Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research xxx (xxxx) xxx

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease

Salama A. Mostafa [a,*], Aida Mustapha [a], Mazin Abed Mohammed [b], Raed Ibraheem Hamed [c], N. Arunkumar [d], Mohd Khanapi Abd Ghani [e], Mustafa Musa Jaber [f], Shihab Hamad Khaleefah [g]

[a] *Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Johor, Malaysia*
[b] *Planning and Follow-Up Department, University Headquarter, University of Anbar, Anbar, Iraq*
[c] *Department of IT, College of Science and Technology, University of Human Development, Sulaymaniyah KRG, Iraq*
[d] *School of EEE, Sastra University, Thanjavur, India*
[e] *Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia*
[f] *Dijlah University College, Baghdad, Iraq*
[g] *Al Maarif University College, 31001 Anbar, Iraq*

Received 16 October 2018; received in revised form 26 November 2018; accepted 9 December 2018

Abstract

An accurate diagnosis of Parkinson's disease by specialists involves many neurological, psychological and physical examinations. The specialists investigate a number of symptoms and signs when examining the nervous system conditions of a person. The diagnosis involves reviewing the medical history and genetic factor of the person. The recent diagnosis methodology to Parkinson's disease relies on voice disorders analysis. This methodology entails extracting feature sets of a recorded person's voice then utilizing a machine learning technique to identify the healthy and Parkinson's cases from the voice. This paper attempts to improve the diagnoses of Parkinson's disease by testing multiple feature evaluation and classification machine learning methods based on the voice disorders analysis. The aim of this paper is to find the optimal solution to the problem by (i) proposing a new Multiple Feature Evaluation Approach (MFEA) of a multi-agent system (ii) implementing five independent classification schemas which are Decision Tree, Naïve Bayes, Neural Network, Random Forests, and Support Vector Machine on the Parkinson's diagnosis before and after applying the MFEA, and (iii) evaluating the diagnosis accuracy of the results. The methodology of the tests encompasses 10-fold cross-validation to evaluate the learning of methods and track variation in their performance. The test results show that the MFEA of the multi-agent system finds the best set of features and improves the performance of the classifiers. The average rate of improvement in the diagnostic accuracy of the classifiers are Decision Tree 10.51%, Naïve Bayes 15.22%, Neural Network 9.19%, Random Forests 12.75%, and Support Vector Machine 9.13%. These results show that the MFEA makes a significant improvement to the classifiers' diagnosis results.

Keywords: Parkinson's disease; Multi-agent system; Feature evaluation; Classification

* Corresponding author.
E-mail addresses: salama@uthm.edu.my (S.A. Mostafa), aidam@uthm.edu.my (A. Mustapha), raed.alfalahy@uhd.edu.iq (R.I. Hamed), khanapi@utem.edu.my (M.K. Abd Ghani).

https://doi.org/10.1016/j.cogsys.2018.12.004
1389-0417/

1. Introduction

Parkinson's disease is considered a common disease among people over 50 years old (Davie, 2008). The central nervous system of the human's body exposes to degenerative disorder at this age (Kaya, Findik, Babaoglu, & Arslan, 2011). This disorder results from the death of dopamine-generating cells in the parts of the midbrain knows as *substantia nigra* (Ramayya, Misra, Baltuch, & Kahana, 2014). Parkinson's disease is still incurable but early diagnosis to this disease helps to mitigate its effects on patients (Kaya et al., 2011). The diagnosis of Parkinson's disease encompasses many neurological, psychological and physical examinations. The inspection of the neurological symptoms for it is a complex procedure that entails advanced observation technologies to the nervous system and brain (Gupta, Sundaram, Khanna, Hassanien, & de Albuquerque, 2018; Little, McSharry, Roberts, Costello, & Moroz, 2007). The inspection of the psychological symptoms includes observation of the depression, dementia, and emotional conditions of the patient. While the inspection of the physical symptoms more focuses on movement difficulties problems such as slowness shaking, and rigidity. Fig. 1 shows the neurological symptoms of the disease in the brain.

Researchers found that up to 90% of Parkinson's disease patients have a disorder of voice which are known as *dysphonia* (Little et al., 2007, 2009). The *dysphonia* is a pathological or functional problem that can be diagnosed according to the constellation of vocal symptoms including vocal impairment symptoms (Das, 2010). Little et al. (2007) propose a speech analysis method to examine the effect of voice disorders on Parkinson's disease patients. This method extracts fractal scaling and recurrence features to discriminate the disordered voice from the normal voice. It outcomes a Parkinson's dataset that contains a wide range of voice disorder classes. Subsequently, Arjmandi and Pooyan (2012) improves the speech analysis method by including pathological voice quality assessment algorithm. The algorithm performs a linear discriminant analysis to identify different voice disorders of the vocal folds. Tatu, Albuquerque, Eisemann, Schneidewind, Theisel, Magnor, and Keim (2009) carry out an automated data analysis and visualization to the Parkinson's dataset of (Little et al., 2007). The aim of this analysis is to find linear and nonlinear correlations and clusters between the pairwise of the dataset. The dataset shows high complexity figure and manifests a robust classification scheme.

