Artificial Intelligence In Medicine xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Symptoms and medications change patterns for Parkinson's disease patients stratification

Anita Valmarska[a,b,*], Dragana Miljkovic[a], Spiros Konitsiotis[c], Dimitris Gatsios[d], Nada Lavrač[a,b], Marko Robnik-Šikonja[e]

[a] Jožef Stefan Institute, Jamova 39, Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Jamova 39, Ljubljana, Slovenia
[c] University of Ioannina, Medical School, Department of Neurology, Ioannina, Greece
[d] University of Ioannina, Department of Biomedical Research, Ioannina, Greece
[e] University of Ljubljana, Faculty of Computer and Information Science, Slovenia

ARTICLE INFO

Keywords:
Parkinson's disease
Analysis of disease progression
Multitask learning
Analysis of medications treatment
Symptoms impact

ABSTRACT

Quality of life of patients with Parkinson's disease degrades significantly with disease progression. This paper presents a step towards personalized management of Parkinson's disease patients, based on discovering groups of similar patients. Similarity is based on patients' medical conditions and changes in the prescribed therapy when the medical conditions change. We present two novel approaches. The first algorithm discovers symptoms' impact on Parkinson's disease progression. Experiments on the Parkinson Progression Markers Initiative (PPMI) data reveal a subset of symptoms influencing disease progression which are already established in Parkinson's disease literature, as well as symptoms that are considered only recently as possible indicators of disease progression by clinicians. The second novelty is a methodology for detecting patterns of medications dosage changes based on the patient status. The methodology combines multitask learning using predictive clustering trees and short time series analysis to better understand when a change in medications is required. The experiments on PPMI data demonstrate that, using the proposed methodology, we can identify some clinically confirmed patients' symptoms suggesting medications change. In terms of predictive performance, our multitask predictive clustering tree approach is mostly comparable to the random forest multitask model, but has the advantage of model interpretability.

1. Introduction

Data mining algorithms have been successfully used to learn predictive models and to discover insightful patterns in the data. Predictive and descriptive data mining approaches have been successfully used also in medical data analysis. The use of data mining methods may improve diagnostics, disease treatment and detection of causes of diseases. In personalized healthcare [16], data mining can be used to improve drug recommendations and medical decision support, leading to reduced costs of medical treatment. The discovered patterns can provide the clinicians with new insights regarding the status of the treated patients and can support decisions regarding therapy recommendations.

Parkinson's disease is the second most common neurodegenerative disease (after Alzheimer's disease) that affects many people worldwide. Due to the death of nigral neurons, patients experience both motor and non-motor symptoms, affecting their quality of life. The reasons for the cell death are still poorly understood, and there is currently no cure for Parkinson's disease. Physicians try to manage patients' symptoms by introducing medications therapies, using antiparkinson medications. Physicians need to carefully prescribe medications therapies since the prolonged intake—in particular of higher dosages of antiparkinson medications—can have significant side-effects.

Changes of the status of Parkinson's disease patients through time is a result of the natural progression of the disease and the medications that the patients are prescribed in order to keep their status stable as long as possible. Physicians follow the guidelines for therapy prescription and the response of patients to medications is usually recorded in clinical studies using simple statistical methods. For example, in our previous work [43], we describe the disease progression of a patient who starts with a good status and recieves only one type of antiparkinson medications (MAO-B inhibitors). As the disease progressed

* Corresponding author at: Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia.
E-mail addresses: anita.valmarska@ijs.si (A. Valmarska), dragana.miljkovic@ijs.si (D. Miljkovic), skonitso@uoi.gr (S. Konitsiotis), dgatsios@cc.uoi.gr (D. Gatsios), nada.lavrac@ijs.si (N. Lavrač), marko.robnik@fri.uni-lj.si (M. Robnik-Šikonja).

https://doi.org/10.1016/j.artmed.2018.04.010
Received 31 October 2017; Received in revised form 26 April 2018; Accepted 30 April 2018

and the patients' motor symptoms worsened, the clinician started the treatment with another type of medications (dopamine agonists) and was successful in keeping the motor symptoms as tremor, bradykinesia, and rigidity stable for about two years. As the effectiveness of these medications wore off, the clinician was forced into introducing the third group of medications (levodopa).

To the best of our knowledge, data mining techniques have not yet been used for analyzing clinicians' decisions of changing drug prescription as a reaction to the change of patients' symptoms when using antiparkinson medications through prolonged periods of time. A possible reason for little data mining research in the field of Parkinson's disease progression may be the unavailability of a monotone measure/test that determines the stages of Parkinson's disease, as the currently used Hoehn and Yahr scale [15] determines the stages of Parkinson's disease through a subjective evaluation of clinicians and response of patients to the prescribed medications. This paper uses multitask learning with predictive clustering trees [4] on short time series data—describing the patients' status at multiple time points—in order to determine the symptoms that trigger the physicians' decisions to modify the medications therapy. We consider trigger symptoms to be the symptoms that a patient cannot tolerate and the physician is pressed to change the medications therapy in order to control them. The proposed methodology addresses the task of determining subgroups of patients with similar symptoms and therapy. As each patient usually receives drugs from several different groups of medications, predicting their changes with multitask learning can lead to improved control over drug interactions.

This work significantly extends the conference paper [42] by extending the experiments, results, and their medical interpretation. We introduce a novel algorithm for determining the symptoms that have the highest influence on the change of the patients' status, which extends the methodology used to determine the status of Parkinson's disease patients based on an extensive set of symptoms [43,44]. We present a solution to the problem of feature ranking with the aim of finding the most influential symptoms affecting the changed status of patients, which may help the clinicians to focus on a small set of the most important symptoms, whose medications treatment would lead to a more stable status of the patient. Our research provides references to the already known findings in Parkinson's disease literature, as well as references to findings about possible influential symptoms that have only recently started being discussed in the Parkinson's disease medical community as early indicators of Parkinson's disease progression. We significantly extend the experiments with PCT models, analyze different sets of attributes, and discuss reasons for particular medications dosage change patterns from the medical perspective. The consulting clinician takes into account trigger symptoms from the trees as well as the patients' overall status concerning their motor and non-motor symptoms.

This paper is structured into six sections. After presenting the background and related work in Section 2, Section 3 describes the Parkinson's Progression Markers Initiative (PPMI) symptoms data set [24], together with the data describing the medications used for symptoms control, available from the so-called PPMI concomitant medications log data set. Section 4 outlines our methodology. In Section 4.1 we present a new algorithm for determining the most influential symptoms. Section 4.2 proposes a methodology for analyzing Parkinson's disease symptoms by learning predictive clustering trees from short data sequences. Results are presented in Section 5. Section 5.1 presents the most influential symptoms, while Section 5.2 describes the results of applying the proposed methodology to the detection of changes in symptoms-based clustering of patients, connected to the changes in medications therapies and finding patterns of symptoms which trigger therapy modifications. In Section 5.3 we explore the influence of the above-mentioned symptoms on clinicians' decisions regarding the modification of dosages of prescribed medications. Finally, Section 6 presents the conclusions and plans for further work.

2. Background and related work

Our work is related to several subareas of data analysis. We first present approaches to Parkinson's disease data analysis in Section 2.1 and Parkinson's disease progression in Section 2.2. In Parkinson's disease management, several groups of medications are used together. We apply multitarget modeling with predictive clustering trees to capture their joint effects and discuss related work from this area in Section 2.3. We are interested in the importance of symptoms affecting the overall status of the disease, which is a problem addressed in feature ranking/evaluation research. We compare and contrast the algorithm we propose with existing approaches in Section 2.4.

2.1. Parkinson's disease data analysis

Data mining research in the field of Parkinson's disease include classification of Parkinson's disease patients, detection of Parkinson's disease symptoms (computational assessment from e.g., wearable sensors), and detection of subtypes of Parkinson's disease patients, as discussed below.

Due to the overlap of Parkinson's disease symptoms with other diseases, only 75% of clinical diagnoses of Parkinson's disease are confirmed to be idiopathic Parkinson disease at autopsy [17]. Classification techniques offer decision support to specialists by increasing the accuracy and reliability of diagnosis and reducing possible errors. Gil and Johnson [13] use Artificial Neural Networks (ANN) and Support Vector Machines (SVM) to distinguish Parkinson's disease patients from healthy subjects. Ramani and Sivagami [29] compare the effectiveness of different data mining algorithms in the diagnosis of Parkinson's disease patients, where the data set consists of 31 people, 23 of which are Parkinson's disease patients.

Tremor is a symptom strongly associated with Parkinson's disease. Several approaches to computational assessment of tremor have been proposed. Methods such as time series analysis [41], spectral analysis [33], and non-linear analysis [33] have addressed tremor detection and quantification. Many recent works are based on body fixed sensors (BFS) for long-term monitoring of patients [26].

Parkinson's disease is a heterogeneous neurodegenerative condition with different clinical phenotypes, genetics, pathology, brain imaging characteristics and disease duration [11]. This variability indicates the existence of disease subtypes. Using k-means clustering, Ma et al. [23] identify four groups of Parkinson's disease patients which is consistent with the conclusions from [22,31]. This division of Parkinson's patients into homogeneous subgroups was done on symptoms data recorded only once for each patient. It does not take into account the progression of the disease and changes in the patients' status due to the medications treatment. Our analysis uses a different data set (see Section 3) which allows us to take these issues into account.

Classification and clustering models usually focus on diagnosing new patients. None of the listed methods follow the progression of the disease, and to the best of our knowledge, no data mining research in the field of Parkinson's disease analyzed the development of the disease in combination with the medications that the patients receive. Identification of groups of patients based on how they react to a certain therapy can be helpful in the assignment of personalized therapies and more adequate patient treatment. To this end, we propose a methodology for determining trigger symptoms, which influence the physician's decision about therapy modification. In addition, our methodology aims to uncover the side-effects of the modified therapy.

