Fundamenta Informaticae 176 (2020) 167–181
DOI 10.3233/FI-2020-1969
IOS Press

Comparison of Different Data Mining Methods to Determine Disease Progression in Dissimilar Groups of Parkinson's Patients

Andrzej W. Przybyszewski[*†]
Polish–Japanese Academy of Information Technology
00-097 Warsaw, Poland
przy@pjwstk.edu.pl

Stanislaw Szlufik
Medical University of Warsaw
03-242 Warsaw, Poland
stanislaw.szlufik@gmail.com

Dariusz M. Koziorowski
Medical University of Warsaw
03-242 Warsaw, Poland
dkoziorowski@esculap.pl

Artur Chudzik
Polish–Japanese Academy of Information Technology
00-097 Warsaw, Poland
artur.chudzik@pjwstk.edu.pl

Piotr Habela
Polish–Japanese Academy of Information Technology
00-097 Warsaw, Poland
piotr.habela@pjwstk.edu.pl

Abstract. Parkinson's disease (PD) is the second after Alzheimer's most popular neurodegenerative disease (ND). Cures for both NDs are currently unavailable. *OBJECTIVE:* The purpose of our study was to predict the results of different PD patients' treatments in order to find an optimal one. *METHODS:* We have compared rough sets (RS) and others, in short, machine learning (ML) models to describe and predict disease progression expressed as UPDRS values (Unified Parkinson's Disease Rating Scale) in three groups of Parkinson's patients: 23 BMT (Best Medical Treatment) patients on medication; 24 DBS patients on medication and on DBS therapy (Deep Brain Stimulation) after surgery performed during our study; and 15 POP (Postoperative) patients who had had surgery earlier (before the beginning of our research). Every PD patient had three visits approximately every six months. The first visit for DBS patients was before surgery. On the basis of the following condition attributes: disease duration, saccadic eye movement parameters, and neuropsychological tests: PDQ39 (Parkinson's Disease Questionnaire - disease-specific health-related quality-of-life questionnaire), and Epworth Sleepiness Scale tests we have estimated UPDRS changes (as the decision attribute). *RESULTS:* By means of RS rules obtained

[*]Also affiliated at: Department of Neurology, UMass Medical School, Worcester, MA 02135, USA.
[†]Address for correspondence: PJAIT, Koszykowa 86, 00-097 Warsaw, Poland.

for the first visit of BMT/DBS/POP patients, we have predicted UPDRS values in the following year (two visits) with global accuracy of 70% for both BMT visits; 56% for DBS, and 67%, 79% for POP second and third visits. The accuracy obtained by ML models was generally in the same range, but it was calculated separately for different sessions (MedOFF/MedON). We have used RS rules obtained in BMT patients to predict UPDRS of DBS patients; for the first session DBSW1: global accuracy was 64%, for the second DBSW2: 85% and the third DBSW3: 74% but only for DBS patients during stimulation-ON. ML models gave better accuracy for DBSW1/W2 session S1(MedOFF): 88%, but inferior results for session S3 (MedON): 58% and 54%. Both RS and ML could not predict UPDRS in DBS patients during stimulation-OFF visits because of differences in UPDRS. By using RS rules from BMT or DBS patients we could not predict UPDRS of POP group, but with certain limitations (only for MedON), we derived such predictions for the POP group from results of DBS patients by using ML models (60%). *SIGNIFICANCE:* Thanks to our RS and ML methods, we were able to predict Parkinson's disease (PD) progression in dissimilar groups of patients with different treatments. It might lead, in the future, to the discovery of universal rules of PD progression and optimise the treatment.

Keywords: Neurodegenerative disease, rough set, decision rules, granularity.

1. Introduction

This publication is an extension of the original communication given at PReMI 2017 and published in LNCS [1]. Only very experienced PD neurologists are successful in implementing individually adjusted therapy. In general doctors have very limited time for each patient and they have different academic backgrounds which may lead to introduction of variable treatments of patients and it may lead to confusions and ineffective therapy. We propose to improve doctor's approach by adding more automatic measurements and intelligence symptom classification [2, 3] that is similar to that was found in the visual system for the complex objects recognition [4].

It is important to estimate patient's disease stage as it determines different sets of therapies. Typical neurological standards are based on the Hoehn and Yahr and the UPDRS (Unified Parkinson's Disease Rating) scales. The last one is more precise and it will be mainly used in this study. We would like to estimate disease progression in different groups of patients that were tested during three visits every half-year.

This study is using several different AI methods in order to predict disease progression of Parkinson's patients with different treatments and in different disease stages. The ultimate purpose of our distinctive data mining methods is to find direction(s) to the optimal therapy for different patients with Parkinson's disease (PD). It is not easy as 'each PD is different' as effect of many compensatory individual processes in response to dying brain cells.

Our method may lead to the introduction of a more precise follow-up and even can be extended to telemedicine by using results of neuropsychological tests with parameters of the reflexive eye movements in order to predict UPDRS (disease progression) [1-3]. The primary Parkinson's disease symptoms are motor tested by neurologists as UPDRS III (Unified Parkinson's Disease Rating Scale, motor part III). Generally, it is difficult to measure them in a doctor-independent, automatic way. However, measurements of the gait as well as eye movements (EM) can be automatised [5]. Our fast (saccadic) EM tests are natural. When we notice a new object that appears in our visual field, we automatically,