Recently, different studies are conducted on the diagnosis of Parkinson's disease based on vocal symptoms. These studies apply various machine learning methods such as a Decision Tree, Naïve Bayes, Neural Network, Support Vector Machine, Rotation Forest and Regression in the diagnosing process. Many of which use the Parkinson's dataset of Little et al. (2007) for testing and evaluation including (Das, 2010; Hariharan, Polat, & Sindhu, 2014; Kaya et al., 2011; Mandal & Sairam, 2013; Mostafa, Mustapha, Khaleefah, Ahmad, & Mohammed, 2018; Ozcift & Gulten, 2011). These methods inspect the patterns of the provided data of a person and predict the corresponding healthy or Parkinson's class of the person (Gnanapriya, Suganya, Devi, & Kumar, 2010). Most recent attempts to improve the diagnosis consider optimizing the feature vector of the dataset as in Kaya et al. (2011) and Ozcift and Gulten (2011), the classification methods as in Can (2013) and Das (2010) or both of them as in (Hariharan et al., 2014; Mandal & Sairam, 2013). The results show that the feature selection process is found to be very important in improving the diagnosis' accuracy. Subsequently, this paper presents another attempt to

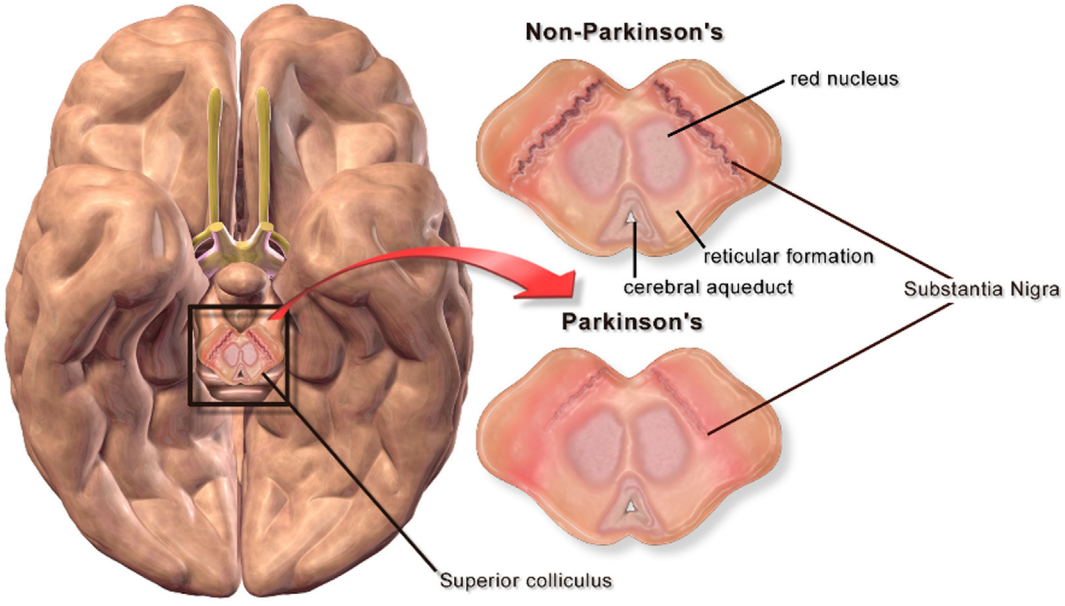

Fig. 1. "Parkinson's Disease" by BruceBlaus. License: CC BY 3.0 (Parkinson's Disease-Symptoms et al., 2018).

Please cite this article as: S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research, https://doi.org/10.1016/j.

improve the diagnosis of Parkinson's disease. The main contributions of this work are proposing a new Multiple Feature Evaluation Approach (MFEA) of a multi-agent system in the Parkinson's diagnosis problem and evaluating its performance by applying five classification methods.

The following section reviews the literature on machine learning based on Parkinson's disease diagnosis. Section 3 presents the research methodology, the used methods and the testing dataset of Parkinson's disease diagnosis. It illustrates the feature selection strategy of a multi-agent system. Section 4 presents the analysis of the testing results. Finally, Section 5 concludes the paper and propose future work.

2. Related work

Machine learning literature introduces a number of data analytics and learning classifiers (Mostafa et al., 2018; Tatu et al., 2009). The classifiers have been utilized to solve complex and critical classification problems including Computer-Aided Diagnosis (CAD) applications (Ozcift & Gulten, 2011). This section presents several works that attempt to apply classification methods for Parkinson's disease diagnosis.

Can (2013) propose a Parallel Distributed Neural Network classifier to improve the diagnostic accuracy rates of Parkinson's disease. The Neural Network has a back-propagation learning algorithm and majority voting scheme technique in its design. The classifier is applied in the Parkinson's dataset. It consistently achieves up to 90.00% accuracy rate.

Das (2010) examines a number of classification methods in the Parkinson's dataset diagnosis accuracy. The methods are a Neural Network, Decision Tree, Regression and DMneural. The Neural Network yields the best classification result among the four classifiers with a score of 92.90%. Das arbitrary test different combinations of features and selects the best set of features among them. The data is randomly divided into training and testing sets and the best division among which is selected.

Ozcift and Gulten (2011) study the effects of adjusting the feature diminution on classification results of Rotation Forest classifier. They use Correlation-based Feature Selection (CFS) to evaluate the Parkinson's dataset features. The CFS reduces the dataset from 23 to 11 features. The study results show that the Rotation Forest accuracy increases by 2.7% when using the 11 features over the 23 features.

Mandal and Sairam (2013) implement a Parkinson's disease diagnosis inference system. The system diagnoses and estimates the severity of Parkinson's disease. It includes Support Vector Machines and ranker search algorithm for feature selection and Neural Networks and Supports Vector Machines for classification. The system is tested via applying the Parkinson's dataset. The test results record high diagnostic accuracy of 97.64%.

Finally, Hariharan et al. (2014) propose hybrid combinations of feature selection and classifications for the diagnosis of Parkinson's disease. The combinations include Sequential Backward Selection (SBS), Principal Component Analysis (PCA), linear discriminant analysis (LDA) and Sequential Forward Selection (SFS) methods for feature selection and Support Vector Machine Neural Networks for classifications. The system is tested in the Parkinson's dataset and the diagnostic accuracy of the test reaches up to 100% accuracy.

In conclusion, some of Parkinson's disease diagnosis literature based on the voice disorders analysis consider applying classification methods to the original Parkinson's dataset. The obtained classification results are considerably high. Others, went farther and apply feature evaluation methods prior to the classification process and reduce the features of the dataset. The results show that features reduction in general increases the accuracy of the classifiers. However, applying combinations of feature evaluation methods might over-reduce or overweight the features. This issue may negatively affect the classification generalization and prone the classifiers to overfitting (i.e., the curse of features dimensionality problem) (Can, 2013; Ozcift & Gulten, 2011).

