Classification of Huntington's disease stage with support vector machines: A study on oculomotor performance

Ângela Miranda[1] · Rui Lavrador[1] · Filipa Júlio[1] · Cristina Januário[1,2] ·
Miguel Castelo-Branco[1] · Gina Caetano[1]

Abstract Alterations in oculomotor performance are among the first observable physical alterations during presymptomatic stages of Huntington's disease (HD). Quantifiable measurements of oculomotor performance have been studied as possible markers of disease status and progression in presymptomatic and early symptomatic stages of HD, on the basis of traditional analysis methods. Whether oculomotor performance can be used to classify individuals according to HD disease stage has yet to be explored via the application of machine-learning methods. In the present study, we report the application of the support vector machine (SVM) algorithm to oculomotor features pooled from a four-task psychophysical experiment. We were able to automatically distinguish control participants from presymptomatic HD (pre-HD) participants with an accuracy of 73.47 %, a sensitivity of 74.31 %, and a specificity of 72.64 %; to distinguish control participants from HD patients with an accuracy of 81.84 %, a sensitivity of 76.19 %, and a specificity of 87.48 %; and to distinguish pre-HD participants from HD patients with an accuracy of 83.54 %, a sensitivity of 92.62 %, and a specificity of 74.45 %. These results demonstrate that the application of supervised classification methods to oculomotor features is a valuable and promising approach to the automatic detection of disease stage in HD.

✉ Gina Caetano
  caetanogina@gmail.com; gcaetano@fmed.uc.pt

[1] Institute for Biomedical Imaging and Life Sciences/Faculty of Medicine, University of Coimbra, Coimbra, Portugal

[2] Coimbra University Hospital, Coimbra, Portugal

Huntington's disease (HD), previously known as "Huntington's chorea", is an autosomal-dominant neurodegenerative disorder that affects on average 5.70 per 100,000 people in the Western world (Pringsheim et al., 2012). HD is caused by a gene mutation, identified as a CAG triplet repeat expansion in the *huntingtin* (HTT) gene, located on chromosome 4, which encodes the huntingtin (HTT) protein. The mutant form of the HTT protein, with toxic properties, is expressed ubiquitously in the brain, and the pathogenesis from neural dysfunction to neural death has yet to be fully established (La Spada, Weydt, & Pineda, 2011; Phillips, Shannon, & Barker, 2008). HD is classically characterized by striatal degeneration, but thinning of the cortical mantle and overall brain volume loss are contingent and widespread throughout disease progression (Paulsen et al., 2014; Tabrizi et al., 2009).

HD is characterized by motor abnormalities, deterioration of cognitive function, and psychiatric disturbances. The onset of clinical HD is determined on the basis of alterations in motor functioning (Huntington Study Group, 1996); however, the order of symptoms' presentation is variable. It is well known that changes in oculomotor performance, as measured in laboratory settings, are one of the earliest detectable manifestations in HD and have the potential to output objective measures of disease status and progression (Anderson & MacAskill, 2013; André-Thomas, Abely, de Ajuriaguerra, & Eullien, 1945; Deurex, 1945; Grabska et al., 2014; Peltsch, Hoffman, Armstrong, Pari, & Munoz, 2008). The search for such objective measures has been a research hallmark, aimed at the validation and evaluation of novel therapeutic approaches that could be based on such disease-stage markers

(Paulsen et al., 2014; Tabrizi et al., 2009). For example, quantitative measurements of eye movements such as saccade latency and duration, as well as errors in oculomotor tasks, have been studied as possible biomarkers of presymptomatic and early symptomatic stages of HD (Ali, Michell, Barker, & Carpenter, 2006; Blekher et al., 2006; Golding, Danchaivijitr, Hodgson, Tabrizi, & Kennard, 2006; Patel, Jankovic, Hood, Jeter, & Sereno, 2012). The results from clinical HD patients have consistently shown significantly increased saccadic latencies (Ali et al., 2006; Antoniades, Xu, Mason, Carpenter, & Barker, 2010; Avanzini, Girotti, Caraceni, & Spreafico, 1979; Blekher et al., 2006; Golding et al., 2006; Patel et al., 2012; Rupp et al., 2012), saccadic durations (Ali et al., 2006; Avanzini et al., 1979), and errors in saccadic direction (Blekher et al., 2006; Patel et al., 2012; Rupp et al., 2012). Also, increased saccadic latencies (Blekher et al., 2006; Golding et al., 2006; Robert et al., 2009) and errors in saccadic direction have been reported in presymptomatic stages (Blekher et al., 2006). These results suggest that measures of saccadic movements are a possible indicator of disease stage in HD (Blekher et al., 2009). However, whether oculomotor performance parameters can be used to classify individuals according to their disease stage remains to be seen. Recently, this question has been studied in other neurodegenerative and neuropsychiatric disorders (Lagun, Manzanares, Zola, Buffalo, & Agichtein, 2011; Tseng et al., 2013), but it still remains unexplored in HD. A variety of experimental paradigms and several machine-learning algorithms have been applied to resolve these classification problems: (1) support vector machines (SVMs; Coco & Keller, 2014; Jin, Cheng, Zhou, & Li, 2012; Lagun et al., 2011; Rivera et al., 2012; Xiang, Yan, & Chen, 2009), (2) support vector machines with recursive feature elimination (SVM-RFE; Tseng et al., 2013), (3) Fisher linear discriminant analysis (Jin et al., 2012), (4) the multinomial log-linear neuronal networks model (Jin et al., 2012), (5) least-square angle regression (Coco & Keller, 2014; Jin et al., 2012), (6) naive Bayes (Lagun et al., 2011), and (7) logistic regression (Lagun et al., 2011).

Among the algorithms listed, SVM-RFE and SVM have demonstrated the best overall performance. For example, Tseng et al. (2013) achieved an accuracy of 89.6 % with Parkinson's disease, and Lagun et al. (2011) of 87 % with mild cognitive impairment. The variety of paradigms used is also reflected in these studies: The former (Tseng et al., 2013) developed a novel approach to analyze eye movements recorded in natural viewing settings, whereas in the latter (Lagun et al., 2011) a visual paired recognition task was applied, proving its sensitivity in identifying functional changes in medial temporal lobe structures.

