Combining Results of Different Oculometric Tests Improved Prediction of Parkinson's Disease Development

Albert Śledzianowski[1(✉)], Artur Szymanski[1], Aldona Drabik[1], Stanisław Szlufik[2], Dariusz M. Koziorowski[2], and Andrzej W.Przybyszewski[1]

[1] Polish-Japanese Academy of Information Technology, 02-008 Warsaw, Poland
albert.sledzianowski@gmail.com

[2] Neurology, Faculty of Health Science, Medical University of Warsaw, 03-242 Warsaw, Poland

Abstract. In this text we compare the measurement results of reflexive saccades and antisaccades of patients with Parkinson's Disease (PD), trying to determine the best settings to predict the Unified Parkinson's Disease Rating Scale (UPDRS) results. After Alzheimer's disease, PD statistically is the second one and until today, no effective therapy has been found. Luckily, PD develops very slowly and early detection can be very important in slowing its progression. In this experiment we examined the reflective saccades (RS) and antisaccades (AS) of 11 PD patients who performed eye-tracking tests in controlled conditions. We correlated neurological measurements of patient's abilities described by the Unified Parkinson's Disease Rating Scale (UPDRS) scale with parameters of RS and AS. We used tools implemented in the Scikit-Learn for data preprocessing and predictions of the UPDRS scoring groups [1]. By experimenting with different datasets we achieved best results by combining means of RS and AS parameters into computed attributes. We also showed, that the accuracy of the prediction increases with the number of such derived attributes. We achieved 89% accuracy of predictions and showed that computed attributes have 50% higher results in the feature importance scoring than source parameters. The eye-tracking tests described in this text are relatively easy to carry out and could support the PD diagnosis.

Keywords: Parkinson's Disease · Reflexive saccades · Antisaccades · Eye tracking · Data mining · Machine learning · UPDRS

1 Introduction

The eyes make projectile movements called saccades when watching surroundings and fixing on important elements. The reason of this behaviour is in the specifics of the fovea which is the part of retina responsible for extracting details from the available view. The fovea is able to receive information only from a small field,

N. T. Nguyen et al. (Eds.): ACIIDS 2020, LNAI 12034, pp. 517–526, 2020.
https://doi.org/10.1007/978-3-030-42058-1_43

so eyes must be often moved from between different locations to scan the visible area [5]. Saccades occur both involuntarily and intentionally and are varied in terms of range, as we make small saccades when we read and longer when we look around ourselves. Saccades are common in most of our daily activities and its mechanism seems to be complex as many brain areas take part in the analysis of visual information. Therefore saccades seem to be sensitive for neurodegenerative changes and it is sensible perform tests based on this kind of eve move during diagnostic connected to diseases like the PD. Several types of tests to study saccades have been developed, usually consisting calculation of the parameters while subject moves his eyes from the fixation point to the peripheral target. One of them is a test for the RS, based on simply move from the fixation point towards the stimuli and the AS test, based on the reverse action. The AS is an eye move in the opposite direction of appearing stimulus. The idea is in suppressing reflexive transfer of the focus to the emerging target and forcing eyes to look into to the mirror location of the target. As the reflexive eye movement have to be inhibited, this test is generally more difficult and usually takes more time than the RS [8]. In terms of the PD disease, various studies have shown that patients have impaired executive function, including deficits in attention, movement initiation, motor planning and decision making [10]. It usually leads to impairments in control of suppression of involuntary behavior, as variety of neurological diseases result in dysfunctions and errors in this mechanism [13]. It was found that the degree of advancement of the PD significantly increases mean latency and error rate in the AS tasks and significantly decreases the velocity [9,11].

In this experiment we used results of both RS and AS tests of PD patients, combining them with their results of neurological tests expressed in the UPDRS, which is most common rating scale of the PD progression. Th UPDRS consists clinician-scored and monitored personal behavior, mood, mental activity, activities of daily living, motor evaluation and the evaluation of complications during the therapy [7]. Unfortunately, most of the patients diverge in their combinations of symptoms, which leads to constant need of development in methods of evaluation of the disease progression.

