classification. Fig. 4 illustrates the mean and standard deviation of the normalized feature space and shows significant differences across several features, with varying levels of significance indicated by asterisks.

Next, we evaluate the ability of several machine learning models to distinguish between HC and PD participants, using both the statistical features extracted from saccadic latencies and velocity, as well as the general fixation and saccade features (Sec. 4). For each model, we use Sequential Feature Selector from scikit-learn[65] to identify optimal feature subsets. Classification performance is assessed using stratified k-fold cross-validation with randomized splits (four independent seeds) to ensure robustness. We evaluate nine representative machine learning models: Logistic Regression, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) with a linear and radial basis function (RBF) kernel, Random Forest, Gradient Boosted Trees (GBT), k-nearest neighbors (KNN), and perceptron. Tab. 1 reports performance metrics for all classifiers, including weighted accuracy, precision, recall, F1 score, and ROC AUC, based on classification results using both generic and task-evoked features. We

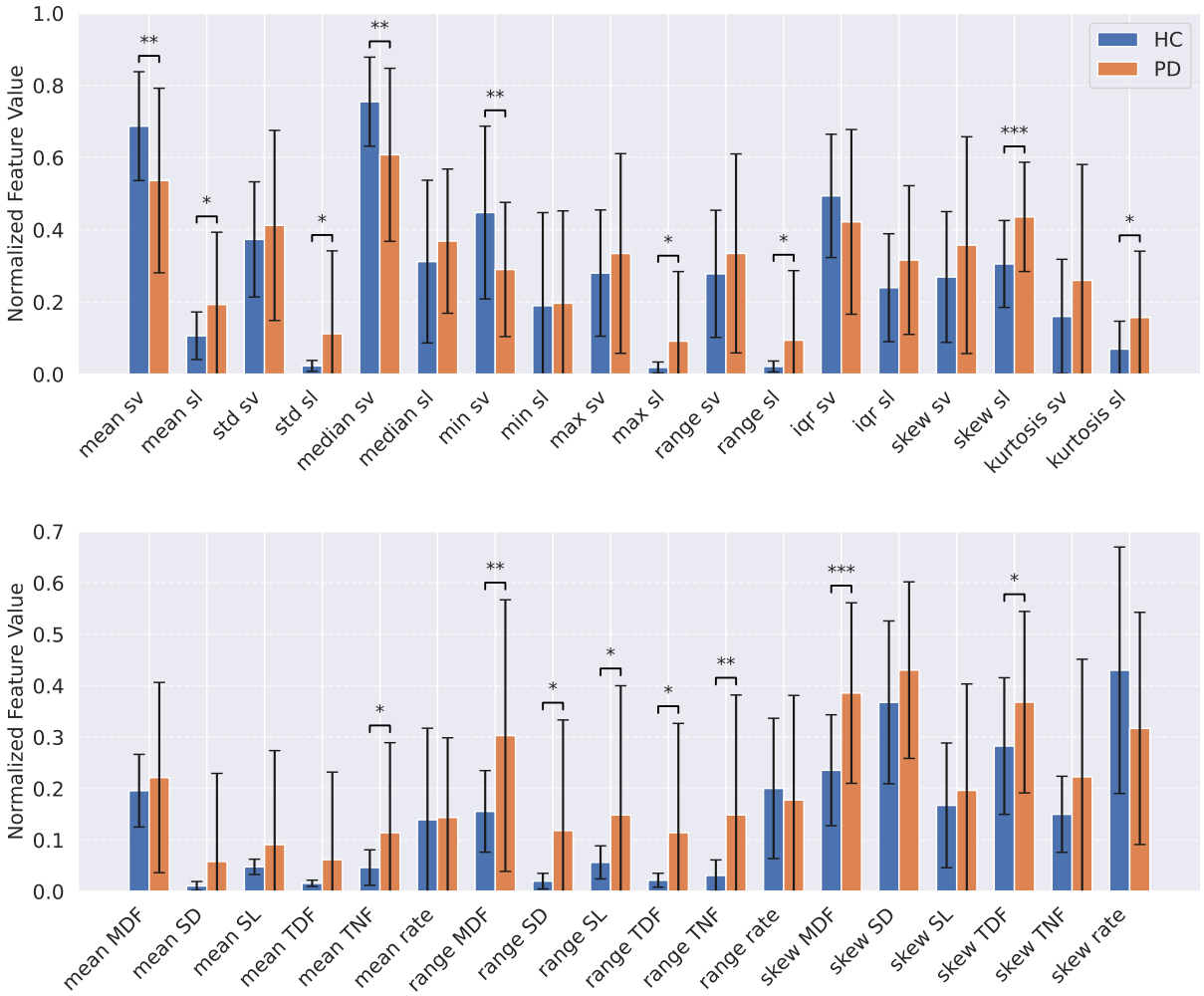


Fig. 4. Mean and standard deviation of normalized (task-evoked and general) feature spaces for Parkinson's disease and healthy controls. Significance levels are based on t-test p-values across features, indicating statistical differences: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

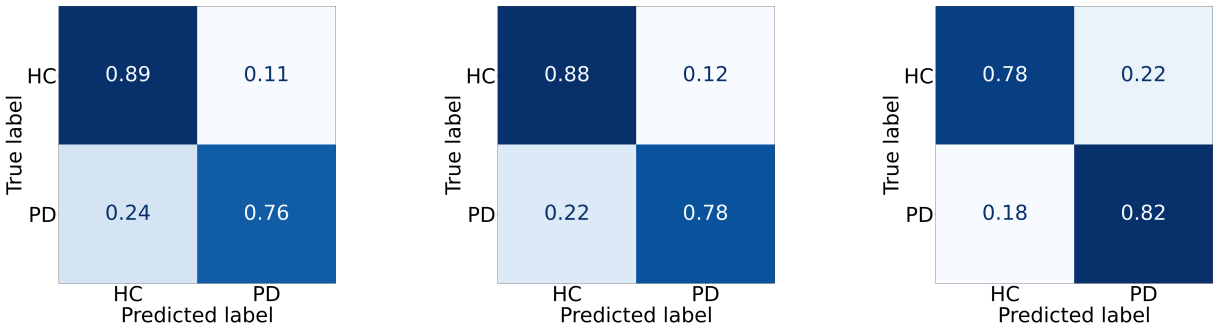
additionally report weighted accuracy without the higher-order statistical features (denoted as Accuracy*). To further analyze classifier performance, we visualize the confusion matrices for the three best-performing models in Fig. 5. These show well each model distinguishes between PD and HC participants, highlighting common misclassification patterns.

To assess how well the learned latent representations separate classes, we apply the Naive Bayes classifiers to the deep features extracted from a deep neural network (compare Sec. 4.3). We evaluate the latent representations as complementary information by concatenating the latent vectors $z \sim q_\theta(z|x)$ with the most informative handcrafted features $f \in \mathbb{R}^d$, selected through sequential feature selection. The increased accuracy observed in Tab. 1 for Naive Bayes^{a+emb} shows that the learned feature space is able to capture additional discriminative information not present in classical features. In parallel, we evaluate the performance of the full neural architecture in an *end-to-end supervised learning* setup, which demonstrates perfor-

Table 1. Performance comparison of models using sequentially selected features. Superscripts indicate the specific set of selected features used by each model, and "Accuracy*" reports the model's classification accuracy without the use of higher statistical features. Here, *sv* refers to saccadic velocity and *sl* to saccadic latency. The tag *vis* marks features computed only when the object was visible; features without this tag are calculated across the entire dataset.

