Technical report
(available at www.ailab.si/parkinsoncheck)

ParkinsonCheck

A Decision Support System for Tremor Detection

Aleksander Sadikov · Jure Žabkar · Martin Možina · Vida Groznik · Dejan Georgiev · Ivan Bratko

the date of receipt and acceptance should be inserted later

Abstract The paper details the construction of a clinical decision support system which is the integral part of the Parkinson*Check* application. Parkinson*Check* is an app for smart phones built to detect signs of Parkinsons disease (PD) and essential tremor (ET), which is the main differential diagnosis from PD in the early stage of the disease. The application is based on spirography (spiral drawing) and uses the built-in decision support system to detect potential signs of the disease. The intention is to warn potential patients of their signs as early as possible and guide them to seek medical help. The decision support system is what separates Parkinson*Check* from similar apps which rely on sending the users data to a medical facility to be assessed by a neurologist. It enables our application to be fully automated. The focus of this paper are the inner workings of the decision support system which is used in two ways: (1) to detect the signs of PD and ET, and (2) to differentiate between PD and ET.

Keywords Parkinson's disease · essential tremor · decision support system · mobile devices · early diagnosis · spirography

A. Sadikov
University of Ljubljana,
Faculty of Computer and Information Science
Večna pot 113, SI – 1000 Ljubljana, Slovenia
Tel.: +386 1 4798 210
E-mail: aleksander.sadikov@fri.uni-lj.si

J. Žabkar, M. Možina, V. Groznik, I. Bratko
University of Ljubljana,
Faculty of Computer and Information Science
Večna pot 113, SI – 1000 Ljubljana, Slovenia

D. Georgiev
University Medical Centre Ljubljana,
Department of Neurology
Zaloška cesta 2, SI – 1000 Ljubljana, Slovenia

1 Introduction and motivation

In this paper we describe the development of a medical tool for early detection of Parkinsons disease, implemented as a mobile application intended to be used by general public. The application is intended to detect the disease from a user-drawn spiral on a touch screen. We present in detail the process of and techniques for the acquisition of the decision knowledge through Machine Learning. Before we go into technical details, we start the paper with the medical significance and background of this application.

Parkinsons disease (PD) is a chronic, progressive neurodegenerative disease. It is the second most common neurodegenerative disorder after Alzheimer's disease. PD is estimated to affect between four and six million individuals over the age of 50, and that number is expected to double by the year 2030 [33]. The costs associated with 1.2 million PD patients in European Union were estimated at 7,029 million EUR for direct medical costs, 5,519 million EUR for direct nonmedical costs, and 1,386 million EUR for indirect costs for a total of 13,934 million EUR in the year 2010 [15].

Complicating matters is the difficulty of accurately diagnosing and effectively treating the disease. Clinical manifestations of PD are diverse. The main motor symptoms of PD are (1) tremor, (2) bradykinesia (slowness of movements), (3) rigidity (increase in muscle tone) and (4) impaired postural reflexes. Each patient experiences the disease differently, given that the duration and severity of the symptoms in each motor state is unique to a given individual.

Essential tremor (ET) is the most prevalent movement disorder. Although distinct clinical entities, ET is very often misdiagnosed as Parkinsonian tremor (PT) [30]. Results from clinical studies show that ET is correctly diagnosed in 50-63% of all cases, whereas PT in 76% of the cases. Co-existence of both disorders is also possible [26]. In addition, PT can be very often observed when the upper limbs are stretched (postural tremor) and even during limb movement (kinetic tremor), which further complicates the differential diagnosis of the tremors.

Digitalised spirography is a relatively new computer-assisted method for detection and evaluation of tremors. It has been used to evaluate different types of tremors, dyskinesias and general drawing impairment. The system for acquisition and analysis of spiral drawings is commonly composed of a computer, a tablet for digital acquisition of the signal and a stylus. The task of the patient is to draw an Archimedean spiral on the tablet. Different quantitative parameters are provided by spirography; these are discussed in subsection 4.1. However, the existing systems for digitalised spirography are not equipped with a decision support system and are unable to provide the physician with more than a visualisation support.

The present paper details the inner workings of our application Parkinson*Check* for smart phones. The application enables the user to take a spirographic test on his or her smart phone with the added value that the results of the test (spiral drawings) are automatically assessed. This is facilitated by

the built-in decision support system which looks for tell-tale signs of PD and ET in the drawn spirals. The construction and the workings of the decision support system are the main topic of this paper.

The main goal of Parkinson*Check* application is to reliably detect signs of PD or ET in the users spirals and to suggest an appointment with the neurologist at the onset of the disease. There are indications that early diagnosis is beneficial in terms of treatment, and there is no disagreement that observing patients in the early stages of the disease is crucial from the scientific point of view [1, 22]. This both leads to improved healthcare.

Parkinson*Check* application can also be used by the neurologists to replace the more expensive and non-mobile setup for digitalised spirography. This increases the availability of spirography while simultaneously reducing the costs of tremor diagnosis. It is of note that the neurologists have so far mostly used spirography as a confirmation test. Our application, perhaps after some revisions when more data are available for learning (which is ongoing), is aimed to potentially be used as a diagnostic test.

Another important aspect is the privacy of users data. There are existing systems for spirographic testing on mobile phones, however, all these systems send the data to a clinical repository where it is analysed by the physician. Having a built-in decision support system sets aside the need to transmit the sensitive data as the spiral assessment is performed on their smart phones. This often neglected aspect is very important to some users.

The organisation of the paper is as follows. We first give the overview of the related work in the next section. Then, in Section 3 we give the problem definition, and describe the learning data. In Section 4 the details of the Parkinson*Check* decision support system are given, including the background knowledge, the most important attributes, and the evaluation of the system. This section also includes a brief overview of the application implementation. The future work and discussion concludes the paper.

2 Related work

Movement disorders are one of the most common signs in patients with Parkinson's disease. Therefore collecting and analysing the motor data can give objective assessment of the severity of the disease.

Researchers are tackling this problem in various ways. Some are developing systems for monitoring and automatic recognition of gait [20, 24], some are monitoring and assessing disorders in speech and voice [16–18, 31], and some are monitoring tremors and bradykinesia [3, 11, 19, 28].

Since there are long queues and waiting times to be assessed by a neurologist, there is a need for a simple, affordable and fast at-home evaluation tool for tremors. Therefore researchers dedicate a lot of their time and effort into developing mobile [32] and wearable systems [6, 23, 27, 35] which could be used for home monitoring.

The majority of these systems only collect data which are then sent to a central repository. The data is afterwards remotely evaluated by the computer software of by the researchers. In these systems, the user's privacy issues can sometimes be of question. The systems do send the data over secured channels, but people can be distrustful towards the technology they are not familiar with and are easily convinced not to use them. To the best of our knowledge, our system is the only system which collects and assesses the data without sending it to the remote repository.