In this experiment we researched different datasets containing results of RS and AS of PD and tried to find the best features system to predict clinically-scored attribute, the UPDRS Total. The UPDRS Total represents all sections of the UPDRS evaluation, therefore it is considered as a good generalization of the PD progression. We are investigating methods of predictions that could extend information about patients, could be automated and be available for everyone on personal devices like PC or Smart-Phone.

2 Methods of the Experiment

We performed tests with 11 patients in one of neurosurgical clinic. Patients differed in terms of the disease treatment. Also our data distinguished patients who were treated with pharmaceuticals (BMT - Best Medical Treatment) and patients who underwent an electrode implantation. (DBS - Deep Brain Stimulation). Implanted electrode stimulates patient's Subthalamic Nucleus (STN), an area of the brain that has a strong effect on the dopaminergic system, damaged as a result of the PD. Patients qualified for the neurosurgery are mostly characterized by the low sensitivity to the stabilizing effects of the L-Dopa [14]. Variants of patients session are presented in the list below:

- Sesion type 0 : BMT off, DBS off
- Sesion type 1 : BMT off, DBS on
- Sesion type 2 : BMT on, DBS off
- Sesion type 3 : BMT on, DBS on

We performed test with a head-mounted eye-tracker, the JAZZ-Novo. This device works in frequency of 1000 Hz, thus provides very high spatial and temporal resolution needed to carry out this kind of experiments. The location of the sensor on the subject's forehead allowed for compensation of a head and body movements, which is very important during eye-tracking measurements of the PD patients because of involuntary quivering movements (the tremor). Experimental attempts consisted of tracking the light marker, moving in horizontal directions behind patient's eyes. Both the RS and the AS attempts started from fixating in the initial position (0°). When fixation point disappeared at the same moment targets of the RS (green eclipses) or the AS (red ellipses) revealed to the patient 10° to the left or right. Patient task was to move eyesight on the appearing target (RS) or in the opposite direction (AS) with the highest precision and shortest latency. Next attempt started after 100 ms without any break ("no-gap" model introduced by Saslow) [15]. Schema on Fig. 1 presents both models of the RS and the AS trials.

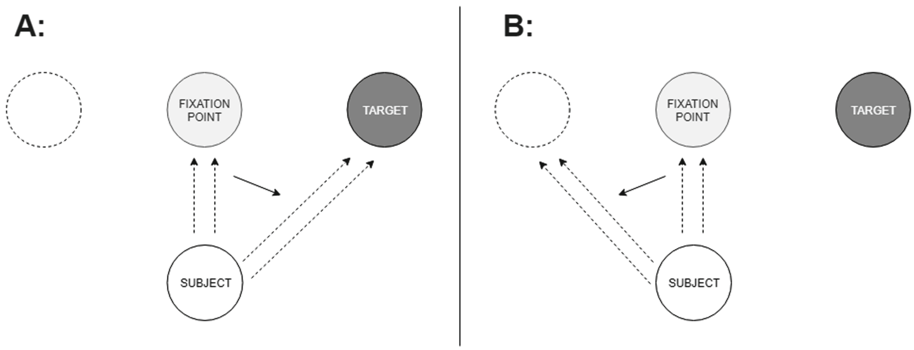

Fig. 1. The models of oculometric tests: A: The reflexive saccade, B: The antisaccade.

All tests were conducted in the same room with the same lighting. From the eye-tracking data (time series of eye positions) we algorithmically calculated different parameters of the RS and the AS - mean delay, mean duration and maximum speed. We used windows as a search space of the eye moves with the beginning and the end designated by appearance and disappearance of the peripheral target. Depending on the test type, algorithm expected straight move between fixation point and the target (RS) or to the opposite direction (AS) - both below the latency of 500 ms. All records not passing the requirements of move direction and latency, including missed records (blinks, head movements, etc.) have not been qualified for further analysis.