SVM is a machine-learning method based on the construction of a separating hyperplane as a decision boundary, so as to maximize the margin of separation among the classes (Burges, 1998; Rizk-Jackson et al., 2011). More specifically, given a training data containing two distinct classes, the SVM algorithm constructs a model such that data from individuals belonging to one class are on one side of the hyperplane, while those of the other class are on the other side of the hyperplane. The optimal separating hyperplane is found by maximizing the distance between each class of the data points and the decision boundary, while the amount of classification error is minimized. The minimization of training errors depends on the cost (C) parameter. A large value of the C parameter corresponds to assigning a higher penalty to the classification error; that is, the SVM algorithm constructs a rigid-margin hyperplane and forces the creation of a more accurate model. On the other hand, a small value of the C parameter implies that the SVM algorithm searches for a larger-margin-separating hyperplane (soft margin) allowing for a greater number of misclassifications (Rizk-Jackson et al., 2011; Zhu & Ji, 2005).

In this study, we used linear SVM to classify eye tracking data for pre-HD, HD, and control groups, and to predict unknown participants on the basis of features provided to the classifier. We hypothesized that, on the basis of oculomotor performance-derived features, we could develop binary classification models able to accurately assign a label to individuals, analyze the feature–accuracy relationship, and compare classifier performance for the different oculomotor tasks.

Experiment

Participants

Fifty participants recruited through the Neurology department (Movement Disorder Unit) of Coimbra University Hospital and the Huntington's Portuguese Association gave written informed consent, according to the Declaration of Helsinki and approved by the local ethics committee (Faculty of Medicine, University of Coimbra), to take part in the study. Participants were assigned to three groups: the CTRL group—22 healthy individuals with no known neurological disorder and not at risk for HD; the pre-HD group—14 presymptomatic HD participants, composed of individuals who presented soft signs of motor abnormalities; and the early symptomatic HD group—14 participants diagnosed with clinical HD who demonstrated evident signs of motor disorders. The demographic characteristics of these groups are shown in Table 1.

The clinical diagnoses of the participants were established according to the motor subscale of the Unified Huntington's Disease Rating Scale-99 (Huntington's Study Group, 1996). Participants with a rating from 0 to 4 were classified as pre-HD, whereas individuals with a rating of ≥5 were classified as HD. Exclusion criteria included a history of substance abuse/dependence, severe ophthalmic disease, or concurrent neurological illness.

Table 1 Demographic characteristics of the three groups: Mean (*SE*)

	Gender F:M	Age	CAG Repeats	Disease Duration (years)	Years to Onset[2]	UHDRS (Motor)
Controls (N = 22)	15:7	36.18 (2.25)	N/A[1]	N/A	N/A	N/A
Presymptomatic HD (N = 14)	8:6	35.79 (2.63)	41.29 (0.52)	N/A	21.01 (2.92)	1.93 (0.46)
Symptomatic HD (N = 14)	10:4	44.86 (4.15)	43.64 (0.71)	4.57 (0.76)	2.54 (2.60)	26.50 (3.70)

N/A: not applicable. [1] Control participants either did not have a family history of HD or had a gene negative status (≤32 repeats) when family history was present. [2] Estimated time to HD clinical onset (Langbehn, Brinkman, Falush, Paulsen, & Hayden, 2004).
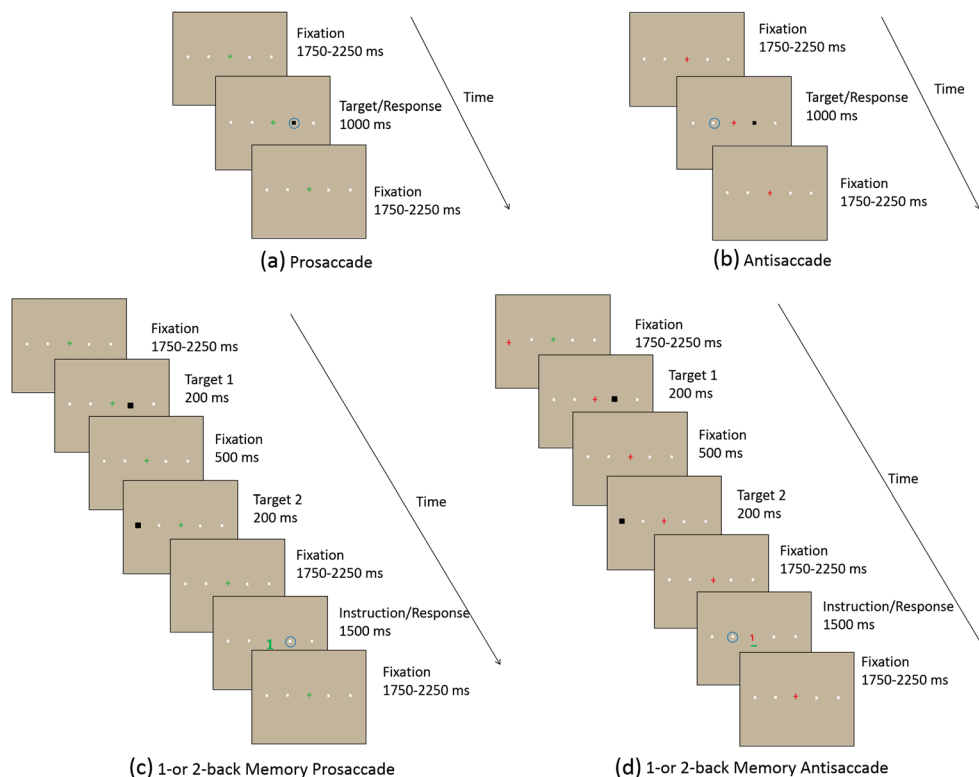
Tasks

The participants completed four horizontal saccadic tasks: prosaccade, antisaccade, 1- or 2-back memory prosaccade, and 1- or 2-back memory antisaccade.

As is shown in Fig. 1, the fixation position (cross, 1° diameter in visual angle) at the center of the screen display was common to all tasks. Peripheral visual targets (black squares, subtending 0.6° of visual angle) appeared randomly at ±6° or ±12° of visual angle, horizontally from the central cross. At each of the target positions were placed small position cues (* symbols, subtending 0.24° of visual angle, light gray in color), which remained throughout the experiments. Each task was preceded by a training period in which the participant was

verbally instructed to ensure that the goal of each task was completely understood.

The prosaccade (PS) task consisted of 60 trials. Each trial comprised a fixation phase, defined between 1,750 and 2,250 ms—or until the participant had had a gaze fixation for at least 500 ms—followed by the presentation of a peripheral stimulus for 1,000 ms (Fig. 1a). The participant was instructed to fixate the green cross during the fixation period and to look as soon as possible at the peripheral stimulus once it had appeared, and then to return to the fixation position (central green cross). The central fixation cross remained visible throughout the entire experiment.