with a delay, look at it (perform a saccade) or sometimes look in the opposite direction (perform an antisaccade). Parameters of these fast EM are important PD biomarkers. Our antagonistic oculomotor actions are mainly suppressed by signals that substantia nigra (SN) sends to superior colliculus [6]. As SN is affected by Parkinson's disease and in less extent by ageing, the reflexive saccades delay is an important parameters related to the PD progression as we have demonstrated previously [2, 3]. The new aspect of this study is related to an extension of our previous results for patients with similar treatments [2, 3]. We would like to predict, using data mining methods, disease progression of patients with different therapies: on only medication, with short term DBS (deep brain stimulation), or more advanced patients with long term DBS. In the original communication [1], we have used only rough set theory methods and our predictions were not possible for all groups of patients. In this work, we have also used other ML models that helped to improve our estimates. Contents of each section are the following: the Methods section has two subsections: Rough Set and Machine Learning; the Results section has the following subsections: Comparing longitudinal UPDRS changes, BMT patients disease progression prediction, DBS and POP patients: rules for estimation of disease progression, Disease progression of DBS patients estimated by rules from BMT patients, Disease progression of POP patients estimated by rules from DBS patients; the following sections: Discussion and Conclusions have no subsections.

2. Methods

All 62 PD patients were divided into three groups: BMT patients (medication only), and patients on medication and with implanted electrodes in the STN (subthalamic nucleus [3]) during our study: DBS group or before our study: POP group.

The Deep Brain Stimulation (DBS) surgery was performed in the Institute of Neurology and Psychiatry, Warsaw Medical University. PD patients were tested in the following sessions: MedON/ MedOFF sessions (sessions with or without medication). The other groups: DBS and POP patients were also tested in StimON/StimOFF session were DBS stimulation was switched ON or OFF. All combinations gave four sessions: 1) MedOFFStimOFF; 2) MedOFFStimON; 3) MedONStimOFF; 4) MedONStimON. Details of these procedures were described earlier [3]. Neurologists from WMU performed the UPDRS and neuropsychological tests. Fast eye movements - reflexive saccades were recorded as described in detail before [2, 3]. The following parameters of saccades were measured: the delay (latency) related to time difference between the beginning of the light spot movements and the beginning of the eye movement; saccade's amplitude in comparison to the light spot amplitude; max velocity of the eye movement; duration of saccade defined as the time from the beginning to the end of the saccade.

2.1. Rough set

Our data mining analysis follows rough set (RS) theory after Zdzislaw Pawlak [7] because RS gave previously the best results in PD symptoms classifications in comparison to other methodologies [3]. Our data are represent as a decision table where rows represented different measurements (may be obtained from the same or different patients) and columns are related to different attributes. An

information system [7] is as a pair *S = (U, A)*, where *U, A* are finite sets: *U* is the universe of objects; and *A* is the set of attributes. The value *a(u)* constitutes a unique element of *V* (where *V* is a value set) for $a \in A$ and $u \in U$.

A decision table for S is the triplet: *S = (U, C, D)* where: *C, D* are condition and decision attributes [8]. Each row of the information table gives a particular rule that connects condition and decision attributes for a single measurement of a particular patient. As there are many rows related to different patients and sessions, they gave many particular rules. Rough set approach allows generalising these rules into universal hypotheses that may determine optimal treatment options for an individual PD patient. Different rules' granularities (abstraction) are similar to complex objects recognition [4] and may simulate association processes of the "Golden Neurologist".
In the present study, we are trying to use data from different groups of patients for training and testing. The purpose was to find limits of rules that may predict symptoms development of patients with different treatments at different disease stages.
We have used the RSES 2.2 (Rough System Exploration Program) [9] with implementation of RS rules to process our data. By means of RSES we have generated rules using four different methods: exhaustive algorithm, genetic algorithm, covering algorithm, or LEM2 algorithm. In each case, we selected a particular algorithm that gave the shortest set of rules.

In rough set approach our discretisation was based on the cut values. We have replaced original attributes with new, binary attributes which indicate whether actual attribute value for an object is greater or lower than c (see [8]), we define c as a cut. This algorithm is implemented in RSES 2.2.

In RSES classification we have used 6-fold method, which means that the population was divided into 6 random subgroups and 5 were used for training and one for testing six times by changing in every trial-testing group. Global accuracy and coverage were the means of accuracy and coverage from all individual tests.

We have used standard RSES parameters: type of classifier: Decision rules (with 6 folds as described above), exhaustive algorithm of rules computation with shortening ratio: 1.0, and conflicts resolved by standard voting. These parameters gave the best global accuracy.

2.2. Machine learning

For the evaluation of the results acquired by RSES we have also used additional methods of a supervised learning process. Because our data contains a set of N training samples of the form (x_1, y_1), ..., (x_n, y_n) such that x_i is the feature vector of the i_{th} sample and y_i is its class, it is possible to use a supervised learning algorithm which seeks for a function $g : X \rightarrow Y$, where the X is the input space, and the Y is the output space. The g function is an element of some space of possible functions G, known as the hypothesis space. There is a slight difference in this approach in the comparison to RSES because supervised learning does not allow us to create the "none of the above" labels per se. In the circumstances of dealing with medical data, it is often precisely the class that is under-represented in the data, the disease or potential fault, that the network should detect [10], and it would be introduced in future research. Furthermore, we cannot generalise rules into universal hypotheses that may

determine optimal treatment options for an individual PD patient, so they had been prepared for each group individually.

We have prepared an automated framework, written in Python and based on the scikit-learn [11], which was responsible for the evaluation of different supervised learning algorithms and provided the details of the result of the most accurate algorithm for a given dataset. The challenging task was to create decision classes, which were comparable with RSES. First of all, some values were missing, so we applied a strategy of imputation using the statistics (specifically: mean values). Then, to achieve similarity with the number of classes produced by RSES, we have applied unsupervised learning for clustering of the training set to obtain normalised classes (considered as different UPDRS ranges). We have observed that different strategies provide better results when considering different subsets of patients. Thus, we used two popular clustering approaches. For the DBS and POP patients' the best results we have achieved by K-Means [12] parametrised with a number of clusters = 4 and with a 'k-means++' method for initialisation including 10 runs and 300 maximum iterations. For BMT patients, better (closer to results of RSES for the sake of comparison) clustering was achieved by the Mean Shift [13] with an estimated value of bandwidth used in the RBF kernel based on standard quantile = 0.30.