Although several products for home use do exist, they are mainly used for monitoring the progression of the disease and not for actually providing one with the preliminary diagnosis. To the best of our knowledge, Parkinson*Check* is the only system that is able to give users immediate feedback on whether they have signs of Parkinsonian or essential tremor or not. However, we do agree that objective monitoring and assessment of the tremor is of high importance. To this end, we are planing to upgrade our application to be able to monitor tremors as well.

3 Data and problem definition

Our dataset consisted of 159 subjects which were diagnosed at the University Medical Centre Ljubljana. Since our aim was to predict whether a person is healthy or define the type of tremor he or she has, it was very important that we have as accurate data as possible. Therefore neurologists experienced in diagnosing movement disorders examined and diagnosed all subjects with either PD, ET, without tremor (control), or other tremor types. The patients with other types of tremors were excluded from our study. This left us with 143 subjects.

The distribution of the diagnoses was: 38 patients diagnosed with PD, 19 with ET and 86 control subjects. When building the final decision model, we excluded the patients who have administered L-Dopa prescription drug less than three hours before taking the test. The final dataset consisted of 124 subjects; 17 patients with ET, 21 patients with PD and 86 control subjects. Initial and final distribution of the data are presented in Table 1.

		Essential tremor	Parkinsonian tremor	Controls
Initial dataset	n	19	38	86
	Proportions	13.29%	26.57%	60.14%
Final dataset	n	17	21	86
	Proportions	13.71%	16.94%	69.35%

Table 1: Distribution of the diagnoses, after excluding subjects with "other tremor" diagnosis.

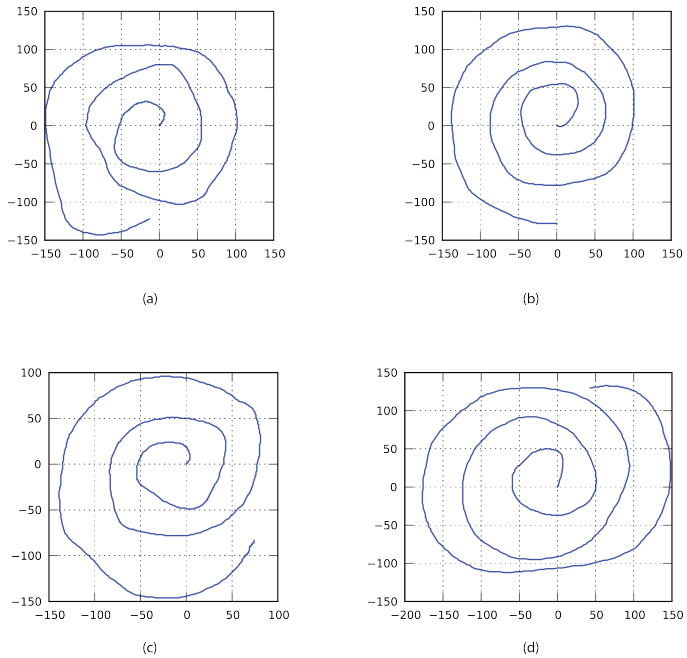

Fig. 1: Example of spirals (a) left hand – template, (b) right hand – template, (c) left hand – freehand and (d) right hand – freehand.

3.1 Data acquisition using spirography

Digitalized spirography [10,25] is a method for detection and evaluation of different types of tremors. The system is usually composed of a computer with a dedicated software, a graphic tablet and a stylus. The digitalized spirography enables us to store the exact timestamp of each point in a two-dimensional area. This is useful for further analyses of the time series (e.g. frequency analysis), which in our case was needed for building the decision model. For our purposes, the spirography test was made using mobile phones with touch screens. Unlike the standard digitalized spirography, the subjects were asked to draw the spirals using their fingers and not a stylus. The reason was that only a minority of people own styluses for touch screen phones.

Every subject had to draw four spirals on a mobile phone (see Figure 1). The first task was to follow the line of an already drawn Archimedean spiral (a template) with a right hand finger. The second task was to draw an Archimedean spiral without a template using the right hand finger. Tasks three and four were the same as one and two, the only difference being that they had to be performed using a left hand finger.

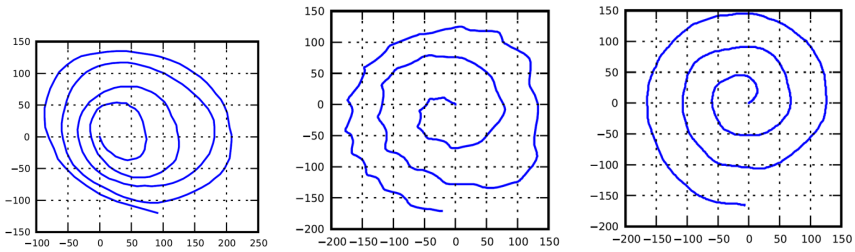

Fig. 2: Typical spirals for PT, ET and control subjects.

3.2 Problem definition

Our main goal was to automatically detect a potential patient with PD or ET using spirography on mobile phones. After a person has drawn the four required spirals, our application should be able to detect potential signs in the spirals indicating either PD or ET. Furthermore, if the test is positive, the application should be able to decide, whether the spirals (or their features) speak in favour of PD or ET. This problem amounts to have a machine learned model that can classify a tested subject from the spirals alone. The learning problem to get such a model was formulated as:

1. Given a set of classified subjects (healthy, PD, ED), each described with four spirals;
2. Learn a classification model that can detect whether a subject has tremor or not, and
3. Learn a classification model the can separate/differentiate between PT and ET.

Figure 2 shows three typical spirals drawn by a PD patient, an ET patient and a healthy subject, respectively. Different types of tremors will usually reflect differently in a drawn curve. For example, tremor of a PD patient has lower frequency, hence, the pattern of tremor will be spread over a longer arc of a spiral: in our case, the spiral became asymmetric. On the other hand, the tremor of an ET patient has a higher frequency which is manifested with a regular fluctuation along the drawn curve. As can be seen, the person from the control group drew a nice, regular spiral.

However, the learning problem is not that easily separable as one could mistakenly assume given the Figure 2. It is not rare for healthy people to draw irregular spirals, which can happen due to various reasons. On the other hand, a person with a PD or ET might still draw an almost flawless spiral. For example, some variants of Parkinson's disease result in tremor only if the person is not moving. When the hand is being moved, as it is during drawing, the tremor disappears. Figures 3 and 4 show examples of almost perfect spirals drawn by patients with essential tremor and PD, respectively.

Note that in the learning problem we did not distinguish between different types of error: misclassifying a healthy person as a potential patient carried

equal weight as if a sick person was classified as healthy. We were instructed to do so by the medical experts, since they were unable to put the weight on one or the other side. The question is problematic due to ethical reasons. If a healthy person is classified as a potential patient, we might induce ungrounded fear in people, while failing to detect a sick person means their medical treatment will be delayed.

4 Decision support system

The decision support system is the heart of the Parkinson*Check* application. In this section we describe the background knowledge that eased the construction of the system, the most important attributes for describing the spiral data, and the model for the detection of tremor signs. We also evaluate the accuracy of the model, and discuss briefly the implementation of the application.