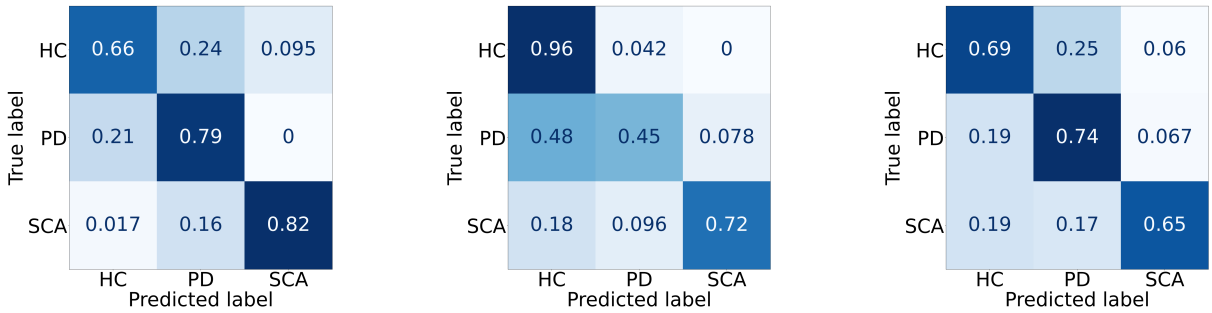
Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	Accuracy*
Naive Bayes [a]	0.824	0.913	0.756	0.808	0.873	0.713
Random Forest [b]	0.804	0.829	0.824	0.818	0.836	0.758
GBT [c]	0.772	0.823	0.780	0.780	0.821	0.754
LDA [d]	0.748	0.796	0.713	0.737	0.737	0.697
SVM RBF Kernel [e]	0.743	0.868	0.614	0.703	0.715	0.723
KNN [f]	0.710	0.798	0.643	0.692	0.679	0.683
Logistic Regression [g]	0.681	0.733	0.671	0.681	0.771	0.646
SVM Linear Kernel [h]	0.667	0.756	0.571	0.624	0.722	0.679
Perceptron [i]	0.648	0.746	0.637	0.647	0.724	0.558
Naive Bayes [a+emb]	0.831	0.859	0.821	0.816	0.871	—
End-to-end	0.741	0.773	0.738	0.731	0.768	—

[a] mean sv, max sv, median TNF vis, range MDF vis, kurtosis rate vis, range rate
[b] min sl, iqr sv, kurtosis sv, iqr TDF vis, median TNF vis, median TDF, mean SL, mean SD, median SD
[c] min TNF vis, mean SL, max SD
[d] std sl, kurtosis sv, mean TNF vis, median TNF vis, range MDF vis, median SL vis, iqr TNF, kurtosis MDF
[e] median sv, iqr ms, kurtosis sl, mean SL, median SL, max SL, range SL, kurtosis SD
[f] median SL
[g] median TNF vis, kurtosis TNF
[h] range sl, max MDF vis, range MDF vis, range rate vis, mean TDF, median SL
[i] std sl, skew sl, kurtosis sl vis, range SD vis, min rate vis, iqr rate vis, mean TDF, kurtosis TNF



(a) Naive Bayes **(a)** Naive Bayes + Emb **(c)** Random Forest

Fig. 5. Confusion matrices for the top three binary classifiers. Each matrix summarizes the classification performance for distinguishing between HC and PD participants.



(a) LDA **(b)** Naive Bayes **(c)** Random Forest

Fig. 6. Confusion matrices for the top three classifiers in the multiclass case. Each matrix summarizes the classification performance for distinguishing between healthy controls, Parkinson's disease and Ataxia participants. Diagonal elements represent correctly classified instances, while off-diagonal entries indicate misclassifications.

mance competitive with classical approaches. We believe that with a larger training dataset, this potential could be further leveraged, leading to greater performance gains. To interpret the learned embedding space, we project latent vectors to two dimensions using PCA. As shown in Fig. 3(b), samples cluster clearly by diagnostic label, highlighting that the latent space encodes disease-specific behavioral patterns.