2.2. Parkinson's disease progression

There are no specific medical tests to determine the progression of Parkinson's disease for an individual patient. Currently, the clinicians commonly use the Hoehn and Yahr scale system [15] to describe the progression of Parkinson's disease symptoms. This evaluation can be

seen as the clinicians' aggregate evaluation of the patient's motor status. Patient's status changes through time and even though the status of the patient is going to get worse during their treatment, there are periods where carefully prescribed medications therapies can cause an improvement of the patient's overall status. This improvement can be reflected in both the patient's *motor* and *non-motor* symptoms.

In our earlier work [43,44], we first used unsupervised learning (k-means clustering) to divide Parkinson's disease patients from the PPMI study into three groups with similar severity of their *motor* and *non-motor symptoms*. We then applied supervised classification rule learning techniques to obtain descriptions for each of the obtained groups. The results suggested that these groups can be described with the aggregated severity of their motor symptoms. In addition, the rules also contained the information about the status of their non-motor symptoms.

The three groups of patients were ordered according to the sum of evaluation values for their motor symptoms from MDS-UPDRS Part III (NP3SUM). The first cluster (*cluster 0*) consisted of patients whose motor symptoms were considered as normal, and the sum of MDS-UPDRS Part III was below 22. The second cluster (*cluster 1*) contained patients whose motor symptoms were slightly worse, and the sum of MDS-UPDRS Part III was between 22 and 42. In the third cluster (*cluster 2*) there were the patients whose sum of evaluation symptoms values from MDS-UPDRS Part III were higher than 42. Note that based on the sum of motor symptoms, the status of patients from *cluster 2* is worse than the status of patients from *cluster 0* and *cluster 1*.

The patients' symptoms are recorded regularly (on their visit to the clinicians) and based on these symptoms, at each visit, the patients are assigned to a cluster. Assignments to clusters may change during different visits to the clinicians. Following these assignments to clusters through their recorded visits to the clinicians gives an overview of the changes in the overall status and how the disease progresses through time.

The separation of patients into three groups provides information about the patients' status based on their aggregate score for the motor symptoms. Unfortunately, it does not provide any information about the symptoms that are particularly bothersome for the patients, and whose change would have the strongest impact on the assignment of patients into a given cluster.

The identification of symptoms that strongly influence the change of the patient's overall status (the patient's assignment to one of the clusters), can help clinicians to focus their attention to a smaller set of symptoms when deciding possible treatment modifications[1] of the patients. Using the real world data, our aim is to reveal the symptoms that are the most susceptible to improvement or decline when the overall status of the patient changes. When deciding on the modification of patient's treatment, the clinicians may consider these symptoms in order to keep the patient's status stable as long as possible. We present the algorithm for identification of the most impactful symbols in Section 4.1.

2.3. Multitask learning

In multitask learning (MTL), multiple related tasks are learned simultaneously on a shared attribute space. Compared to single-task learning, MTL can improve model generalization and prevent overfitting [6]. This is achieved by transfer of intermediate knowledge between jointly learned tasks, e.g., constructed relevant paths in tree-based models or important joint subconcepts in neural networks. In this way, the learning does not focus on a single task (thus preventing overfitting) and what is learned for one task can help other tasks (thus improving generalization).

[1] A treatment modification is any change in the overall LEDD (levodopa equivalent daily dosage) (change of frequency intake, change of medications group etc.).

Caruana et al. [7] use knowledge from the future to rank patients according to their risk to die from pneumonia. The shared attribute space consists of patients' symptoms at the time they are admitted to the hospital. The multiple tasks which are learned by the model are a set of hospital tests performed to determine whether the patients are of a risk of dying of pneumonia. Zhou et al. [51] use multitask learning to model Alzheimer's disease progression. They use two clinical/cognitive measures, Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-Cog) as multiple evaluations to determine the progression of the disease. Zhang et al. [50] propose a multitask model for prediction of multiple regression and classification variables in Alzheimer's disease, which takes advantage of the multi-modal nature of patient's symptoms. Similarly to Parkinson's disease patients, Alzheimer's patients can be described by symptoms collected from multiple sources. All of these approaches use quantitative data "from the future" (values of tests taken in the future) to determine how the disease progresses. The authors take historical data and use multitask learning to predict the two years in the future results of two tests (the MMSE and the ADAS-Cog questionnaire). Using the baseline MRI, FDG-PET, and CSF data they estimate the disease progression by predicting these two values and predicting the conversion of patients with a mild cognitive disorder (MCI) to patients with Alzheimer's disease (AD). Unfortunately, there are no tests to appropriately measure the progression of Parkinson's disease. None of the above-mentioned methods look at the medications patients are receiving to decelerate the disease progression.

We use multitask learning with the aim to simultaneously predicting the values of several target attributes (medications in our case). We use a supervised learning method called predictive clustering trees (PCTs) [3,4]. This method adapts the basic top-down induction of decision trees with clustering and allows for multitask learning. The PCT learning algorithm used is implemented in the CLUS data mining framework [4]. We obtain multitask decision trees, simultaneously predicting three target variables: change of levodopa dosage, change of dopamine agonists dosage, and change of MAO-B inhibitors dosage, referring to three most important medication groups used in Parkinson's disease patient management. The PCT-based approach is described in Section 4.2, and evaluated in Sections 5.2 and 5.3.

2.4. Feature evaluation

Feature subset selection can improve the accuracy, efficiency, applicability, and comprehensibility of a learning process and its resulting model [2]. For this reason, many feature subset selection approaches have been proposed. In general, three types of feature selection methods exist: wrapper, filter, and embedded methods. Wrapper methods use the performance of a given learning algorithm as the criterion to include/exclude attributes. Embedded methods use feature selection as an integral part of their learning process. Filter methods introduce some external criterion independent of the predictor. They evaluate features according to that criterion, which allows for ranking of features and selection of a suitable subset. This is fit for our purpose.

Our approach to determining the importance of symptoms for the overall disease progression is strongly related to the well-known Relief family of algorithms [19,34,32]. These algorithms evaluate attributes based on their ability to distinguish between similar instances with different class values. Contrary to the majority of feature evaluation heuristics (e.g., information gain, gini index, etc.) that assume conditional independence of attributes w.r.t. the target variable, the Relief approaches do not make this assumption and are suitable for problems that involve feature interaction. The Relief algorithms randomly select an instance and find the nearest instance from the same class and nearest instances from different classes. When comparing feature values of near instances the algorithm rewards features that separate instances with different class values and punishes features that separate instances with the same class value. The whole process is repeated for large

Table 1
Characteristics of the questionnaire data used in the analysis.

Questionnaire	Number of questions	Answers value range	Ordered values	Higher value indicates higher symptom severity
MDS-UPDRS Part I	6	0–4	Yes	Yes
MDS-UPDRS Part Ip	7	0–4	Yes	Yes
MDS-UPDRS Part II	13	0–4	Yes	Yes
MDS-UPDRS Part III	35	0–4	Yes	Yes
MDS-UPDRS Part IV	6	0–4	Yes	Yes
MoCA	11	0–1	Yes	No
PASE	7	1–2	Yes	No
SCOPA-AUT	21	0–3	Yes	Yes
COGCAT	4	0–1	Yes	Yes
QUIP	4	0–1	Yes	Yes
Total	114			

enough sample. The approach we propose also uses similar instances but uses cluster membership as a criterion for similarity instead of a distance in the feature space. When updating the importance of features our approach assesses joint transitions from one cluster to another or from better patient status to a worse one, while Relief algorithms use similarities in target variable.

Some recent feature selection approaches try to explore the interconnection between the features by exploring the similarity graph of features [30,38]. Other approaches pose feature selection as an optimization problem, for example, Sun et al. [40] use optimization in combination with a game theory based method. Our approach also uses a graph of transitions between clusters to assess similarity of patients, but we work in an unsupervised scenario and use time order of patients' visits as links between nodes. Details are explained in Sections 4.1 and 4.2.

3. Parkinson's disease data set

In this paper we use the PPMI data collection [24] gathered in the observational clinical study to verify progression markers in Parkinson's disease. In Section 3.1 we present the PPMI symptoms data sets and in Section 3.2 we present the medications data used in the experiments. As there are altogether 114 attributes in the described data sets, in their everyday practice, physicians focus on a subset of chosen symptoms to follow the development of the disease and decide when to intervene with medication modifications. The symptoms which are in the focus of physician's attention are discussed in Section 3.3.

3.1. PPMI symptoms data sets

The medical condition and the quality of life of a patient suffering from Parkinson's disease is determined using the Movement Disorder Society (MDS)-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [14]. This is a four-part questionnaire consisting of 65 questions concerning the development of the disease symptoms. Part I consists of questions about the "non-motor experiences of daily living". These questions address complex behaviors, such as hallucinations, depression, apathy, etc., and patient's experiences of daily living, such as sleeping problems, daytime sleepiness, urinary problems, etc. Part II expresses "motor experiences of daily living". This part of the questionnaire examines whether the patient experiences speech problems, the need for assistance with the daily routines such as eating or dressing, etc. Part III is referred to as the "motor examination", while Part IV concerns "motor complications", which are mostly developed when the main antiparkinson drug levodopa is used for a longer time period. Questions from the MDS-UPDRS represent symptoms characteristic for Parkinson's disease, while their answers indicate the symptom's severity that a patient is experiencing. Each answer is given on a five-point Likert scale, where 0 = normal (patient's condition is normal, the symptom is not present), and 4 = severe (symptom is present and severely affects the independent functioning of the patient).

Cognitive state of a patient is determined using the Montreal Cognitive Assessment (MoCA) [8] questionnaire consisting of 11 questions (maximum 30 points), assessing different cognitive domains. In addition to the MoCA data, physicians also use the Questionnaire for Impulsive-Compulsive Disorders (QUIP) [48] to address four major and three minor impulsive-compulsive disorders.

Scales for Outcomes in Parkinson's disease—Autonomic (SCOPA-AUT) is a specific scale to assess autonomic dysfunction in Parkinson's disease patients [45]. Physical Activity Scale for the Elderly (PASE) [46] is a questionnaire which is a practical and widely used approach for physical activity assessment in epidemiologic investigations. Cognitive Categorization (COGCAT) is a questionnaire filled in by clinicians evaluating the cognitive state and possible cognitive decline of patients. The above data sets are periodically updated to allow the clinicians to monitor patients' disease development through time. Answers to the questions from each questionnaire form the vectors of attribute values.

