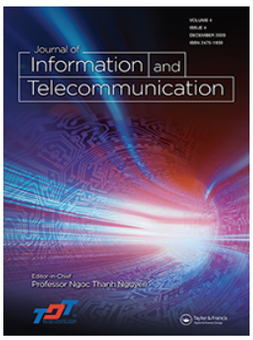

Taylor & Francis
Taylor & Francis Group

Journal of Information and Telecommunication

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjit20

Parkinson's disease development prediction by c-granule computing compared to different AI methods

Andrzej W. Przybyszewski & Albert Śledzianowski

To cite this article: Andrzej W. Przybyszewski & Albert Śledzianowski (2020) Parkinson's disease development prediction by c-granule computing compared to different AI methods, Journal of Information and Telecommunication, 4:4, 425-439, DOI: 10.1080/24751839.2020.1749410

To link to this article: https://doi.org/10.1080/24751839.2020.1749410

Published online: 22 Apr 2020.

JOURNAL OF INFORMATION AND TELECOMMUNICATION
2020, VOL. 4, NO. 4, 425–439
https://doi.org/10.1080/24751839.2020.1749410

Parkinson's disease development prediction by c-granule computing compared to different AI methods

Andrzej W. Przybyszewski [a,b] and Albert Śledzianowski[a]

[a]Polish-Japanese Academy of Information Technology, Warszawa, Poland; [b]Department of Neurology, UMass Medical School, Worcester, USA

ABSTRACT

Both rough set theory (RST) and fuzzy rough set theory (FRST) are related to intelligent granular computing (GrC) primarily with the help of static granules. Our granules are sets of attributes measured from Parkinson's disease (PD) patient in a certain moment of his/her disease. Our complex granule (c-granule) approach was used to model longitudinal PD development. With RST/FRST we were looking for similarities between attributes of patients in different disease stages to more advanced PD patients. We have compared group (G1) of 23 PD with attributes measured three times (visits V1–V3) every half of the year (G1V1, G1V2, G1V3) to the other group of 24 more advanced PD (G2V1). By means of RST/FRST, we have found rules describing symptoms of G2V1 and applied them to G1V1, G1V2, and G1V3. With RST (FRST), we've got the following accuracies: G1V1 −59 (38)%; G1V2 – 68 (54)%; G1V3 – 86 (61)%, but global coverage for FRST was better. We also tried to compare results with several different machine learning methods, obtaining accuracies of G1V1 – 59%, G1V2 – 73%, and G1V3 – 78%. In summary, several different methods confirmed that generally one group of PD patients during disease development become more similar to a different group of more advanced PD.

ARTICLE HISTORY
Received 27 January 2020
Accepted 27 March 2020

KEYWORDS
Granular computing; similarity; aggregation; disease progression; disease model

1. Introduction

Our goal was to simulate Parkinson's disease (PD) development in time with the help of granular computing (GrC) methods (Pawlak, 1991; Zadeh, 2002). As PD-related neurodegeneration (ND) starts ∼20 years before the first symptoms and during this period of time, the ND process is effectively compensated by brain plasticity, each patient's PD progressions are different.

In this work, we have used intelligent GrC based on the principle of complex object classifications from the visual brain (Przybyszewski, 2008;, 2019). As stated in the schematic (Figure 1), properties of the unknown object p are represented as α and compared with the model α_M (in the brain – possible objects (Przybyszewski, 2008), here symptoms of more

CONTACT Andrzej W. Przybyszewski przy@pjwstk.edu.pl Polish-Japanese Academy of Information Technology 02-008 Warszawa, Poland; Department of Neurology, UMass Medical School, Worcester, MA 01655, USA

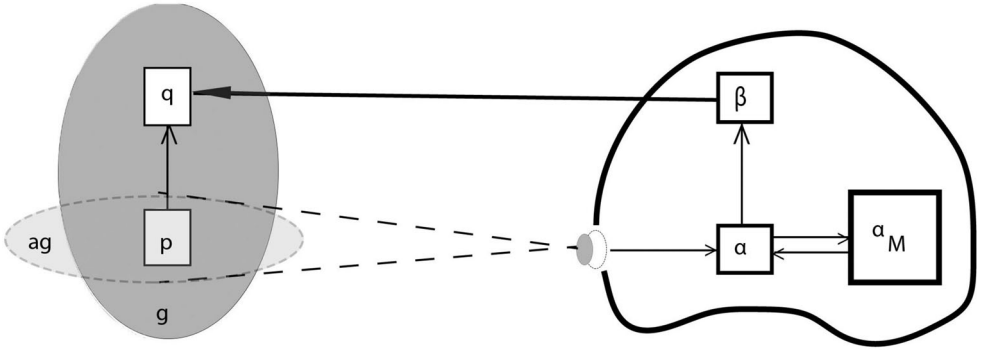

Figure 1. It is a principal schematic of the intelligent GrC based on the brain intelligence. We observe a limited part of c-granule *g* that is generally subpart ag of environment env in a time interval $[t - \Delta, t]$. Interaction between env and ag in time *t* during ΔInt_g,t,Δ (env, ag) represents *a* and results that rule $(a \cap a_M, \gamma) -> \beta$ is learned by ag, where γ represents properties of the structure *g*, and *a* property of the interaction process, and *β* describes unknown, expected properties, the other part of *g* that might be reason for future changes into the disease; a_M it is the model of the world that interacts with *a* in order to extract its significant features (modified after Skowron & Dutta, 2018).

advanced PDs). Its results rule *β* that determine new object's properties or PD time development.

2. Methods

Our data mining analysis is based on GrC implemented in RST (rough set theory) proposed by Pawlak (1991) and FRTS (fuzzy rough set theory) by extending RST indiscernibility with concepts of the tolerance (Zadeh, 2002).