The antisaccade (AS) task (Fig. 1b) had the same number of trials, delay intervals, and durations for the fixation and task

Fig. 1 Experimental designs of the four horizontal saccadic tasks

phases as the PS task. In this task, the participant was instructed to look as soon as possible at the opposite position from the visual target, and then to return to the fixation position (central red cross). The central fixation cross remained visible throughout the entire experiment.

The 1- or 2-back memory prosaccade (MPS) task, schematized in Fig. 1c, contained 96 trials. Each trial comprised a fixation phase, ranging between 1,750 and 2,250 ms, followed by the presentation of two stimuli sequentially—visual targets were presented for 200 ms each, with an interval in-between of 500 ms. Afterward, a new fixation phase occurred (1,750–2,250 ms or until the fixation criteria were met), followed by an instruction phase of 1,500-ms duration, during which the participant had to perform the task. Participants were requested to hold fixation on the central green cross during the whole trial, with the exception of the instruction phase. During the instruction/task phase, the green cross was replaced by a digit. If the digit was 1, the participant was asked to look as soon as possible at the remembered position of the first visual target. If the digit was 2, the participant was asked to look at the remembered position of the second visual target.

The 1- or 2-back memory antisaccade (MAS) task (Fig. 1d) had the same number of trials and duration for each phase as in the MPS task. In the instruction/task phase, if the digit was 1 (or 2), the participant was asked to look as soon as possible at the opposite position from the remembered position of the first (or second) visual target. The participant was instructed to return to the central red cross after the task.

Data acquisition

The eye movement data were recorded in a psychophysical environment using an iView X Hi-Speed (1.06, SensoMotoric Instruments, Teltow) noninvasive eye tracking system, with a sampling rate of 240 Hz. Stimuli were presented on a 17-in. monitor at a resolution of 1,024 × 768 pixels while participants were seated 52 cm from the computer screen, with their heads placed in a stable chinrest. Before each eye tracking recording session, a nine-point calibration was performed for the dominant eye of the participant. The tasks were performed in a fixed order, with an increase in memory and executive load: prosaccade, antisaccade, 1- or 2-back memory prosaccade, and 1- or 2-back memory antisaccade. The experiment took between 40 min and 1 h to complete.

Method

Data preprocessing

The eye tracking data recorded during each task were viewed, analyzed, and exported using the BeGaze software (3.4, SensoMotoric Instruments, Teltow).

BeGaze experiments were created for each task and participant, on the basis of saccade detection, with the following criteria: a peak velocity threshold of 40°/ms; a minimum fixation duration of 50 ms, with all fixations below the limit rejected; and velocities of 15°/ms and 85°/ms to identify saccade initiation and termination, respectively. In addition, regions of interest (ROIs; ±2.5° × ±4° of visual angle) were defined around the possible positions for the peripheral visual targets and around the fixation position.

The raw data were visually inspected with the Scan Path, Line Graph, and Heat Map tools of the BeGaze software (See Fig. S1).

We identified and excluded participants for whom the following problems occurred: The participant closed the eye for long time periods, moved the head consistently, or simply did not perform the task. Finally, the data on saccades, fixations, and blinks were exported for further analysis.

Identification of valid trials The identification of valid trials, for each task and participant, was performed with MATLAB (Version R2013a). Criteria were defined for each task trial, as follows: (1) The data should contain a primary saccade, performed in the correct direction, according to the task instructions; (2) the primary saccade was initiated within the limits defined for the central fixation position; (3) the primary saccade had a latency higher than 80 ms—saccades below this latency were considered anticipatory, and the full trial was discarded; (4) the total saccadic movement finished within the ROI surrounding the visual target, as defined by the task—so long as no direction and latency errors were present; (5) the trial did not contain blinks prior to eye gaze reaching the target position, as defined by the task instructions; and (6) the primary saccade was initiated prior to 700 ms (PS and AS tasks) or 1,000 ms (MPS or MAS tasks)—latencies above the indicated thresholds were considered long-latency errors, and the whole trial was disregarded.

The trials that did not meet the above requirements (invalid trials, comprising anticipatory errors, direction errors, long-latency errors, and final ROI errors) were excluded from the analysis. Participants who had more than 85 % of their trials invalid—that is, less than nine valid trials (out of a total of 60) for the AS and PS tasks, and less than 14 valid trials (out of a total of 96) for the MAS and MPS tasks—were excluded (see Table 2).

Feature extraction

Once valid trials were defined, for each task and participant a set of nine features were extracted:

- *Percentage of anticipatory saccade errors (AE):* percentage of anticipatory saccades in relation to the total number of trials.

Table 2 Number of included and excluded participants after identification of valid trials, per task

	Excluded[a]	Excluded[b]	Included
PS task	0 CTRL	0 CTRL	22 CTRL
	0 pre-HD	0 pre-HD	14 pre-HD
	0 HD	0 HD	14 HD
AS task	0 CTRL	0 CTRL	22 CTRL
	0 pre-HD	0 pre-HD	14 pre-HD
	1 HD	6 HD	7 HD
MPS task	0 CTRL	1 CTRL	21 CTRL
	0 pre-HD	0 pre-HD	14 pre-HD
	3 HD	2 HD	9 HD
MAS task	0 CTRL	1 CTRL	21 CTRL
	0 pre-HD	1 pre-HD	13 pre-HD
	5 HD	1 HD	8 HD

[a] Participant did not perform the task. [b] Participants with less than 15 % valid trials

- *Percentage of direction errors (DE):* percentage of reflexive saccades performed in the opposite direction from the correct hit in relation to the total number of trials.
- *Percentage of errors, ROI (ER):* percentage of trials in which the primary saccade was executed in the correct direction but in which the target position did not correspond to the target ROI, in relation to the total number of trials.
- *Percentage of errors, long latency (EL):* percentage of errors identified in which the primary saccade had a latency higher than 700 ms (PS/AS task) or 1,000 ms (MAS/MPS task) post-visual-target presentation.
- *Average number of saccades (NS):* average number of saccades, for valid trials, needed to reach the target.
- *Saccade start latency (SL):* median of latency values corresponding to the time from stimulus appearance to the onset of the primary saccade (milliseconds).
- *Dispersion of saccade start latencies (DSL):* dispersion of the latency values corresponding to the interquartile range—that is, the difference between the 3rd and 1st quartiles. The calculation of the dispersion included only valid trials (milliseconds).
- *Dispersion of saccade duration (DSD):* dispersion of the duration values of primary saccades in valid trials (milliseconds). The dispersion corresponded to the interquartile range—that is, the difference between the 3rd and 1st quartiles.
- *Dispersion of saccadic end latencies (DSL_end):* dispersion of the values corresponding to the time at which the correct saccadic movement attained the target ROI (milliseconds).