When training data and labels were prepared for each set, we were able to apply a wide range of supervised algorithms and choose the best one on the basis of the highest accuracy score (in scikit-learn estimators have a 'score' method providing a default evaluation criterion for the problem they are designed to solve). The framework challenged every two pairs of the training set (e.g. POPW1) against validation dataset (e.g. BMTW1) and attempted to predict the UPDRS class of every patient. All scikit-learn classifiers are capable of multiclass classification, so the selection of presented algorithms has been based on the strategy. We have chosen well-known models that apply: inherently multiclass behavior (e.g. GradBoost), multiclass as One-Vs-One (e.g. SVC), multiclass as One-Vs-All (e.g. logit), multilabel support (e.g. Decision Tree), and with multiclass-multioutput support (e.g. Random Forest). Among others, we have evaluated the algorithms listed below; the selection was made both empirically and based on previous research [3].

- Logistic Regression (logit) - applied with 'multinomial' option with the 'lbfgs' solver and 'balanced' weights associated with classes,
- Linear Supported Vector Classification (LinearSVC) - with a 'crammer_singer' multi-class strategy;
- Gradient Boosting Classifier (GradBootst) - chosen loss function was 'deviance' with learning rate = 0.1;
- Gaussian Naive Bayes (GaussianNB) - used without prior probabilities of the classes;
- Decision Tree - with Gini impurity metric;
- K-Neighbors Classifier (KNC) - with a number of neighbours to get = 5 and uniform weights (all points in each neighbourhood were weighted equally);
- Random Forest - applied with 10 trees in the forest and Gini impurity metric;
- C-Support Vector Classification (SVC) - with the 'rbf' kernel type, 'ovo' (one-versus-one) multi-class strategy, penalty parameter $C = 1.0$ and $\gamma = 0.001$.

As the Results section shows, there was no most accurate model and most of them were found to be optimal only for a specific subset.

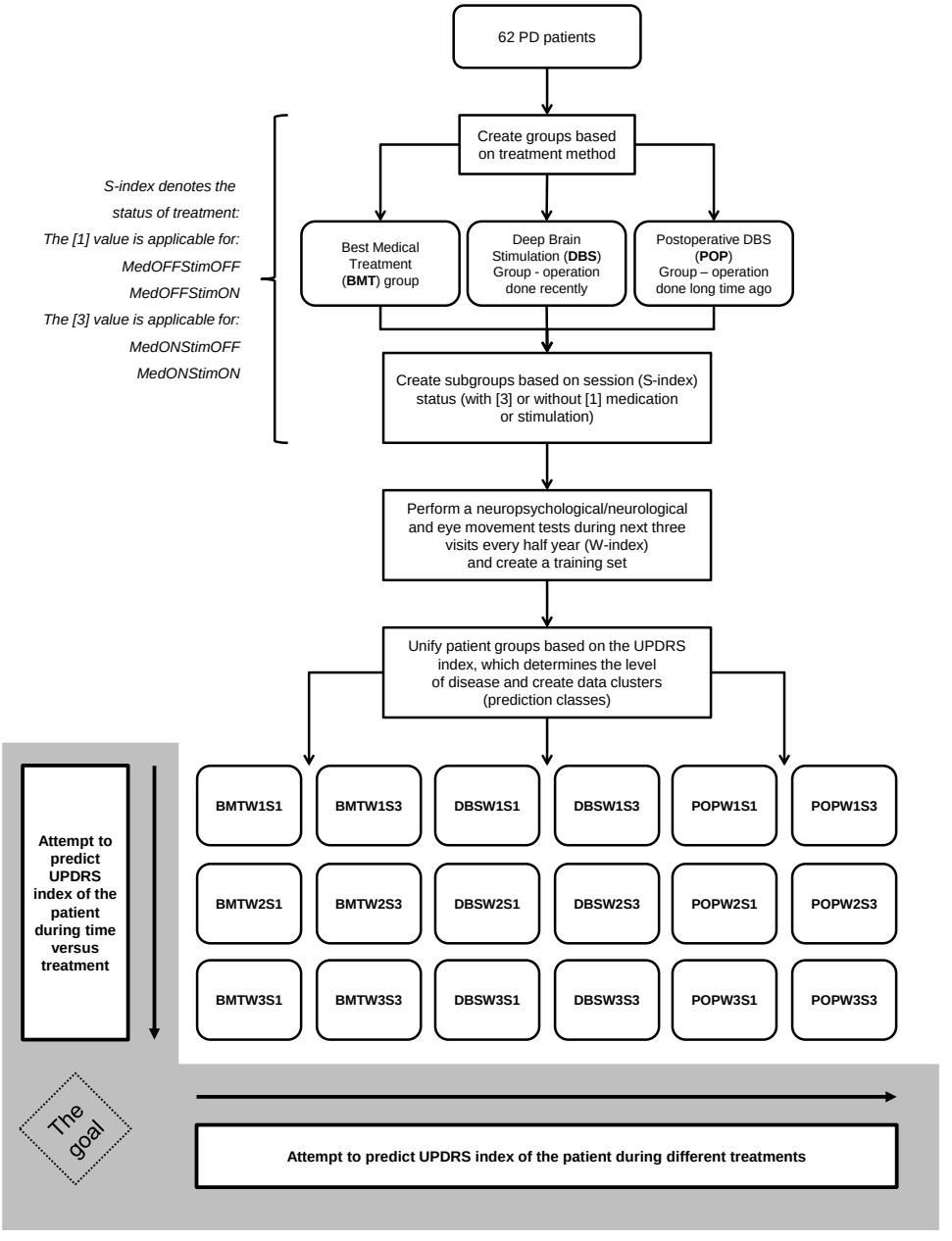

Figure 1: The diagram shows a flow model of the data mining process presented in the article. From 62 PD patients, we distinguished three groups based on treatment (BMT / DBS / POP). Each group was split further into subgroups based on the session (S) status and visit index (W). We aimed to predict the UPDRS index in two ways: during single treatment and among different treatments.