5.2. *Discrimination of Neurodegenerative Diseases*

To evaluate the scalability of our approach, we extend the classification task to a three-class setting by incorporating spinocerebellar ataxias (SCA) as an additional target class. SCAs are a group of hereditary neurodegenerative disorders whose hallmark is the neurodegeneration of Purkinje cells in the cerebellum, as well as the deep cerebellar nuclei and brainstem, leading to discoordination of oculomotion, speech, standing, gait, and fine motor skills. The resulting task requires models to distinguish between oculometric signatures across three groups: HC, PD, and SCA. We adopt the same feature selection and cross-validation strategy used in the binary task, and account for class imbalance by using balanced class weights due to the smaller number of SCA participants. As shown in Tab. 2 and in Fig. 6, all models exhibit a decrease in absolute classification metrics, reflecting the increased difficulty of multi-class discrimination. Nevertheless, classifiers continue to perform above chance, confirming that the extracted features retain discriminative power across heterogeneous movement disorders.

6. Discussion

We highlight the feasibility of using consumer-grade VR headsets for automated neurological assessment, focusing on Parkinson's disease (PD) classification from eye tracking data. Our approach combines a controlled VR gaze task with a diverse feature extraction strategy that includes both statistical metrics and learned representations, addressing the limitations of noisy, low-resolution hardware and a limited dataset. We demonstrate that combining general oculomotor metrics with task-evoked feature computation of saccadic latency and velocity

Table 2. Multiclass performance comparison of models using sequentially selected features. Superscripts indicate the specific set of selected features used by each model, where *sv* refers to saccadic velocity and *sl* to saccadic latency. *vis* marks features computed only when the object was visible.

Model	Accuracy	Precision	Recall	F1 Score
LDA [a]	0.759	0.804	0.733	0.734
Naive Bayes [b]	0.708	0.774	0.642	0.629
Random Forest [c]	0.693	0.740	0.667	0.661
Gradient Boosted Trees [d]	0.692	0.751	0.692	0.686
SVM (RBF Kernel) [e]	0.591	0.551	0.579	0.523
Logistic Regression [f]	0.583	0.653	0.613	0.598
SVM (Linear Kernel) [g]	0.558	0.623	0.625	0.597
Perceptron [h]	0.516	0.589	0.525	0.510

[a] median sl, kurtosis sv, std TDF vis, median TNF vis, median MDF vis, skew MDF vis, max SD vis, iqr TDF, median SD, skew rate
[b] kurtosis v, median TDF vis, iqr TDF vis, kurtosis MDF vis, kurtosis TDF, min TNF
[c] median TDF vis, max MDF vis, kurtosis MDF vis, mean SL vis, mean SD vis, median SD vis, std rate, max rate, range rate
[d] max MDF vis, kurtosis MDF vis, mean SD vis, min TNF
[e] mean sl, median sl, median TDF vis, min TDF vis, max MDF vis, range MDF vis, median&min SD vis, mean MDF, median MDF
[f] median sl, min sl, min TDF vis, median TNF vis, range MDF, skew MDF, kurtosis MDF, skew rate
[g] median TDF vis, min TDF vis, skew MDF vis, skew SL vis, skew TDF, skew MDF, min SD
[h] mean sl, skew MDF vis, std SD vis, min SD vis, min rate vis, std rate, min rate, max rate, iqr rate

results in high discriminative performance. In particular, features based on higher-order statistics such as skewness and kurtosis contribute meaningfully to classification, indicating that distributional shape carries relevant diagnostic information not captured by simple averages. Deep learning has also proven to be a valuable tool for extracting additional features, and we believe that with a larger dataset, its contribution could be even more substantial. Finally, we show that our experimental setup is scalable, achieving robust performance not only in binary classification but also in more complex multi-class scenarios.

Note, that this approach does not yet provide sufficiently reliable confidence estimates to serve as a standalone diagnostic tool and remains a research prototype that can offer supportive measures. Integration into clinical practice requires privacy-compliant handling of eye-tracking data, clear communication to patients about possible insurance implications of early PD diagnosis, and approval of consumer VR devices as medical devices, which demands transparent data curation and close collaboration between clinicians and data scientists. Its main potential lies in the quantitative monitoring of PD symptoms, for example, to assess current symptom levels, track treatment response, and to assist differential diagnosis.