Our data are converted to the decision table, where rows were related to different measurements and columns represent different attributes. An information system (Pawlak, 1991) is a pair $S = (U, A)$, where *U, A* are non-empty finite sets called the universe of objects *U* and the set of attributes *A*. If $a \in A$ and $u \in U$, the value $a(u)$ is a unique element of *V* (where *V* is a value set).

We define, as in Pawlak (1991), for RST the indiscernibility relation of any subset *B* of *A* or IND *(B)* as: $(x, y) \in \text{IND}(B)$ or $x I(B) y$ iff $a(x) = a(y)$ for every $a \in B$, where the value of $a(x) \in V$. It is an equivalence crisp relation $[u]_B$ that we understand as a *B*-elementary granule. The family of $[u]_B$ gives the partition *U/B* containing *u* will be denoted by *B(u)*. The set $B \subset A$ of information system *S* is a reduct IND (*B*) = IND (*A*) and no proper subset of *B* has this property (Pawlak, 1991). In most cases, we are only interested in such reducts that are leading to expected rules (classifications). On the basis of the reduct, we have generated rules using four different ML methods (RSES 2.2): exhaustive algorithm, genetic algorithm, covering algorithm, or LEM2 algorithm.

A *lower approximation* of set $X \subseteq U$ in relation to an attribute *B* is defined as all elements have *B* attribute: $\underline{B}X = \{u \in U: [u]_B \subseteq X\}$. The upper approximation of *X* is defined as some elements have *B* attribute: $\overline{B}\,X = \{u \in U: [u]_B \cap X \neq \phi\}$. The difference between *B X* and *B X* is the boundary region of *X* that we denote as $BN_B(X)$. The difference of $\overline{B}$ and $\underline{B}$ is empty,

then set X is exact with respect to B; otherwise, if $BN_B(X)$ is not empty and X is not rough with respect to B.

A decision table (training sample in ML) for S is the triplet: $S = (U, C, D)$ where: C, D are condition and decision attributes (Pawlak, 1991). Each row of the decision table gives a particular rule that connects condition and decision attributes for a single measurement, RST generalizes these particular rules into universal hypotheses (object or disease classification).

Dubois and Prade (1990) have generalized RST to FRTS by extending RST indiscernibility with concepts of tolerance after Zadeh's membership degrees in fuzzy sets (Zadeh, 2002).

As a consequence, 'crisp' dependences were replaced by a fuzzy tolerance or similarity relations Ra (x, y) as a value between two observations x and y. As Ra (x, y) is a similarity relation, it must be reflexive, symmetric, and transitive. As summarized in Riza et al. (2014), there are several tolerance relationships such as the normalized difference (so-called equation (1)) or Gaussian or exponential differences (Riza et al., 2014). There are also formulas related to normalized differences between pairs of attributes. The most common are *Łukasiewicz* and *t.cos t.norm* $-\ \tau$ (Riza et al., 2014). As decision attributes are nominative, we used crisp relations between them.

We define B-lower and B-upper approximations for each observation x in FRST as following: *B-lower approximation* as: $(R_B \downarrow X)(x) = \inf_{y \in U} I(R_B(x, y), X(y))$, where I is an implicator (Riza et al., 2014). The B-lower approximation for the observation x is then the set of observations, which are the most similar to the observation x and it can predict the decision attribute with the highest confidence, based on conditional attributes B.

The *B-upper approximation* is defined by $(R_B \uparrow X)(x) = \sup_{y \in U} \tau(R_B(x, y), X(y))$, where τ is the t.norm. The B-upper approximation is a set of observations for which the prediction of decision attribute has the smallest confidence (Riza et al., 2014).

Notice that rules in FRST have dissimilar formation than in RST. They are based on the tolerance classes and appropriate decision concepts. The fuzzy rule is a triple (B, C, D), where B is a set of conditional attributes that appear in the rule, C stands for fuzzy tolerance class of object, and D stands for decision class of object.

We have used RST algorithms implemented as the RSES 2.2 (logic.mimuw.edu.pl/~rses/get.html) Exploration Program Rough System and FRST implemented as Rough Set package in R (Riza et al., 2014).

2.1. Measured attributes

We have tested two groups of PD patients: the first group (G1) of 23 patients was measured three times every half of the year (visits were numbered as V1, V2, V3) and the second group (G2) had more advanced 24 patients and were a reference model of disease progression in the first group. Both groups of patients were only on medication. The major medication in this group was L-Dopa that increases the concentration of the transmitter dopamine in the brain which is lacking in Parkinson's patients. In most cases, PD starts with ND in substantia nigra that is responsible for the release of dopamine.

All patients were measured in two sessions: MedOFF (session S#=1 without medication) and MedON (session S#=2 patients on medications). In addition, all patients have the following procedures: neuropsychological tests: PDQ39 (quality of life), Epworth (sleepiness test); neurological tests: eye movements and standard PD test: UPDRS (Unified Parkinson's

Disease Rating Scale). All tests were performed in Brodno Hospital, Department of Neurology, Faculty of Health Science, Medical University Warsaw, Poland. In the present work, we have tested and measured fast eye movements: reflexive saccades (RS) as described in our previous publications (Przybyszewski et al., 2016, 2018a). In summary, every subject was sitting in a stable position without head movements and watching a computer screen before him/her. At the beginning, he/she has to fixate in the centre of the screen and to keep on moving light spot. This spot was jumping randomly, 10° to the right or 10° to the left. The patient has to follow movements of the light spot and the following parameters were measured: latency (RSLat) – time difference between the beginning of the spot and eyes movements, saccade duration (RSDur), saccade amplitude (RSAmp), and saccade velocity (RSVel).