Table 1 presents a summary of the symptoms data sets used in our research.[2] It lists the number of considered questions from each questionnaire, the range of attribute values, and the nature of the attribute values. Answers to the questions from questionnaires presented in Table 1 represent the combined set of symptoms used in our research to determine the status of Parkinson's disease patients. The total number of symptoms from the mentioned questionnaires is 114.

Answers to the considered questions are ordered values and, with the exception of MoCA and PASE questions, larger values suggest higher symptom severity and decreased quality of life for Parkinson's disease patients.

3.2. PPMI concomitant medications log

The PPMI data collection offers information about all of the concomitant medications that the patients used during their involvement in the study. We concentrate on whether a patient receives a therapy with antiparkinson medications and which combination of antiparkinson medications she/he received between two consecutive time points when the MDS-UPDRS and MoCA tests were administered. The three main families of drugs used for treating motor symptoms are levodopa, dopamine agonists, and MAO-B inhibitors [25].

The medications therapy for Parkinson's disease patients is highly personalized. Patients take different medications with personalized plans of intake. In order to be able to compare different therapies, dosages of Parkinson's disease medications are translated into a common Levodopa Equivalent Daily Dosage (LEDD).

[2] We do not have permission to share the data. Access to data can be obtained on the PPMI website: http://www.ppmi-info.org/access-data-specimens/download-data/.

Table 2
Description of *motor* (upper part) and *non-motor* (lower part) symptoms used by Parkinson's disease physicians in everyday practice to estimate patient's quality of live. The values intervals (*normal* and *problematic*) are defined by the clinician.

Symptom	Data set	Question number	Normal values interval	Problematic values interval
bradykinesia	MDS-UPDRS Part III	3.14	0–1	2–4
tremor	MDS-UPDRS Part II and III	mean value	0	1–4
gait	MDS-UPDRS Part III	3.10	0–1	2–4
dyskinesia	MDS-UPDRS Part IV	4.3	0–1	2–4
ON/OFF fluctuations	MDS-UPDRS Part IV	4.5	0	1–4
daytime sleepiness	MDS-UPDRS Part I	1.8	0–1	2–4
impulsivity	QUIP	SUM	0–1	≥2
depression	MDS-UPDRS Part I	1.3	0–1	2–4
hallucinations	MDS-UPDRS Part I	1.2	0–1	2–4
cognitive disorder	MoCA	SUM	26–30	< 26

3.3. Experimental symptoms data selected by clinicians

In their everyday practice, physicians use a vector of chosen symptoms to follow the development of the disease and decide when to intervene with medication modifications. They focus their attention on both *motor* and *non-motor* aspects of patients' quality of life. Physicians evaluate the motor aspect of patient's quality of life using the following symptoms: *bradykinesia*, *tremor*, *gait*, *dyskinesia*, and *ON/OFF fluctuations*. The *non-motor* aspect of patient's quality of life is determined using *daytime sleepiness*, *impulsivity*, *depression*, *hallucinations*, and *cognitive disorder*. In addition to *motor* and *non-motor* symptoms, physicians also consider epidemiological symptoms which include *age*, *employment*, *living alone*, and *disease duration*. According to the collaborating clinicians, physicians are inclined to change the therapy of younger patients (younger than 65[3]), who are still active, who live alone, and for the patients diagnosed with Parkinson's disease for a shorter time (less than 8 years). For these patients, physicians will try more changes to the therapy in order to find the most suitable therapy, rather than therapy prolongation with increased medications dosage strategy which is applied to older Parkinson's disease patients.

In modifying the patient's medications based on the numerical evaluation of symptoms, the physicians decide whether the symptom is *problematic* and needs their immediate attention or not. Table 2 presents the *motor* and *non-motor* symptoms influencing the physicians' decisions for medications modifications, the data sets they are part of, and the intervals of values that are considered *normal* or *problematic* for Parkinson's disease patients. For example, the value of *tremor* is defined as the mean value of all questions concerning tremor from MDS-UPDRS Part II and Part III. Intervals of *normal* and *problematic* values are determined by the clinical expert. For all UPDRS items, value 0 is normal, value 1 is slight or minor, value 2 is mild, 3 is moderate and 4 is severe. Thus, in most cases and given the progressive nature of Parkinson's disease, values 0 and 1 of symptoms are not problematic and are baring for the patients, but become annoying and hampering when they progress in the range 2–4: this leads to distinguishing between values *normal* and *problematic* [14]. The selection of these 10 *motor* and *non-motor* symptoms, and *age* as an *epidemiological* symptom, constituted the subset of attributes considered in the experiments presented in Section 5.2. The reason for excluding *employment*, *living alone*, and *cognitive disorder*, which could be important *epidemiological* attributes, is that the PPMI data collection does not have data about patients' employment and living arrangements. We omitted the *cognitive disorder* attribute due to its values in the database, which were either *normal* or *missing*[4].

For each patient in the data set, the *motor* and *non-motor* symptoms data were obtained and updated periodically (on each patient's visit to the clinician's), providing the clinicians with the opportunity to follow the development of the disease. The data set contains 897 instances, containing information about 368 PPMI patients. Most of the considered patients have records about two or three visits to the clinician. The maximum number of visits is 4.

4. Methodology

In this section, we present two methodologies: a methodology for patients' symptoms impact on the Parkinson's disease progression and a methodology for detecting medications dosage change patterns as a result of the patient's symptoms. Section 4.1 outlines an algorithm for determining which symptoms have the strongest impact on the patients' overall status. The patients' overall status is determined by the severity of a large set of symptoms (see Section 3.1). This methodology is closely related to our previous research on Parkinson's disease progression, shortly summarized in Section 2.2 as well as the work done on feature evaluation (Section 2.4). Results from this methodology, i.e. a list of symptoms that our algorithm finds to have the strongest impact on the change of patients' overall status are presented in Section 5.1.

Section 4.2 presents our methodology for detecting medications dosage change patterns as a result of the patient's symptoms. This methodology serves two aims: detecting patterns of medications dosage changes based on the patient's overall status as well as identifying clinically confirmed symptoms suggesting medications change. Our methodology is related to the work done on multitask learning (Section 2.3). Results from the evaluation of the methodology on the set of symptoms data selected by clinicians are presented in Section 5.2.

4.1. Symptoms' impact on Parkinson's disease progression

This section outlines a pseudo code of the algorithm which estimates the impact of symptoms on the change of patients' overall status—their change of clusters. The most important symptoms found by this algorithm are presented in Section 5.1.

The *getAttrChangeProbabilities* function, presented in Algorithm 1, is a supervised approach that estimates the probabilities that feature (symptom) values changed when the patients' overall status also changed (i.e. when the patients have crossed clusters) or stayed the same (the patients have not changed clusters between two consecutive visits).

Algorithm 1. Assessment of feature impact on cluster changes.

As the input Algorithm 1 takes **F**, patients' symptoms data described in Section 3.1, the index data set **I**, and the assigned cluster labels **c**. The patients' symptoms data **F** contains the information about the patients' symptoms values at different visits to the clinicians. It is a matrix of dimension n (number of instances) times $|\mathbf{A}|$ (number of considered symptoms). The features data set **F** contains the information on 114 *motor* and *non-motor* symptoms of Parkinson's disease patients. **F** rows

[3] Retirement age for men (https://en.wikipedia.org/wiki/Retirement_age).

[4] We explored the option of handling 'structurally missing' data, where the *cognitive disorder* attribute was kept in the final analysis. Across all attributes a new attribute value *missing* was introduced. The generated model had a lower classification accuracy than the model presented in Section 5.2.

```
getAttrChangeProbabilities(F,I,c):
  Input      : F – concatenated view (feature data set);
               A – attribute space;
               I – indices of patient-visit combinations;
               c – assigned cluster labels;
  Parameters : K – number of clusters in c;
  Output     : attrChangeProbability;
               attrSameProbability;
  // Count for each cluster crossing. Matrix noOfCrossings^{K×K} is initialized to 0.
  noOfCrossings = {1:K, 1:K} ← 0
  // Number of value changes for each attribute and each cluster crossing.
  // Matrix attrChangeNo^{K×K×|A|} is initialized to 0.
  attrChangeNo = {1:K, 1:K, 1:|A|} ← 0
  // Probability of value changes for each attribute and each cluster crossing.
  // Matrix attrChangeProbability^{K×K×|A|} is initialized to 0.
  attrChangeProbability = {1:K, 1:K, 1:|A|} ← 0
  // Number of unchanged values for each attribute and each cluster crossing.
  // Matrix attrSameNo^{K×K×|A|} is initialized to 0.
  attrSameNo = {1:K, 1:K, 1:|A|} ← 0
  // Pobability of unchanged values for each attribute and each cluster crossing.
  // Matrix attrSameProbability^{K×K×|A|} is initialized to 0.
  attrSameProbability = {1:K, 1:K, 1:|A|} ← 0

  for patient in [1:p] do
      // consecutive visits for a given patient
      patientsVisits ← I[patient,1:allVisits[patient]]
      for v_j, v_{j+1} in patientsVisits do
          prevCluster ← c[patient][v_j]
          currCluster ← c[patient][v_{j+1}]
          incrementByOne(noOfCrossings[prevCluster, currCluster])
          prevFeatures ← F[v_j]
          currFeatures ← F[v_{j+1}]
          for attr in A do
              if differs(prevFeatures[attr],currFeatures[attr]) then
                  incrementByOne(attrChangeNo[prevCluster, currCluster][attr])
              end
              else
                  incrementByOne(attrSameNo[prevCluster, currCluster][attr])
              end
          end
      end
  end

  // Determine the probability of changed/unchanged attribute values
  // for each cluster crossing.
  for c1 in [1:K] do
      for c2 in [1:K] do
          clusterCrosses = noOfCrossings[c1,c2]
          for attr in attrChangeNo[c1,c2] do
              attrChanges = attrChangeNo[c1,c2][attr]
              attrChangeProbability[c1,c2][attr] = attrChanges / clusterCrosses
          end
          for attr in attrSameNo[c1,c2] do
              attrSame = attrSameNo[c1,c2][attr]
              attrSameProbability[c1,c2][attr] = attrSame / clusterCrosses
          end
      end
  end
  return attrChangeProbability, attrSameProbability
```

represent the instances (patient p_i on visit v_{ij}), and the columns present patients' symptoms. The index data set **I** holds the instance indexes represented as a combination of patients and their visits. Vector **c** holds the information about the cluster to which a patient in a certain visit has been assigned to (i.e. c_{ij} marks the cluster patient p_i was assigned to on visit v_{ij}, see Section 2.2).