4.1 Initial visualisation of data

Previous to our project, the data (spirography) were collected without the decision support system. They obtained the data using graphic tablets. Each spiral was visualised with six basic plots that helped the doctor determine —

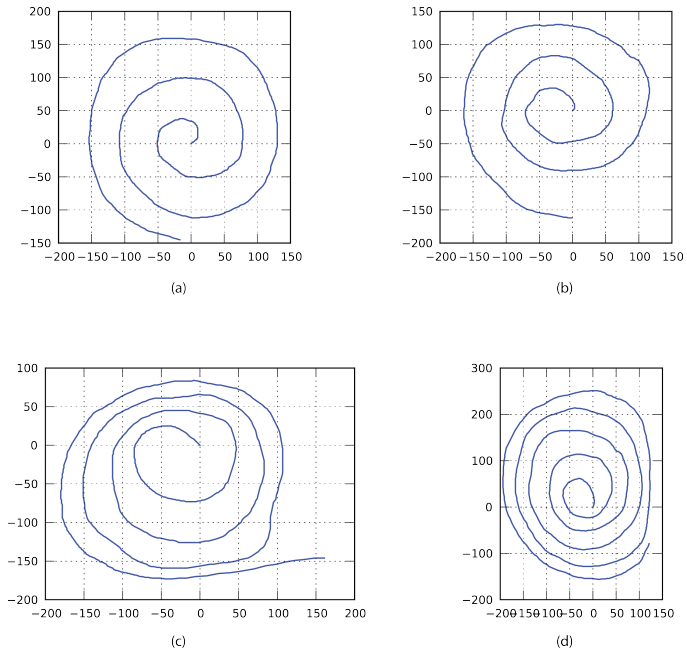

Fig. 3: Example of non-typical ET spirals (a) left hand – template, (b) right hand – template, (c) left hand – freehand and (d) right hand – freehand.

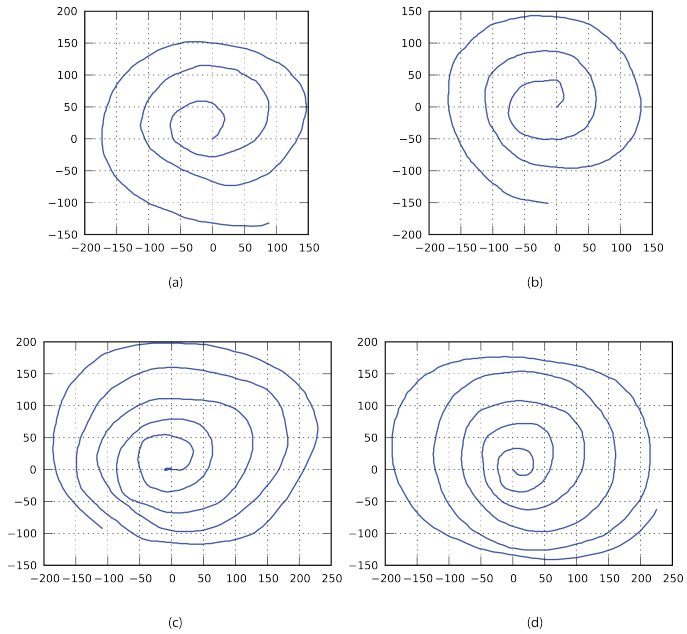

Fig. 4: Example of non-typical PT spirals (a) left hand – template, (b) right hand – template, (c) left hand – freehand and (d) right hand – freehand.

or in most cases just confirm — the diagnosis. It should be noted that the assessment was strictly visual. These plots represent a starting point in our modelling process, particularly for attribute construction.

The plot of **a spiral in cartesian coordinate system** (Fig. 5a) provides the first visual impression of the patient's trial. It helps detecting the most obvious cases, which are reflected in typical irregularities, e.g. the wave-like pattern of ET patients or skewness of the spiral in PD patients. However, in trying to give their best, the patients often reduce the drawing speed or increase the pressure of the pen/finger upon the drawing surface, which results in spirals that dim the actual disease.

Absolute speed (Fig. 5b) measures the actual speed of drawing. Taking two consecutive points, (x_i, y_i) and (x_{i+1}, y_{i+1}) the absolute speed v is the distance between them divided by the difference in timestamps for the two points:

$$v = \frac{\sqrt{(x_{i+1} - x_i)2 + (y_{i+1} - y_i)2}}{t_{i+1} - t_i}.$$

The patients with severe movement disorders may be able to draw the spirals that look very much like the ones drawn by a healthy person yet they may need up to ten times longer to do so. Observing such trials suggests that they also press the pen or finger much stronger against the drawing surface, but we were not able to measure the force reliably with the smart phones.

The Archimedean spiral in **polar coordinates** (Fig. 5c) is a straight line, which makes visual comparison with the drawn spiral converted to polar coordinates much easier. In the context of the next two plots — radial and angular velocity — it focuses on smoothness of the spiral.

Radial velocity (Fig. 5d) measures the change in radius over time:

$$v_r = \frac{r_{i+1} - r_i}{t_{i+1} - t_i},$$

and reflects the smoothness of the spiral. Ideally, the radius should monotonically increase over time of drawing, but the v_r should remain constant. Certain movement disorders (e.g. ET) prevent the patients to control this, which results in different non-monotonic patterns.

Angular velocity (Fig. 5f) measures the change in angle over time:

$$\omega = \frac{\varphi_{i+1} - \varphi_i}{t_{i+1} - t_i}.$$

The angular velocity usually decreases during drawing.

To denoise the above signals we smoothed them using the median filter with window size 5.

Frequency analysis (Fig. 5e) plots the power spectrum density of the absolute speed signal. The absolute speed is first denoised, and then also normalised to the interval [0,1] by dividing each value with the maximum observed speed. The normalisation proved crucial for the calculation of useful attributes. This plot can reveal rhythmic motion at different frequencies which is the hallmark of various tremors.

4.2 Background knowledge

Essential tremor (ET) is one of the most prevalent movement disorders [9]. It is characterized by postural and kinetic tremor with a frequency between about 6 and 12 Hz. Although it is regarded as a symmetrical tremor, ET usually starts in one upper limb and then spreads to the other side affecting the contralateral upper limb, consequently spreading to the neck and vocal cords, giving rise to the characteristic clinical picture of the disorder. However, there are many deviations from this classical presentation of ET, e.g. bilateral tremor onset, limb tremor only, head tremor only, isolated voice tremor. Parkinsonian tremor (PT), on the other hand, is a resting tremor classically described as "pill rolling" tremor with a frequency between 4 and 6 Hz. It is one of the major signs of Parkinson's disease (PD), which also includes bradykinesia, rigidity and postural instability. PT is typically asymmetrical, being more pronounced on the side more affected from the disease onset.