2.2. Comparison with different methods

We also wanted to compare our approach with other methods of classifications using the same data. We developed a classification ranking procedure iterating over different scaling and classification methods implemented in scikit-learn, which is Python module integrating classic machine learning algorithms (Pedregosa et al., 2011). We have performed two types of tests: classification on train test (trainset G2V1, test set G2V1) and using datasets for which we wanted to check predictions (trainset G2V1, test sets G1V1–3). The purpose of the comparison was not only to find out the accuracy of classification but also to check whether the prediction results for individual G1V1–3 test sets would represent a similar upward trend. The similarity of test set G1–3 to trainset G2V1 should increase with each subsequent visit number, which should be expressed in the results of classifications.

With discretized data sets, we started to iterate over different pairs of scaling-transforming and classification methods (listed below) trying to estimate and score predictions finding the best values for the hyper-parameters. For this purpose, we used the GridSearchCV (GSCV), the method implemented in the scikit-learn which exhaustively searches best values off hyper-parameters using cross-validation. The cross-validation divides the data set into the number of folds and performs the train-test procedure for each one with the remaining folds, averaging the evaluation results (Pedregosa et al., 2011).

We used default 5 folds split, because we didn't see any improvement in the results by increasing the value of this parameter. For predictions and estimations of the prediction score, we used cross-validated methods implemented in the scikit-learn ('cross_val_predict', 'cross_val_score') (Pedregosa et al., 2011). The ranking method worked according to the diagram shown in Figure 2.

Different pairs of train and test sets were iterated using combinations of scalers and classifiers tested after tuning of hyper-parameters. We decided to use scalers because of the different value scales represented by our features. In the example, UPDRS varies between 0 (no disability) and possible maximum of 199 points (total disability). In the PDQ-39 scale, different parts of the survey are grouped into eight scales and scored by sums of scores in a percentage ranging from 0 and 100. The Epworth scale varies between 0 and 18. The same for oculometry parameters, the latency have its minimum around 200, duration around 40, while amplitude varies between 5 and 20. There are different methods to get all attributes values to into the same scale and we decided to check the results of the predictions using different scaling approaches:

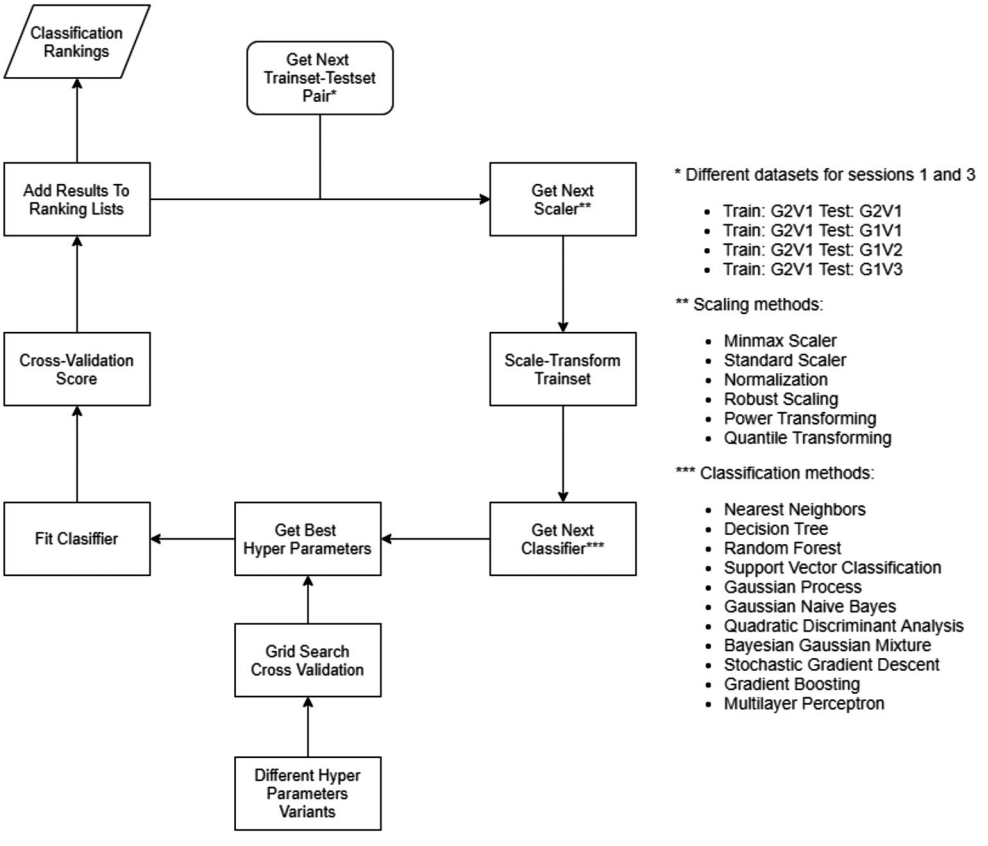

Figure 2. Schema of classification ranking procedure.

- Min–max scaling – where a feature values are shifted and rescaled by subtracting the minimum and dividing by the difference between maximum and the minimum, so that they end up ranging from 0 to 1.
- Standard scaling (standardization) – where a feature mean values are subtracted and divided by the variance, so resulting distribution has unit variance (standard normal distribution) untied from the predefined range in contrast to the min–max scaling.
- Normalization – where a feature values, by each data unit are scaled independently of the other to have unit norm.
- Robust scaling – where the feature medians are removed, then centred and scaled according to the quantile range by computing the relevant statistics on the samples, which creates distribution that are robust to outliers.
- Power transforming – where a feature values are power transformed to make distribution more like Gaussian by applying zero-mean and unit-variance normalization and by estimating the optimum values for variance stabilization and minimal skewness by the maximum likelihood
- Quantile transformation (non-linear) – where a feature values are mapped to uniform distribution by estimation of the cumulative distribution function and associated quantile function, spreading out the most frequent values reducing the impact of outliers.