The output of the algorithm is two matrices, *attrChangeProbability* and *attrSameProbability*, of dimension $K \times K \times |A|$ (K is the number of clusters), which hold the probabilities that an attribute will change value or stay the same for a certain cluster crossing, respectively.

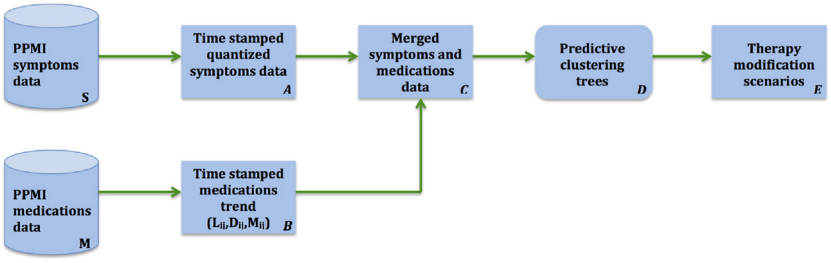

Fig. 1. Outline of the methodology for determining medications change patterns in PPMI data using predictive clustering trees.

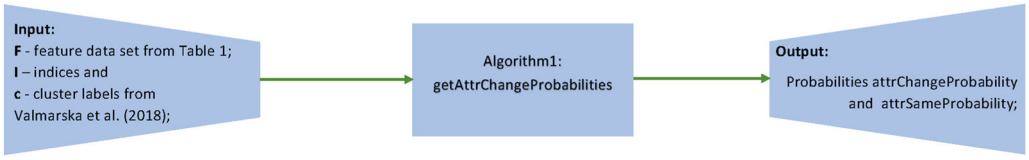

Fig. 2. A flowchart presenting the input, output, and method used for determining the most influential symptoms. Details about the input data can be found in Table 1 and [43,44].

The algorithm first initializes its working spaces and storage matrixes (lines 2–6). For each patient and for each two consecutive visits, the algorithm compares the assigned cluster labels for each instance (for each combination of (p_i, v_{ij}, c_{ij}) and $(p_i, v_{ij+1}, c_{ij+1})$) in lines 8–24. For each cluster change combination, the algorithm also takes note of what happens to the symptoms' values—whether they changed or stayed the same (lines 15–21). The recorded changes of symptoms and clusters are normalized with the total number of cluster crossing (lines 25–37) and the resulting probabilities are returned (line 38).

As a result of Algorithm 1, we get probabilities which reflect the impact of the attributes on cluster changes. This can serve in inference on the disease progression but also to select only the most influential attributes and thereby decrease the dimensionality of attribute space. We discuss the use of Algorithm 1 in Section 5.1.

4.2. Medications dosage change patterns

Our goal is to support physicians in their decisions regarding the patients' therapies. The physicians have several groups of medications at their disposal with which they try to preserve the good quality of patient's life. They use and switch between different groups of drugs and their dosages to treat different symptoms (e.g., levodopa is used for *motor* symptoms), and also to prevent overuse of any specific drug in order to reduce side-effects and undesired drug interactions. Our multitask learning approach based on Predictive Clustering Trees (PCTs) [4] (introduced in Section 2.3) allows for modeling of all medication groups simultaneously. By simultaneously predicting several target variables, the model allows physicians to observe the interactions between different groups of medications, which is not possible with univariate models. As training data, we use time-stamped symptoms and medications data. Fig. 1 outlines the proposed five-step methodology, which uses symptoms data collected over time (i.e. over several patient's visits) and respective changes in medications therapies. Our goal is to identify symptoms scenarios for which the physicians need to consider modifications of therapies.

The input to the methodology are PPMI data sets of patient symptoms (described in Section 3.1) and the PPMI medications log data set (described in Section 3.2). The output of the methodology are patterns of patients' symptoms for which particular changes of medications were administered by the clinicians.

In step ***A*** we construct a time-stamped symptoms data set consisting of the symptoms (attributes) described in Section 3.1. This data set consists of patient-visit pairs (p_i, v_{ij}) describing the patients and their visits to the clinician.

In step ***B*** we construct a data set of medications changes which are represented with (p_i, m_{ij}, m_{ij+1}) tuples, where m_{ij} and m_{ij+1} are medication therapies of patient p_i in two consecutive visits, v_{ij} and v_{ij+1}. A patient receives a therapy which is any combination of levodopa, dopamine agonists, and MAO-B inhibitors. For each of the three medications groups, we determine whether its dosage in the time of visit v_{ij+1} has changed (*increased* or *decreased*) or remained unchanged with respect to the dosage at visit v_{ij}. The output of step ***B*** is a data set of medications changes, presented as tuples (L_{ij}, D_{ij}, M_{ij}), indicating whether between visits v_{ij} and v_{ij+1} a change of dosage in levodopa (L), dopamine agonist (D), or MAO-B inhibitors (M) took place.

In step **C** we concatenate the data sets obtained in steps ***A*** and ***B*** into a merged data set of symptoms and medications data. We use patient-visit pairs (p_i, v_{ij}) describing patient's symptoms at visit v_{ij} and the changes of medications in the same visit with respect to the next visit v_{ij+1}. These data consist of a set of attributes describing the condition of the patient, and three attributes (levodopa, dopamine agonists, and MAO-B) indicating the changes in their dosage, respectively. The set of symptoms describing the condition of the patient can be preselected by clinicians, automatically selected, or a combination of both approaches. The merged data set is used in step ***D*** to determine medications change patterns. The three medications groups are used as multitask variables (multiple classes) in the predictive clustering trees learning approach. We want to determine which symptoms influence decisions of physicians to modify the therapies that patients receive. The discovered therapy modifications patterns are analyzed by the physician in step ***E***.

Models produced by the PCT approach serve three aims: determining patterns of medications dosage changes, identification of Parkinson's disease symptoms suggesting medications dosage changes, and discovering groups of similar patients. These aims depend on the interpretation of the PCTs. Patterns of medications dosage changes are found in the leaves of the tree. Branches from the root of the tree to its leaves identify the symptoms influencing a particular pattern of medications dosage change, while patients experiencing these symptoms and medications dosage changes construct groups of patients that are similar based on both their symptoms and their medications therapy modifications.

We test the proposed methodology in two experimental settings, using two different symptoms data sets described in more detail below. In the first experimental setting, (Section 5.2) we use symptoms which were selected by our consulting clinician. In the second experimental setting (Section 5.3), we test the proposed methodology for determination of the symptoms' impact (see Section 4.1) and form a merged data set with symptoms selected by the clinician and the most influential symptoms according to Algorithm 1. We analyze symptom patterns for which the physicians modified the patients' therapies. We use the changes of the three medications groups as the target classification variables. Changes in dosage (increase or decrease) are marked with the class label *yes*, while unchanged drug dosages are marked with the class label *no*.

5. Evaluation

We split the evaluation of the proposed methodology into three parts. In Section 5.1 we use Algorithm 1 to find the most influential symptoms. In Section 5.2 we analyze the medications dosage change patterns detected from symptoms selected by clinicians (see Section 3.3). The most influential symptoms from Section 5.1 together with the symptoms selected by clinicians form a new data set and are analyzed in Section 5.3.

5.1. The most influential symptoms

When patients change clusters between two consecutive visits, this change can be considered as *positive* or *negative*. A *positive* cluster change occurs when between two consecutive visits a patient has crossed from a cluster with higher index (e.g., *cluster 2*) to a cluster with lower index (e.g., *cluster 1*). Given the cluster descriptions from [44], this change indicates that the overall status of the patient concerning her/his motor symptoms has improved (indicated with lower MDS-UPDRS values). Contrarily, when in two consecutive visits a patient moves from a cluster with lower index (e.g., *cluster 1*) to a cluster with

Table 3
List of most influential symptoms according to Algorithm 1. The symptoms are ordered according to their average rank of positive and negative impact.

PPMI attribute	Attribute description	PPMI data set	Attribute importance for cluster change
NP3BRADY	Bradykinesia	MDS-UPDRS Part III	0.314
NP3TTAPL	Toe tapping (left)	MDS-UPDRS Part III	0.297
NP3RTCON	Constancy of rest	MDS-UPDRS Part III	0.291
NP3FACXP	Facial expression	MDS-UPDRS Part III	0.282
NP3FTAPL	Finger tapping (left)	MDS-UPDRS Part III	0.273
NP3FTAPR	Finger tapping (right)	MDS-UPDRS Part III	0.255
NP3PRSPL	Hand pronation/supination (left)	MDS-UPDRS Part III	0.244
NP3TTAPR	Toe tapping (right)	MDS-UPDRS Part III	0.239
NP3PRSPR	Hand pronation/supination (right)	MDS-UPDRS Part III	0.203
NP1SLPN	Sleep problems (night)	MDS-UPDRS Part Ip	0.155
NP1SLPD	Daytime sleepiness	MDS-UPDRS Part Ip	0.147
NP2HWRT	Handwriting	MDS-UPDRS Part II	0.144
NP1FATG	Fatigue	MDS-UPDRS Part Ip	0.138
NP1URIN	Urinary problems	MDS-UPDRS Part Ip	0.134
NP1PAIN	Pain and other sensations	MDS-UPDRS Part Ip	0.117
MCATOT	MoCA total score (cognition)	MoCA	0.097

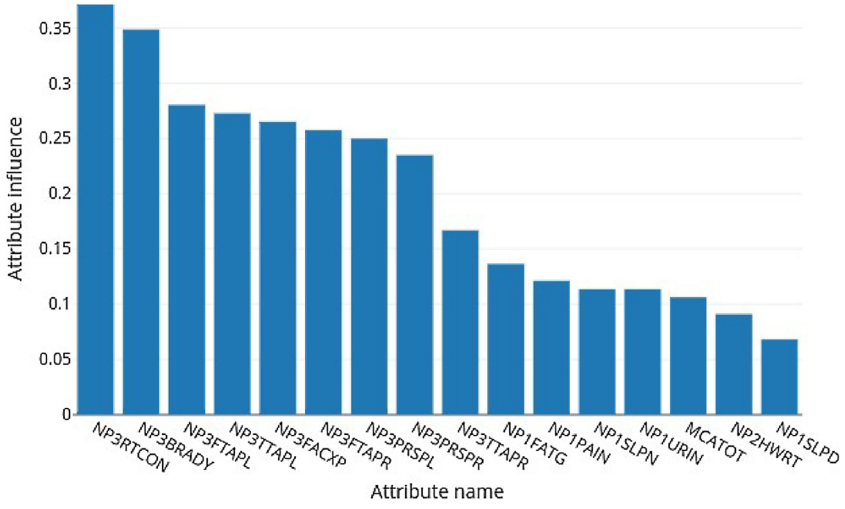

Fig. 3. Symptoms whose values improved most frequently when the overall status of patients improved. The acronyms are explained in Table 3.