Moreover, we took into account the fixation periods in-between tasks and extracted two additional features:

- *Percentage of fixation errors (FE):* percentage of detected fixations outside of the ROI defined for the fixation period (fixation at cross displayed in the center of the screen), in relation to the total number of fixations. The first fixation detected outside the central ROI was discarded for all participants, to minimize the inclusion of fixations on the target position from the former task period.
- *Fixation duration (FD):* average duration for fixations within the ROI defined for fixation periods in-between task trials.

Automatic classification

Training and test classification algorithm In the present study, we used the LIBSVM ("Library for Support Vector Machines") toolkit (Chang & Lin, 2011) on the MATLAB platform to train the classifier and then predict the labels of unknown participants.

The process of distinguishing between the CTRL and pre-HD participants, CTRL and HD participants, and pre-HD and HD participants, for each task, included the following steps:

(i) *Running SVM classification several times between two groups.* Each group had a different number of participants; therefore, to prevent biased classification (i.e., different numbers of participants between classified groups or unmatched ages), we determined all possible combinations of n elements (the number of participants in the smaller group) for the group with the higher number of participants, in each of the binary classifications. From the possible combinations of n elements, only subgroups matched for age were considered (Mann–Whitney test).

Within the set of possible combinations, we selected 10 % for the classification between two groups; this was preceded by testing classification results once considering 100 %, 30 %, and 10 % of the combinations, and ensuring that classification results did not vary between the full set and the random selection of subgroup combinations. The performances of the classifiers obtained when we used 10 % or 100 % of the possible subgroups were similar (no statistically significant differences).

(ii) *Selection of training and testing samples.* The training and testing samples were defined using the cross-validation (CV) method "leave one out". A participant from each class was randomly selected for the test sample, while the remaining participants were selected for the training sample. This process was repeated 100 times.

(iii) *Classification of the participants.* The classification process included two stages: the training stage and the prediction stage. In the first, a set of training data (features) and the respective participants' labels were provided to create the classification model. In the second, the classification

model and a new set of the features (testing data) were included, and the most likely class of the testing data (e.g., CTRL, pre-HD, or HD participant) was obtained.

To calculate accuracy estimates, the classification process described above was repeated for each of the subsets of the training and testing data obtained from each subgroup in CV.

Evaluation of the performance of the classifier The evaluation measures of the classifier's performance are frequently defined from a matrix composed of the samples correctly and incorrectly classified for each class (Costa, Lourena, Carvalho, & Freitas, 2007)—that is, an $m \times m$ matrix, where m is the number of classes and an M_{ij} entry indicates the number of classes belonging to the class i that have been classified as belonging to the class j. This matrix is named the *confusion matrix* or *contingency table*. Therefore, in order to evaluate the classification algorithm's performance, we computed a confusion matrix for hold-out test data during CV.

The confusion matrix is composed of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) for each binary classification.

The final results of the classification algorithm performance are reported, for each task, determining the mean values of a priori defined evaluation metrics (Costa et al., 2007; Lagun et al., 2011; Sokolova & Lapalme, 2009).

- *Accuracy:* proportion of the participants correctly classified in relation to the total number of participants in the test set.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- *Sensitivity:* rate of participants belonging to the positive class who were correctly predicted as positive.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- *Specificity:* proportion of participants classified as negative who were really negative.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Results

Figure 2a–d display the mean accuracy values obtained in the classification, for each task, when using the SVM algorithm with only one of the oculomotor performance features. The mean values of accuracy, sensitivity, and specificity for each feature are reported in Tables S.1, S.2, S.3, and S.4 of the supplementary material. The performance of the classifiers

for the best combinations of features and the whole set of features are reported in Tables 3, 4, 5, and 6 for the PS, AS, MPS, and MAS tasks, respectively. The performance of classifiers when using features from different tasks is reported in Tables S.5 and S.6 (supplementary material), and a summary of standard oculomotor results (mean and dispersion measures) for each of the tasks and groups can also be found in the supplementary material (Tables S.7, S.8, S.9, and S.10).

Classification results for the PS task

The best performance was obtained for the CTRL vs. HD and pre-HD vs. HD classifiers when using the FE feature (accuracies ≥ 75 %), whereas for the CTRL vs. pre-HD classification, the DSD feature achieved an accuracy of 59.14 %. When considering only features related to the task period, the best performance of the CTRL vs. HD classifier was obtained for EL (accuracy of 72.88 %), and the best performance of the pre-HD vs. HD classifier was achieved with the AE feature (accuracy of 72.00 %).

Adding features to the SVM algorithm did not result in improved classifier performance (Table 3). The combination of features AE, EL, and FE proved to be the best for the CTRL vs. HD classifier (accuracy of 76.88 %), whereas for pre-HD vs. HD, the best combination of features was AE, DE, and EL (accuracy of 72.50 %).

Classification results for the AS task

The best performance was achieved for the CTRL vs. pre-HD and CTRL vs. HD classifiers with the FE feature (accuracies ≥ 67 %), whereas for the pre-HD vs. HD classifier it was obtained when the SVM algorithm used the DSL feature, with an accuracy of 74.53 %. When considering only the trials in which the participant was performing the task, the best performance of the CTRL vs. pre-HD and CTRL vs. HD classifiers was obtained for the DE feature (accuracies of 64.57 % and 70.73 %, respectively). Adding features to the classification only improved the performance of the CTRL vs. pre-HD classifier (Table 4), with the best combination of features being DE, DSL, and FE (accuracy of 69.27 %).

Classification results for the MPS task

The best performance for the CTRL vs. HD and pre-HD vs. HD classifiers was achieved using the DSL_end feature (accuracies > 77 %). On the other hand, the CTRL vs. pre-HD classifier did not reach chance level.