Each of scalers was tested in pair with one of the classifiers. We list all tested methods with a short description below:

- Nearest Neighbours (NN) – acts as a uniform interface to three different NN algorithms: BallTree, KDTree, and Brute-Force, determining the best approach from the training data by the 'auto' mode.
- Decision Tree (DT) – where the classification process is based on learning using simple if-then-else decision rules inferred from the data features.
- Random Forest (RF) – where the number of randomized decision trees are fit on various subsamples of the dataset averaging results on the output.
- Support Vector Classification (SVC) – implementation of support vector machines (SVM) where a subset of training points creates vectors determining the decision boundaries.
- Gaussian Process (GP) – where the classification process is based on the Laplace approximation.
- Gaussian Naive Bayes (GNB) – where the likelihood of the features is assumed to be Gaussian.
- Quadratic Discriminant Analysis (QDA) – where the likelihood of the features is determined by the quadratic plot of decision boundaries (more flexible than the linear plot).
- Bayesian Gaussian Mixture (BGM) – implements variational Bayesian estimation of a function comprising several Gaussians.
- Stochastic Gradient Descent (SGD) – fits a linear SVM learning with the gradient of the loss estimated each sample at a time updating the model with a decreasing strength schedule.
- Gradient Boosting (GB) – builds an additive model where in each stage, trees are fit on the negative gradient of the deviance loss function.

In order to compare different results, in addition to the Accuracy, we have also used the F1 metrics. The F1 provides a single score that balances in one value two concerns suitable for classification of imbalanced datasets – precision and recall. Precision is the ratio of correctly predicted positive examples divided by the total number of positive examples that were predicted (True Positives/True Positives + False Positives). It quantifies the number of true positive predictions and, therefore, calculates the accuracy for the minority class. The recall is the ratio of correctly predicted positive examples made out of all positive predictions (True Positives/(True Positives + False Negatives)). Unlike precision which reveals correct positive predictions out of all positive predictions, recall provides a measure of missed positive predictions. In this way, recall might be understood as a coverage of the positive class. Therefore, we define F1 as follows:

$$\mathrm{F1} = 2{*}(\mathrm{Precision}{*}\mathrm{Recall})/\ (\mathrm{Precision} + \mathrm{Recall}) \tag{1}$$

3. Results

For the first group of PD patients, we have performed three tests, every half-year, whereas for the second group of more advanced PD, we have measured only one time. The mean age of the first group (G1) was 57.8 ± 13 (SD) years with disease duration 7.1 ± 3.5 years; UPDRS MedOff/On was 48.3 ± 17.9 and 23.6 ± 10.3 for the first visit (V1); 57.3 ± 16.8 and

27.8 ± 10.8 for the second visit (V2), and 62.2 ± 18.2 and 25 ± 11.6 for the third visit (V3). The second group (G2) of patients was more advanced with the mean age 53.7 ± 9.3 years and disease duration 10.25 ± 3.9 years; UPDRS MedOff/On was 62.1 ± 16.1 and 29.9 ± 13.3 measured one time only. Data were placed in four information tables: G1V1, G1V2, G1V3, and G2V1.

Table 1 has 46 rows: 23 patients measured in two sessions each. Condition attributes patient number P#, S# session number, t_dur – disease duration, PDQ39, Epworth (as above), RS parameters (above). The decision attribute is UPDRS that is proportional to the disease progression; it increases from G1V1 to G1V3 and it will be referred to G2V1.

3.1. Rough set approach

In the next step, Table 1 is discretized by RST and part of the table for G1V1 patients in Table 2. Notice that some less significant attributes were, by the algorithm of RSES 2.2, discarded: RSDur, RSAmp, and RSVel – duration, amplitude, and velocity of RS.

By means of the discretization RSES software RSES 2.2 (see Methods), UPDRS was divided into four ranges: '(−Inf, 33.5)', '(33.5, 43.0)', '(43.0, 63.0)', and '(63.0, Inf)'. All other attributes, except symbolic attributes P# (number given to each patient) and S# (session number) were also discretized (Table 2).

Cross-validation (6-fold) based on the decomposition tree of the first visit G1V1 data gave the global accuracy 0.896 and global coverage 0.35. Prediction, based on rules from G1V1, of UPDRS in G1V2 and G1V3 gave global accuracy 0.7 with coverage 1, and these results do not indicate time-related disease progression.

For G2V1 group, rules from G1V1 gave global accuracy 0.64 and coverage 0.5. However, it was more interesting to estimate G1V1–G1V3 from the other more advanced model group of patients G2V1.

This way we can follow our c-granular approach (Figure 1) where the model are granules of attributes of G2V1 that might predict PD time: G1V1, G1V3, G1V3 development. With the help of RSES, we have found rules describing relationships between condition and decision attributes in G2V1 and we are using these rules to predict disease symptoms in the G1 group for each visit V1, V2, and V3. If the disease progression has direction going to the model (G2V1) group, then predictions should increase with the time of the disease. We demonstrate our predictions in Tables 3–5.

In table 3, predictions of the UPDRS for the first visit group of patients (G1V1) are given. Notice that in this and Tables 4 and 5, we could not predict UPDRS between 33.5 and 43. The accuracy in Table 3 was <60%, but it increases for each following visit: G1V2 has global

Table 1. Part of the decision table for three G1V1 patients.

P#	Ses	tdur	PDQ39	Epworth	RSLat	RSDur	RSAmp	RSVel	UPDRS
10	1	5.3	90	17	205	51	9.8	343	58
10	2	5.3	90	17	182	56	10	333	35
11	1	15	122	8	245	55	12	503	57
11	2	15	122	8	266	55	12	431	40
12	1	5.5	20	3	178	54	10	421	25
12	2	5.5	20	3	161	58	13	505	15
13	1	4.8	68	9	299	59	13	472	46
13	2	4.8	68	9	234	57	11	367	26

Table 2. Discretized-table (Table 1) for three G1V1 patients.