As we mentioned before, neurologists can to a certain extent visually appraise the drawn spirals from the graphs described in the previous section. Experts' background knowledge was based on their experience with digitalised spirography, which they have been using in clinical practice for several

years. The majority of this knowledge was successfully elicited in our previous study [14]. We were aware of many characteristics of the spirals associated with certain diagnoses, e.g. saw-like patterns, sharp square-like edges, presence of peaks and harmonics in the spectral density graph, compressions and asymmetry of the spirals on one side, asymmetry between left/right hand spirals, etc. Some of these characteristics can be seen in Figure 2. The relevant background knowledge is listed alongside each of the constructed attributes in the subsequent section.

4.3 Attributes for describing the spirals

In the following subsections, the most important attributes for describing the spirals are explained. With the exception of the last group, all the attributes are calculated for all four trials: freehand and template drawing with both the left and the right hand. The user's handedness is also recorded, and taken into account — the trials are separated into those with the dominant hand and those with the non-dominant hand.

4.3.1 Symmetry group

The attributes in this group describe the symmetry of the drawn spirals in two ways: (a) the symmetry of the spiral itself, and (b) the symmetry between spirals drawn with the left and the right hand.

Single spiral asymmetry steams from the disturbed movement physiology during spiral drawing (e.g. tremor, muscle rigidity and slowness of movements in PD patients). This in turn skews the spiral in one direction or another and creates asymmetry. On the other hand, the asymmetry between drawing with the left and the right hand is present in most people as the majority are not ambidextrous. However, this left/right hand asymmetry gets more pronounced in PD patients and in patients with other movement disorders i.e. dystonia. It is of special importance that the tremor-onset and tremor-presentation is typically unilateral in PD and bilateral in ET, making asymmetry a potential attribute for differentiating ET and PD.

Attributes `Xsymm` and `Ysymm` measure the asymmetry in X and Y coordinates, respectively. The spiral always starts at $(0, 0)$. `Xsymm` is calculated as the difference between the distance of the leftmost point of the spiral from the Y axis and the distance of the rightmost point of the spiral from the Y axis, divided by the sum of these two distances. It is divided by the sum of the distances as to normalise with the overall size of the spiral in the given direction. `Ysymm` is calculated analogously. Besides X and Y, other directions of asymmetry could be checked as a further improvement.

Please note that the Parkins*Check* application requires the user to start the spiral in the centre of the screen and a deviation from that is automatically detected prompting the user to redraw the spiral. This prevents spirals from being skewed due to lack of the drawing surface in any given direction.

The left/right hand asymmetry is measured by the attribute `angVsRad.template.distance`. This attribute is calculated on left/right hand spirals drawn on templates as these can be compared. The template ensures that the spirals are of the same size as it is exactly the same for both hands.

The calculation is performed on polar representation of the two spirals. The first half turn is disregarded as the start of the drawing is usually not reliable. The remaining length (turns) of the two spirals are then sampled evenly (the angle is monotonically increasing) until the end of the shorter spiral; in all 300 points are sampled. At sampled points (angles) the radius is interpolated for both left and right hand spirals. These two interpolated radius series are compared using root mean squared error (RMSE). The RMSE is the value of the attribute `angVsRad.template.distance`.

4.3.2 Extrema group

The attributes in this group are quite rudimentary and can be applied to most time series data. They were the first programmed with the intention to quickly get a working prototype with at least some classification ability. It was believed that these attributes would later be superseded by the more involved ones, however, it turned out that they remained quite important in the final model.

The initial idea behind these attributes was to describe the number, size, and distribution of the peaks in the three speed series (absolute, radial, and angular). The attributes were motivated by the saw-like patterns in these series 5(b,d,f). Such patterns involve a number of direction changes in the series manifesting themselves as oscillations or local extrema. Physiologically, the saw-like patterns occur due to tremor as the speed and direction of movement rapidly change. This is especially noticeable with ET as its frequency is usually higher, manifesting itself in many smaller peaks (extrema) over a shorter time frame.

The sequence of measurements is first analysed for local extremes (minima and maxima). These are extracted along with the first and the last point of the sequence. The absolute differences between each two neighbouring extremes are calculated (and referred to as $\Delta Peaks$ in the continuation).

Attribute `nPeaks` is the number of $\Delta Peaks$ (direction changes) normalised with the length of the sequence. The attributes `avgP` and `stdevP` are the mean and standard deviation of $\Delta Peaks$, respectively. The attribute `t.rho` represents the Spearman's rank correlation coefficient between the given sequence of measurements and the timestamps for the sequence.

Attribute `incDev` describes whether and how the standard deviation of $\Delta Peaks$ changes with time. Before computing it, the input sequence is pruned at the beginning and the ending to remove the occasionally unreliable parts; 5% of the sequence is removed from both sides. The calculation uses a sliding window and compares the standard deviations over 30 overlapping windows. It reports the correlation with time using the Spearman's rank correlation coefficient.

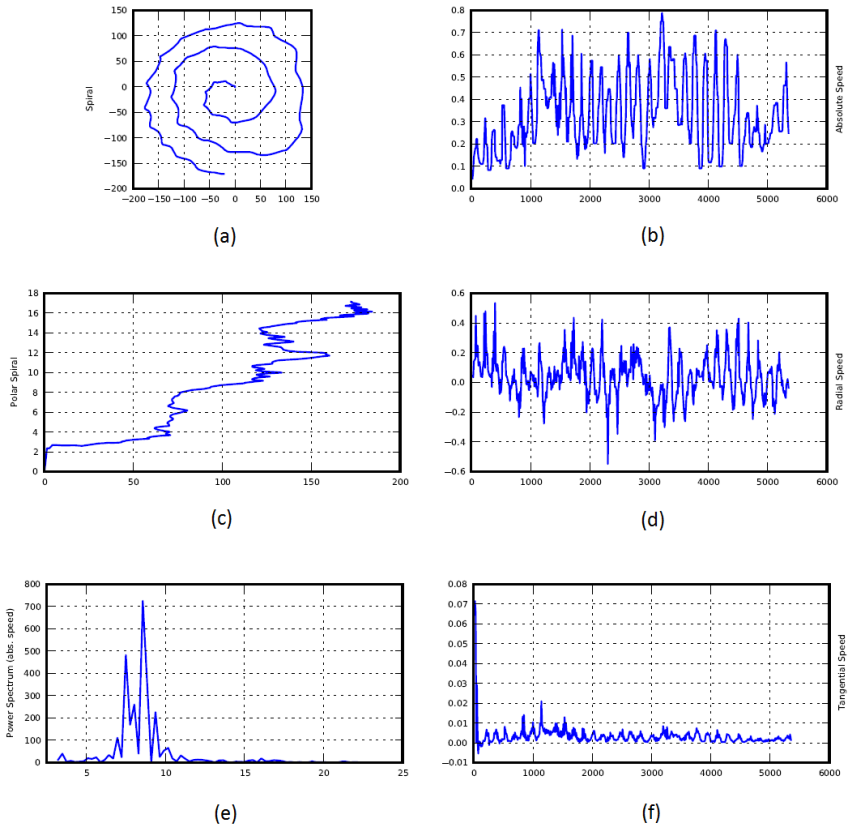

Fig. 5: Six basic plots of spirography: (a)spiral in cartesian coordinates, (b)absolute speed, (c)spiral in polar coordinates, (d)radial speed, (e)frequency graph, and (f)tangential speed.