3. Results

All 62 PD patients were divided into three groups: BMT patients (medication only), and patients on medication and with implanted electrodes in the STN (subthalamic nucleus [2, 5]) during our study: DBS group or before our study: POP group.

In 23 patients of BMT group the mean age was 57.8+/- 13 (SD) years; disease duration was 7.1+/- 3.5 years, UPDRS was 36.1+/-19.2. In 24 patients of DBS group the mean age of 53.7+/- 9.3 years, disease duration was 10.25+/- 3.9 years (stat. diff. than BMT-group: $p < 0.025$), UPDRS was 62.1+/- 16.1 (stat. diff. than BMT-group: $p < 0.0001$). In 15 patients of POP group the mean age was 56.2+/- 11.3 (SD) years and disease duration was 13.5+/- 3.6 years (stat. diff. than DBS-group: $p < 0.015$), UPDRS was 59.2+/-24.5 (stat. diff. than BMT-group: $p < 0.0001$).

These statistical data are related to the data obtained during the first visit for each group: so-called BMTW1 (visit one), DBSW1 (visit one) and POPW1 (visit one).

3.1. Comparing longitudinal UPDRS changes

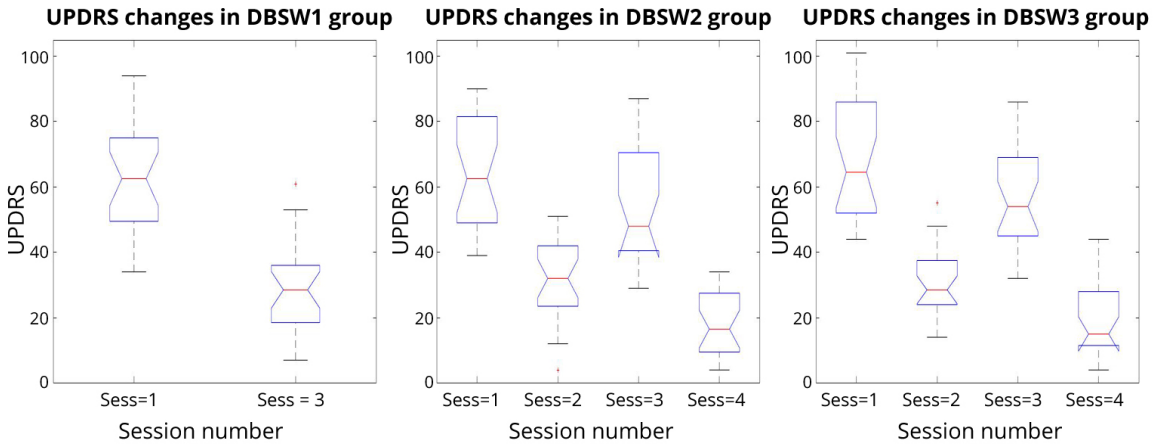

Figure 2: The boxplot shows that the median of UPDRS in different sessions and visits; W1, W2, W3 of the DBS group. There are minimum and maximum values in each session. Since the notches in the box plot do not overlap, you can conclude, with 95% confidence that the true medians do differ.

The first plot from the right presents UPDRS of DBSW1 group. There are only two sessions as patients were before the surgery. In the first session (described as Ses1) mean UPDRS was 62.2+/- 16.1, in the second (described as Ses3) was 29.9+/-13.3 strongly ($p < 0.0001$) different from Ses1 and it represents the effect of medication. Plot in the middle of Fig. 2 represents DBSW2 after the surgery and UPDRS in Ses1 is larger than that before the surgery 65.3+/- 17.6 but there are no statistically significant differences, however UPDRS in Ses1 of DBSW3 is 68.7+/- 17.7 and statistically significant ($p < 0.03$) than in W2. Effects of different therapies (session numbers) are significantly different in W1, W2, and W3, but not different between the same session numbers in different visit (with the exception of the Ses3 in W1 as after the surgery the dosage of medication is reduced).

In POP-group UPDRS values are similar. There is an increase of the UPDRS Ses1 from W1: 63.1 +/- 18.2 to W2: 68.9+/-20.3 to W3: 74,2+/- 18.4 but there were smaller differences for Ses4 (MedOnDBSOn) W1: 21 +/- 11.3 to W2: 23.3+/-9.5 to W3: 23,8+/- 10.7. Therefore, we have assumed that groups DBS and POP are similar.

In BMT group UPDRS in Ses1 was W1: 48.3+/-17.9; W2: 57.3+/-16.8 ($p < 0.0005$ different than W1); W3: 62.2+/-18.2 ($p < 0.05$ different than W2). In Ses3 UPDRS was W1: 23.6+/- 10.3; W2: 27.8+/-10.8; W3: 25+/-11.6 (no statisticall difference between visits for Ses3).