P#	Ses	tdur	PDQ39	Epworth	RSLat	RSDur	RSAmp	RSVel	UPDRS
10	1	'(50.5,Inf)'	'(−Inf,5.65)'	'(14,Inf)'	'(−Inf,264)'	*	*	*	'(43,63)'
10	2	'(−Inf,5.65)'	'(50.5,Inf)'	'(14,Inf)'	'(−Inf,264)'	*	*	*	'(33.5,43)'
11	1	'(5.65,Inf)'	'(50.5,Inf)'	'(−Inf,14)'	'(−Inf,264)'	*	*	*	'(43,63)'
11	2	'(5.65,Inf)'	'(50.5,Inf)'	'(−Inf,14.)'	'(264,Inf)'	*	*	*	'(33.5,43)'
12	1	'(−Inf,5.65)'	'(−Inf,50.5)'	'(−Inf,14)'	'(−Inf,264)'	*	*	*	'(−Inf,33.5)'
12	2	'(−Inf,5.65)'	'(−Inf,50.5)'	'(−Inf,14)'	'(−Inf,264)'	*	*	*	'(−Inf,33.5)'
13	1	'(−Inf,5.65)'	'(50.5,Inf)'	'(−Inf,14.0)'	'(264,Inf)'	*	*	*	'(43,63)'
13	2	'(−Inf,5.65)'	'(50.5,Inf)'	'(−Inf,14.)'	'(−Inf,264)'	*	*	*	'(−Inf,33.5)'

accuracy 68% and G1V3 – 86%. Therefore, patients' symptoms become more similar to the G2V1 group with time.

An important part in these estimations is to find rules that are enough general to be patient independent (there are different patients in the G1 and G2 groups) and not too general in order to find differences between different visits.

There were altogether 71 rules, e.g.

$$(\text{Ses} = 2)\&\ (\text{PDQ39} = '(\text{-Inf, 50.5})') = >(\text{UPDRS} = \nwarrow(\text{-Inf, 33.5})'\ [10]) \tag{2}$$

$$\begin{aligned}&(\text{Ses} = 2)\&\ (\text{Epworth} = '(\text{-Inf, 14.0})')\&\ (\text{RSLat} = '(264.0, \text{Inf})')\\ &= >(\text{UPDRS} = '(63.0, \text{Inf})'[4])\end{aligned} \tag{3}$$

$$\begin{aligned}&(\text{dur} = '(5.65, \text{Inf})')\&\ (\text{Ses} = 2)\&\ (\text{RSLat} = '(\text{-Inf, 264.0})')\\ &= >(\text{UPDRS} = '(\text{-Inf, 33.5})'[14])\end{aligned} \tag{4}$$

$$\begin{aligned}&(\text{Ses} = 1)\&\ (\text{PDQ39} = '(\text{-Inf, 50.5})')\&\ (\text{RSLat} = '(264.0, \text{Inf})')\\ &= >(\text{UPDRS} = '(63.0, \text{Inf})'[1])\end{aligned} \tag{5}$$

Equations (2–4) were for *Ses=2* (patient on medication) and they were fulfilled by 10 (2), 4 (3), and 14 (4) cases, whereas equation (5) was for one case only.

We can read Equation (1) as: for patients on medication (*Ses=2*) and with *PDQ39* (quality of life test result) <50.5 then his/her *UPDRS will be <33.5*.

3.2. Fuzzy rough set approach

We have obtained our predictions using the generalized fuzzy rough set rules (GFRS) with aggregation by the t.norm **Łukasiewicz**, similarity expressed as tolerance equation (3) (modified Gaussian from Riza et al. (2014)), and implicator – Łukasiewicz; α precision

Table 3. Confusion matrix for UPDRS of G1V1 patients by rules obtained from G2V1 patients with RST.

	Predicted				
Actual	'(63.0, Inf)'	'(33.5, 43.0)'	'(43.0, 63.0)'	'(−Inf, 33.5)'	ACC
'(63.0, Inf)'	2.0	0.0	0.0	0.0	1.0
'(33.5, 43.0)'	1.0	0.0	1.0	1.0	0.0
'(43.0, 63.0)'	6.0	0.0	1.0	0.0	0.14
'(−Inf, 33.5)'	3.0	0.0	2.0	17.0	0.77
TPR	0.17	0.0	0.25	0.94	

TPR, true positive rates for decision classes; ACC, accuracy for decision classes: the global accuracy was 0.59 and global coverage was 0.74.

Table 4. Confusion matrix for UPDRS of G1V2 patients by rules obtained from G2V1 patients with RST.

	Predicted				
actual	'(63.0, Inf)'	'(33.5, 43.0)'	'(43.0, 63.0)'	'(−Inf, 33.5)'	ACC
'(63.0, Inf)'	3.0	0.0	1.0	0.0	0.75
'(33.5, 43.0)'	0.0	0.0	2.0	1.0	0.0
'(43.0, 63.0)'	4.0	0.0	0.0	0.0	0.0
'(−Inf, 33.5)'	0.0	0.0	1.0	16.0	0.94
TPR	0.43	0.0	0.0	0.94	

TPR, true positive rates for decision classes; ACC, accuracy for decision classes: the global accuracy was 0.68 and global coverage was 0.61.

Table 5. Confusion matrix for UPDRS of G1V3 patients by rules obtained from G2V1 patients with RST.