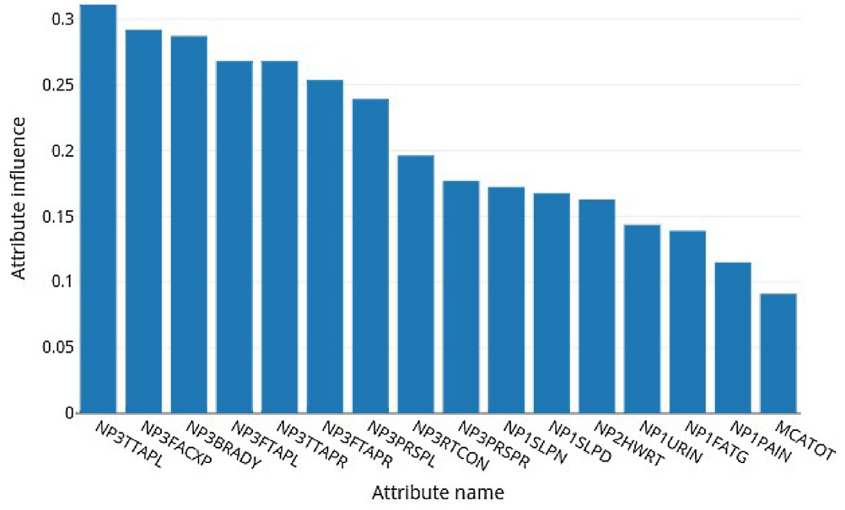

Fig. 4. Symptoms whose values worsen most frequently when the overall status of patients degraded. The acronyms are explained in Table 3.

higher index (e.g., *cluster 2*) her/his overall status has worsen (as indicated by the sum of the motor symptom).

We ran the Algorithm 1 twice, the first time using numerical scores of symptoms (values 0–4 for MDS-UPDRS symptoms), and the second time using discretized values of the symptoms (*normal* and *problematic*). A flowchart presenting the input, output, and method used in this experimental setting is presented in Fig. 2. For each run, the algorithm returned the probabilities of symptom changes and symptoms staying

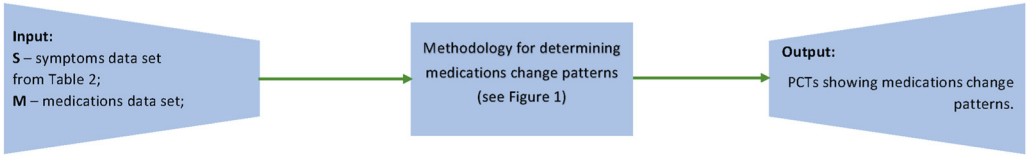

Fig. 5. A flowchart presenting the input, output, and method used for determining medications dosage change patterns detected from symptoms selected by clinicians. Details about the used symptoms data can be found in Table 2.

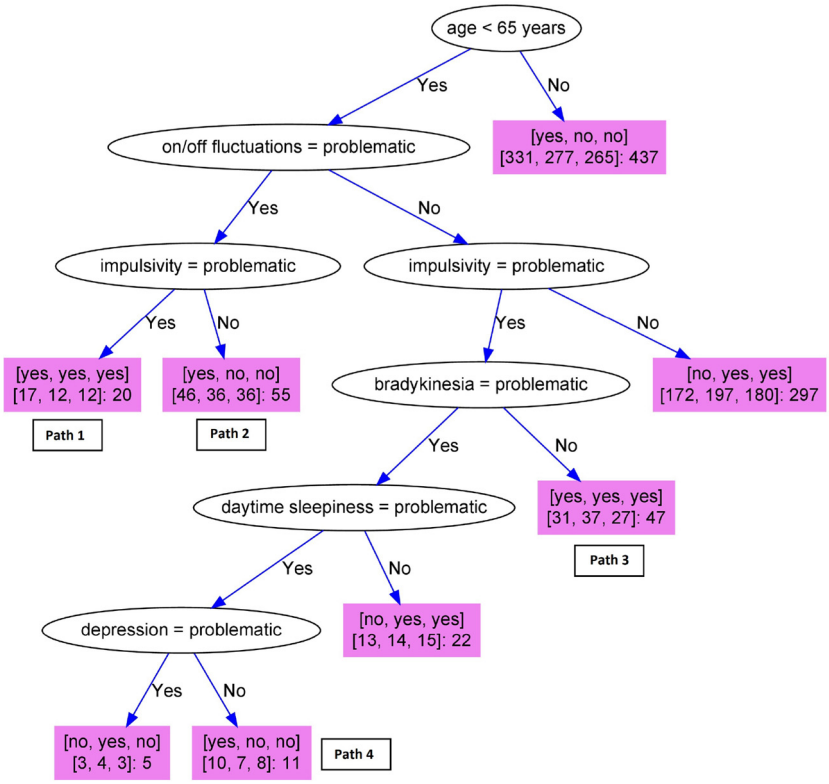

Fig. 6. Pruned predictive clustering tree modeling dosage changes for three groups of medications. Medication dosage changes are modeled by patients' symptoms.

unchanged. Ranking the symptoms by the decreased probability of symptom changes and intersecting the top 25[5] features we get a list of symptoms that have the strongest impact on cluster changes. These are symptoms that have most frequently changed values, and whose change of values brought significant improvement (from *problematic* to *normal*) or decline (from *normal* to *problematic*).

Table 3 presents the intersection of lists obtained by two runs of the algorithm, for symptoms whose values have changed most frequently when a cluster change has occurred. The symptoms are presented with their code names from the PPMI data collection and with their descriptions. The results are ordered according to the decreased probability of cluster changes (weighted *positive* and *negative* changes).

We can note that the upper part of Table 3 is populated with the motor symptoms from MDSUPDRS Part III. This is not surprising since as we mentioned above, the obtained clusters were ordered in accordance with the aggregate score of their motor symptoms from MDS-UPDRS Part III. In addition to the influential motor symptoms, the algorithm finds also a subset of influential non-motor symptoms whose values vary as the overall status of the patient's changes.

In practice, *positive* and *negative* changes are not treated equally and may not be caused by the same symptoms. Clinicians try to avoid *negative* changes and actively promote *positive* changes. We first report symptoms indicating positive changes, followed by the symptoms indicating negative changes.

Fig. 3 presents the symptoms whose values improve most frequently when the patients make a positive cluster change (their overall status between two consecutive visits improves). The results suggest that in over 37% of cases when the patient's status improves, also the value of their constancy of rest improves (NP3RTCON). The second most frequently improved symptom is bradykinesia (NP3BRADY), followed by the finger tapping in the left hand (NP3FTAPL).

Fig. 4 presents the results for the symptoms whose values degrade most frequently when the patients make a negative cluster change and their overall status between two consecutive visits worsens. The results suggest that in over 30% of cases when the patients' status worsens, they experience problems with toe taping, facial expression, and bradykinesia.

Rigidity is a relevant and bothersome symptom for patients that was not detected by Algorithm 1. A reason for this omission may be the fact that rigidity is reported through five questions from MDS-UPDRS Part III (both hands, both legs, and neck). Patients can experience rigidity problems on different parts of the body and each of this parts may not be statistically strong enough to be ranked high by Algorithm 1. A way to alleviate this problem would be to combine answers from multiple questions concerning the same underlying symptom before running Algorithm 1. We plan such detailed analysis for our further work as well as a selection of two separate lists of symptoms which improve or decline most frequently. This could lead to distinguishing the symptoms for which change of medications dosage is the most effective, as well as those who are most inclined to worsening as the disease progresses.

Similarly to our approach to determining the importance of

[5] The number of top-ranked features was set experimentally so that the length of the intersection list is sufficiently informative and manageable for clinicians.

Table 4
Comparison of the classification accuracy obtained by the default model, the pruned multitask model, the pruned single-task models, and the random forest models.

Medications group	Default model	Multitask PCT model	Single-task PCT model	Random forest ensembles
Levodopa	0.637	0.685	0.686	0.695
Dopamine agonists	0.501	0.642	0.642	0.622
MAO-B	0.518	0.602	0.586	0.601

symptoms for the overall disease progression, the feature evaluation algorithms Relief and ReliefF [19,34] also compare feature values of similar instances from the same class and similar values from a different class. Relief and ReliefF reward the features that separate instances with different class values and punish the features that separate the instances with the same class value. If applied to our problem, these algorithms cannot take into account temporal progress of patients, i.e. they cannot track individual patients on their consecutive visits to the clinician. In effect, they show which attributes influence the initial assignment of patients into clusters, but reveal no information about attributes which are the most influential for changes of patients' overall status (i.e., for crossing of clusters). We note that the assignment to clusters was done based on the patients' overall status represented with sums of attributes values from the respective questionnaires presented in Section 3.1. Nevertheless, we evaluated the symptoms using the ReliefF algorithm [34,35]. Out of the best 16 symptoms as evaluated by the ReliefF algorithm, 9 were selected into the top 16 most most influential symptoms (see Table 3) by Algorithm 1 (MCATOT, NP1SLPD, NP1URIN, NP3RTCON, NP1SLPN, NP1FATG, NP3PRSPL, NP3TTAPR, NP3TTAPL). Symptoms—such as bradykinesia—that are strong indicators of the disease progression were evaluated as insignificant by the ReliefF algorithm. For this reason, the results of ReliefF for symptom evaluation are not included.