The extremes represent the change in direction (increase/decrease) in the sequence. For example, in the sequence representing absolute speed of drawing, the extremes represent the time points where the user started to accelerate or decelerate. Frequent bursts and changes manifest themselves as saw-like patterns, especially short changes. This is the idea behind using the average and standard deviation along with the normalised number of $\Delta Peaks$. The reasoning behind the `t.rho` attribute is the observation that several sequences should be increasing with time; this is certainly true for radius and angle (measured continuously), and is mostly true for the speed of drawing as well as the outer turns of the spiral get bigger and people usually draw them faster. The motivation for the `incDev` attribute is the speculation that the deviations could be more pronounced on the more convoluted parts of the spiral.

The attributes is this group are applied to various input sequences: all three different speeds, and also to the change of radius and angle over time.

4.3.3 Error (RMSE) group

Measured data consists of a timestamp t and the x, y screen coordinates at a given timestamp.

From the medical perspective, movement disorders present with increased disability to follow the template while drawing. To capture the deviations of the drawn spiral from the template — the ideal Archimedean spiral — we constructed several attributes that measure root mean squared error (RMSE) between the two.

Attributes `err`, `err0`, and `errBF` measure the error in polar coordinate space. First, we transform the original $(x(t), y(t))$ coordinates from Cartesian to polar coordinate system by: $r = \sqrt{x(t)^2 + y(t)^2}$ and $\varphi(t) = atan2(y(t), x(t))$. To respect the continuity of $\varphi(t)$ we add an offset of 2π to $\varphi(t)$ when $\varphi(t-1) > \varphi(t) + \pi$.

To construct the attribute `err`, we perform a least squares fit to the points (φ, r) and compute the RMSE. In a similar way, we construct the attribute `err0` — the only difference is that we constrain the fit line to pass through $(0, 0)$ as the ideal Archimedean spiral should.

However, the spiral may be perfect, but translated from $(0, 0)$. To measure the deviation due to translation, we compute the origin of the user's spiral, O_u, translate it to $(0, 0)$ and compute RMSE as in the case of the attribute `err0`. O_u is the origin of the spiral that is the best fit to the user's spiral. We use differential evolution algorithm [29] to determine O_u. The attribute `errBF` measures the error due to the translation of the spiral.

A visual observation of the spirals suggests that curvature may be an important attribute. The curvature of the Archimedean spiral decreases monotonically with increasing φ, while this is usually not true for the spirals drawn by the users, neither healthy nor those with movement disorders. Instead of the curvature, we calculated the approximations of radii of curvature by fitting circles to subsequences of x, y of different window sizes: 10, 20, 30, 50, 75, 100, 125, 150, and 175; a wider window size implies more smoothing. For example, we calculated the radius of curvature R_i by fitting a circle to a set of points $(x_j, y_j)|j = i, \ldots, i + wSize$, where $wSize$ is the size of the window. Finally, we fitted a straight line to the set of radii R_i and measured the RMSE. The attributes `rmse100` and `rmse150` — 100 and 150 denoting the window size — turned out as important.

Also important were the attributes `rmseSum` and `rmseMin`, representing the sum and the minimum of all window sizes, respectively. These two attributes capture the overall imperfection in curvature with respect to smoothing. The intuitive idea behind these attributes is that no level of smoothing would cover up the curvature imperfection caused by movement disorders, while it would cover up the imperfections by the healthy user.

4.3.4 Frequency group

Tremor is a movement disorder characterized by rhythmic, oscillatory and alternate contraction of agonist and antagonist muscles [4]. It is a common symptom and sign of different neurological disorders and can be a separate nosological entity in its own merits, as in the case of ET. ET tremor is typically symmetrical, bilateral postural tremor with a frequency between 4 and 12 H. In contrast, PT tremor is described as resting tremor with lower frequency (between 4 and 9 Hz) than ET. Hence, the frequencies of ET and PT overlap considerably. In addition, resting tremor can be present in ET in as high as 18% of the patients [7], and PD patients also have a postural tremor in almost 60% of the cases [2].

The attributes in this group describe the frequency profile of the absolute speed of drawing. The other two speed graphs were also analysed in this manner, but yielded no significant benefit to the model.

The attribute `freq.dev` represents the standard deviation of the power spectrum density of the absolute speed (see Figure 5 e the one showing the frequency graph) in the range between 3 and 22.5 Hz. The intuitive explanation for the added value of this attribute is that it captures whether there are peaks in the spectrum or not, and the peaks represent rhythmic movement.

The attribute `peakAmpXOut` describes the combined magnitude and outlier level of the main peak in the power spectrum density. The main function of this attribute is to detect whether there is a prominent (and how prominent) peak in the power spectrum.

The power spectrum density of absolute speed (normalised with the maximum speed) is first calculated by Welch's average periodogram method. The input is divided into blocks of length 128. The sampling frequency is automatically detected for each device from the obtained data. The resulting power spectrum density is limited to the band between 4.5 and 22.5 Hz due to Nyquists law and some observed unreliability of very low frequencies. The power (amplitude) of the largest peak and its absolute deviation from the sample mean in units of the sample standard deviation are calculated (the sample being the power spectrum in the given band). The value of `peakAmpXOut` attribute is the product of these two values (the power and the outlier level). In this was this attribute takes both measures equally into account. It is quite surprising that the frequency at which the main peak occurs, did not feature in the final model.

4.3.5 Radial group

The motivation behind this group was the observation that movement disorders often result in frequent changes of direction when drawing the spirals. These changes are also often at relatively sharp angles (rotations). Apart from the spiral itself, such changes/rotations are also visible on the graph of radial speed.

This group features only a single attribute appearing in the final model. The attribute `percNeg` measures the percentage of the spiral length when the user is drawing towards the centre. This is simply the percentage of the length when the radial speed is below zero. When drawing the ideal Archimedean spiral, the radius is constantly increasing, thus the drawing is never towards the centre. Obviously, even the healthy users occasionally have to correct the direction and draw towards the centre, however, this is much less frequent than with users with impaired movement.

4.3.6 Multi-test group

All the attributes described in the other groups describe some characteristic of a single spiral, i.e. they are calculated given just one drawing as an input. However, just a single arm can be affected by the movement disorder and even that can manifest itself only in free drawing (or template drawing). This was the motivation behind this group of attributes. We calculated minimum, maximum, average, and the sum of any attribute previously described over all four drawings (left/right template or freehand) and included these statistics as further attributes.

4.4 Classification model for detecting signs of PD or ET

An important step of the expert system is to automatically detect whether a drawn spiral contains patterns that are typical for Parkinsonian disease or the essential tremor. We tested several machine learning methods and selected a logistic regression model as the one to be implemented in the system. In the remaining of this section, we describe the experiments and the results that led us to choose the logistic regression.