3.2. BMT patients disease progression prediction

In these sections we have compared two different approaches: in 3.2.1 we have found RSES rules describing attributes for the first visit BMTW1 and have predicted on their basis disease progression in the same patients half (BMTW2) and one year later (BMTW3). In section 3.2.2 we have predicted disease progression in the same time periods, but using Gaussian Nave Bayes.

3.2.1. BMT patients: RS rules for the disease progression

The BMT patients (only on medication) were tested in two sessions (session 1: without, and session 3: with medication) three times every half-year.

We have used ML and rough set theory [9] in order to obtain rules determining decision and condition attributed for the first visit BMTW1. We have obtained these rules using Exhaustive algorithm and additionally rules filter that removes rules with support 1 (a single case rules) [9]. In the most cases Exhaustive algorithm gave the shortest set of rules and the largest support (for more cases).

On the basis of these rules we have predicted the UPDRS values obtained during the second (half - year later W2 - BMTW2) and the third (one year later BMTW3) visits. UPDRS was optimally divided by RSES into 4 ranges: $"(-Inf, 24.0)", "(24.0, 36.0)", "(36.0, 45.0)", "(45.0, Inf)"$ for both visits (W2 and W3) the global coverage was 1.0 and the global accuracy was 0.7. Example of rules from BMTW1:

$$(Ses = 3)\&(PDQ39 = "(-Inf, 50.5)") \Rightarrow (UPDRS = "(-Inf, 33.5)"[12])12 \quad (1)$$

$$(dur = "(-Inf, 5.65)")\&(Ses = 3)\&(Epworth = "(-Inf, 14.0)") \Rightarrow (UPDRS = "(-Inf, 33.5)"[7])7 \quad (2)$$

$$(dur = "(5.65, Inf)")\&(Ses = 3)\&(Epworth = "(14.0, Inf)") \Rightarrow (UPDRS = "(-Inf, 33.5)"[4])4 \quad (3)$$

In the first rule (1) if the session number 3 and *PDQ39 = "(-Inf, 50.5)"* then UPDRS was *(-Inf, 33.5)* in 12 cases. The second rule (2) was fulfilled in 7 cases and the third one (3) in 4 cases. There were altogether 70 rules.

3.2.2. BMT patients: ML models for the disease progression

As mentioned above, every ML model was independently trained for every session separately, to enhance its prediction results.

For session 1, the UPDRS ranges of BMTW1 were optimally divided into four classes with the middle intervals in points: $C_0 = 26.0, C_1 = 45.5, C_2 = 58.67, C_3 = 104.0$. The scopes of 4 ranges were: $"(0, 35.0)", "(36.0, 52.0)", "(53.0, 81.0)", "(82.0, 114.0)"$. On the basis of these scopes, we have predicted the UPDRS values obtained during the second (half -year later W2 - BMTW2) and the third (one year later BMTW3) visits. In this session, the Gaussian Naive Bayes (GaussianNB) achieved the best results. The prediction on BMTW2 set had shown the accuracy of 0.61 with the coverage 1.0. For BMTW3 the accuracy was 0.57 with the coverage = 1.0. For the session 3, as the basis for learning model, we had chosen BMTW2 to predict classes of UPDRS in BMTW3. UPDRS was normalised and clustered into new scopes: $"(0.0, 19.0)", "(20.0, 28.0)", "(29.0, 40.0)", "(41, 114.0)"$ with centroids: $C_0 = 13.87, C_1 = 25.00, C_2 = 32.67, C_3 = 47.50$. On the basis of those rules, we were able to predict the UPDRS classes of BMTW3 with the accuracy = 0.61 (Table 1).

Table 1: Confusion matrix for UPDRS of BMTW3 (Session 3) by model obtained from BMTW2 predicted by Decision Tree. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.61.

		Predicted				
		"(20.0, 28.0)"	"(0.0, 19.0)"	"(29.0, 40.0)"	"(41.0, 114.0)"	ACC
Actual	"(20.0, 28.0)"	5	1	1	0	0.71
	"(0.0, 19.0)"	1	7	2	0	0.70
	"(29.0, 40.0)"	0	1	1	2	0.25
	"(41.0, 114.0)"	1	0	0	1	0.50
	TPR	0.71	0.78	0.25	0.33	

3.3. DBS and POP patients: rules for estimation of disease progression

In these sections we have compared two different approaches: in 3.3.1 we have found RSES rules describing attributes for the second visit DBSW2 and have predicted on their basis disease progression in the same patients half year later (DBSW3). We have performed similar predictions for POP patients. On the basis of the first visit POPW1 we have predicted disease progression in two later visits: POPW2 and POPW3. I section 3.3.2 we have used the Logic Regression algorithm in order to predict disease progression in POPW3 on the basis of POPW2 group.

3.3.1. DBS patients: RS rules for estimation of disease progression

As DBSW1 had only 2 sessions (before surgery) we could only predict session DBSW3 on the basis of DBSW2 (half a year earlier) (Tab.2). We have predicted UPDRS for visits POPW2 and POPW3 on the basis of visit POPW1 with total accuracy: 0.667 and 0.793 with a coverage: 1 and 0.967.

Table 2: Confusion matrix for UPDRS of DBSW3 by rules obtained from DBSW2. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.562.