3. Research framework

This work intends to find the best combination of features that improve the Parkinson's classification accuracy results and maintain balanced feature selection. It includes applying multiple feature evaluation and classification methods for improved diagnoses of Parkinson's disease. In this work, the key conducted activities represented by data preparation, feature evaluation, methods setting, classification, and classifiers evaluation. The work encompasses processing the original Parkinson's dataset and outcomes a new filtered dataset. The sequence of the research activities and outcomes are shown in Fig. 2.

The Parkinson's dataset consists of a number of features that describe healthy and Parkinson's persons. The Parkinson's dataset preparation activities include performing a number of preprocessing steps to check the dataset quality and make it ready for the process. The feature evaluation includes applying the MFEA that contain a number of feature evaluation and ranking algorithms to weight the worth of the features and select the best set of features. The evaluation outcomes are the Parkinson's original dataset that its classes contain 23 features and the filtered dataset that its classes contain 11 features. The methods for setting activities include implementing five well-known classification methods which are Decision Tree, Naïve Bayes, Neural Network, Random Forests, and Support Vector Machine (SVM) respectively. The classifiers are evaluated by applying the 10-fold cross-validation technique to the dataset before and after feature evaluation. The main activities and methods of this work are detailed in the following sections.

Please cite this article as: S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research, https://doi.org/10.1016/j.

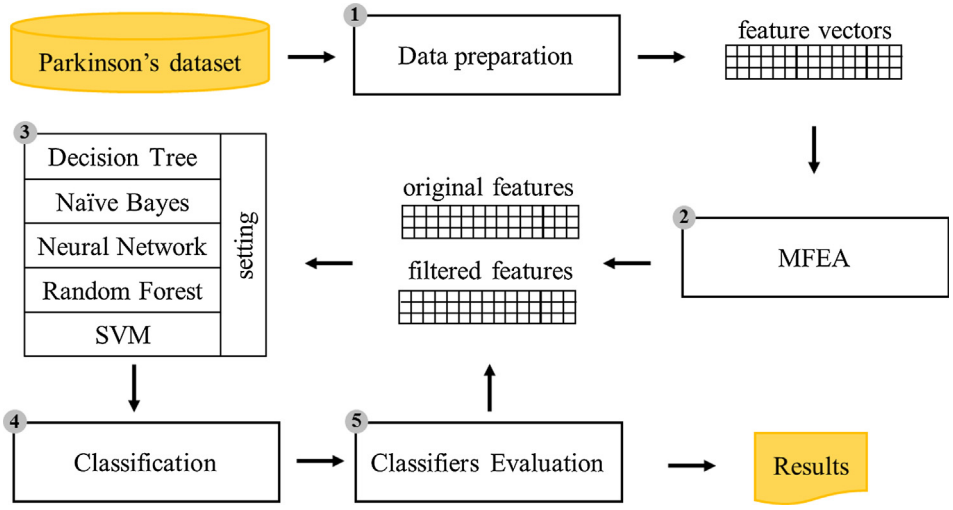

Fig. 2. The flow diagram of the research activities.

3.1. Data preparation

Little et al. (2007) use several methods in the process of extracting the Parkinson's dataset. After getting the physical signals, they analyse and filter the signals to 5923 signals (Ramayya et al., 2014). The novelty of their work is they update the density functions to force the data to become further normally distributed in order to manifest stable UPDRS features for the regression scheme. They applied several linear signals processing methods such as short-time autocorrelation and nonlinear methods such as classification and regression tree (CART) in a vowel phonation (*dysphonia*) measures to obtain the characteristics of the signals. The obtained characteristics of the features extraction process form the Parkinson's dataset. Non-parametric statistical tests are used in the comparison between the measures and the UPDRS including Spearman correlation coefficient and Gaussian kernels to ensure accurate assessment. Subsequently, they map the extracted features by regression methods in order to obtain the final Unified Parkinson's Disease Rating Scale (UPDRS). The regression analysis uses a statistical process to check the dependency of the features in order to estimate the relationships among the features. In the feature selections process, to construct the class model, they applied a Bayesian information criterion and Akaike information criterion (Can, 2013). These two methods can handle a different number of parameters (features) with different sequence and measure the best fit of the model. They reject and pass the input features that are not related to the target output to the following nodes of the test (Parkinson's Disease-Symptoms, Stages and Life Expectancy, Pathology, Lecturio, 2018). Meanwhile, in the work of (Tsanas, Little, McSharry, & Ramig, 2010), they utilize the Least Absolute Shrinkage and Selection Operator (LASSO) as a feature selection method. However, both techniques shared the same results which did not affect the features number and dimensions (Tsanas et al., 2010). The process outcomes Parkinson's dataset that has 23 features of numerical real numbers. The first two features are the patient name and recording number which are excluded in the classification process. Table 1 shows the Parkinson's dataset key features and their description.

Finally, they used the cross-validation statistical method for model generalization. The model generalization performed by applying unused data in the model's parameters setting. The dataset is divided into two training and testing subsets to calculate the Mean Absolute Error (MAE) of the classification result. This division to the data decreases the computation time of the tests (Parkinson's Disease-Symptoms et al., 2018). The overall results of the tests show acceptable prediction accuracy.

Our close examination to the Parkinson's dataset reveals that (1) all the features are numerical of real numbers (2) it is complete and there are no noisy or non-numeric data, and (3) it does not have observable patterns. However, the Parkinson's dataset is imbalanced in which 75% are Parkinson's classes and 25% are healthy classes. This issue increases the complexity of the classification process and decreases the robustness of the classification results (Can, 2013; Kaya et al., 2011).