4.4.1 Experimental setup

We experimented with four machine learning methods: logistic regression (LR), naive bayes classification (NB), support vector machines (SVM), and random forests (RF). We selected the first two methods, as they are both linear and can be easily explained and validated against our background knowledge of the domain. We included the latter two methods in case some nonlinear relevant patterns were present. The methods were implemented and evaluated using the Orange data-mining library [8]. The logistic regression is from a ported version of the *liblinear* library for large linear classifications [12] and was used with the default parameters. In the naive bayesian classifier, the m parameter of the m-estimate of probability [5] was set to 20 to reduce over-fitting. The parameters of SVM were automatically optimized with internal cross validation. The number of trees in the random forest was set to 150.

As most of the attributes derived from the spirals have continuous values, we were faced with the question whether we should discretise them or not.

There were two reasons for doing so: (1) setting-up thresholds is a common practice in medicine since there usually are "normal" and "abnormal" values (e.g. temperature higher than 37 degrees implies sickness), and (2) the learning data is relatively small, hence discretisation can lead to less over-fitting. On the other hand, the discretisation reduces information and we expected that some methods would suffer from it; especially methods such as logistic regression that are inherently built for continuous attributes. To answer this question we tested every method twice:

without discretisation: where the attributes were left as they were, and
with discretisation: where the entropy-based discretisation [13] that minimizes the class-entropy of training examples was used. Note that this approach also acts as a feature subset selection, as it will filter out an attribute, if its discretisation does not sufficiently decrease the entropy.

The classification in the application is divided into two steps. The first step analyses the drawn spirals for any signs indicating a disease. The second step, invoked if the first one is positive, distinguishes between Parkinsonian disease and essential tremor. Our question was: Is it better to use one model for both steps or would it be better to learn a separate model for the first step, where learning examples (patients) having Parkinsonian disease or essential tremor would be combined into one class? These two different approaches were then tested on the two-class problem (first step):

learning from two classes: where a classification model is learned from the data classified in two classes (healthy versus PD+ET), and
learning from three classes: where a classification model is learned using all three classes (healthy, PD, and ET) and tested on the two-class data. The probability of having PD or ET was computed as the sum of predicted probabilities for PD and ET.

All the methods were evaluated with three measures: classification accuracy (CA), area under curve (AUC) and Brier score (Brier).

4.4.2 Evaluation

Table 2 compares the selected methods with and without discretisation. The results show that all methods benefit from discretisation, which is interesting and somehow counter-intuitive. As already mentioned, we expected that some methods, especially logistic regression, should perform worse with discretisation.

We decided to further explore this surprising result. Namely, one could argue that it is the inherent feature subset selection (FSS) of the discretisation that helps and not the discretisation itself. Table 3 disputes such an argument. The first two rows contain logistic regression without and with discretisation. In the third row, the results of logistic regression with forced discretisation without FSS are given: all attributes were discretised and included, even if the entropy decrease was not sufficient. In the last row, logistic regression learned

Table 2: Comparison of selected methods (with and without discretization) on the three class problem.

	Without discretization			With discretisation		
Method	CA	AUC	Brier	CA	AUC	Brier
Majority	0.694	0.500	0.472	0.694	0.500	0.472
LR	0.799	0.837	0.390	0.835	0.930	0.235
NB	0.784	0.872	0.342	0.822	0.900	0.304
RF	0.790	0.872	0.309	0.813	0.925	0.263
SVM	0.786	0.889	0.303	0.812	0.914	0.261

Table 3: Why discretisation helps? Comparison of logistic regression (LR), logistic regression with discretisation (LR + disc), logistic regression with forced discretisation of all attributes (LR + forced disc), and logistic regression where discretisation was used as feature subset selection, while the attributes remained continuous (LR + FSS only).

Method	CA	AUC	Brier
LR	0.799	0.837	0.390
LR + disc	0.835	0.930	0.235
LR + forced disc	0.806	0.926	0.282
LR + FSS only	0.777	0.870	0.325

Table 4: Comparison of selected methods on the two class problem. Methods were learned from data classified in either two or three classes. Attributes were discretised.

	Learning from two classes			Learning from three classes		
Method	CA	AUC	Brier	CA	AUC	Brier
Majority	0.694	0.500	0.472	0.694	0.500	0.472
LR	0.889	0.932	0.172	0.902	0.951	0.153
NB	0.894	0.903	0.196	0.895	0.938	0.185
RF	0.888	0.917	0.197	0.909	0.942	0.172
SVM	0.894	0.920	0.174	0.896	0.926	0.162

on continuous attributes is given, however taking only attributes that were kept by the entropy-based discretisation. Therefore, only the FSS part of the discretisation was used.

Given the results from Table 3, we concluded that discretisation without FSS helps, as having forced discretisation is better than learning on continuous attributes (according to all three measures). Likewise, we also concluded that FSS helps, as forced discretisation is worse than the FSS-enabled discretisation. Nevertheless, it is the discretisation that brings the larger portion to improvement, because logistic regression with forced discretisation is better than logistic regression with FSS only.

Table 4 shows surprising results on the two-class problem. It turned out that all models learned on the three-class data are better than the models learned on the two-class data, when tested on the two-class data. A plausible explanation might be that the spirals drawn by Parkinsonian patients are dissimilar to the spirals of essential tremor patients, therefore grouping them into the same class is not beneficial. These results speak against having two models for the two classifications steps in our application.

4.4.3 Selected model description

Our selected model was the logistic regression classification model on discretised attributes. This model is used for both classification tasks in the application.

Table 5 gives a glimpse of the logistic regression model. As logistic regression is designed for two-class problems, we needed to learn three models and combine their predictions by normalising their probability predictions. The three models were: healthy vs PD+ET, ET vs PD+healthy, and PD versus ET+healthy. Table 5 shows the first eight factors and their corresponding betas (linear coefficients) of the healthy versus PD+ET logistic regression model.

The most important attribute turned out to be `LF.err`: the distance between the optimal spiral and the spiral drawn freely (without a template) by the non-leading hand (LF). Interestingly, most of the important factors (six out of eight) come from the free drawings. In the future this could potentially enable us to use only freehand drawings when expanding the application for the purpose of monitoring the level of medication. In such a setting the user would need to take the test several times a day, and using only freehand drawings could mean a significant speed-up.

Table 5: Eight most important factors in logistic regression model for healthy vs having tremor (ET or PT).