		Predicted					
		”(46.0, 72.0)”	”(38.0, 46.0)”	”(19.5, 38.0)”	”(72.0, Inf)”	”(-Inf, 19.5)”	ACC
Actual	”(46.0, 72.0)”	12	5	2	5	1	0.48
	”(38.0, 46.0)”	2	5	1	2	2	0.42
	”(19.5, 38.0)”	0	4	13	3	7	0.48
	”(72.0, Inf)”	4	0	0	12	0	0.75
	”(-Inf, 19.5)”	0	0	4	0	12	0.75
	TPR	0.67	0.4	0.65	0.55	0.6	

3.3.2. POP patients: ML models for estimation of disease progression

We were only able to predict POPW3 in Session 3 by model obtained from POPW2 with the usage of Logistic Regression, which achieved the accuracy of 0.6. UPDRS was normalised and clustered into scopes: $”(0.0, 15.0)”, ”(16.0, 22.0)”, ”(23.0, 32.0)”, ”(33.0, 114.0)”$ with centroids: $C_0 = 12.33, C_1 = 18.00, C_2 = 27.00, C_3 = 38.00$ (Tab. 3).

Table 3: Confusion matrix for UPDRS of POPW3 (Session 3) by model obtained from POPW2 by Logistic Regression. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.60.

		Predicted				
		”(16.0, 22.0)”	”(33.0, 114.0)”	”(0.0, 15.0)”	”(23.0, 32.0)”	ACC
Actual	”(16.0, 22.0)”	1	2	1	0	0.25
	”(33.0, 114.0)”	2	2	1	0	0.40
	”(0.0, 15.0)”	0	0	4	0	1.00
	”(23.0, 32.0)”	0	0	0	2	1.00
	TPR	0.33	0.50	0.67	1.00	

3.4. Disease progression of DBS patients estimated by rules from BMT patients

In these sections we have compared different approaches: in 3.4.1 by using RSES we have found rules describing attributes for the first visit BMTW1 and have predicted on their basis disease progression in different patients with different therapy in three visits: DBSW1, DBSW2, and DBSW3. In session 3.4.2 in order to predictions of DBS patients disease progression in the BMT group we have used several different algorithms: Gradient Boosting Classifier, C-Support Vector Classification, and LinearSVC algorithms.

3.4.1. Disease progression of DBS patients estimated by RS rules from BMT patients

As BMT patients had only two sessions (S1 - MedOff, and S3 - MedON) and DBS patients four sessions (see Methods) we have divided them into two sets: one with StimON set-up and another one with StimOFF set-up. We were not successful in the prediction of StimOFF sessions as DBS patients

were in a more advanced stage than BMTW1 group. Our UPDRS predictions for DBSW1 had global accuracy 0.64 (coverage 0.5); for DBSW2 - global accuracy was 0.85 (coverage 0.3); for DBSW3 - global accuracy was 0.74 (coverage 0.6).

3.4.2. Disease progression of DBS patients estimated by ML models from BMT patients

For S1, we were able to predict the UPDRS of DBSW1 by model obtained from BMTW3 based on Gradient Boosting Classifier with the accuracy 0.54. C-Support Vector Classification shown the accuracy of prediction as high as 0.88 for DBSW2 and DBSW3 when it was trained on the basis of BMTW3 (coverage 1.0). However, this classification should be treated with caution because there is a significant difference between UDPRS scores distribution between the classes, and their scopes and centroids are shifted (for example: the first calculated class of model based on BMTW3 has a range which is very extensive: "$(0.0, 55.0)$" and every UPDRS score from DBSW2 is lower than 51.0).

For S3, the prediction of UPDRS scores in DBSW2 has an accuracy of 0.58 and 0.54 for DBSW3 when trained on BMTW2 by LinearSVC (coverage 1.0). We were unable to predict classes of DBSW1 based on BMT.

3.5. Disease progression of POP patients estimated by rules from DBS patients

In these sections we have compared different approaches: in 3.5.1 we have no success to predict POP patient disease progression on the basis of DBS patients. In 3.5.2 by using Random Forest algorithm we could predict POPW1 disease progression from DBSW2 patients.

3.5.1. Disease progression of POP patients estimated by RS rules from DBS patients

We could not predict UPDRS of POP patients from rules obtained from DBS patients probably because many years of DBS might influence some brain synaptic connections. In case of POP patients responses to MedON/OFF are not consistent with responses in DBS patients.

3.5.2. Disease progression of POP patients estimated by ML models from DBS patients

However, in this section we are looking for prediction of UPDRS only in one session (MedON). The UPDRS data in subset of DBSW2 (Session 3 - MedON) was optimally divided with centroids:

Table 4: Confusion matrix for UPDRS of POPW1 (Session 3) by model obtained from DBSW2 by Random Forest. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.60.

		Predicted				
		"(12.0, 18.0)"	"(0.0, 11.0)"	"(27.0, 114.0)"	"(19.0, 26.0)"	ACC
Actual	"(12.0, 18.0)"	4	1	1	0	0.67
	"(0.0, 11.0)"	0	2	0	0	1.00
	"(27.0, 114.0)"	3	0	2	0	0.40
	"(19.0, 26.0)"	0	0	1	1	0.50
	TPR	0.57	0.67	0.50	1.00	

$C_0 = 6.71, C_1 = 16.12, C_2 = 21.60, C_3 = 32.17$ and with ranges of: "$(0.0, 11.0)$", "$(12.0, 18.0)$", "$(19.0, 26.0)$", "$(27.0, 114.0)$". On the basis of the model from DBSW2 - session 3, we were able to predict the UPDRS classes of POPW1 session 3 with the accuracy of 0.60 (Tab.4). It is worth noting that it was true only for particular datasets: DBSW2 and POPW1 and the particular session - MedON.

Table 5: Performance comparison among different ML models.