Adding features to the SVM algorithm did not result in performance improvements for the classifiers (Table 5). The best performance for pre-HD vs. HD was obtained using the combination of features AE, DE, and ER, whereas the combination of features DE, ER, and DSL proved to be the best for the CTRL vs. HD (accuracy of 72.82 %). For CTRL vs. pre-
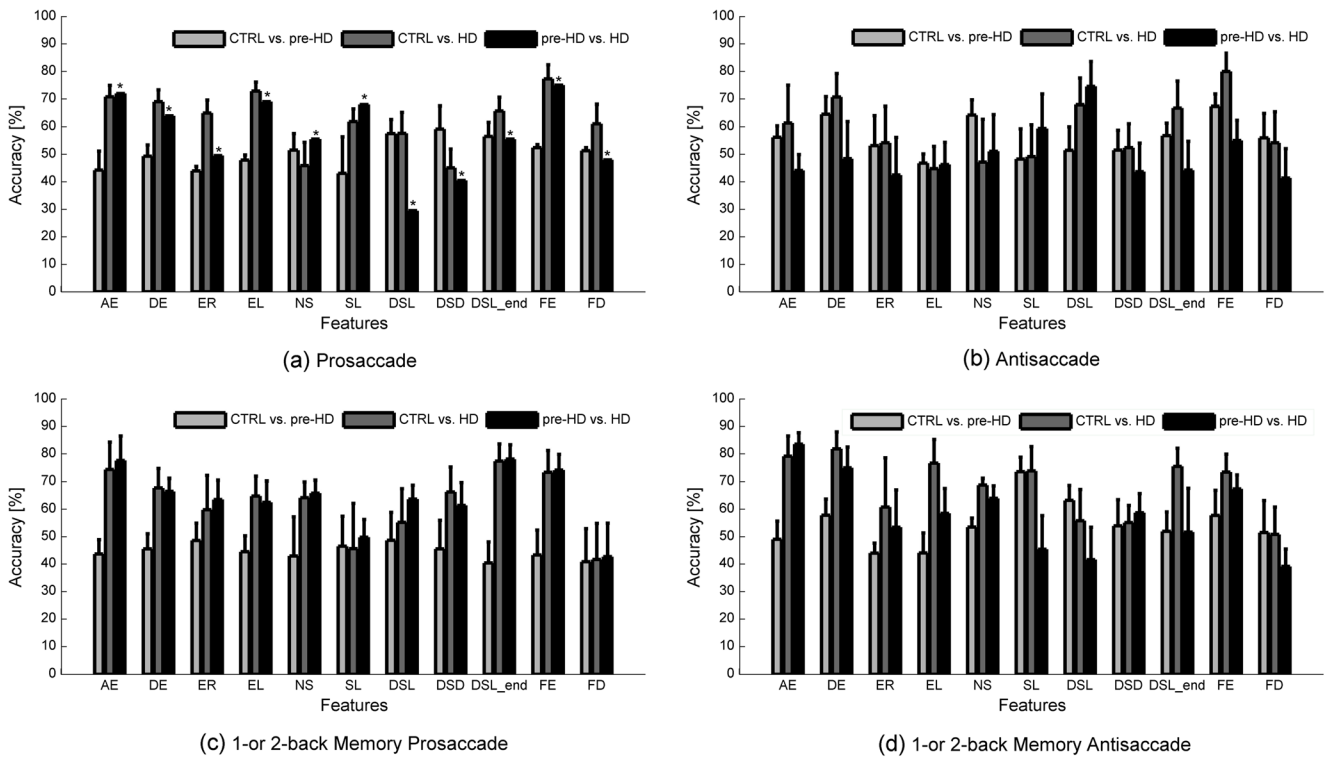
(a) Prosaccade

(b) Antisaccade

(c) 1-or 2-back Memory Prosaccade

(d) 1-or 2-back Memory Antisaccade

Fig. 2 Mean accuracy (and *SD*) values for the CTRL vs. pre-HD, CTRL vs. HD, and pre-HD vs. HD classifiers, for each of the oculomotor tasks and oculomotor features, respectively. *The *SD* associated with the accuracy values could not be calculated for the pre-HD vs. HD

classifier in the prosaccade task, since after application of the exclusion criteria (Table 2), only one possible combination of pre-HD and HD groups was possible

HD the classification did not reach chance level, independent of the combinations of features.

Classification results for the MAS task

The best performance of the CTRL vs. HD classifier was achieved using the DE feature (accuracy of 81.84 %), whereas for the pre-HD vs. HD classifier, the best performance was obtained with the AE feature (accuracy of 83.54 %). For the CTRL vs. pre-HD classifier, the best performance was obtained with the SL feature (accuracy of 73.47 %). Adding features to

the SVM algorithm did not result in performance improvements for the classifiers (Table 6). The combination of the features AE, DE, and FE proved to be the best for the CTRL vs. pre-HD and CTRL vs. HD (accuracies ≥61 %). For pre-HD vs. HD, the best performance was obtained for the combination of features AE, DE, and EL (accuracy of 74.35 %).

Classification results for the combinations of features from different tasks

For the combination of features from the PS and AS tasks (Table S.5 of the supplementary material), the classifier

Table 3 Classifier's performance with different combinations of features for the prosaccade (PS) task: Mean (*SD*)

Classifier	CTRL vs. pre-HD			CTRL vs. HD			pre-HD vs. HD		
Features	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
AE + EL + FE	0.53	0.05	1.00	0.77	0.82	0.71	0.66	0.71	0.61
	(0.01)	(0.02)	(0.00)	(0.07)	(0.07)	(0.10)	(*)	(*)	(*)
AE + DE + EL	0.45	0.75	0.15	0.74	0.87	0.61	0.73	0.87	0.58
	(0.06)	(0.12)	(0.11)	(0.05)	(0.06)	(0.08)	(*)	(*)	(*)
All features	0.52	0.04	1.00	0.71	0.72	0.70	0.50	0.46	0.54
	(0.01)	(0.02)	(0.00)	(0.07)	(0.08)	(0.09)	(*)	(*)	(*)

The standard deviations (*SD*s) associated with the mean accuracy (Acc.), sensitivity (Sens.), and specificity (Spec.) of each classifier were calculated for the 10 % of possible combinations randomly selected for the PS task.

Table 4 Classifier's performance with different combinations of features for the antisaccade (AS) task: Mean (*SD*)

Classifier	CTRL vs. pre-HD			CTRL vs. HD			pre-HD vs. HD		
Features	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
DE + DSL + FE	0.69	0.73	0.65	0.68	0.69	0.67	0.66	0.67	0.66
	(0.07)	(0.10)	(0.07)	(0.13)	(0.13)	(0.15)	(0.14)	(0.14)	(0.16)
AE + ER + DSL_end	0.38	0.43	0.33	0.73	0.74	0.72	0.73	0.74	0.73
	(0.08)	(0.09)	(0.12)	(0.10)	(0.15)	(0.10)	(0.09)	(0.13)	(0.08)
All features	0.58	0.60	0.56	0.64	0.66	0.61	0.61	0.58	0.63
	(0.08)	(0.10)	(0.08)	(0.13)	(0.15)	(0.16)	(0.16)	(0.16)	(0.20)

The standard deviations (*SD*s) associated with the mean accuracy (Acc.), sensitivity (Sens.), and specificity (Spec.) of each classifier were calculated for the 10 % of possible combinations randomly selected for the AS task.

performance only improved for the pre-HD vs. HD classification and when the FE feature was selected in both tasks (accuracy of 79 %). When considering only features extracted from periods in which participants were requested to perform a task, the classifiers' performance did not increase in comparison to the best result obtained from each task separately. Similarly, no improvement in the classifiers' performance was achieved when combining features from the MPS and MAS tasks (supplementary material, Table S.6), in comparison with the best results obtained from each task separately. In this case, the best performance was achieved for the CTRL vs. HD classifier when using the features DSL_end and DE in the MPS and MAS tasks, respectively (accuracy of 79 %).