Acknowledgments

This work has been funded by the Ministry of Culture and Science North Rhine-Westphalia under grant number PB22-063A (InVirtuo 4.0: Experimental Research in Virtual Environments). Lio Schmitz and Zorah Lähner have been funded by the Federal Ministry of Research, Technology and Space of Germany and the state of North Rhine-Westphalia as part of the Lamarr Institute for Machine Learning and Artificial Intelligence.

References

1. N. Snegireva, W. Derman, J. Patricios and K. Welman, Eye tracking technology in sports-related concussion: a systematic review and meta-analysis, *Physiological measurement* **39**, p. 12TR01 (2018).
2. N. Snegireva, W. Derman, J. Patricios and K. E. Welman, Awareness and perceived value of eye tracking technology for concussion assessment among sports medicine clinicians: a multinational study, *The Physician and Sportsmedicine* **48**, 165 (2020).
3. C. N. Feller, M. Goldenberg, P. D. Asselin, K. Merchant-Borna, B. Abar, C. M. C. Jones, R. Mannix, K. Kawata and J. J. Bazarian, Classification of comprehensive neuro-ophthalmologic measures of postacute concussion, *JAMA network open* **4**, e210599 (2021).
4. N. Snegireva, W. Derman, J. Patricios and K. Welman, Eye tracking to assess concussions: an intra-rater reliability study with healthy youth and adult athletes of selected contact and collision team sports, *Experimental brain research* **239**, 3289 (2021).
5. J. H. Yoo, C. Kang, J. S. Lim, B. Wang, C.-H. Choi, H. Hwang, D. H. Han, H. Kim, H. Cheon and J.-W. Kim, Development of an innovative approach using portable eye tracking to assist adhd screening: a machine learning study, *Frontiers in Psychiatry* **15**, p. 1337595 (2024).
6. L. Merzon, K. Pettersson, E. T. Aronen, H. Huhdanpää, E. Seesjärvi, L. Henriksson, W. J. MacInnes, M. Mannerkoski, E. Macaluso and J. Salmi, Eye movement behavior in a real-world virtual reality task reveals adhd in children, *Scientific reports* **12**, p. 20308 (2022).
7. L. Chetcuti, K. J. Varcin, M. Boutrus, J. Smith, C. A. Bent, A. J. Whitehouse and K. Hudry, Feasibility of a 2-minute eye-tracking protocol to support the early identification of autism, *Scientific Reports* **14**, p. 5117 (2024).
8. R. K. Wang, K. Kwong, K. Liu and X.-J. Kong, New eye tracking metrics system: the value in early diagnosis of autism spectrum disorder, *Frontiers in Psychiatry* **15**, p. 1518180 (2024).
9. S. Hutton and C. Kennard, Oculomotor abnormalities in schizophrenia: a critical review, *Neurology* **50** (1998).
10. E. Pretegiani and L. M. Optican, Eye movements in parkinson's disease and inherited parkinsonian syndromes, *Front Neurol.* (2017).
11. P. Diotaiuti, G. Marotta, F. Di Siena, S. Vitiello, F. Di Prinzio, A. Rodio, T. Di Libero, L. Falese and S. Mancone, Eye tracking in parkinson's disease: A review of oculomotor markers and clinical applications, *Brain Sciences* **15**, p. 362 (2025).
12. C. A. Antoniades and M. Spering, Eye movements in parkinson's disease: from neurophysiological mechanisms to diagnostic tools, *Trends in Neurosciences* **47**, 71 (2024).
13. T. J. Anderson and M. R. MacAskill, Eye movements in patients with neurodegenerative disorders, *Nature Reviews Neurology* **9**, 74 (2013).
14. S.-H. Kim, S.-U. Lee, B.-H. Cho, K.-H. Cho, S. Yu, B.-J. Kim and J.-S. Kim, Analyses of head-impulse tests in patients with posterior circulation stroke and vestibular neuritis, *Neurology* **100**, e2374 (2023).
15. L. Chen, M. Todd, G. M. Halmagyi and S. Aw, Head impulse gain and saccade analysis in pontine-cerebellar stroke and vestibular neuritis, *Neurology* **83**, 1513 (2014).
16. G.-S. Nam, H.-J. Shin, J.-J. Kang, N.-R. Lee and S.-Y. Oh, Clinical implication of corrective saccades in the video head impulse test for the diagnosis of posterior inferior cerebellar artery infarction, *Frontiers in Neurology* **12**, p. 605040 (2021).
17. C. Antoniades and C. Kennard, Ocular motor abnormalities in neurodegenerative disorders, *Eye* **29**, 200 (2015).
18. K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik and A. Torralba, Eye tracking for everyone, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE, 2016).