	Predicted				
Actual	'(63.0, Inf)'	'(33.5, 43.0)'	'(43.0, 63.0)'	'(−Inf, 33.5)'	ACC
'(63.0, Inf)'	3.0	0.0	0.0	0.0	1.0
'(33.5, 43.0)'	0.0	0.0	1.0	2.0	0.0
'(43.0, 63.0)'	0.0	0.0	0.0	0.0	0.0
'(−Inf, 33.5)'	0.0	0.0	2.0	16.0	0.94
TPR	1.0	0.0	0.0	1.0	

TPR, true positive rates for decision classes; ACC, accuracy for decision classes: the global accuracy was 0.86 and global coverage was 0.48.

was 0.05. As the decision attribute must be nominal, we have chosen classes that are similar to already used in our previous section: '(−Inf, 33.5)' = '1'; '(33.5, 43.0 ' ='2'; '(43.0, 63.0)' ='3',: '(63, Inf)' = '4'. The examples of FRST rules are below:

$$(\text{Ses} = 1)\& \ (\text{Epworth} = '16')\& \ (\text{RSLat} = '192') => (\text{UPDRS} = '3') \tag{6}$$

$$(\text{Ses} = 2) => (\text{UPDRS} = '1') \tag{7}$$

$$(\text{Ses} = 1)\& \ (\text{Epworth} = '1')\& \ (\text{RSLat} = '289') => (\text{UPDRS} = '4' \ ) \tag{8}$$

We can read Equation (6) as for patients without medication (Ses=1) and with Epworth (quality of sleep test result) ∼16 and saccade latency about 192 then his/her UPDRS will be ∼3 (between 34 and 63).

FRST rules have some similarities to RST rules, but there are not 'crisp', there are fuzzy and more difficult to interpret as their fuzziness are not given directly because they are dependent on aggregation, tolerance, and implicator equations. Their advantage to RST rules is that they cover all cases with the global coverage =1. As described earlier, we have found FRST rules for our model group of patients G2V1 and applied these rules to other groups G1V1, G1V2, and G1V3 (Tables 6–8).

Accuracies of our FRST predictions were inferior in comparison to RST predictions, but as before accuracy was increasing with each visit: G1V1 – accuracy was <40%, for G1V2 – 54%, and for G1V3 visit was >60%. Also, notice that we did not get right predictions for UPDRS nominal values 2 and 3 that were between (33.5 and 43) and between (43 and 63), but we have got relatively good predictions for classes 1 and 4 where accuracy for decision classes ACC was almost for all estimations near 1.

Table 6. Confusion matrix for UPDRS of G1V1 patients by rules obtained from G2V1 patients with FRST.

	Predicted				
Actual	'1'	'2'	'3'	'4'	ACC
'1'	19	0	0	5	0.79
'2'	2	0	0	4	0.0
'3'	2	0	0	11	0.0
'4'	0	0	0	3	1.0
TPR	0.826	0.0	0.0	0.13	

TPR, true positive rates for decision classes; ACC, accuracy for decision classes: the global accuracy was 0.38 and global coverage was 1.

Table 7. Confusion matrix for UPDRS of G1V1 patients by rules obtained from G2V1 patients with FRST.

	Predicted				
Actual	'1'	'2'	'3'	'4'	ACC
'1'	18	0	0	0	1.0
'2'	4	0	0	2	0.0
'3'	2	0	0	13	0.0
'4'	0	0	0	7	1.0
TPR	0.75	0.0	0.0	0.32	

TPR, true positive rates for decision classes; ACC, accuracy for decision classes: the global accuracy was 0.54 and global coverage was 1.

Table 8. Confusion matrix for UPDRS of G1V3 patients by rules obtained from G2V1 patients with FRST

	Predicted				
Actual	'1'	'2'	'3'	'4'	ACC
'1'	17	0	0	0	1.0
'2'	5	0	0	3	0.0
1	2	0	0	8	0.0
'4'	0	0	0	11	1.0
TPR	0.74	0.0	0.0	0.5	

TPR, true positive rates for decision classes; ACC, accuracy for decision classes: the global accuracy was 0.54 and global coverage was 1.

3.3. Results comparison with different AI methods

Our additional purpose was to compare our results with different methods of classifications using the same, initial dataset. To achieve it, we have discretized G2V1 and G1V1–3 in the same way as described above, in Section 3.1. At first, we have evaluated data sets without division into different sessions. Our results were quite good (Accuracy: G1V1 = 0.70, G1V2 = 0.75, G1V3 = 0.71, F1:G1V1 = 0.72, G1V2 = 0.74, G1V3 = 0.72); however, they did not show the previously found upward trend, although the results of UPDRS patients had this trend. Our results on classification self-test (G2V1 as train and test dataset) were always much worst and stayed mostly <0.70. Finally, we executed our ranking procedure on data sets divided into different sessions.

Table 9 presents results obtained for Session 1. The new results have reflected the upward trend; however, it seems that their accuracy values were different from that obtained with RSES (G1V1 = 0.59, G1V2 = 0.68, G1V3 = 0.86) where we can see a large accuracy shift between visits 2 and 3. They seem to be more balanced. Unfortunately, for the second session, the classification didn't perform the same way and most of the best results were overfitted. Table 10 presents the results obtained for the second session. The solution

Table 9. Classification results: trainset = G2V1, test set = G1V1–3, Session 1.

	G1V1	G1V2	G1V3
Accuracy	0.59	0.73	0.78
Best classification method:	Random Forest Robust Scaler	SGD Classifier Quantile Transformer	Nearest Neighbours Min–Max Scaler
F1	0.57	0.70	0.78
Best classification method:	SGD Classifier Standard Scaler	SVC Standard Scaler	Extra Trees Normalizer

Table 10. Classification results: trainset = G2V1, test set = G1V1–3, Session 2.

	G1V1	G1V2	G1V3
Accuracy	1.00	0.87	0.87
Best classification method	Decision Tree Standard Scaler	SGD Classifier Standard Scaler	Random Forest Min–Max Scaler
F1	1.00	0.83	0.86
Best classification method	Decision Tree Standard Scaler	SGD Classifier Standard Scaler	Random Forest Robust Scaler

to the overfitting problem could be in additional pre-processing techniques, i.e. in oversampling, which would offset the effects of uneven samples. However, the purpose was to compare different methods using the same initial data input and here RSES performs much better with differences in the number of class representations. Results of our earlier experiments showed that RSES is much better than other data sets in dealing with predicates based on a small number of samples, which may be the reason for its results (Śledzianowski et al., 2018).