Discussion

Our results highlight the potential use of oculomotor performance features to classify individuals within Huntington's disease stages. We demonstrated that the classifiers' performance is directly influenced by the selected features within each oculomotor task. Hence, the features to feed into the SVM algorithm that allow the highest accuracy estimates depend on the oculomotor task and HD stage.

For all classifiers, the best performance was achieved when one feature was selected within the MAS task (accuracy, specificity, and sensitivity > 70 %). Namely, SL proved to be the best feature to distinguish controls from pre-HD participants, whereas to distinguish control or pre-HD participants from HD patients, DE and AE, respectively, proved to be the best features. We emphasize that even within the PS task, the classifier was able to distinguish control and pre-HD participants from HD patients with accuracies above 70 %—by using either the AE feature or the FE feature. These results demonstrate the potential to integrate this methodological approach in clinical environment, where with a relatively simple task one might aim at obtaining a clear indication (predictive value and specificity) of disease stage and progression.

The choice of the SVM algorithm, with a linear kernel, has revealed the potential to classify and predict unknown individuals' labels according to HD stage. As a matter of fact, former studies by Coco and Keller (2014), Jin, Cheng, Zhou, and Li (2012), and Lagun, Manzanares, Zola, Buffalo, and Agichtein (2011) had explored the use of several machine-learning algorithms, reaching the conclusion that SVMs provided the overall best accuracy, specificity, and sensitivity. However, we have to consider that other algorithms may have performed better. Additionally, our results seem to

Table 5 Classifier's performance with different combinations of features for the 1- or 2-back memory prosaccade (MPS) task: Mean (*SD*)

Classifier	CTRL vs. pre-HD			CTRL vs. HD			pre-HD vs. HD		
Features	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
AE + DE + ER	0.41	0.39	0.44	0.70	0.78	0.62	0.74	0.77	0.71
	(0.09)	(0.28)	(0.20)	(0.09)	(0.09)	(0.12)	(0.08)	(0.09)	(0.09)
DE + ER + DSL	0.38	0.43	0.33	0.73	0.74	0.72	0.73	0.74	0.73
	(0.11)	(0.17)	(0.12)	(0.07)	(0.09)	(0.08)	(0.06)	(0.09)	(0.08)
All features	0.39	0.36	0.42	0.62	0.59	0.65	0.61	0.59	0.63
	(0.11)	(0.13)	(0.12)	(0.10)	(0.14)	(0.10)	(0.07)	(0.10)	(0.08)

The standard deviations (*SD*s) associated with the mean accuracy (Acc.), sensitivity (Sens.), and specificity (Spec.) of each classifier were calculated for the 10 % of possible combinations randomly selected for the MPS task.

Table 6 Classifier's performance with different combinations of features for the 1- or 2-back memory antisaccade (MAS) task: Mean (*SD*)

Classifier	CTRL vs. pre-HD			CTRL vs. HD			pre-HD vs. HD		
Features	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
AE + DE + EL	0.52	0.57	0.47	0.75	0.78	0.72	0.74	0.75	0.74
	(0.09)	(0.13)	(0.12)	(0.10)	(0.12)	(0.11)	(0.05)	(0.09)	(0.05)
AE + DE + FE	0.61	0.63	0.59	0.81	0.83	0.78	0.73	0.75	0.71
	(0.08)	(0.09)	(0.12)	(0.08)	(0.10)	(0.09)	(0.06)	(0.10)	(0.07)
All features	0.55	0.54	0.56	0.71	0.74	0.67	0.58	0.64	0.53
	(0.10)	(0.11)	(0.12)	(0.09)	(0.14)	(0.08)	(0.10)	(0.12)	(0.11)

The standard deviations (*SD*s) associated with the mean accuracy (Acc.), sensitivity (Sens.), and specificity (Spec.) of each classifier were calculated for the 10 % of possible combinations randomly selected for the MAS task.

indicate that combining features from different oculomotor tasks does not necessarily imply a better prediction of single individuals' classes. As such, classifiers' performance did not improve in comparison to the best results achieved using features from within one single task. One might refute this by considering that a multikernel approach would have been best suited to combine features from different oculomotor tasks—one kernel for each task. However, despite reducing users' influence on kernel choice, given the number of participants in this study and the features we explored, we think the overall performance would not surpass what we have reported.

When taking into account the known neuropathology in Huntington's disease and reported oculomotor deficits in manifest HD, our results seem to be in agreement with these factors. Both the DE and AE features reflect abnormalities of oculomotor performance in manifest HD, which parallels certain neurodegenerative processes—for example, marked deficits in voluntary saccade initiation and saccadic suppression, impersistent gaze, and increased incidence of anticipatory saccades (Ali et al., 2006; Antoniades et al., 2010; Blekher et al., 2006; Golding et al., 2006). HD's hallmark is well known to be striatal neural loss and atrophy (Aylward, 2007; Aylward et al., 2011), with altered functioning of the cortico–basal ganglia–thalamocortical circuitry (Joel, 2001); neurodegeneration spreads throughout the brain with disease progression (Ross et al., 2014). Findings suggest that through an initial loss of medium-sized spiny neurons, a major impact occurs in the function of the basal ganglia indirect pathway (Calabresi, Picconi, Tozzi, Ghiglieri, & Di Filippo, 2014; Galvan, André, Wang, Cepeda, & Levine, 2012; Hikosaka, Takikawa, & Kawagoe, 2000), although other mechanisms that would have an early impact on the direct pathway's function are also being unraveled (Calabresi et al., 2014; Galvan et al., 2012). Thus, reduced inhibition of the superior colliculus is indicated as impairing saccade suppression, leading to some of the usual HD oculomotor patterns.

DE relates to impaired suppression of saccades, which better differentiates the manifest HD group—the classifiers' performance reflects the disease neuropathology. Furthermore, both features present accuracies above 70 % in other tasks than the MAS task.