19. B. J. Hou, Y. Abdrabou, F. Weidner and H. Gellersen, Unveiling variations: A comparative study of vr headsets regarding eye tracking volume, gaze accuracy, and precision, in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, (IEEE, 2024).
20. I. Rigas, L. Friedman and O. Komogortsev, Study of an extensive set of eye movement features: Extraction methods and statistical analysis, *Journal of Eye Movement Research* **11**, 10 (2018).
21. Y. Terao, H. Fukuda and O. Hikosaka, What do eye movements tell us about patients with neurological disorders?—an introduction to saccade recording in the clinical setting—, *Proceedings of the Japan Academy, Series B* **93**, 772 (2017).
22. L. Graham, R. Vitorio, R. Walker, G. Barry, A. Godfrey, R. Morris and S. Stuart, Digital eye-movement outcomes (demos) as biomarkers for neurological conditions: a narrative review, *Big Data and Cognitive Computing* **8**, p. 198 (2024).
23. Y. Xu, C. Zhang, B. Pan, Q. Yuan and X. Zhang, A portable and efficient dementia screening tool using eye tracking machine learning and virtual reality, *NPJ Digital Medicine* **7**, p. 219 (2024).
24. T. W. Frazier, M. Strauss, E. W. Klingemier, E. E. Zetzer, A. Y. Hardan, C. Eng and E. A. Youngstrom, A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism, *Journal of the American Academy of Child & Adolescent Psychiatry* **56**, 546 (2017).
25. M. R. MacAskill, C. F. Graham, T. L. Pitcher, D. J. Myall, L. Livingston, S. van Stockum, J. C. Dalrymple-Alford and T. J. Anderson, The influence of motor and cognitive impairment upon visually-guided saccades in parkinson's disease, *Neuropsychologia* **50**, 3338 (2012).
26. S. Stuart, R. A. Lawson, A. J. Yarnall, J. Nell, L. Alcock, G. W. Duncan, T. K. Khoo, R. A. Barker, L. Rochester, D. J. Burn *et al.*, Pro-saccades predict cognitive decline in parkinson's disease: Icicle-pd, *Movement Disorders* **34**, 1690 (2019).
27. J. Waldthaler, L. Stock, C. Krüger-Zechlin, Z. Deeb and L. Timmermann, Cluster analysis reveals distinct patterns of saccade impairment and their relation to cognitive profiles in parkinson's disease, *Journal of Neuropsychology* **17**, 251 (2023).
28. S. Marx, G. Respondek, M. Stamelou, S. Dowiasch, J. Stoll, F. Bremmer, W. H. Oertel, G. U. Höglinger and W. Einhäuser, Validation of mobile eye-tracking as novel and efficient means for differentiating progressive supranuclear palsy from parkinson's disease, *Frontiers in behavioral neuroscience* **6**, p. 88 (2012).
29. J. E. Choi, P. A. Vaswani and R. Shadmehr, Vigor of movements and the cost of time in decision making, *Journal of neuroscience* **34**, 1212 (2014).
30. M. O. Al-Moteri, M. Symmons, V. Plummer and S. Cooper, Eye tracking to investigate cue processing in medical decision-making: A scoping review, *Computers in Human Behavior* **66**, 52 (2017).
31. K. Blondon, R. Wipfli and C. Lovis, Use of eye-tracking technology in clinical reasoning: a systematic review, *Digital Healthcare Empowering Europeans* , 90 (2015).
32. A. Van der Gijp, C. Ravesloot, H. Jarodzka, M. Van der Schaaf, I. Van der Schaaf, J. P. van Schaik and T. J. Ten Cate, How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology, *Advances in Health Sciences Education* **22**, 765 (2017).
33. G. E. Juantorena, F. Figari, A. Petroni and J. E. Kamienkowski, Web-based eye-tracking for remote cognitive assessments: The anti-saccade task as a case study, *bioRxiv* , 2023 (2023).
34. M. Daniol, D. Hemmerling, J. Sikora, P. Jemiolo, M. Wodzinski and M. Wojcik-Pedziwiatr, Eye-tracking in mixed reality for diagnosis of neurodegenerative diseases, in *Proceedings of the 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, 2024).