3.2. Feature evaluation

The Multiple Feature Evaluation Approach (MFEA) operates by a multi-agent system. The Multi-agent system is described as a set of intelligent agents that have an interaction with each other within an environment (Doan & Horiguchi, 2004; Mostafa, Mustapha, Hazeem, Khaleefah, & Mohammed, 2018). These agents function jointly to solve problems which they cannot be solved independently (Ganapathy et al., 2013; Mostafa et al., 2015, 2018). The agents' application in a system helps to increase the flexibility of the system by segregating its functionality and enabling interaction with its modules during runtime (Farahnakian & Mozayani, 2009; Mostafa, Ahmad,

Please cite this article as: S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research, https://doi.org/10.1016/j.

Table 1
The features of Parkinson's dataset (Hariharan et al., 2014; Little et al., 2007).

Features	Description
MDVP: Fo(Hz), Fhi(Hz), and Flo(Hz)	Average, maximum and minimum vocal fundamental frequencies
MDVP: Shimmer, Shimmer (dB), APQ3, APQ5, APQ and DDA	Measures of variation in amplitude
MDVP: Jitter (%), Jitter (Abs), RAP, PPQ and DDP	Measures of variation in fundamental frequency
NHR and HNR	Measures of the ratio of noise to tonal components in the voice
Spread 1, Spread2, and PPE	Non-linear measures of fundamental frequency variation
RPDE and D2	Non-linear dynamical complexity measures
DFA	The signal fractal scaling exponent

Table 2
The feature evaluation of the Parkinson's dataset.

i	x_i	Feature evaluation				
		α_1	α_2	α_3o	α_4	α_5
0	Name-ASCII	–	–	–	–	–
1	MDVP:Fo (Hz)	13	1	13	3	–
2	MDVP:Fhi (Hz)	11	2	7	10	–
3	MDVP:Flo (Hz)	–	3	1	6	14
4	MDVP:Jitter (%)	–	–	–	–	7
5	MDVP:Jitter (Abs)	1	–	14	4	4
6	MDVP:RAP	–	–	6	12	5
7	MDVP:PPQ	14	–	–	–	1
8	Jitter:DDP	2	4	8	9	6
9	MDVP:Shimmer	–	–	10	7	10
10	MDVP:Shimmer (dB)	–	–	12	13	8
11	Shimmer:APQ3	–	–	–	–	3
12	Shimmer:APQ5	–	14	11	8	9
13	MDVP:APQ11	3	5	3	5	12
14	Shimmer:DDA	4	11	–	14	2
15	NHR	5	6	5	–	11
16	HNR	6	–	–	–	–
17	status	–	–	–	–	–
18	RPDE	7	13	–	–	–
19	D2	9	9	–	–	–
20	DFA	8	12	–	–	13
21	spread1	12	7	2	2	–
22	spread2	–	8	9	11	–
23	PPE	10	10	4	1	–

Mustapha, & Mohammed, 2017; Oliinyk, Oliinyk, & Subbotin, 2012).

There are several agent-based feature selection approaches in the literature and some of which are reviewed in Ganapathy et al. (2013). As examples, Doan and Horiguchi (2004) propose a feature selection approach to text mining. The agents are presented as categories of features and each agent represent a subset of features. The feature selection is dynamically performed according to the agents' observations and a pre-determined threshold. Farahnakian and Mozayani (2009) apply a C4.5 features selection algorithm of a decision tree in a multi-agent soccer simulation. The simulation uses a data mining technique to analyze the scoring behaviour. The test results show that the reduction of the features dimension improves the learning accuracy and the execution speed.

The MFEA encompasses five agents in which each agent operates a particular feature evaluation methods. The agents evaluate and rank the features and produce subsets of feature vectors via implementing search algorithm. The process includes weighting each feature and ordering subsets of the feature according to their individual evaluation results. The agents based on the subsets of features perform feature filtering process and produce a number of preliminary copies of feature vectors. Then they collaborate with each other to generate an optimized feature vector. Fig. 3 presents an overview of the MFEA.

The five agent-based feature evaluation operators are described as follows:

- The first agent, α_1, operates an Autocorrelation feature evaluator, ∂_1. The ∂_1 evaluates the worth of a subset of features by determining its pitch period. A feature is considered effective if its correlation coefficient scores less than 0.95, otherwise, it is omitted.
- The second agent, α_2, operates a CFS feature evaluator, ∂_2. The ∂_2 evaluates the worth of a subset of features by measuring the predictive ability of each feature and the redundancy degree between the all features.
- The third agent, α_3, operates a Gain Ratio feature evaluator, ∂_3. The ∂_3 evaluates the worth of a subset of features by measuring the gain ratio with respect to the corresponding class.
- The fourth agent, α_4, operates an Info Gain feature evaluator, ∂_4. The ∂_4 evaluates the worth of a subset of features by measuring the information gained with respect to the corresponding class.
- The fifth agent, α_5, operates a SVM evaluator, ∂_5. The ∂_4 evaluates the worth of a subset of features by using its classifier.

The agents collaborate to measure the frequency of the features appearance and the mean score value of the features rank. The selected features are the features that have a higher frequency of appearance and lower mean rank values. The agent then combines subsets of the selected features to form the filtered feature vector. This approach ensures selecting the best features that have higher ranks and at the same time prevents over-reducing or overweighting the features. It omits or eliminates the nominal features and produce a feature vector that fit different types of classifiers. All the operators might differently rank the features and cooperatively omit the weak features. Hence, the MFEA is meant to improve the classification process and at the same time avoid the global minimum learning of

Please cite this article as: S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research, https://doi.org/10.1016/j.