Attribute	Beta
Prior	3.359
`LF.err` > 18.894	-1.532
`RF.rmse100` < 7.566	+1.061
`RF.rmseSum` < 206.47	+1.043
`LF.ang.avgP` > 1.290	+0.897
`LT.ang.peaks` > 0.017	-0.821
`RF.percNeg0` > 0.405	-0.800
`angVsRad.template.distance` > 236.36	-0.764
`LF.Ysymm` > 0.234	-0.735

4.5 Implementation

The decision support system was implemented into an application for mobile phones with touch screens called Parkinson*Check* (Figs. 6, 7). Our goal was to make it available for as many mobile platforms as possible. Therefore we developed an application for Android, iOS, BlackBerry and Windows Phone platforms and the application is freely available in the corresponding app stores. Moreover, we developed a HTML 5 web service[1], which allows users of any mobile platform to use the application in a web browser. At the moment the

[1] The Slovenian version of a web service is available at http://www.parkinsoncheck.net/pc

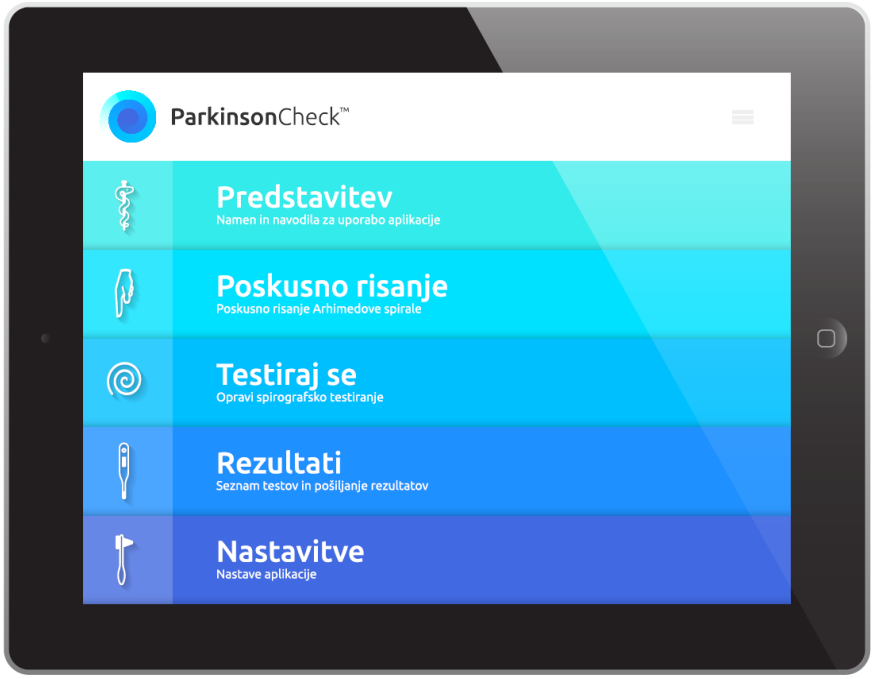

Fig. 6: Main menu of the Parkinson*Check* application.

application is available in Slovenian language and is being translated into an English version.

Parkinson*Check* application is able to perform the analysis of the spirals locally on mobile devices. The users are therefore not obliged to send their data to a remote server to get the results of the testing. However, we do encourage users to send us the data so we can develop better decision models in the future. The data is anonymised and sent using security protocols and cannot be tracked back to the users if in any way intercepted.

5 Discussion and conclusions

The initial classification problem, as defined in section 3.2, is essentially a time-series classification problem. Such problems involve training a classifier, where a case is represented by a set of ordered continuous values - spirals in our case. Several publications suggest that for this type of problems the k-nearest neighbour with dynamic time warping (kNN-DTW) method is very hard to beat [34]. However, our initial experiments with kNN-DTW lead to considerably worse results when compared to attribute-based machine learning. It seems that the available background knowledge, encoded in the designed attributes, enables considerably more information to a learning method than just

Fig. 7: Displaying diagnostic results in the Parkinson*Check* application.

the raw data representing a spiral. Due to the initial results, we did not pursue to experiment with time-series classification methods any further. Instead, we focused on the implementation of new attributes describing spirals.

As the saying goes, a picture is worth a thousand words. Our quest to feature construction started by examining the plots of spirals commented by clinical doctors. However, coding new meaningful attributes turned out to be much harder than observing an irregularity in an image of a spiral. The process was iterative - we constructed a new feature, visualised it and made several versions of it, varying its parameters. Finally, we let the FSS algorithm make the final choice.

The choice of which attributes to implement was generally driven by our background knowledge of the domain. However, in some cases, this selection was also driven by the errors produced by our selected machine learning methods. Here we will briefly describe the process how machine learning led us to implement new attributes.

We started by learning new models (with selected machine learning methods) using the current set of attributes. Then we identified instances that were misclassified by all models and called these examples the *problematic examples*. The question was: can we design such attributes that would help machine learned models to correctly predict these examples. To tackle the problem we involved a neurological expert and asked him to diagnose the problematic pa-

tients given the spirals only. If the doctor made the same mistake as the learned models, the instance was deemed as "not explainable" by spirals only and not a concern any more. However, when the doctor correctly diagnosed a patient, we tried to incorporate his explanations as new attributes. This process is similar to a learning loop in the argument-based machine learning [21].

In section 4.4.1 we mentioned that logistic regression and naive bayesian classifier were preferred because they are linear and can be easily explained. With linear models we can analyse which of the implemented attributes are deemed more important by the model, which enabled us, on several occasions, to spot anomalies such as bugs in attribute implementations or irregularities in the data. In cases where important attributes were not included in the model, we explored the causes and fixed it accordingly. The choice of a linear model also simplified the design architecture of the application. Currently, the logistic regression weights are stored in a separate file that is read by the application. When new data will arrive and new models will be built, the application will be upgraded with a new linear model by simply replacing that file.

The linearity also fits well in our future plans, when we intend to provide explanations of computer decisions. We intend to improve the user experience by visualizing the factors that are most influencing the automatic prediction (whether you have PD or ET) of the selected model. In a linear model, selecting the most influencing factors is usually a trivial task.

Acknowledgements The authors would like to thank Prof. Zvezdan Pirtošek, MD for helping with data acquisition and medical advice, and Ciril Bohak for designing the figures. The research was funded by Slovenian Research Agency (ARRS), Ministry of Education, Science and Sport and European Regional Development Fund.

References

1. Alberio T, Fasano M (2011) Proteomics in Parkinson's disease: An unbiased approach towards peripheral biomarkers and new therapies. Journal of Biotechnology 156:325-337
2. Bain PG (2007) Tremor. Parkinsonism & Related Disorders 13, Supplement 3(0):S369-S374, DOI 10.1016/S1353-8020(08)70032-4, proceedings of the XVII WFN World Congress on Parkinson's Disease and Related Disorders
3. Barroso MC, Esteves GP, Nunes TP, Silva LM, Faria AC, Melo PL (2011) A telemedicine instrument for remote evaluation of tremor: design and initial applications in fatigue and patients with Parkinson's disease. BioMedical Engineering OnLine 10(1):1–17, DOI 10.1186/1475-925X-10-14
4. Buijink AW, Contarino MF, Koelman JH, Speelman JD, van Rootselaar AF (2012) How to tackle tremor — systematic review of the literature and diagnostic work-up. Frontiers in Neurology 3(146), DOI 10.3389/fneur.2012. 00146
5. Cestnik B (1990) Estimating probabilities: A crucial task in machine learning. In: Proceedings of the Ninth European Conference on Artificial Intelligence, pp 147–149
6. Chen B, Patel S, Buckley T, Rednic R, McClure D, Shih L, Tarsy D, Welsh M, Bonato P (2011) A web-based system for home monitoring of patients with Parkinson's disease using wearable sensors. Biomedical Engineering, IEEE Transactions on 58(3):831–836, DOI 10.1109/TBME.2010.2090044
7. Cohen O, Pullman S, Jurewicz E, Watner D, Louis E (2003) Rest tremor in patients with essential tremor: Prevalence, clinical correlates, and electrophysiologic characteristics. Archives of Neurology 60(3):405–410, DOI 10.1001/archneur.60.3.405