Reference	Training Set	Test Set	RS Accuracy	RS Coverage	ML Accuracy	ML Coverage	ML Model
3.2.1	BMTW1S{1,3}	BMTW2S{1,3}	0.70	1.00			
3.2.1	BMTW1S{1,3}	BMTW3S{1,3}	0.70	1.00			
3.2.2	BMTW1S1	BMTW2S1			0.61	1.00	Gaussian Naive Bayes
3.2.2	BMTW1S1	BMTW3S1			0.57	1.00	Gaussian Naive Bayes
3.2.2	BMTW2S3	BMTW3S3			0.61	1.00	Decision Tree
3.3.1	DBSW2S{1,3}	DBSW3S{1,3}	0.56	1.00			
3.3.2	POPW1S{1,3}	POPW2S{1,3}	0.67	1.00			
3.3.2	POPW1S{1,3}	POPW3S{1,3}	0.79	0.97			
3.3.3	POPW2S3	POPW3S3			0.60	1.00	Logistic Regression
3.4.1	BMTW{1,2,3}S{1,3}	DBSW1S{1,3}	0.64	0.50			
3.4.1	BMTW{1,2,3}S{1,3}	DBSW2S{1,3}	0.85	0.30			
3.4.1	BMTW{1,2,3}S{1,3}	DBSW3S{1,3}	0.74	0.60			
3.4.2	BMTW3S1	DBSW1S1			0.54	1.00	Gradient Boosting
3.4.2	BMTW3S1	DBSW2S1			0.88	1.00	C-Support Vector
3.4.2	BMTW3S1	DBSW3S1			0.88	1.00	C-Support Vector
3.4.2	BMTW2S3	DBSW2S3			0.58	1.00	LinearSVC
3.4.2	BMTW2S3	DBSW3S3			0.54	1.00	LinearSVC
3.5.2	DBSW2S3	POPW1S3			0.60	1.00	Random Forest

4. Discussion

There are novel technologies and data constantly improving PD patients' treatments, but there are also still doubts if the actual procedures are optimal for a particular case. Our long time purpose is to use data mining and machine learning in order to compare different neurological protocols and their effectiveness. We believe that the best future approach will be to perform all tests automatically at home, process them with intelligent algorithms and to submit results to the doctor for his/her decision (compare with telemedicine based methods [14]). Another, more advanced approach that we were testing in this work, would be to create standard treatment for each new case on the basis of already successfully treated patients and correct treatment as symptoms are developing in time (see technology in PD [15]). We have demonstrated that it is relatively easy to estimate symptoms and their time development in populations treated in different ways (e.g. only medication treatment) (compare to motor automaticity [16]). This result may give the basic (locally optimal) follow-up PD symptoms. If the patient is doing significantly worse than others (rules), their treatment is not optimal and should be changed. In the next step, we may use rules obtained from different clinics to make them even more universal and optimal. Our new approach is related not only to longitudinal study but also to test different patient population with different treatments. Could our results lead to finding optimal procedures for different cases and diverse cares? It seems to be a good direction, but there are many particular issues, for example when patients are in different stages of the disease. In our case, the second group of patients (DBS group) were in a more advanced stage of disease so it was not possible

to get 100% coverage like in the first case. The second group with longitudinal study had a new treatment (brain stimulation) that started from the second visit. We have tested if the same treatment in different populations gives similar results. Patients got two treatments: medication (medication ON and OFF) and electric brain stimulation (ON and OFF). We have analysed these treatments as two different sets: 1) StimOFF: medication ON and OFF; 2) StimON: medication ON and OFF. As a result, it was not possible to obtain sufficient accuracy in the first situation (1) as patients with DBS (DBS and POP groups) were more advanced in their disease stages and had different UPDRS ranges: mean UPDRS in BMT group was 36 and in DBS/POP groups was 54/56 respectively (see Results section). Therefore, we can interpret brain stimulation (DBS) as resetting symptoms in time of disease - patients with 10 years disease duration have symptoms similar to patients with 7 years disease duration and we could predict their symptoms with reasonable accuracy.

But this assumption was not exact for patients with long period of brain stimulation (DBS) - POP group (with over 13 years disease duration). This group was different than the other two as we did not succeed in forming a good prediction by rules obtained by other groups BMT or DBS. It might be related to the longer period of brain stimulation (DBS) that has changed some central mechanisms [17]. It is an important negative result that needs further study. In the near future, we may look for additional condition attributes in order to improve the global accuracy. In the present paper in comparison to the previous one [1] we have introduced additional to RS machine learning (ML) models [10-11].

We have compared intelligent methods based on rough set theory with several different machine learning algorithms: Gaussian Nave Bayes, Decision Tree, Logistic Regression, Gradient Boosting, C-Support Vector, LinearSVC, and Random Forest. Generally rough set method gave better accuracy but not as good coverage as other algorithms (see Table 5).

In most cases these new methods gave similar results to our RS rules. However, in contrast to RS rules, ML models could predict results of one session (session 3 - MedON) in POPW1 patients on the basis of DBSW2 session 3 results. It is a very limited case, but showing some common mechanisms between DBS and POP groups.

Furthermore, as shown in the Table 5, Rough Set-based approach allows creating more general rules without the necessity of additional data splitting (in this case: into different sessions), which was required in the ML modelling. The principal conclusion which comes from that comparison is an observation that RS is much more universal method when considering medical data, which constitutes a confirmation of the findings from [3].

In summary, we have demonstrated that the DBS (electric STN stimulation) procedure revoked and improved rules that became similar to rules of early stage Parkinson's disease patients.