4. Discussions

In the first part of our work, we have used c-granular computing to estimate disease progression in time (longitudinal study) of patients with PD. As in each individual, PD symptoms and their developments are different (No two people face Parkinson's in quite the same way), we would like to know if we could predict a particular patient progression by looking to more advanced group of patients.

We have used GrC with RST and FRST. RST looks into 'crisp' granules and estimates objects by upper and lower approximations that determine the precision of the description as dependent from properties of granules. Therefore, RST can give very precise estimation but not for all objects (measures). That can be seen in our results where we can precisely predict symptoms (measured as UPDRS) of patients, but not all of them (global coverage <1). If we make our granules fuzzy (not crisp), they can describe properties of all objects (measures) with global coverage =1, but less precisely. Our present results might be a good example of these differences and comparison with the statistical data (Szlufik et al., 2019).

In the second part, we have compared the GrC approach with the scikit-learn methods. It has revealed a problem in the inability to classify data of the second session without the overfitting. One possible reason might be the differences between the sizes of patient groups. In data sets, both session groups were almost equal in size, but varied in numbers of objects in different UPDRS classes. We find out that as the main source for overfitting, as for the second session, the lowest UPDRS class ($\leq$33.5) were

overrepresented and contained ∼70% of all samples. In contrast, the largest UPDRS group in data sets of session 1 was never larger than ∼50%, which could explain its healthier results. This effect can be seen especially for dataset G1V1 session 2; however, for the two remaining data sets, results in F1 metrics (which is more suitable for data sets imbalanced in class numbers) look similar to the expected results for last two visits. The results increased from 83% to 86% giving the same final result (last visit) as when using the c-granule approach. This prompts us to conclude that the GrC/RST/FRST is just better than other tested methods at deducing results for unequal groups of observations and also for small groups in numbers. Classifications using c-granule in such cases allow us to get around the oversampling, which sometimes can have a bad effect on the authenticity of the results, as it creates synthetic observations based only on similarities to known representations.

Another important property of c-granule approach is similarities between findings all-important aspects (symptoms) of the disease and recognition of the complex object (Figure 1). In the visual brain, we are trying to infer not clear visible object's properties from attributes we have classified from α to β and back to a new part (q) of the object (Figure 1). However, there is a very important neurological principle of our vision – the Model (Przybyszewski, 2008, 2019). It consists what we have learned from the world, all particular environments, and known objects. We are able to precisely classify a complex, unknown object as we are tuning and comparing it particular attributes in many different levels (and even with different logics (Pawlak, 1991)). The model is an important part of our approach.

Our model is determined by attributes of the more advanced group of PD patients (in our case, G2V1 group). As it is illustrated in Figure 3, granules describing different disease stages might develop or stay constant. Blobs represent symptoms of all our patients who are at the actual disease stage. They all together represent c-granule of all patients (top of the figure). Disease development is related to the changes in the size of different blobs and also relationships (interactions) between symptoms, which is represented as the change of the granule shape (lower part of the figure). In each disease stage, we are comparing actual symptoms with the model and look for the similarities. We have demonstrated on a group of over 20 patients that even if each one has different symptoms, their path (c-granule) is going in the direction to our model. Changing treatment, it might push patient symptoms to the different path. In order to test such options, we need several models and to measure how to change the treatment to direct patients to a different model. By testing several different patients' groups, we have demonstrated that the certain long-lasting treatments can change disease develop to new directions that are not similar to classical medication treatments (Przybyszewski et al., 2018b). In such cases, one possible solution is to increase the number of granules (dimension of attributes) but adding new attributes that might 'sense' new direction of the disease development. We are actually testing the influence of the depression on the direction of the symptoms changes, as depression is characteristic not only for PD but also for more common Alzheimer's disease (AD) where late (after 65 years of age) onset AD (LOAD) is in 50% related to depression.

In summary, we have demonstrated that by using an approach similar to the visual brain intelligence might give us a new way of looking into similarities between different groups of patients. In addition, we might see longitudinal studies as c-granules and measure symptoms by distance to the Model (advance stage of the disease).

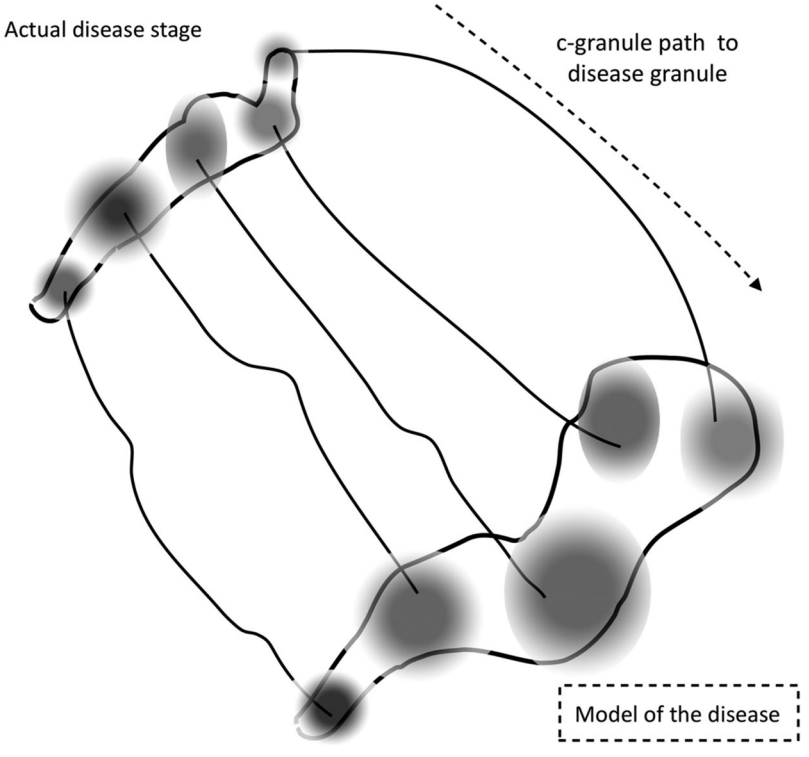

Figure 3. C-granule path in PD disease development from the actual stage (top) to the advanced disease stage, which in our approach was a Model towards which patients' symptoms were moving.