3 Computational Basis

As mentioned before, the subject of the calculation from the eye tracker data were mean latency, mean duration and max speed (because we think that max speed is better for overviewing the eye performance than the average speed). We calculated the latency as a difference in time between showing the target and starting the eye movement and the beginning of a duration from the starting point when eyes began to follow the target and when simultaneously the eye speed started to rise. The end of the duration was determined by the eye movement inhibition. The means of those two parameters were calculated arithmetically for each patient. The max speed was counted as the maximum value determined in the period of the eye move duration.

After performing oculometric tests and calculations we created the dataset from parameters of the RS and AS data performed in different sessions. As previously mentioned, the neurological data was expressed in the attribute UPDRS TOTAL. The dataset contained 52 records (26 RS/26 AS) collected from 11 patients. Patients differed between methods of the treatment, so it was not possible to collect data of the same patient in each of the session type, this why.

The oculometric parameters were calculated for each eye separately. The dataset contained additional features: standard deviations of the mean parameters (latency and duration) and attributes representing the number of frames on the basis of which given oculometric parameter was calculated (for particular record). The Table 1 presents the examples of rows of the initial dataset.

Table 1. The example of dataset before preprocessing

id	updrs	sess	type	mtre	stre	mspre	mrdur	stdredur	delay_r_cn	ms_r_cnt	dur_r_cnt
1	55	0	RS	0.265	0.043	5.068	1.951	1.224	2	8	4
1	55	0	AS	0.296	0.055	5.427	4.372	0.785	5	30	15
2	52	1	RS	0.422	0.073	5.240	4.014	1.703	3	17	8
2	52	1	AS	0.376	0.071	5.641	2.965	1.760	7	19	9

Where columns: "id" is a patient number, "sess", "type" describes types of the session and test, "mtre", "stre" - mean and standard variance for latency, "mrdur", "stdredur" - mean and standard variance for duration, mspre, max speed and delay_r_cn, ms_r_cnt dur_r_cnt are frame counters corresponding to the parameters. Presented eye move parameters were calculated separately for each eye

In the next phase, the dataset has been preprocessed by several methods implemented in the scikit-learn [1]. First we obtained matrix ("X") containing all the feature values of all instances in the dataset, excluding the target attribute UPDRS. The target attribute data ("y") was discretized using pandas "cut" method into 5 bins (classes) [2]. Then in the matrix "X" we aligned the scales of attribute values using scikit-learn's "Standard Scaler" which standardizes features by removing the mean and scaling to unit variance [1]. As discretized "y" class bins differed in number of records (from 4 to 17 items), we used Synthetic Minority Over-sampling Technique (SMOTE) to oversample the dataset with number of neighbours depending on number of items in the smallest group [1]. After the oversampling the "X" contained an equal (17) number of records for each of the five classes of discretized UPDRS. The number of records has increased by 38% from 52 to 85. Next, because we wanted to compare different combinations of dataset in terms of predictions accuracy, we generated 2 additional matrices separating the RS records from the AS. We also created 2 copies of initial "X" matrix in order to add new features. We wanted to check the accuracy of prediction when we average the results of particular patient - regardless the session and the test type. In order to do that, we calculated and combined means for particular patients on the basis of the eye move parameters and added it to first copy of the matrix in number of 4, and to the second copy in number of 8 new features. We also unified eye move parameters for both eyes in those two datasets. Table 2 presents the differences in the features between different datasets. In the next stage we split each of the 5 versions of the dataset into training and test/validation sets using Stratified Shuffle Split (SSS). The SSS returns stratified and randomized folds with preserved percentage of samples for each class. In stratified sampling the data is divided into homogeneous subgroups ("stratas") and the right number of instances is sampled from each stratum to guarantee that the test and training sets are representative [1].