5.1.1. Medical interpretation of the results

According to the consulting clinician, in general, the computed symptom importance is in accordance with the medical literature on Parkinson's disease [10,1]. Below we present some further interesting findings.

Cognitive decline, as depicted by the MoCA total score, and bradykinesia are very important factors when considering changing patients' medications [10,1]. Braykinesia is a score combining toe tapping [18,14] (for lower limbs bradykinesia assessment), hand pronation/supination, and finger tapping (for upper limbs bradykinesia assessment) [21]. As confirmed by the expert, the constancy of rest tremor and pain are symptoms which are important for some patients who find these symptoms particularly bothersome and demand an intervention with medications. Dyskinesia and fluctuations are important symptoms not ranked at the top of the list according to our Algorithm 1. The reason is that the PPMI database includes many newly diagnosed and early-stage patients, for who these symptoms do not change values often.

The importance of handwriting is an interesting finding of the study and confirms recent studies [9] suggesting that handwriting could be a useful marker for disease diagnosis [36] and progression [27]. Our further analysis of patients with problematic handwriting revealed that these patients experience more problems with their motor symptoms (reflected by the sum of symptoms from MDS-UPDRS Part III), and also suffer from bradykinesia, pain, and rigidity with higher severity than patients who do not have problems with handwriting. Results of our analysis also suggest that patient's handwriting sensitively reflects improvements and worsening of patients' motor symptoms.

5.2. Medications dosage change patterns detected from symptoms selected by clinicians

For this set of evaluations, we use the data set composed of symptoms selected by clinicians (see Section 3.3). A flowchart with the input, output, and method used in this experimental setting is presented in Fig. 5. A pruned predictive clustering tree (PCT) model of medications changes based on the patient's status is shown in Fig. 6. The PCT models the dosage changes of all three antiparkinson medication groups simultaneously, allowing for the detection of drug interactions based on the patient's status. Notice that in PCT construction, the user can decide how to prune the tree. In our experiments, we used the default pruning method, called C4.5 [39,28].

The leaves of the predictive clustering tree hold information about the recorded therapy modifications. The components of the lists presented in each tree leaf predict dosage changes of levodopa, dopamine agonists, and MAO-B inhibitors, respectively. The list of numbers in the leaves represents the total number of instances that are described by the symptoms from the root to the leaf. The number of instances for which

Table 5
Description of the *motor* (upper part) and *non-motor* (lower part) symptoms which are reported as the most influential by Algorithm 1. Toe tapping, finger tapping, and hand pronation/supination were generated as the maximum value of the basic symptom on the patient's left and right side. The values intervals (*normal* and *problematic*) were defined by the clinician. The three symptoms marked with bold typeface were independently selected by clinicians as the most important.

Symptom	Data set	Question number	Normal values interval	Problematic values interval
bradykinesia	MDS-UPDRS Part III	3.14	0–1	2–4
toe tapping	MDS-UPDRS Part III	max(3.7a, 3.7b)	0–1	2–4
constancy of rest	MDS-UPDRS Part III	3.18	0–1	2–4
facial expression	MDS-UPDRS Part III	3.2	0–1	2–4
finger tapping	MDS-UPDRS Part III	max(3.4a, 3.4b)	0–1	2–4
hand pronation/supination	MDS-UPDRS Part III	max(3.6a, 3.6b)	0–1	2–4
sleep problems	MDS-UPDRS Part Ip	1.7	0–1	2–4
daytime sleepiness	MDS-UPDRS Part I	1.8	0–1	2–4
handwriting	MDS-UPDRS Part II	2.7	0–1	2–4
fatigue	MDS-UPDRS Part Ip	1.13	0–1	2–4
urinary problems	MDS-UPDRS Part Ip	1.10	0–1	2–4
pain and other sensations	MDS-UPDRS Part Ip	1.9	0–1	2–4
cognitive disorder	MoCA	SUM	26–30	< 26

Table 6
Extended symptoms data set consisting of symptoms handpicked by the expert and the most influential symptoms ranked by Algorithm 1. Details about the symptoms can be found in Tables 2 and 5.

Motor symptoms	Non-motor symptoms	Epidemiological symptoms
bradykinesia	*daytime sleepiness*	*age*
tremor	*impulsivity*	*disease duration*
gait	*depression*	
dyskinesia	*hallucinations*	
ON/OFF fluctuations	*sleep problems*	
toe tapping	*handwriting*	
constancy of rest	*fatigue*	
facial expression	*urinary problems*	
finger tapping	*pain and other sensations*	
hand pronation/supination		

the proposed medications dosage change has actually happened are written in the square brackets. For example, the list [yes, yes, yes] presented in the first leaf on the left (Path 1), indicates that the dosages of levodopa, dopamine agonists, and MAO-B changed. The total number of covered instances is 20. Out of these 20 patients, for 17 the dosage of levodopa changed, for 12 the dosage of dopamine agonists changed and for 12 the dosage of MAO-B inhibitors changed.

The attributes of instances (patient-visit pairs) influencing this change are presented along the path from the tree root to the respective leaf. In this example, these are the patients who are younger than 65, have problems with ON/OFF fluctuations, and have problems with their impulsivity. The leaf [yes, no, no] on the right (Path 2) suggests that physicians only considered changes in levodopa. These dosage changes can be justified by the patients' symptoms, i.e. patients have problems with ON/OFF fluctuations and no problems with their impulsivity. Moreover, in younger patients without ON/OFF fluctuation problems but with other problematic symptoms: impulsivity, bradykinesia and daytime sleepiness, the physicians also change only levodopa dosages (Path 4 in Fig. 6). This might reflect the current clinical practice—many patients want treatment of their motor symptoms first, which usually improve based on increased levodopa dosages.

Path 4 in Fig. 6 shows that for younger patients with problematic impulsivity and without problems with ON/OFF fluctuations and bradykinesia, physicians change the dosages of all three medication groups. These results are in accordance with the literature on Parkinson's disease [47] and were confirmed by the clinical expert.

The results reveal that if the patient experiences ON/OFF fluctuations problems (left subtree in Fig. 6), physicians will react with the change of dosage of levodopa medications [12]. If the patients experience non-motor symptoms (e.g., impulsivity, depression), physicians will react by modifying the dosages of dopamine agonists [37]. This is in accordance with the literature on Parkinson's disease and was confirmed by the expert. Increased dosages of dopamine agonists can produce non-motor related side-effects. Physicians will react by lowering the dosage of dopamine agonists (consequently increasing the dosage of levodopa). This was revealed in our post analysis, where we followed the actual changes of levodopa and dopamine agonists. In this post analysis the target variables (levodopa and dopamine agonists) had three values: *increase*, *decrease*, and *unchange*. The PCT model was built on the symptoms data presented in Section 3.3 in combination with the newly generated target features.

While prediction is not the ultimate goal of the developed methodology, reasonably high classification accuracy on a separate data set can increase clinicians' trust in using the model. Table 4 presents the classification accuracy of four models: (i) the default model (predicting the most probable value for each target), (ii) the multitask PCT classification model, iii) the single-task decision tree models constructed separately for each medications group, and (iv) the multitask random forest model [5,20]. The results are obtained using 10-fold cross-validation. The results show that random forests generate models that have slightly better classification accuracy for levodopa. However, the PCT models yield better classification accuracy for dopamine agonists. Multitask PCT model and random forests return comparable classification accuracy for MAO-B. The advantage of using the multitask tree approach is the ability to observe the interactions between the targets.

We employed the Wilcoxon [49] paired test to examine whether there are statistical differences between the performance of the multitask approach and the single-task approach, and the multitask approach and random forest. Results showed that there are no statistical differences at the level of significance $\alpha = 0.05$ for any pair of the above pairs. We used the same folds across all approaches.

5.2.1. *Impact of symptoms history*

We analyzed the temporal aspect of the proposed approach. Our data set offers certain time-related information (1–4 observations are available for each patient in the data set, one for each visit). So far we only analyzed the changes in symptoms and dosage between two consecutive visits. According to our consulted clinicians, the current state of the patient is all that matters to the clinician when considering their therapy, so this makes sense. However, the question remains if taking into account more than one historical event can improve models.

To adequately answer this comment, we conducted a separate set of experiments. We looked further back into patients' history to see how their medications have changed, based on symptoms from more distant visits. Comparison of 10-fold cross validation classification accuracy on models with different spans of look-back showed that the classification accuracy of models decreases as we include references to more distant visits. The highest accuracy is achieved when we consider only the actual state of the patients (these results are presented in this section).

5.3. *Medications dosage change patterns detected from extended symptoms data*

The PCT model from Section 5.2 was generated on a set of symptoms selected based on the expert's choice. In this Section, we explore the model for dosage change of antiparkinson medications if in addition to the symptoms that are pre-selected by the clinicians we include also the most influential attributes from Table 3. We present the description of the newly introduced symptoms, the predictive clustering tree model generated on this extended data set, short interpretation of the tree, the classification accuracy of the models, and a short discussion on the differences between the original model (Section 5.2) and the revised model (Section 5.3.2).