On the other hand, SL is the feature that best differentiated pre-HD individuals from controls, where other features resulted in much poorer performance. It is known that the saccade latency can be affected by numerous factors—for example, the design of a paradigm and its elicited brain activity (e.g., Peltsch et al., 2008); parietal eye field projections to the superior colliculus, with lesions to the former resulting in increased latencies; lesions or dysfunction of frontal cortical areas, such as frontal eye field, also increase saccade latencies (Anderson & MacAskill, 2013; Leigh & Kennard, 2004); degeneration or dysfunction of the substantia nigra pars reticulata, the superior colliculus, or the brainstem itself can also result in increased latencies (Hikosaka & Wurtz, 1985a, b); cerebellar cortex interplays and also performs parallel processing of visuomotor information, and damage can result in increased latencies (Filippopulos, Eggert, & Straube, 2013). All of these can be contributing factors in more advanced stages of neurodegeneration. However, the pattern detected indicates faster latencies in HD gene carriers for the MAS task (Table S.9). Hence, given the inherent load in spatial working memory and executive functioning, with inhibition of saccades, we suggest that the inhibitory pathway may play a major role. The neural correlates, function, and compensatory mechanisms can be further unraveled with full cognitive characterization (Júlio, Caetano, Januário, & Castelo-Branco, manuscript in preparation) and through neuroimaging studies.

The relatively small number of individuals per group was one major limitation of this study, although the proof of concept per se remains. Moreover, the pre-HD group was mostly composed of participants estimated to be far from disease onset; thus, we expect that smaller effect sizes were present, which might have further hindered the CTRL vs. pre-HD classifiers' performance. The use of other oculomotor

features—such as gain, velocity, saccade duration, or express saccades from gap or delayed paradigms, and in our case, distinguishing between 6° and 12° trials—could have implied better classifier performance. Nonetheless, the replicability of measures within short time spans is important. The study by Blekher et al. (2009) has shown that saccade latency and the percentage of errors have moderate to high replicability in control and premanifest HD groups (see also Klein & Fischer, 2005), whereas gain and velocity only present moderate reliability. Moreover, the replicability of the latter did not reach significance in a simple visually guided (prosaccade) task. In addition, distinguishing between 6° and 12° trials would not have added information to our classification between groups, but might possibly have refined the DSD feature, where intragroup differences reached significance in three situations (see the supplementary material, Table S.11). Finally, gap and delayed paradigms might increase the classifiers' performance for the percentage of express saccades feature (latencies between 80 to 130 ms), when considering premanifest HD individuals close to disease onset and manifest HD patients. These paradigms are known to increase the number of express saccades, suggested as being linked to release mechanisms when a fixation target disappears—which is affected in HD—with direct functional involvement of the rostral pole of the superior colliculus (Dickov & Morrison, 2006). Our experimental paradigm was originally aimed at underpinning the interference of working memory and executive load with the inhibition of saccades, in parallel with a full cognitive characterization of the participants (Júlio, Caetano, Januário, & Castelo-Branco, manuscript in preparation).

The use of machine-learning methods in HD oculomotor data will improve substantially with a higher number of participants, with the capacity to distinguish and classify between different premanifest stages being one of the most relevant potential applications. Additionally, heterogeneities in the development of symptoms at the earliest manifest stages of the disease may benefit from automatic detection of oculomotor patterns.

All in all, this study has pioneered the application of machine-learning methods to eye movement data in HD, demonstrating that SVMs are a valuable option to automatically detect disease stage at the individual level.

References

Ali, F. R., Michell, A. W., Barker, R. A., & Carpenter, R. H. S. (2006). The use of quantitative oculometry in the assessment of Huntington's disease. *Experimental Brain Research, 169,* 237–245.

Anderson, T. J., & MacAskill, M. R. (2013). Eye movements in patients with neurodegenerative disorders. *Nature Reviews Neurology, 9,* 74–85.

André-Thomas, M., Abely, X., de Ajuriaguerra, J., & Eullien, L. (1945). Troubles de l'élévation des globes oculaires dans un cas de chorée de Huntington. *Revue Neurologique, 77,* 248–250.

Antoniades, C. A., Xu, Z., Mason, S. L., Carpenter, R. H. S., & Barker, R. A. (2010). Huntington's disease: Changes in saccades and hand-tapping over 3 years. *Journal of Neurology, 257,* 1890–1898.

Avanzini, G., Girotti, F., Caraceni, T., & Spreafico, R. (1979). Oculomotor disorders in Huntington's chorea. *Journal of Neurology, Neurosurgery and Psychiatry, 42,* 581–589.

Aylward, E. H. (2007). Change in MRI striatal volumes as a biomarker in preclinical Huntington's disease. *Brain Research Bulletin, 72,* 152–158.

Aylward, E. H., Nopoulus, P. C., Ross, C. A., Langbehn, D. R., Pierson, R. K., Mills, J. A., … PREDICT-HD Investigators and Coordinators of the Huntington Study Group. (2011). Longitudinal change in regional brain volumes in prodromal Huntington disease. *Journal of Neurology, Neurosurgery and Psychiatry, 82,* 405–410.

Blekher, T., Johnson, S. A., Marshall, J., White, K., Hui, S., Weaver, M., … Foroud, T. (2006). Saccades in presymptomatic and early stages of Huntington disease. *Neurology, 67,* 394–399.

Blekher, T., Weaver, M. R., Cai, X., Hui, S., Marshall, J., Jackson, J. G., … Foroud, T. M. (2009). Test–retest reliability of saccadic measures in subjects at risk for Huntington disease. *Investigative Ophthalmology and Visual Science, 50,* 5707–5711.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2,* 121–167.

Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V., & Di Filippo, M. (2014). Direct and indirect pathways of basal ganglia: A critical reappraisal. *Nature Neuroscience, 17,* 1022–1030.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2,* 27:1–27:27.

Coco, M. I., & Keller, F. (2014). Classification of visual and linguistic tasks using eye-movement features. *Journal of Vision, 14*(3), 11:1–18. doi:10.1167/14.3.11

Costa, E. P., Lourena, A. C., Carvalho, A. C. P. L. F., & Freitas, A. A. (2007). *A review of performance evaluation measures for hierarchical classifiers*. In: *Evaluation methods for machine learning II: papers from the AAAI-07 workshop (pp. 1–6)*. Palo Alto, CA: AAAI Press.

Deurex, J. (1945). Chorée chronique et paralysie hypertonique du regard. *Revue Neurologique, 77,* 207–208.

Dickov, L. A., & Morrison, J. D. (2006). Effects of uncertainty and target displacement on the latency of express saccades in man. *Vision Research, 46,* 2505–2512.