Table 2. The comparison of the features for different datasets

RS	AS	RS + AS	RS + AS + 4	RS + AS + 8	Features
N	N	Y	Y	Y	Contains oculometric test type parameter
Y	Y	Y	N	N	Separated parameters for left/right eye
N	N	N	Y	Y	Unified parameters for both eyes
N	N	N	Y	Y	Calculated attributes from source parameters

Where RS is the dataset containing only saccade and AS only antisaccade data, RS + AS, +4/+8 contain both types of test and respectively 4 and 8 new attributes computed from means of parameters for the patient, regardless oculometric test type and session.

4 Results

With prepared 5 versions of datasets we started supervised learning using different classifiers implemented in the scikit-learn, as we decided to check predictions using different methods:

– Nearest Neighbors
– Support Vector Machine
– Decision Tree
– Random Forest
– Multi-Layer Perceptron
– Naive Bayes
– Quadratic Discriminant Analysis
– Gaussian Process

For each of listed classifiers we used the GridSearchCV (GSCV), the feature implemented in scikit-learn for searching possible best values of hyper-parameters. The GSCV performs exhaustive search over parameters values optimized by the cross-validation returning various types of scoring and also best values of parameters [1]. For ranking different input variants we used GSCV "best score" which is a calculation of cross-validated mean score of the best estimation achieved for particular classifier [1]. Table 3 shows score ranking for different datasets and classifiers.

Table 3. The prediction score ranking for different datasets

Dataset type	Best classifier	Score
Only RS	QDA	0.68
Only AS	QDA	0.75
RS + AS	Decision tree	0.61
RS + AS + 4 combined features	Decision tree	0.77
RS + AS + 8 combined features	Decision tree	0.89

As we can see in Fig. 3 the Quadratic Discriminant Analysis (QDA) performed the highest accuracy predictions for smaller and simpler datasets containing only the RS or only the AS. Discriminant analysis is research method where the criterion (or the dependent variable) is divided into categories and the predictor (or the independent variable) is an interval in nature. The quadratic decision boundary determining the division criterion is set by Bayes rule which fits conditional densities of a class to the data. The QDA have no assumption that the covariance of each of the classes is identical, also doesn't require any hyper-parameter tuning. For larger datasets consisting both the RS and the AS, the best predictions were obtained by the Decision Tree. As can bee seen in the Table 3, its accuracy rises linearly, dependently on number of added features computed on basis of the oculometric parameters (described in the previous section). Figure 2 is showing the confusion matrix of this dataset, obtained with the Decision Tree classifier [1].

We also examined feature importance between different datasets. Figure 3 shows 5 most important features for the Decision Tree (the best ranked classifier)

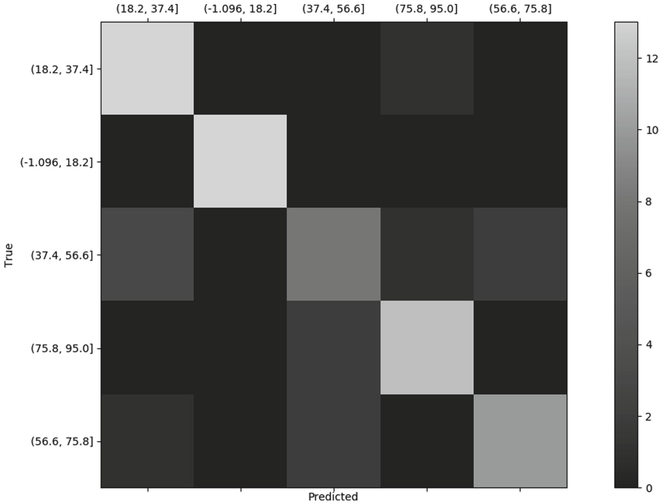

Fig. 2. The confusion matrix of the total UPDRS.

in 3 different variants of large dataset (containing both the RS and the AS data). As can be seen on the charts, computed features occupy a high places in the ranking of top 5 important features and their position and number increases with the number of such parameters added into the dataset.