5.3.1. *Extended data set*

As already mentioned, when monitoring the patient's status and deciding about the modification of their medications therapy, clinicians think in terms whether the symptom's severity is *normal* for a Parkinson's disease patient or it is *problematic* and a change of dosage of

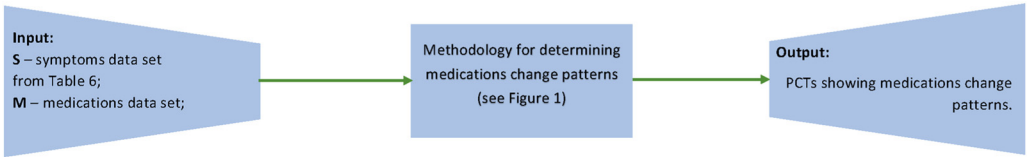

Fig. 7. A flowchart presenting the input, output, and method used for determining patterns of medications dosage change from the extended symptoms data. The extended symptoms data set is presented in Table 6.

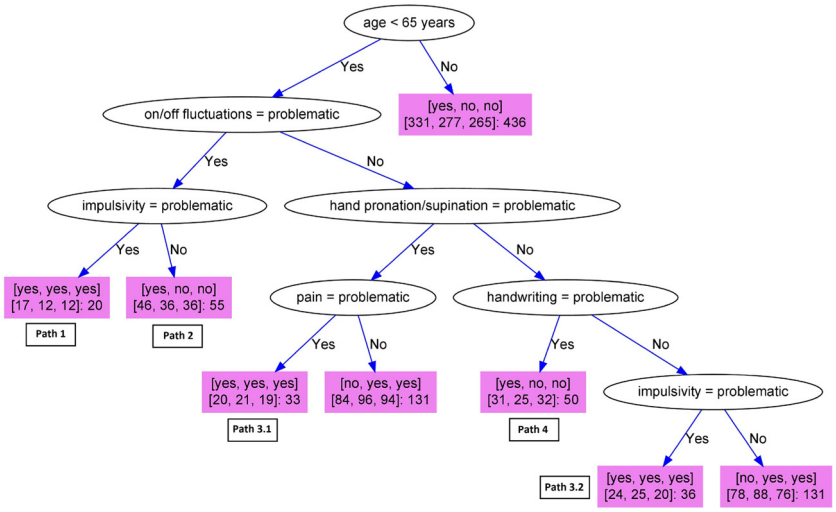

Fig. 8. Pruned predictive clustering tree modeling the dosage changes for three groups of medications. The model is generated on the extended set of Parkinson's disease patients symptoms. For improved model readability, the minimal number of covered instances is set to 20.

Table 7
Comparison of the classification accuracy obtained by the default model, the pruned multitask PCT model, the pruned single-task PCT models, and the random forest ensemble model on the extended symptoms data set.

Medications group	Default model	Multitask PCT model	Single-task PCT model	Random forest ensemble
Levodopa	0.637	0.657	0.671	0.683
Dopamine agonists	0.501	0.631	0.630	0.642
MAO-B	0.518	0.583	0.572	0.615

antiparkinson medications is needed. Table 5 presents the most influential *motor* and *non-motor* symptoms according to Algorithm 1 and the intervals for their quantization into *normal* and *problematic* symptoms values. Six of the symptoms from Table 3 were merged into three new (revised) symptoms. These six symptoms were pairs of three underlying symptoms, each concerning a different side of the body (left or right), and were therefore paired into three new symptoms. The three new symptoms are: toe tapping, finger tapping, and hand pronation/supination. The values of the newly constructed symptoms are obtained as the maximum of the two basic symptoms values (left and right).

The extended symptoms data set used in the experiments below is presented in Table 6. These symptoms are *motor*, *non-motor*, and *epidemiological*, consisting of the symptoms that were pre-selected by our consulting expert (see Section 3.3) and the most influential symptoms returned by Algorithm 1 (Table 5). We decided to omit the *cognitive disorder* attribute due to the fact that its only values present in the database were *normal* and missing. The reason for this is that in this analysis we only consider patients with included medications data.

Note that out of the 16 symptoms that were top-ranked by Algorithm 1, our consulting clinician reported 3 as the symptoms they consider when deciding about the change of Parkinson's disease patient's therapy. These symptoms are *cognition*, *daytime sleepiness*, and *bradykinesia* (marked in bold in Table 5).

5.3.2. Revised results and discussion

The extended symptoms data set was used as an input to our methodology for determining medications dosage change patterns in PPMI data using predictive clustering trees (presented in Section 4.2). A flowchart outlining the input, output, and method used in this experimental setting is presented in Fig. 7. The obtained model for symptoms scenarios that caused clinicians reaction with medications dosage change is presented in Fig. 8.

The revised model for dosage changes is slightly different from the original model presented in Fig. 6. The roots of the trees are the same, i.e. the clinician's decision about modifying the patient's medications treatment is mostly influenced by the age of the patient. For younger patients, the decision is influenced also by their on/off fluctuations and impulsivity. The right-hand side of the subtree concerning younger patients is different. In this subtree, the symptoms that influence the dosage change of antiparkinson medications are the newly introduced symptoms: pain and other sensations, hand pronation/supination, and handwriting. Path 1 and Path 2 are the same in both models. Path 4 in both models reveals medications change pattern [yes, no, no], indicating that based on the symptoms patterns (the paths from the root to the leaf of the tree) the clinicians consider changing the dosage of levodopa, and leave the dosages of dopamine agonists and MAO-B inhibitors unchanged. Paths 3.1 and 3.2 suggest that the clinician should consider updating the dosages of all antiparkinson medications which is similar to Path 3 of Fig. 6.

For each of the medications groups, Table 7 presents the classification accuracy of the default model, revised multitask PCT model, revised single-task PCT models, and random forest multitask model. Results are obtained using 10-fold cross-validation. As it is the case with the model from Fig. 6, the accuracy values obtained by the multitask PCT model are comparable to those obtained by the single-task PCT model and are better when compared to the default model. The multitask random forest ensemble returned the best classification accuracy for all targets. This model also has an improved classification accuracy for dopamine agonists and MAO-B compared to models from Section 5.2 (see Fig. 6 and Table 4). This improvement can be explained with the additional information available in the extended set of attributes and non-trivial interactions between different targets, which can be captured by ensembles. The main disadvantage of the ensemble multitask models is their lack of model interpretability. The Wilcoxon [49]

paired test revealed that for the target variable levodopa the random forest ensemble performs significantly better than our multitask approach ($\alpha = 0.05$, p-value = 0.012). For the target variable MAO-B, the multitask approach performs significantly better than the single-task approach ($\alpha = 0.05$, p-value = 0.018). Other differences were not significant. We used the same folds across approaches.

The classification accuracy of both the revised multitask PCT model and the revised single-task PCT models are lower than the accuracies of the models generated on the original symptoms data set (Table 2). A reason for this difference may be the fact that our models are trained on and reflect the history of clinicians' decisions, and do not necessarily reflect the actual symptoms clinicians should react to.

5.3.3. Medical evaluation of the results

For patients covered by rules from Path 1 and Path 2 it is reasonable to introduce levodopa and try to provide the optimal dosage even in younger patients (average age of 53 years) when they have on/off fluctuations (i.e. disease is rapidly progressing). The presence of impulsivity dictates the medications dosage changes the clinician should make. Path 3.1 covers younger patients (average age of 52.42 years) who suffer from severe bradykinesia (NP3BRADY = 1.94 ± 0.84[6]). Their overall motor symptoms are severe, i.e. the sum of MDS-UPDRS Part III (NP3SUM) is 34.03 ± 11.99. Patients' quality of daily living is affected, i.e. the sum of MDS-UPDRS Part II (NP2SUM) is 13.85 ± 6.36. Along with the presence of pain, many changes in medications dosages are done in an effort to better manage the advanced disease severity. Patients who do not have problems with pain and are treated with [no,yes,yes] medications dosage change pattern are patients who also have severe motor symptoms (NP3SUM = 31.06 ± 0.84 and disturbing bradykinesia (NP3BRADY = 1.71 ± 0.84)). However, their overall status is slightly better and the mild problems with pain lead to more dosage changes of dopamine agonists and MAO-B inhibitors, and a stable treatment with levodopa.

Patients covered by Path 4 are overall in a better condition that patients mentioned in previous paths. Their motor symptoms are less severe (NP3SUM = 20.39 ± 9.15), they do not have problems with on/off fluctuations, they have mild bradykinesia, and have no cognitive problems. Their handwriting seems to be a useful marker of disease progression which leads to dosage changes in levodopa. Changes of medication dosages for patients covered by Path 3.2 are imposed by the problematic impulsivity. Dosages of dopamine agonists are lowered to stabilize impulsivity, while levodopa is increased in order to control the motor symptoms. Younger patients who do not have problems with impulsivity (nor problems with on/off fluctuations, hand pronation/supination, handwriting) and are treated with [no,yes,yes] medications dosage change pattern are patients who are in better condition than all the other patients included in the predictive clustering tree from Fig. 8. Reasonably, only dopamine agonists and MAO-B inhibitors are modified in an effort for better management of the disease. Levodopa is either not prescribed or only low dosages are prescribed. Older patients (average age of 68.45 ± 4.87 years) have problems with many symptoms. The disease is managed with levodopa, and an optimal regime is sought through changes.

6. Conclusions

We present the methodology to detect trigger symptoms for change of medications therapy of Parkinson's disease patients. We consider trigger symptoms to be the ones which press the physicians to make modifications of the treatment for their patients. We test the developed methodology on a chosen subset of time-stamped PPMI data. The data set offers an insight into the patients' symptoms progression through time, as well as the response of physicians following problematic states of either motor or non-motor symptoms. We identify clinically confirmed patients' symptoms indicating the need for medication changes.

The proposed approach allows identifying patient subgroups for which certain medications modifications have either a positive or a negative effect. By post analysis of the patients who respond well to the medications modification and those who do not, and the underlying characteristics of each group, we may be able to assist the physicians with the therapy modifications for a given patient by narrowing the number of possible medication prescriptions scenarios.

We also present an algorithm for determining the symptoms which have the largest influence on the change of the Parkinson's disease patients' overall status. These are the symptoms that change most frequently as the status of the patient improves/declines. We relate this work with our previous work, where we developed a methodology for determining groups of patients with similar severity of symptoms and establishing how the disease progresses in terms of the severity of several groups of symptoms.