35. J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan and H. Devos, Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease, *IEEE transactions on visualization and computer graphics* **23**, 1302 (2017).
36. W. Mehringer, M. Stoeve, D. Krauss, M. Ring, F. Steussloff, M. Güttes, J. Zott, B. Hohberger, G. Michelson and B. Eskofier, Virtual reality for assessing stereopsis performance and eye characteristics in post-covid, *Scientific Reports* (2023).
37. I. B. Adhanom, P. MacNeilage and E. Folmer, Eye tracking in virtual reality: a broad review of applications and challenges, *Virtual Reality* **27**, 1481 (2023).
38. A. W. Przybyszewski, A. Śledzianowski, A. Chudzik, S. Szlufik and D. Koziorowski, Machine learning and eye movements give insights into neurodegenerative disease mechanisms, *Sensors* **23**, p. 2145 (2023).
39. A. Chudzik, A. Śledzianowski and A. W. Przybyszewski, Machine learning and digital biomarkers can detect early stages of neurodegenerative diseases, *Sensors* **24**, p. 1572 (2024).
40. A. W. Przybyszewski, M. Kon, S. Szlufik, A. Szymanski, P. Habela and D. M. Koziorowski, Multimodal learning and intelligent prediction of symptom development in individual parkinson's patients, *Sensors* **16**, p. 1498 (2016).
41. A. Sledzianowski, A. Szymanski, A. Drabik, S. Szlufik, D. M. Koziorowski and A. W. Przybyszewski, Measurements of antisaccades parameters can improve the prediction of parkinson's disease progression, in *Proceedings of the Asian Conference on Intelligent Information and Database Systems*, (Springer, 2019).
42. D. C. Brien, H. C. Riek, R. Yep, J. Huang, B. Coe, C. Areshenkoff, D. Grimes, M. Jog, A. Lang, C. Marras *et al.*, Classification and staging of parkinson's disease using video-based eye tracking, *Parkinsonism & Related Disorders* **110**, p. 105316 (2023).
43. É. de Villers-Sidani, P. Voss, D. Guitton, J. M. Cisneros-Franco, N. A. Koch and S. Ducharme, A novel tablet-based software for the acquisition and analysis of gaze and eye movement parameters: a preliminary validation study in parkinson's disease, *Frontiers in Neurology* **14**, p. 1204733 (2023).
44. N. A. Koch, P. Voss, J. M. Cisneros-Franco, A. Drouin-Picaro, F. Tounkara, S. Ducharme, D. Guitton and É. de Villers-Sidani, Eye movement function captured via an electronic tablet informs on cognition and disease severity in parkinson's disease, *Scientific Reports* **14**, p. 9082 (2024).
45. Y. R. Sun, S. B. Beylergil, P. Gupta, F. F. Ghasia and A. G. Shaikh, Monitoring eye movement in patients with parkinson's disease: what can it tell us?, *Eye and Brain* , 101 (2023).
46. Y. Mao, Y. He, L. Liu and X. Chen, Disease classification based on eye movement features with decision tree and random forest, *Frontiers in Neuroscience* **14**, p. 798 (2020).
47. K. Vodrahalli, M. Filipkowski, T. Chen, J. Zou and Y. J. Liao, Predicting visuo-motor diseases from eye tracking data, in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022*, (World Scientific, 2021).
48. G. Uribarri, S. E. von Huth, J. Waldthaler, P. Svenningsson and E. Fransén, Deep learning for time series classification of parkinson's disease eye tracking data, *arXiv preprint arXiv:2311.16381* (2023).
49. J. Reiner, L. Franken, E. Raveh, I. Rosset, R. Kreitman, E. Ben-Ami and R. Djaldetti, Oculometric measures as a tool for assessment of clinical symptoms and severity of parkinson's disease, *Journal of Neural Transmission* **130**, 1241 (2023).
50. M. Jiang, Y. Liu, Y. Cao, S. Xia, F. Teng, W. Zhao, Y. Lin and W. Liu, Diagnosis of parkinson's disease by eliciting trait-specific eye movements in multi-visual tasks, *Journal of Translational Medicine* **23**, p. 65 (2025).
51. I. Schuetz and K. Fiehler, Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability, *Journal of Eye Movement Research* **15**, 10 (2022).