A part of this study was published earlier (Przybyszewski, 2019) and it has discussed during International Conference on Computational Collective Intelligence 4–6 September in Henday, France. We have extended previous communication by several different data mining methods and compared their results: added to the Abstract, to the Methods section as paragraph 2.2, added to the Results section as paragraph 3.3 as well we have added another paragraph to the Discussions section.

Acknowledgements

The original article's content has been supplemented with the description of the analysis using alternative methods to that described in the previous version, with a comparison of the results. We have added new parts to the following sections: Abstract – concentrated information on new methods and comparison with previous results. Methods – detailed description of new methods of classification applied on the data set, including description of the procedure and new Figure 2 (Section 2.2). Results – comparison of the results obtained with methods described in Section 2.2 including new Tables 9 and 10 (Section 3.3). In the discussion section – the new methods were compared to our original GrC/RST/FRS approach.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Andrzej W. Przybyszewski, MA mathematics, PhD neurophysiology, DSc neuroscience is currently Professor of Informatics at the Polish-Japanese Academy of Information Technology, Warsaw, Poland and Assistant Professor of Neurology at the University of Massachusetts, Worcester, US. The long-term goal of his research is to understand how the brain integrates streams of the sensory information and converts them into actions. Actually, he is involved in the clinical research related to the neurodegenerative diseases(ND). The objective of these projects is to replace qualitative doctors' measurements and intuitions by quantitative and objective methods to create intelligent data base system that becomes a standard and with data mining help may lead to discovery of new ND biomarkers.

Albert Śledzianowski, MBA in "Innovations and Data Analysis" at the Institute of Computer Sciences of the Polish Academy of Sciences, Warsaw (in cooperation with the Utah Valley University, USA). PhD candidate at Polish-Japanese Academy of Information Technology, Warsaw. His research focuses on analyzing signals from eye-trackers and other biometric devices to distinguish different types of emotions and healthy people from patients with neurodegenerative diseases, like in the Parkinson's patients.

ORCID

Andrzej W. Przybyszewski http://orcid.org/0000-0002-0156-7856

References

Dubois, D., & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, *17*(2–3), 191–209. https://doi.org/10.1080/03081079008935107

Pawlak, Z. (1991). *Rough sets – theoretical aspects of reasoning about data*. Kluwer Academic Pub.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Przybyszewski, A. W. (2008). The neurophysiological bases of cognitive computation using rough set theory. In J. F. Peters, A. Skowron, & H. Rybinski (Eds.), *Transactions on rough sets IX* (pp. 287–317). LNCS 5390.

Przybyszewski, A. W., (2019). Parkinson's disease development prediction by C-granule computing. In N. T. Nguyen (Ed.), *ICCCI 2019, LNAI 11683* (pp. 296–306). https://doi.org/10.1007/978-3-030-28377-3_24

Przybyszewski, A. W. (2019). SI: SCA measures – fuzzy rough set features of cognitive computations in the visual system. *Journal of Intelligent and Fuzzy Systems*, *36*, 3155–3167. https://doi.org/10.3233/JIFS-18401

Przybyszewski, A. W., Kon, M., Szlufik, S., Szymanski, A., Habela, P., & Koziorowski, D. M. (2016). Multimodal learning and intelligent prediction of symptom development in individual Parkinson's patients. *Sensors*, *16*(9), 1498. https://doi.org/10.3390/s16091498

Przybyszewski, A. W., Szlufik, S., Habela, P., & Koziorowski, D. M. (2018a). Fuzzy RST and RST rules can predict effects of different therapies in Parkinson's disease patients. In M. Ceci (Ed.), *ISMIS 2018, LNAI 11177* (pp. 409–416). Springer.

Przybyszewski, A. W., Szlufik, S., Habela, P., & Koziorowski, D. M. (2018b). Multimodal learning determines rules of disease development in longitudinal course with Parkinson's patients. In R. Bembenik (Ed.), *Intelligent methods and big data in industrial applications, studies in Big data 40*. https://doi.org/10.1007/978-3-319-77604-0_17

Riza, L. S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Slezak, D., & Benítez, J. M. (2014). Implementing algorithms of rough set theory and fuzzy rough set theory in the R package RoughSets. *Information Sciences*, *287*, 68–89. https://doi.org/10.1016/j.ins.2014.07.029

Skowron, A., & Dutta, S. (2018). Rough sets: Past, present, and future. *Natural Computing*, *17*(4), 855–876. https://doi.org/10.1007/s11047-018-9700-3

Śledzianowski, A., Szymański, A., Szlufik, S., & Koziorowski, D. (2018). Rough set data mining algorithms and pursuit eye movement measurements help to predict symptom development in Parkinson's disease. In N. Nguyen, D. Hoang, T. P. Hong, H. Pham, & B. Trawiński (Eds.), *Intelligent information and database systems. ACIIDS 2018. Lecture Notes in computer Science* (Vol. 10752, pp. 428–435). Springer.

Szlufik, S., Przybyszewski, A., Dutkiewicz, J., Mandat, T., Habela, P., & Koziorowski, D. (2019). Evaluating reflexive saccades and UDPRS as markers of deep brain stimulation and best medical treatment improvements in Parkinson's disease patients: A prospective controlled study. *Polish Journal of Neurology and Neurosurgery*, *53*(5), 341–347.

Zadeh, L. A. (2002). From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions. *International Journal of Applied Mathematics and Computer Science*, *12*, 307–324.