Our results show that some of the most impactful symptoms for changes in the patients' overall status detected by Algorithm 1, are currently not considered by the clinicians when deciding about the change of antiparkinson medication dosages. This requires further study and offers an opportunity for improved disease management in the future.

In future work, we plan to apply model explanation approaches to describe relevant subgroups of patients, therapy change patterns with a positive influence on the control of symptoms, and therapy patterns which are more likely to lead to side effects. There are some open opportunities in the analysis of more than one previous time point. Taking longer history into account we might be able to detect groups of patients which do not react well to the changes in antiparkinson medications.

Acknowledgements

This work was supported by the PD_manager project, funded within the EU Framework Programme for Research and Innovation Horizon 2020 grant 643706. We acknowledge the financial support from the Slovenian Research Agency (research core fundings No. P2-0209 and P2-0103). This research has received funding also from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 720270 (HBP SGA1).

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) (https://www.ppmi-info.org/data). For up-to-date information on the study, visit https://www.ppmi-info.org. PPMI—a public-private partnership—funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners. Corporate Funding Partners are: AbbVie, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, GE Healthcare, GLAXOSMITHKLINE (GSK), Eli Lilly and Company, Lundbeck, Merck, Meso Scale Discovery (MSD), Pfizer Inc, Piramal Imaging, Roche, Sanofi Genzyme, Servier, Takeda, Teva, UCB. Philanthropic Funding Partners: Golub Capital. List of funding partners is found at https://www.ppmi-info.org/fundingpartners.

References

[1] Evidence Based Medicine Publications for Treatment of Motor and Non-motor symptoms of Parkinson's disease. http://www.movementdisorders.org/MDS/Resources/Publications-Reviews/EBM-Reviews.htm [accessed 20.10.17].

[2] Arauzo-Azofra A, Aznarte JL, Benítez JM. Empirical study of feature selection methods based on individual feature evaluation for classification problems. Expert Syst Appl 2011;38(7):8170–7.

[3] Blockeel H, De Raedt L. Top-down induction of first-order logical decision trees. Artif Intell 1998;101(1–2):285–97.

[4] Blockeel H, Raedt LD, Ramon J. Top-down induction of clustering trees. Proceedings of the fifteenth international conference on machine learning (ICML)

[6] In further analysis of the tree leaves in the model from Fig. 8 we calculated average symptom values of covered patients.

1998:55–63.
[5] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[6] Caruana R. Multitask learning. Mach Learn 1997;28(1):41–75.
[7] Caruana R, Baluja S, Mitchell T. Using the future to "sort out" the present: Rankprop and multitask learning for medical risk evaluation. Advances in neural information processing systems 1996:959–65.
[8] Dalrymple-Alford J, MacAskill M, Nakas C, Livingston L, Graham C, Crucian G, et al. The MoCA: well-suited screen for cognitive impairment in Parkinson disease. Neurology 2010;75(19):1717–25.
[9] Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease. Comput Methods Prog Biomed 2014;117(3):405–11.
[10] Ferreira J, Katzenschlager R, Bloem B, Bonuccelli U, Burn D, Deuschl G, et al. Summary of the recommendations of the EFNS/MDS-ES review on therapeutic management of Parkinson's disease. Eur J Neurol 2013;20(1):5–15.
[11] Foltynie T, Brayne C, Barker RA. The heterogeneity of idiopathic Parkinson's disease. J Neurol 2002;249(2):138–45.
[12] Fox SH, Katzenschlager R, Lim S-Y, Ravina B, Seppi K, Coelho M, et al. The movement disorder society evidence-based medicine review update: treatments for the motor symptoms of Parkinson's disease. Mov Disord 2011;26(S3):S2–41.
[13] Gil D, Johnson M. Diagnosing Parkinson by using artificial neural networks and support vector machines. Glob J Comput Sci Technol 2009;9(4):63–71.
[14] Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008;23(15):2129–70.
[15] Hoehn MM, Yahr MD. Parkinsonism onset, progression, and mortality. Neurology 1967;17(5):427.
[16] Holzinger A. Trends in interactive knowledge discovery for personalized medicine: cognitive science meets machine learning. IEEE Intell Inf Bull 2014;15(1):6–14.
[17] Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 1992;55(3):181–4.
[18] Kim J-W, Kwon Y, Kim Y-M, Chung H-Y, Eom G-M, Jun J-H, et al. Analysis of lower limb bradykinesia in Parkinson's disease patients. Geriatr Gerontol Int 2012;12(2):257–64.
[19] Kira K, Rendell LA. The feature selection problem: traditional methods and a new algorithm. Proceedings of the tenth national conference on artificial intelligence, Vol. 2 1992:129–34.
[20] Kocev D, Vens C, Struyf J, Džeroski S. Ensembles of multi-objective decision trees. European conference on machine learning. 2007. p. 624–31.
[21] Lainscsek C, Rowat P, Schettino L, Lee D, Song D, Letellier C, et al. Finger tapping movements of Parkinson's disease patients automatically rated using nonlinear delay differential equations. Chaos: Interdiscip J Nonlinear Sci 2012;22(1):013119.
[22] Lewis S, Foltynie T, Blackwell A, Robbins T, Owen A, Barker R. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. J Neurol Neurosurg Psychiatry 2005;76(3):343–8.
[23] Ma L-Y, Chan P, Gu Z-Q, Li F-F, Feng T. Heterogeneity among patients with Parkinson's disease: cluster analysis and genetic association. J Neurol Sci 2015;351(1):41–5.
[24] Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The Parkinson's progression markers initiative (PPMI). Prog Neurobiol 2011;95(4):629–35.
[25] National Collaborating Centre for Chronic Conditions. Parkinson's disease: national clinical guideline for diagnosis and management in primary and secondary care. London: Royal College of Physicians; 2006.
[26] Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. IEEE Trans Inf Technol Biomed 2009;13(6):864–73.
[27] Pinto S, Velay J-L. Handwriting as a marker for PD progression: a shift in paradigm. Neurodegener Dis Manag 2015;5(5):367–9.
[28] Quinlan JR. C4. 5: programs for machine learning. Elsevier; 2014.
[29] Ramani RG, Sivagami G. Parkinson disease classification using data mining algorithms. Int J Comput Appl 2011;32(9):17–22.
[30] Rana B, Juneja A, Saxena M, Gudwani S, Kumaran SS, Behari M, et al. Graph-theory-based spectral feature selection for computer aided diagnosis of Parkinson's disease using T1-weighted MRI. Int J Imaging Syst Technol 2015;25(3):245–55.
[31] Reijnders J, Ehrt U, Lousberg R, Aarsland D, Leentjens A. The association between motor subtypes and psychopathology in Parkinson's disease. Parkinsonism Relat Disord 2009;15(5):379–82.
[32] Reyes O, Morell C, Ventura S. Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context. Neurocomputing 2015;161:168–82.
[33] Riviere CN, Reich SG, Thakor NV. Adaptive Fourier modeling for quantification of tremor. J Neurosci Methods 1997;74(1):77–87.
[34] Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn 2003;53(1–2):23–69.
[35] Robnik-Šikonja M, Savicky P. Corelearn: classification, regression and feature evaluation, r package version 0.9.45. 2015http://CRAN.R-project.org/package=CORElearn.
[36] Rosenblum S, Samuel M, Zlotnik S, Erikh I, Schlesinger I. Handwriting as an objective tool for Parkinson's disease diagnosis. J Neurol 2013;260(9):2357–61.
[37] Seppi K, Weintraub D, Coelho M, Perez-Lloret S, Fox SH, Katzenschlager R, et al. The movement disorder society evidence-based medicine review update: treatments for the non-motor symptoms of Parkinson's disease. Mov Disord 2011;26(S3).
[38] Shang R, Wang W, Stolkin R, Jiao L. Subspace learning-based graph regularized feature selection. Knowl Based Syst 2016;112:152–65.
[39] Struyf J, Zenko B, Blockeel H, Vens C, Dzeroski S. Clus: User's manual. 2010.
[40] Sun X, Liu Y, Li J, Zhu J, Liu X, Chen H. Using cooperative game theory to optimize the feature selection problem. Neurocomputing 2012;97:86–93.
[41] Timmer J, Gantert C, Deuschl G, Honerkamp J. Characteristics of hand tremor time series. Biol Cybern 1993;70(1):75–80.
[42] Valmarska A, Miljkovic D, Konitsiotis S, Gatsios D, Lavrač N, Robnik-Šikonja M. Combining multitask learning and short time series analysis in Parkinson's disease patients stratification. Proceedings of the conference on artificial intelligence in medicine in Europe. 2017. p. 116–25.
[43] Valmarska A, Miljkovic D, Lavrač N, Robnik-Šikonja M. Analysis of medications change in Parkinson's disease progression data. J Intell Inf Syst 2018.
[44] Valmarska A, Miljkovic D, Robnik-Šikonja M, Lavrač N. Multi-view approach to Parkinson's disease quality of life data analysis. Proceedings of the international workshop on new frontiers in mining complex patterns. 2016. p. 163–78.
[45] Visser M, Marinus J, Stiggelbout AM, Van Hilten JJ. Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. Mov Disord 2004;19(11):1306–12.
[46] Washburn RA, Smith KW, Jette AM, Janney CA. The physical activity scale for the elderly (PASE): development and evaluation. J Clin Epidemiol 1993;46(2):153–62.
[47] Weintraub D, Koester J, Potenza MN, Siderowf AD, Stacy M, Voon V, et al. Impulse control disorders in Parkinson disease: a cross-sectional study of 3090 patients. Arch Neurol 2010;67(5):589–95.
[48] Weintraub D, Mamikonyan E, Papay K, Shea JA, Xie SX, Siderowf A. Questionnaire for impulsive-compulsive disorders in Parkinson's disease-rating scale. Mov Disord 2012;27(2):242–7.
[49] Wilcoxon F. Individual comparisons by ranking methods. Biomet Bull 1945;1(6):80–3.
[50] Zhang D, Shen D, ADNI. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 2012;59(2):895–907.
[51] Zhou J, Liu J, Narayan VA, Ye J, ADNI. Modeling disease progression via multi-task learning. NeuroImage 2013;78:233–48.