Filippopulos, F., Eggert, T., & Straube, A. (2013). Deficits of cortical oculomotor mechanisms in cerebellar atrophy patients. *Experimental Brain Research, 224,* 541–550. doi:10.1007/s00221-012-3332-0

Galvan, L., André, V. M., Wang, E. A., Cepeda, C., & Levine, M. S. (2012). Functional differences between direct and indirect striatal output pathways in Huntington's disease. *Journal of Huntington's Disease, 1,* 17–25.

Golding, C. V. P., Danchaivijitr, C., Hodgson, T. L., Tabrizi, S. J., & Kennard, C. (2006). Identification of an oculomotor biomarker of preclinical Huntington disease. *Neurology, 67,* 485–487.

Grabska, N., Rudzinska, M., Wójcik-Pedziwiatr, M., Michalski, M., Slawekc, J., & Szczudlik, A. (2014). Saccadic eye movements in juvenile variant of Huntington disease. *Polish Journal of Neurology and Neurosurgery, 48*, 236–241.

Hikosaka, O., Takikawa, Y., & Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological Reviews, 80*, 953–978.

Hikosaka, O., & Wurtz, R. H. (1985a). Modification of saccadic eye movements by GABA-related substances: I. Effect of muscimol and bicuculline in monkey superior colliculus. *Journal of Neurophysiology, 53*, 266–291.

Hikosaka, O., & Wurtz, R. H. (1985b). Modification of saccadic eye movements by GABA-related substances: II. Effect of muscimol in monkey substantia nigra pars reticulata. *Journal of Neurophysiology, 53*, 292–308.

Huntington Study Group. (1996). Unified Huntington's Disease Rating Scale: Reliability and consistency. *Movement Disorders, 11*, 136–142.

Jin, Z.-L., Cheng, S.-J., Zhou, M., & Li, L. (2012, December). *Eye movements reveal process of feature integration* Paper presented at the International Conference on Wavelet Active Media Technology and Information Processing, Chengdu, China.

Joel, D. (2001). Open interconnected model of basal ganglia–thalamocortical circuitry and its relevance to the clinical syndrome of Huntington's Disease. *Movement Disorders, 16*, 407–423.

Klein, C., & Fischer, B. (2005). Instrumental and test–retest reliability of saccadic measures. *Biological Psychology, 68*, 201–213.

Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein, E. (2011). Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of Neuroscience Methods, 201*, 196–203.

Langbehn, D. R., Brinkman, R. R., Falush, D., Paulsen, J. S., & Hayden, M. R. (2004). A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clinical Genetics*, 65, 267–277.

La Spada, A. R., Weydt, P., & Pineda, V. V. (2011). Huntington's disease pathogenesis: Mechanisms and pathways. In D. C. Lo & R. E. Hughes (Eds.), *Neurobiology of Huntington's disease: Applications to drug discovery* (pp. 29–54). Boca Raton, FL: CRC Press.

Leigh, R. J., & Kennard, C. (2004). Using saccades as a research tool in the clinical neurosciences. *Brain, 127*, 460–477.

Patel, S. S., Jankovic, J., Hood, A. J., Jeter, C. B., & Sereno, A. B. (2012). Reflexive and volitional saccades: Biomarkers of Huntington disease severity and progression. *Journal of the Neurological Sciences, 313*, 35–41.

Paulsen, J. S., Long, J. D., Ross, C. A., Harrington, D. L., Erwin, C. J., Williams, J. K., … PREDICT-HD Investigators and Coordinators of the Huntington Study Group. (2014). Prediction of manifest Huntington's disease with clinical and imaging measures: A prospective observational study. *Lancet Neurology, 13*, 1193–1201.

Peltsch, A., Hoffman, A., Armstrong, I., Pari, G., & Munoz, D. P. (2008). Saccadic impairments in Huntington's disease. *Experimental Brain Research, 186*, 457–469. doi:10.1007/s00221-007-1248-x

Phillips, W., Shannon, K. M., & Barker, R. A. (2008). The current clinical management of Huntington's disease. *Movement Disorders, 23*, 1491–1504.

Pringsheim, T., Wiltshire, K., Day, L., Dykeman, J., Steeves, T., & Jette, N. (2012). The incidence and prevalence of Huntington's disease: A systematic review and meta-analysis. *Movement Disorders, 27*, 1083–1091.

Rivera, S., Best, C. A. , Yim, H., Martinez, A. M., Sloutsky, V. M., Walther, D. B. (2012). Automatic selection of eye tracking variables in visual categorization for adults and infants. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Building bridges across cognitive sciences around the world: Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 2240–2245). Austin, TX: Cognitive Science Society.

Rizk-Jackson, A., Stoffers, D., Sheldon, S., Kuperman, J., Dale, A., Goldstein, J., … Aron, A. R. (2011). Evaluating imaging biomarkers for neurodegeneration in pre-symptomatic Huntington's disease using machine learning techniques. *NeuroImage, 56*, 788–796.

Robert, M. P. A., Nachev, P. C., Hicks, S. L., Golding, C. V. P., Tabrizi, S. J., & Kennard, C. (2009). Saccadometry of conditional rules in presymptomatic Huntington's disease. *Annals of the New York Academy of Sciences, 1164*, 444–450.

Ross, C. A., Aylward, E. H., Wild, E. J., Langbehn, D. R., Long, J. D., Warner, J. H., … Tabrizi, S. J. (2014). Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nature Reviews Neurology, 10*, 204–216.

Rupp, J., Dzemidzic, M., Blekher, T., West, J., Hui, S., Wojcieszek, J., … Foroud, T. (2012). Comparison of vertical and horizontal saccade measures and their relation to gray matter changes in premanifest and manifest Huntington disease. *Journal of Neurology, 259*, 267–276.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management, 45*, 427–437.

Tabrizi, S. J., Langbehn, D. R., Leavitt, B. R., Roos, R. A. C., Durr, A., Craufurd, D., … TRACK-HD investigators. (2009). Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: Cross-sectional analysis of baseline data. *Lancet Neurology, 8*, 791–801. doi:10.1016/S1474-4422(09)70170-X

Tseng, P.-H., Cameron, I. G. M., Pari, G., Reynolds, J. N., Munoz, D. P., & Itti, L. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *Journal of Neurology, 260*, 275–284.

Xiang, J., Yan, H., & Chen, J. (2009, November). *The study of decoding mental state based eye tracks using SVM*. Paper presented at the International Workshop on Information Security and Application, Qingdao, China.

Zhu, Z., & Ji, Q. (2005). Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision and Image Understanding, 98*, 124–154.