5 Discussion and Conclusions

In the results we can see that computed attributes specifying joined results of the patient are more sensitive in predicting scoring group of the total UPDRS, than standard parameters (Fig. 3). What can be noticed, with the number of computed attributes added, the accuracy of the prediction increases, as well as the importance of those attributes. Attributes "redur_join" being the average of the patient's results in terms of eye move duration and "stre_join" - the average of standard deviations of patient's latency results, played the most important role in the predictions in terms of the most accurate results.

Also importance scoring for those attributes seems to be much higher than for standard ones. Attribute "sess" describing type of the session, previously the most important decision attribute, in the final shape of the dataset plays only the third role in the importance ranking, with a 50% worse scoring result than the best computed attribute. The results obtained using final dataset suggests that we can increase the sensitivity of the UPDRS groups prediction by increasing the dataset with combined attributes, representing the averages of oculometric parameters from different tests. We think that it can be hypothesized that combined attributes create an union between same features from different tests which helps classifiers in prediction and creates bridge between those separated

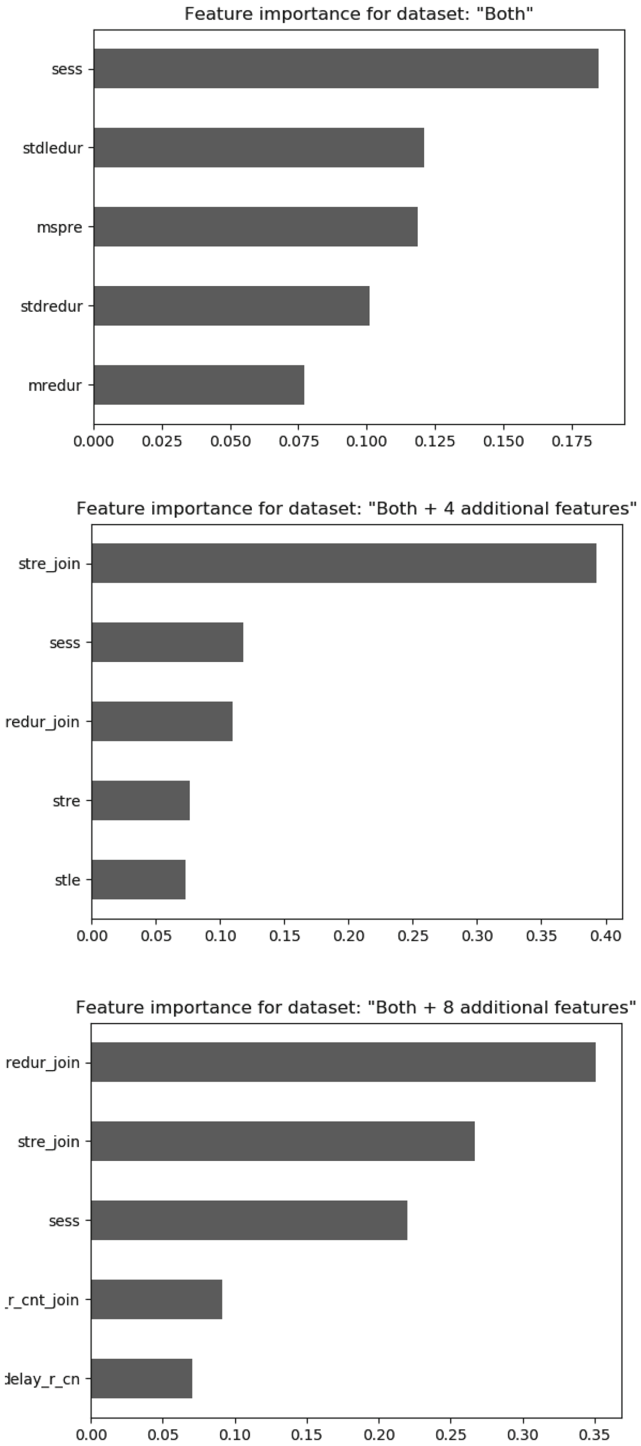

Fig. 3. The decision tree feature importance comparison for different datasets.

results of a patient. Additional attributes have also extended the size of the dataset and larger sizes of samples can improve the accuracy [16].