52. F. Zuo, P. Jing, J. Sun, J. Duan, Y. Ji and Y. Liu, Deep learning-based eye-tracking analysis for diagnosis of alzheimer's disease using 3d comprehensive visual stimuli, *IEEE Journal of Biomedical and Health Informatics* **28**, 2781 (2024).
53. J. Sun, Y. Liu, H. Wu, P. Jing and Y. Ji, A novel deep learning approach for diagnosing alzheimer's disease based on eye-tracking data, *Frontiers in Human Neuroscience* **16**, p. 972773 (2022).
54. L. Rello and M. Ballesteros, Detecting readers with dyslexia using machine learning with eye tracking measures, in *Proceedings of the 12th International Web for All Conference*, (ACM, 2015).
55. I. A. Vajs, G. S. Kvaščev, T. M. Papić and M. M. Janković, Eye-tracking image encoding: Autoencoders for the crossing of language boundaries in developmental dyslexia detection, *IEEE Access* **11**, 3024 (2023).
56. B. Nerušil, J. Polec, J. Škunda and J. Kačur, Eye tracking based dyslexia detection using a holistic approach, *Scientific Reports* **11**, p. 15687 (2021).
57. I. Vajs, V. Ković, T. Papić, A. M. Savić and M. M. Janković, Dyslexia detection in children using eye tracking data based on vgg16 network, in *2022 30th European Signal Processing Conference (EUSIPCO)*, (IEEE, 2022).
58. A. E. El Hmimdi, Z. Kapoula and V. Sainte Fare Garnot, Deep learning-based detection of learning disorders on a large scale dataset of eye movement records, *BioMedInformatics* **4**, 519 (2024).
59. M. Elbattah, J.-L. Guérin, R. Carette, F. Cilia and G. Dequen, Nlp-based approach to detect autism spectrum disorder in saccadic eye movement, in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, (IEEE, 2020).
60. S. Chen and Q. Zhao, Attention-based autism spectrum disorder screening with privileged modality, in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, (IEEE, 2019).
61. M. Jiang and Q. Zhao, Learning visual attention to identify people with autism spectrum disorder, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, (IEEE, 2017).
62. D. Su, Y. Cui, C. He, P. Yin, R. Bai, J. Zhu, J. S. Lam, J. Zhang, R. Yan, X. Zheng *et al.*, Projections for prevalence of parkinson's disease and its driving factors in 195 countries and territories to 2050: modelling study of global burden of disease study 2021, *bmj* **388** (2025).
63. A. Srivastava, R. Sharma, S. K. Sood, G. Shukla, V. Goyal and M. Behari, Saccadic eye movements in parkinson's disease, *Indian journal of ophthalmology* **62**, 538 (2014).
64. G. Buscher, A. Dengel and L. van Elst, Eye movements as implicit relevance feedback, in *CHI'08 extended abstracts on Human factors in computing systems*, 2008 pp. 2991–2996.
65. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, Scikit-learn: Machine learning in python, *the Journal of machine Learning research* **12**, 2825 (2011).