8. Demšar J, Curk T, Erjavec A, Gorup v, Hočevar T, Milutinović M, Možina M, Polajnar M, Toplak M, Starič A, Štajdohar M, Umek L, Žagar L, Žbontar J, Žitnik M, Zupan B (2013) Orange: Data mining toolbox in Python. Journal of Machine Learning Research 14:2349–2353
9. Deuschl G, Wenzelburger R, Löffler K, Raethjen J, Stolze H (2000) Essential tremor and cerebellar dysfunction clinical and kinematic analysis of intention tremor. Brain 123(8):1568–1580, DOI 10.1093/brain/123.8.1568
10. Elble RJ, Sinha R, Higgins C (1990) Quantification of tremor with a digitizing tablet. J Neurosci Methods 32(3):193–8
11. Engin M (2007) A recording and analysis system for human tremor. Measurement 40(3):288–293, DOI 10.1016/j.measurement.2006.05.015
12. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ (2008) Liblinear: A library for large linear classification. Journal of Machine Learning Research 9:1871–1874
13. Fayyad UM, Irani KB (1993) Multi-interval discretization of continuous-valued attributes for classification learning. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp. 1022–1029
14. Groznik V, Guid M, Sadikov A, Možina M, Georgiev D, Kragelj V, Ribarič S, Pirtošek Z, Bratko I (2013) Elicitation of neurological knowledge with argument-based machine learning. Artificial Intelligence in Medicine 57(2):133–144, DOI 10.1016/j.artmed.2012.08.003
15. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, Dodel R, Ekman M, Faravelli C, Fratiglioni L, Gannon B, Jones D, Jennum P, Jordanova A, Jönsson L, Karampampa K, Knapp M, Kobelt G, Kurth T, Lieb R, Linde M, Ljungcrantz C, Maercker A, Melin B, Moscarelli M, Musayev A, Norwood F, Preisig M, Pugliatti M, Rehm J, Salvador-Carulla L, Schlehofer B, Simon R, Steinhausen H, Stovner L, Vallat J, Van den Bergh P, van Os J, Vos P, Xu W, Wittchen H, Jönsson B, Olesen J (2011) Cost of disorders of the brain in Europe 2010. European Neuropsychopharmacology 21:718–779, DOI 10.1016/j.euroneuro.2011.08.008
16. Khan T, Westin J, Dougherty M (2014a) Cepstral separation difference: A novel approach for speech impairment quantification in Parkinson's disease. Biocybernetics and Biomedical Engineering 34(1):25–34, DOI 10.1016/j. bbe.2013.06.001
17. Khan T, Westin J, Dougherty M (2014b) Classification of speech intelligibility in Parkinson's disease. Biocybernetics and Biomedical Engineering 34(1):35–45, DOI 10.1016/j.bbe.2013.10.003
18. Little M, McSharry P, Hunter E, Spielman J, Ramig L (2009) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Biomedical Engineering, IEEE Transactions on 56(4):1015–1022, DOI 10.1109/TBME.2008.2005954
19. Memedi M, Westin J, Nyholm D (2013) Spiral drawing during self-rated dyskinesia is more impaired than during self-rated Off. Parkinsonism Relat Disord 19(5):553–556, DOI 10.1016/j.parkreldis.2013.01.011
20. Moore ST, MacDougall HG, Gracies JM, Cohen HS, Ondo WG (2007) Longterm monitoring of gait in Parkinson's disease. Gait & Posture 26(2):200–207, DOI 10.1016/j.gaitpost.2006.09.011
21. Možina M, Žabkar J, Bratko I (2007) Argument based machine learning. Artificial Intelligence 171(10/15):922–937
22. Murman D (2012) Early treatment of Parkinson's disease: opportunities for managed care. Am J Manag Care 18(7):183–8
23. Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, Akay M, Dy J, Welsh M, Bonato P (2009) Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. Information Technology in Biomedicine, IEEE Transactions on 13(6):864–873, DOI 10.1109/TITB.2009.2033471
24. Pogorelc B, Bosnić Z, Gams M (2012) Automatic recognition of gait-related health problems in the elderly using machine learning. Multimedia Tools and Applications 58:333–354, DOI 10.1007/s11042-011-0786-1
25. Pullman SL (1998) Spiral analysis: A new technique for measuring tremor with a digitizing tablet. Movement Disorders 13(S3):85–89, DOI 10.1002/mds.870131315
26. Quinn NP, Schneider SA, Schwingenschuh P, Bhatia KP (2011) Tremor — some controversial aspects. Movement Disorders 26(1):18–23, DOI 10.1002/mds.23289

27. Rodriguez-Martin D, Perez-Lopez C, Sama A, Cabestany J, Catala A (2013) A wearable inertial measurement unit for long-term monitoring in the dependency care area. Sensors 13(10):14079–14104, DOI 10.3390/s131014079
28. Salarian A, Russmann H, Wider C, Burkhard P, Vingerhoets FJG, Aminian K (2007) Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system. Biomedical Engineering, IEEE Transactions on 54(2):313–322, DOI 10.1109/TBME.2006.886670
29. Storn R, Price K (1997) Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11(4):341–359, DOI 10.1023/A:1008202821328
30. Thanvi B, Lo N, Robinson T (2006) Essential tremor — the most common movement disorder in older people. Age and Ageing 35(4):344–349, DOI 10.1093/ageing/afj072
31. Tsanas A, Little M, McSharry P, Ramig L (2010) Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. Biomedical Engineering, IEEE Transactions on 57(4):884–893, DOI 10.1109/TBME.2009.2036000
32. Westin J, Dougherty M, Nyholm D, Groth T (2010) A home environment test battery for status assessment in patients with advanced Parkinson's disease. Computer Methods and Programs in Biomedicine 98(1):27–35, DOI 10.1016/j.cmpb.2009.08.001
33. World Health Organization (2008), The global burden of disease
34. Xi X, Keogh E, Shelton C, Wei L, Ratanamahatana CA (2006) Fast time series classification using numerosity reduction. In: Proceedings of the 23th International Conference on Machine Learning, pp 1033–1040
35. Yang MH, Sheu YH, Shih YH, Young M (2003) Portable tremor monitor system for real-time full-wave monitoring and analysis. Review of Scientific Instruments 74(3):1303–1309, DOI 10.1063/1.1536259