We found presented results as very indicating with so small group of records and the fact that we were able to increase prediction accuracy for our dataset over 21%. We think that the approach presented in this article can improve the quality of the UPDRS predictions based on eye movement testing. We also think that the eye-move performance tests are good indicators of the motor skills of the PD patients, revealing the scales of disease progression. Our results also proves the sense of further investigations of correlations between parameters of different eye movements tests, computed attributes and the UPDRS evaluations. We hope that development of methods like the one described in this text may help in improving the PD diagnosis. We believe that in the upcoming future, oculometric tests combined with the machine-learning pipelines could improve detection and progression evaluation of the neurodegenerative diseases. If such tests could be available on personal devices, patients could perform tests under different conditions providing more information about the disease.

Ethic Statement. This study was carried out in accordance with the recommendations of Bioethics Committee of Warsaw Medical University with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Bioethics Committee of Warsaw Medical University.

References

1. Pedregosa, F.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
2. McKinney, W., et al.: Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, vol. 445, pp. 51–56 (2010)
3. Hallett, P.E., Adams, B.D.: The predictability of saccadic latency in a novel voluntary oculomotor task. Vision. Res. **20**(4), 329–339 (1980)
4. Purves, D., Augustine, G.J., Fitzpatrick, D., et al. (eds.): Types of Eye Movements and Their Functions. Neuroscience, 2nd edn. Sinauer Associates Inc, Sunderland (2001)
5. Pratt, J., Abrams, R., Chasteen, A.: Initiation and inhibition of saccadic eye movements in younger and older adults an analysis of the gap effect. J. Gerontol.: Psychol. Sci. **52B**(2), 1O3–1O7 (1997)
6. Robinson, D.A.: Control of eye movements. In: Handbook of Physiology, Section I: The Nervous System. American Physiological Society, Bethesda (1981)
7. Bhidayasiri, R., Martinez-Martin, P.: Clinical assessments in Parkinson's disease: scales and monitoring. In: International Review of Neurobiology, vol. 132, pp. 129–182 (2017)
8. Olk, B., Kingstone, A.: Why are antisaccades slower than prosaccades? A novel finding using a new paradigm. Cogn. Neurosci. Neuropsychol. **14**(1), 151–155 (2003)
9. Kitagawa, M., Fukushima, J., Tashiro, K.: Relationship between antisaccades and the clinical symptoms in Parkinson's disease. Neurology **44**(12), 2285 (1994)

10. Hood, A.J., Amador, S.C., Cain, A.E., et al.: Levodopa slows prosaccades and improves antisaccades, an eye movement study in Parkinson's disease. J. Neurol. Neurosurg. Psychiatry **78**, 565–570 (2007)
11. Everling, S., DeSouza, J.F.X.: Rule-dependent activity for prosaccades and antisaccades in the primate prefrontal cortex. J. Cogn. Neurosci. **17**(9), 1483–1496 (2005)
12. Nakamura, T., Funakubo, T., Kanayama, R., Aoyagi, M.: Reflex saccades, remembered saccades and antisaccades in Parkinson's disease. J. Vestib. Res. **4**(6), S32 (1996). (Suppl. 1)
13. Everling, S., Fischer, B.: The antisaccade: a review of basic research and clinical studies. Neuropsychologia **36**(9), 885–899 (1998)
14. Przybyszewski, A., Szlufik, S., Szymanski, A., Habela, P., Koziorowski, D.: Multimodal learning and intelligent prediction of symptom development in individual Parkinson's patients. Sensors **16**(9), 1498 (2016)
15. Saslow, M.G.: Effects of components of displacement-step stimuli upon latency for saccadic eye movement. J. Opt. Soc. Am. **57**(8), 1024–1029 (1967)
16. Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C.: Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. NeuroImage **60**(1), 59–70 (2012)