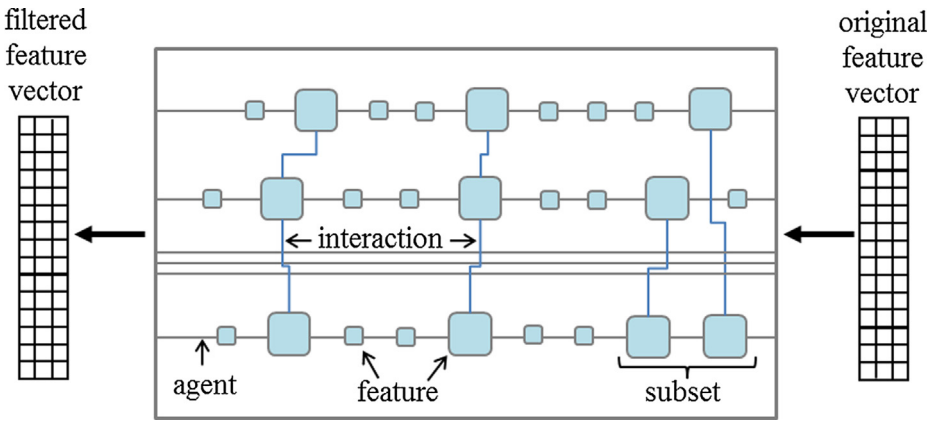

Fig. 3. An overview to the MFEA.

the classifiers that results from overfitting some of the classified features.

Let an initial dataset D has a length of m instances of a feature vector, V. The $V = \{X, Y\}$ in which it has a set of input parameters, $X = \{x_1, x_2, \cdots, x_n\}$ and output parameters $Y = \{y_1, y_2, \cdots, y_m\}$. Ultimately, the D has the following format:

$$D = \begin{bmatrix} x_1 & x_2 & .. & x_n \\ x_1 & x_2 & .. & x_n \\ x_1 & x_2 & .. & x_n \\ x_1 & x_2 & .. & x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ : \\ y_m \end{bmatrix}$$

Each of the agents generates its own evaluation of the features, $\alpha_1 : X \rightarrow x_1 \text{»} \cdots \text{»} x_n$ in which the » denotes the permutation process of X. The collaborative mechanism of the multi-agent system comprises a *filter* function that receives the X and determines the included feature, x^{in}, from the excluded feature, x^{ex} of the X.

$$filter(X) = \begin{cases} x^{in}_{k1}, & f_i \geq \frac{1}{n}\sum_{i=1}^{n} f_i \wedge rank\left(x_i, \frac{1}{f_i}\sum_{j=1}^{f_i} r_{i,j}\right) \leq t \\ x^{ex}_{k2}, & otherwise \end{cases} \quad (1)$$

where x_i is a feature, n is the total number of features of the X and $i = \{1, 2, \cdots n\}$ in which $x^{in} + x^{ex} = X$; $k1$and $k2$ are the indexes for the x^{in} and x^{ex} respectively; f_i is the appearance frequency of x_i and $r_{i,j}$ is the rank of x_i by a ∂_j in which the j is the index of the corresponding feature evaluator agent; *rank* is a function that returns a feature's referenced rank as an integer value and t is the threshold of the required number of the x^{in}.

The t determines the final number of selected features according to their evaluation results. The t can be manually set by a human or autonomously calculated by the agents. The agents follows (3) to determine a particular feature vector length of a D:

$$t = \frac{\frac{1}{k0}\sum_{k0=1}^{d} s_k * n}{m}, t < n \quad (2)$$

where m is the size of a D, n is the number of the original features of the D, d is the training diminutions for a test attempt, $k0$ is an index of the attempt and s is the size of the training dataset for each of the attempts.

The following algorithm shows the MFEA of a multi-agent system in which the $\widetilde{X}$ represents the filtered features.

Algorithm 1: The MFEA

```
begin;
initial parameters;
for each active a_j do
    X ← prepare(a_j, D);
    a_j do i until n
        f_{i,j} ← evaluate(a_j, x_i);
        r_{i,j} ← rank(a_j, x_i);
        x_{i,j} ← search(a_j, x_i);
    end-do
    X̃ ← select(a_j, x_{1,2,...,j});
end-for
X̃ ← collaborate(
        t ← update(a_{1,2,...}, t),
        {x^in, x^ex} ← filter(a_{1,2,...} X̃_{1,2,...})
        );
end;
```

3.3. Classification methods

This work use five independent classification schemas which are Decision Tree, Naïve Bayes, Neural Network, Random Forests, and Support Vector Machine on the Parkinson's diagnosis. The basic information, the types of the classifiers and the setting properties for each of the classifiers are as follows:

- *Decision tree:* It is a well-known and simple decision analysis method. This method has a tree structure of root, leaves and branches (Das, 2010; Gupta et al.,

2018; Pereira et al., 2018). The outer leaf nodes denote the classes, the inner leaf nodes denote the processes and the branches denote the outcomes. The root connects all the tree's components in which the paths from the root to the outer leaf nodes represent the classification rules of the tree (Mostafa et al., 2018). The utilized Decision Tree in this work has an Iterative Dichotomiser 3 (ID3) algorithm that constructs the tree in a top-down manner.