5. Conclusions

This work is a continuation of our previous findings [1, 3], comparing classical approach used by most neurologists and based on their partly subjective experience and intuitions with the intelligent data processing (machine learning, data mining) classifications. This paper represents a new technological and data mining trend in treatments of the neurodegenerative diseases [18, 19, 20]. We have furthermore demonstrated that such additional attributes as parameters of eye movements and

neuropsychological data are significant in predicting longitudinal symptom developments in different therapy related groups of PD patients. We have confirmed that a multidisciplinary approach incorporating neurologists, automatic measurements and intelligent data mining such as granular computing is an example of open-science, collaborative research that is the future of complex healthcare. In the near future, we are planning to involve also other clinics in our project as well as other neurodegenerative diseases such as Alzheimer's disease. Our direction is to increase the number of attributes in order to find significant attributes not only for each disease or treatment, but also for each individual patient. It would lead to a patient-centred and patient-oriented intelligent telemedicine.

We have demonstrated that on the basis of the first group of early PD patients (only on medications), we could predict symptoms of patients in more advanced stages and with different treatments. It means that our rules have some universal properties, which means that in the future we may suggest an optimal treatment for each individual patient.

Acknowledgments

Authors would like to thank English Lecturer Michal Zawistowski for proofreading the article.

Authors' contributions

The first author AWP has written the paper and performed RST data analysis, the second author AC has written part of the paper and performed ML analysis, the third SS and fifth DMK coauthors have performed all tests with PD patients, the fourth coauthor has created the database.

References

[1] Przybyszewski AW, Szlufik S, Habela P, Koziorowski DM. Rough Set Rules Determine Disease Progressions in Different Groups of Parkinson's Patients. B.U. Shankar et al. (Eds.): PReMI 2017, vol. 10597 of LNCS, 2017 pp. 270–275. doi:10.1007/978-3-319-69900-4_34.

[2] Przybyszewski AW, Kon M, Szlufik et al. Data Mining and Machine Learning on the Basis from Reflexive Eye Movements Can Predict Symptom Development in Individual Parkinson's Patients. In Nature-Inspired Computation and Machine Learning; Eds. Gelbukh et al. Springer, vol. 8857 of LNCS, 2014 pp. 499–509. doi:10.1007/978-3-319-13650-9_43.

[3] Przybyszewski AW, Kon M, Szlufik S, Szymanski A, Koziorowski DM. Multimodal Learning and Intelligent Prediction of Symptom Development in Individual Parkinson's Patients. Sensors, 2016. 16(9):1498. doi:10.3390/s16091498.

[4] Przybyszewski AW. The Neurophysiological Bases of Cognitive Computation Using Rough Set Theory. *Transactions on Rough Sets IX* (J.F. Peters et al. Eds.), vol. 5390 of LNCS, 2008 pp. 287–317. doi:10.1007/978-3-540-89876-4_16.

[5] Przybyszewski AW. Applying Data Mining and Machine Learning Algorithms to predict symptoms development in PD. Annales Academiae Medicae Silesiensis, 2014. 68(5):332–349.

[6] Yasuda M, Hikosaka O. To Wait or Not to Wait - Separate Mechanisms in the Oculomotor Circuit of Basal Ganglia. Front Neuroanat. 2017 pp.13, 11:35. doi:10.3389/fnana.2017.00035.

[7] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer, Dordrecht, 1991. ISBN:978-0-7923-1472-1, 978-94-010-5564-2.

[8] Bazan J, Son Nguyen H., Trung T, Nguyen HS, Skowron A, Stepaniuk J. Desion rules synthesis for object classification. In: E. Orlowska (ed.), Incomplete Information: Rough Set Analysis, Physica – Verlag, Heidelberg, 1998 pp. 23–57.

[9] Bazan J, Szczuka M. RSES and RSESlib - A Collection of Tools for Rough Set Computations. W. Ziarko and Y. Yao (Eds.): RSCTC 2000, vol. 2005 of LNAI, 2001 pp. 106–113. doi:10.1007/3-540-45554-X_12.

[10] Marsland S. Novelty detection in learning systems: Neural computing surveys, 2003. 3(2):157–195.

[11] Pedregosa et al. Scikit-learn: Machine Learning in Python: JMLR 2011. 12(85):2825–2830. `http://jmlr.org/papers/v12/pedregosa11a.html`.

[12] MacQueen J. Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1967. 1(14):281–297. ID:6278891.

[13] Comaniciu D, Meer P. Mean Shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. 24(5):603–619. doi:10.1109/34.1000236.

[14] Lindauer A, Croff R, Mincks N, Mattek SJ. Shofner N, Bouranis L, Teri R. It Took the Stress out of Getting Help: The STAR-C-Telemedicine Mixed Methods Pilot. Care Wkly, 2018. 2:7–14. doi:10.14283/cw.2018.4.

[15] Espay AJ, Bonato P, Nahab FB et al. From Movement Disorders Society Task Force on Technology.: Technology in Parkinson disease: Challenges and Opportunities. Mov Disord., 2016. 31(9):1272–1282.

[16] Wu T, Hallett M, Chan P. Motor automaticity in Parkinson's disease. Neurobiol Dis., 2015. 82:226–234. doi:10.1016/j.nbd.2015.06.014.

[17] Umemura A, Oyama G, Shimo Y, et al. Current Topics in Deep Brain Stimulation for Parkinson Disease. Neurol Med Chir (Tokyo), 2016. 56(10):613–625. doi:10.2176/nmc.ra.2016-0021.

[18] Tucker C, Han Y, Nembhard HB et al. A data mining methodology for predicting early stage Parkinson's disease using non-invasive, high-dimensional gait sensor data. Healthcare (Basel) Jun, 2018. 6(2):54.

[19] Dinov ID, Heavner B, Tang M et al. Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations. PLoS One, 2016. 11(8):e0157077. doi:10.1371/journal.pone.0157077.

[20] Gao C, Sun H, Tuo Wang T et al. Model-based and Model-free Machine Learning Techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson's Disease. Sci Rep., 2018. 8:7129. doi:10.1038/s41598-018-24783-4.