- *Naïve Bayes:* it is a probabilistic classifier that has a supervised machine learning approach. The Naïve Bayes applies a conditional probability of Bayes probabilistic theorem to determine the classification outcomes (Can, 2013). Conventionally, the Naïve Bayes represents a problem instance by a feature vector of$probability\ (Class_i|feature_1, feature_2, \cdots)$ (Mostafa et al., 2018). The Naïve Bayes differs from other methods by the naive that evaluates each feature independently from other features (i.e., independent relationships between features) (Mohammed et al., 2018; Pereira et al., 2018). The utilized Naïve Bayes setting include an activation function to approximate its outputs.
- *Neural network:* It is a numerical learning method that simulates the behaviour of the biological neural networks of humans (Can, 2013). It consists of a number of artificial neurons that known as network nodes. The nodes are linked by weighted interconnections and perform linear or nonlinear computational operations. Neural Networks are typically applied to deal with highly complex problems that are difficult for traditional methods. There are many types of Neural Networks that cover many research fields including approximation, blind source separation compression, classification, clustering and filtering (Das, 2010; Ganapathy et al., 2013; Mohammed et al., 2018). The adopted Neural Network in this work has a multi-layer perceptron architecture with one hidden layer and sigmoid neurons.
- *Random forests:* It is a simple and flexible supervised machine learning method. This method encompasses a number of decision trees that are randomly restricted. The Random forests method attempts to solve the training overfitting of the decision tree classifier and obtain accurate and stable results. It finds at the training phase the most fitted class among the generated classes of the decision trees (Mostafa, Mustapha, Mohammed, Ahmad, & Mahmoud, 2018). The Random forests trains by a "bagging" training technique in which it uses the out-of-bag error to estimate classification errors and permutation technique to measure the importance of features.
- *Support vector machine:* The SVM is a supervised machine learning method that plots the training cases as points in space to generalize a separable gap and identify the classification hyperplane (Abd Ghani, Mohammed, Arunkumar, Mostafa, Ibrahim, Abdullah, Jaber, Abdulhay, Ramirez-Gonzalez, & Burhanuddin, 2018; Little, McSharry, Hunter, Spielman, & Ramig, 2009). Then it maps the new cases into that same space to classify them according to their location in that gap. The SVM uses the kernel trick to handle the non-linear cases by projecting these cases to high-dimensional feature spaces. This projection simplifies the generalization of the separable gap for such non-linear cases (Mandal & Sairam, 2013). The SVM in this work has a Radial Basis Function (RBF) kernel and sequential minimal optimization (SMO) training algorithm.

4. Results and evaluation

The first results that are obtained related to the multiple feature evaluation of the Parkinson's dataset. The MFEA calculates the average frequency of appearance of all features and the threshold t for the required number of the x^{in}. This threshold preliminarily includes a maximum of 14 manually set features and excludes 8 features. Table 2 summarizes the features' worth ranks/scores of the five agents according to the MFEA.

As Table 3 shows, the evaluators produce different feature selection and ranking results. The average integer $f$of the features is 3 which represents the threshold for the collaborative selection condition of the multi-agent system. Table 3 shows the final selected 11 features of the Parkinson dataset after considering the conjunction of the frequency of appearance and the mean rank scores by the MFEA.

The feature selection phase outcomes the filtered dataset that its feature vector contains 11 features from the original Parkinson's 23 features. Initially, ZeroR classifier is applied in the Parkinson's original and filtered datasets to find an initial accuracy result for the two datasets. The accuracy and error rates of the ZeroR results are almost similar in both datasets with slide improvement of 1.32% accuracy to the filtered dataset. The initial accuracy rate that is calculated by the ZeroR classifier is 75.01% for the original dataset and 76.32 for the filtered dataset.

Subsequently, Decision Tree, Naïve Bayes, Neural Network, Random Forests, and Support Vector Machine

Table 3
The selected best 11 features of the Parkinson's dataset.

i	x_i	f_i	$\frac{1}{f_i}\sum_{j=1}^{f_i} r_{i,j}$	$rank()$
1	MDVP:Fo(Hz)	4	7.50	8
2	MDVP:Fhi(Hz)	4	7.50	9
3	MDVP:Flo(Hz)	4	6.00	5
5	MDVP:Jitter(Abs)	4	5.75	2
6	MDVP:RAP	3	7.67	10
8	Jitter:DDP	5	5.80	4
13	MDVP:APQ11	5	5.60	1
14	Shimmer:DDA	4	7.75	11
15	NHR	4	6.75	7
21	spread1	4	5.75	3
23	PPE	4	6.25	6

Please cite this article as: S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research, https://doi.org/10.1016/j.

Table 4
The results of the original dataset.

Classifier	Classification results						
	TP	FP	Precision	Recall	ROC	MAE	ACC (%)
Decision Tree	0.864	0.168	0.871	0.864	0.848	0.1356	86.440
Naïve Bayes	0.741	0.168	0.833	0.741	0.877	0.2571	74.111
Neural Network	0.878	0.207	0.882	0.878	0.862	0.1311	87.755
Random Forests	0.867	0.209	0.875	0.867	0.836	0.1694	86.734
SVM	0.863	0.383	0.866	0.863	0.740	0.1371	86.294

are respectively applied in the Parkinson's disease diagnoses to set a benchmark of the MFEA. The tests follow a 10-fold cross-validation to evaluate the classification learning for the methods. The number of the performed tests in total is 100 (i.e., 5 methods * 10-fold cross-validation * 2 datasets). Table 4 shows the overall results of the original dataset including the True Positive (TP), False Positive (FP), Receiver Operating Characteristic (ROC), Mean Absolute Error (MAE) and the accuracy (ACC) results.

The original Parkinson's dataset test results show that the Neural Network yields the highest diagnostic accuracy results of 87.75% when the MAE is the lowest, 0.1311. The Naïve Bayes yields the lowest accuracy results of 74.11% when the MAE is the highest, 0.2571. The Decision Tree, Random Forests and SVM yield an intermediate accuracy of 86.50% on average.

On the other hand, the filtered Parkinson's dataset test results show that the Random Forests yields the highest diagnostic accuracy results of 99.49% when the MAE is the lowest, 0.0284. The Naïve Bayes also yields the lowest accuracy results of 89.34% when the MAE is the highest, 0.1190. The Decision Tree and Neural Network yield intermediate accuracy results of 96.95% on average while the SVM yields a slightly lower accuracy result of 95.43%. Table 5 shows the overall results of the filtered dataset.

Generally, the diagnosis results of the five classifiers are noticeably improved when implementing the filtered dataset. The highest improvement in the accuracy of the results is recorded by the Naïve Bayes which is 15.22%. The lowest improvement in the accuracy of the results is recorded by the SVM which is 9.13%. The average improvement in the diagnosis accuracy results is up to 11.36%. Fig. 4 illustrates the variation between the diagnosis accuracy results of the classifiers in the two datasets.

Principally, an accurate classification entails investigating (i) the best prediction or diagnosis method to the problem or disease (ii) the best feature evaluation and selection algorithm to the extracted features of the problem and (iii) the best setting to the method's parameters according to the selected features. Subsequently, different classification methods have been applied in the Parkinson's dataset in which a relevant classifier identifies healthy and Parkinson's persons. The related work proves that the features reduction frequently enhances the feature vector of the dataset, reduces the complexity of the classifiers and improves the diagnostic accuracy results. For example, the classifiers that accurately deal with complex decision boundaries such as neural networks are more sensitive to feature selection and likely prone to overfitting.

However, there is no fixed rule for feature evaluation and selection in the classification problems. This process depends on the properties of the training data, the complexity of the decision boundaries, and the type of classifiers. This work contributes a new Multiple Feature Evaluation Approach (MFEA) of a multi-agent system. The MFEA implementation in the Parkinson's disease diagnosis shows the importance of feature selection on the diagnosis accuracies for this classification scheme. The MFEA has the merits of adopting a generalized feature selection through applying combinations of collaborative feature evaluation methods. Additionally, the MFEA does not over-reduce and strictly overweight features to avoid overfitting of classifiers. Hence, this work anticipates finding the best combination of features to maintain balanced feature selection and improve the Parkinson's classification accuracy results.

The limitation of this work is that it excludes the runtime from the evaluation process. In addition, the

Table 5
The results of the filtered dataset.

Classifiers	Classification results						
	TP	FP	Precision	Recall	ROC	MAE	ACC (%)
Decision Tree	0.970	0.080	0.970	0.970	0.996	0.0435	96.954
Naïve Bayes	0.893	0.147	0.898	0.893	0.950	0.1190	89.340
Neural Network	0.970	0.024	0.971	0.970	0.986	0.0441	96.950
Random Forests	0.995	0.002	0.995	0.995	1.000	0.0284	99.492
SVM	0.954	0.043	0.957	0.954	0.982	0.0708	95.431

Please cite this article as: S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research, https://doi.org/10.1016/j.

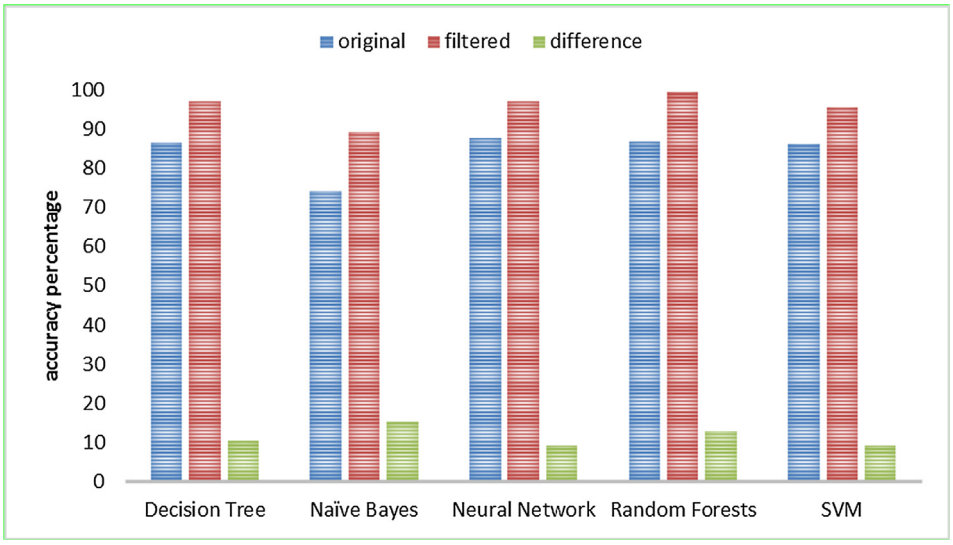

Fig. 4. The variation between the results of the original and filtered datasets.

redundancy of the filtered out features might help to reduce the effect of the implanted data. Naïve Bayes classification results can be a good indication criterion in this dilemma as it processes features of independent relationships. Nevertheless, the Naïve Bayes classification results are improved steadily and there is no observable fluctuation.

5. Conclusion

Parkinson's is a disease that affects elderly people and disturb their life. It is difficult to diagnose as its symptoms are unclear and might be associated with other diseases. Many recent research studies have been conducted for Parkinson's disease diagnosis. Different medical and technical methods have been applied to improve diagnostic accuracy. Recently, studies apply machine learning classifiers on voice disorder detection features to diagnose Parkinson's disease. However, there is no universal classifier that can guarantee high diagnostic accuracy for all the analytical data of the studies. Hence, obtaining optimum solutions require evaluating various ensemble-learning approaches. This paper investigates multiple feature evaluation and classification methods for improving the diagnoses of Parkinson's disease. It contributes deploying a multi-agent system for Multiple Feature Evaluation Approach (MFEA). The MFEA is tested in five independent classification schemas which are Decision Tree, Naïve Bayes, Neural Network, Random Forests, and Support Vector Machine. The total number of the performed tests for both the original and the filtered datasets is 100 tests. Random Forests yields the highest diagnostic accuracy results of 99.49% among the other classifiers. In general, the diagnosis accuracy results of the five classifiers are noticeably improved when utilizing the filtered dataset of the MFEA. The highest improvement in the accuracy of the results is recorded by the Naïve Bayes which is 15.22%. The lowest improvement in the accuracy of the results is recorded by the SVM which is 9.13%. The average improvement in the diagnosis accuracy results for all of the tested classifiers is up to 11.36%. The future work considers linking the dataset properties with the feature evaluation in the MFEA. Subsequently, testing the outcome with datasets that have various properties.

Acknowledgment

This project is sponsored by postdoctoral grant of Universiti Tun Hussein Onn Malaysia (UTHM) under Vot D004.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogsys.2018.12.004.

References

Abd Ghani, M. K, Mohammed, M. A., Arunkumar, N. Mostafa, S. A., Ibrahim, D. A., Abdullah, M. K., Jaber, M. M., Abdulhay, Enas, Ramirez-Gonzalez, Gustavo, & Burhanuddin, M.A. (2018) Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. Neural Computing and Applications. doi: 10.1007/s00521-018-3882-6.

Arjmandi, M. K., & Pooyan, M. (2012). An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomedical Signal Processing and Control, 7*(1), 3–19.

Can, M. (2013). Neural networks to diagnose the Parkinson's disease. *South East Europe Journal of Soft Computing, 2*(1).

Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications, 37*(2), 1568–1572.

Davie, C. A. (2008). A review of Parkinson's disease. *British Medical Bulletin, 86*(1), 109–127.

Doan, S., & Horiguchi, S. (2004). An agent-based approach to feature selection in text categorization. In Proceedings of 2nd International Conference on Autonomous Robot and Agent (pp. 362–366).

Farahnakian, F., & Mozayani, N. (2009). Evaluating feature selection techniques in simulated soccer multi agents system. In Advanced Computer Control, 2009. ICACC'09. International Conference on (pp. 107–110). IEEE.

Ganapathy, S., Kulothungan, K., Muthurajkumar, S., Vijayalakshmi, M., Yogesh, P., & Kannan, A. (2013). Intelligent feature selection and classification techniques for intrusion detection in networks: A survey. *EURASIP Journal on Wireless Communications and Networking, 2013* (1), 271.

Gnanapriya, S., Suganya, R., Devi, G. S., & Kumar, M. S. (2010). Data mining concepts and techniques. *Data Mining and Knowledge Engineering, 2*(9), 256–263.

Gupta, D., Julka, A., Jain, S., Aggarwal, T., Khanna, A., Arunkumar, N., & de Albuquerque, V. H. C. (2018). Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. *Cognitive Systems Research., 52*, 36–48.

Gupta, D., Sundaram, S., Khanna, A., Hassanien, A. E., & de Albuquerque, V. H. C. (2018). Improved diagnosis of Parkinson's disease using optimized crow search algorithm. *Computers & Electrical Engineering, 68*, 412–424.

Hariharan, M., Polat, K., & Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine, 113*(3), 904–913.

Kaya, E., Findik, O., Babaoglu, I., & Arslan, A. (2011). Effect of discretization method on the diagnosis of Parkinson's disease. *International Journal of Innovative Computing, Information and Control, 7*, 4669–4678.

Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering, 56* (4), 1015–1022.

Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine, 6* (1), 23.

Mandal, I., & Sairam, N. (2013). Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system. *International Journal of Medical Informatics, 82*(5), 359–377.

Mohammed, M. A., Al-Khateeb, B., Rashid, A. N., Ibrahim, D. A., Ghani, M. K. A., & Mostafa, S. A. (2018). Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers & Electrical Engineering., 70*, 871–882.

Mohammed, M. A., Ghani, M. K. A., Arunkumar, N., Mostafa, S. A., Abdullah, M. K., & Burhanuddin, M. A. (2018). Trainable model for segmenting and identifying Nasopharyngeal carcinoma. *Computers & Electrical Engineering, 71*, 372–387.

Mostafa, S. A., Ahmad, M. S., Mustapha, A., & Mohammed, M. A. (2017). Formulating layered adjustable autonomy for unmanned aerial vehicles. *International Journal of Intelligent Computing and Cybernetics, 10*(4), 430–450.

Mostafa, S. A., Ahmad, M. S., Annamalai, M., Ahmad, A., & Gunasekaran, S. S. (2015). Formulating dynamic agents' operational state via situation awareness assessment. In Advances in Intelligent Informatics (pp. 545–556). Springer, Cham.

Mostafa, S. A., Mustapha, A., Khaleefah, S. H., Ahmad, M. S., & Mohammed, M. A. (2018). Evaluating the performance of three classification methods in diagnosis of Parkinson's disease. In International Conference on Soft Computing and Data Mining (pp. 43–52). Springer, Cham.

Mostafa, S. A., Mustapha, A., Hazeem, A. A., Khaleefah, S. H., & Mohammed, M. A. (2018). An agent-based inference engine for efficient and reliable automated car failure diagnosis assistance. *IEEE Access, 6*, 8322–8331.

Mostafa, S. A., Mustapha, A., Mohammed, M. A., Ahmad, M. S., & Mahmoud, M. A. (2018). A fuzzy logic control in adjustable autonomy of a multi-agent system for an automated elderly movement monitoring application. *International Journal of Medical Informatics, 112*, 173–184.

Oliinyk, A. O., Oliinyk, O. O., & Subbotin, S. A. (2012). Agent technologies for feature selection. *Cybernetics and Systems Analysis, 48*(2), 257–267.

Ozcift, A., & Gulten, A. (2011). Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Computer Methods and Programs in Biomedicine, 104*(3), 443–451.

Parkinson's Disease-Symptoms, Stages and Life Expectancy, Pathology, Lecturio (2018) <https://www.lecturio.com/magazine/parkinsons-disease/>.

Pereira, C. R., Pereira, D. R., Rosa, G. H., Albuquerque, V. H., Weber, S. A., Hook, C., & Papa, J. P. (2018). Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification. *Artificial Intelligence in Medicine, 87*, 67–77.

Pereira, C. R., Pereira, D. R., Weber, S. A., Hook, C., de Albuquerque, V. H. C., & Papa, J. P. (2018). A survey on computer-assisted Parkinson's Disease diagnosis. *Artificial Intelligence in Medicine*. https://doi.org/10.1016/j.artmed.2018.08.007.

Ramayya, A. G., Misra, A., Baltuch, G. H., & Kahana, M. J. (2014). Microstimulation of the human substantia nigra alters reinforcement learning. *Journal of Neuroscience, 34*(20), 6887–6895.

Tatu, A., Albuquerque, G., Eisemann, M., Schneidewind, J., Theisel, H., Magnor, M., & Keim, D. (2009). Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on (pp. 59–66). IEEE.

Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering, 57* (4), 884–893.

Please cite this article as: S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research, https://doi.org/10.1016/j.