Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bbe

Original Research Article

A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques

Q1 *Mehrbakhsh Nilashi* *, *Othman Ibrahim, Hossein Ahmadi, Leila Shahmoradi* *, *Mohammadreza Farahmand*

Faculty of Computing, Universiti Technologi Malaysia, Johor, Malaysia

ARTICLE INFO

Article history:
Received 7 June 2017
Received in revised form
5 August 2017
Accepted 18 September 2017
Available online xxx

Keywords:
Healthcare
Parkinson Disease diagnosis
UPDRS
Clustering
Dimensionality reduction
ISVR

ABSTRACT

Parkinson's Disease (PD) is a progressive degenerative disease of the nervous system that affects movement control. Unified Parkinson's Disease Rating Scale (UPDRS) is the baseline assessment for PD. UPDRS is the most widely used standardized scale to assess parkinsonism. Discovering the relationship between speech signal properties and UPDRS scores is an important task in PD diagnosis. Supervised machine learning techniques have been extensively used in predicting PD through a set of datasets. However, the most methods developed by supervised methods do not support the incremental updates of data. In addition, the standard supervised techniques cannot be used in an incremental situation for disease prediction and therefore they require to recompute all the training data to build the prediction models. In this paper, we take the advantages of an incremental machine learning technique, Incremental support vector machine, to develop a new method for UPDRS prediction. We use Incremental support vector machine to predict Total-UPDRS and Motor-UPDRS. We also use Non-linear iterative partial least squares for data dimensionality reduction and self-organizing map for clustering task. To evaluate the method, we conduct several experiments with a PD dataset and present the results in comparison with the methods developed in the previous research. The prediction accuracies of method measured by MAE for the Total-UPDRS and Motor-UPDRS were obtained respectively MAE = 0.4656 and MAE = 0.4967. The results of experimental analysis demonstrated that the proposed method is effective in predicting UPDRS. The method has potential to be implemented as an intelligent system for PD prediction in healthcare.

Q2 * *Corresponding authors at*: Faculty of Computing, Universiti Technologi Malaysia, Johor, Malaysia.
E-mail addresses: nilashidotnet@hotmail.com (M. Nilashi), leilashahmoradi@hotmail.com (L. Shahmoradi).

http://dx.doi.org/10.1016/j.bbe.2017.09.002

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

1. Introduction

Parkinson's Disease (PD) [1–3] is a progressive degenerative disease of the nervous system that affects movement control. This disease affects approximately 1% of the population over 60 years [4] with a prevalence of approximately 250 per 100,000 persons. The average age at onset lies between 55 and 65 years. The evidence shows that some environmental factors (e.g., consumption of well water, rural living and pesticide exposure) can increase an individual's risk of developing PD [5]. The PD burden is considerable and its diagnosis is clinical in nature. This disease includes four cardinal features which are: resting tremor (usually in a frequency of 4–7 Hz), rigidity, bradykinesia (slowness of movement), and impaired postural instability [6]. Three hierarchical levels have been defined by the World Health Organization (WHO) which a disease can be assessed. 'Impairment' reflects the basic level and in PD consists of clinical signs, such as tremor, bradykinesia and rigidity. 'Disability' depends on overlapping PD features that cause a functional decrease of activities of daily life. A multitude of assessment scales has been developed for the assessment of PD. The need for scales arose after the introduction of an effective therapy, so many were developed in the sixties and seventies. In 1987 the Unified Parkinson's disease rating scale (UPDRS) [7] was introduced, and in the subsequent years this scale has evolved into the most widely used scale for the assessment of PD. The UPDRS is a useful way to maintain an ongoing record of patient function and to assess disability [8]. According to Rascol et al. [9], UPDRS is the most widely used standardized scale to assess parkinsonism. UPDRS is made up of the (1) Motor sections, (2) ADL, (3) Mentation, Behavior and Mood. Motor-UPDRS and Total-UPDRS of UPDRS which refer respectively to the full range of the metric and the motor section of the UPDRS are used in the diagnosis of PD. The structure and metric properties of the UPDRS has been investigated in several studies on PD [10,11]. An example of a UPDRS chart measuring the change from baseline (mean ± SE) is shown in Fig. 1.

PD is a very complex disorder in which individual motor features vary in their presence and severity over time. The early diagnosis of PD is important. The problem of an early diagnosis of PD has raised an interest of numerous researchers. In addition, the researchers in the previous studies have emphasized that the main challenge in the diagnosis of PD is the correctly recognition of PD affected subjects at the early stage [12]. The early diagnosis of PD can affect on the progression of the disease and the quality of life the patients [13]. It has been emphasized that the main medical challenge is to correctly recognize the Parkinson's disease affected subjects at the early stage [12,14,15].

As early as 1997, the potential of data mining for improving the problems in the medical domain had been identified by World Health Organization (WHO) [16]. The usefulness of knowledge detection from medical data repositories has been emphasized by WHO as it benefits medical diagnosis and prediction.

Data mining is a process of discovering useful knowledge from database to build a structure (i.e., model or pattern) that can meaningfully interpret the data. It has been defined is as a process of discovering interesting patterns and knowledge from large amount of data [17]. It uses the machine learning techniques to discover hidden pattern in the data. These techniques can be in the three main categories which are supervised learning techniques, unsupervised learning techniques and semi-supervised learning techniques [18]. Expert systems developed by machine learning techniques can be used to assist physicians in diagnosing and predicting diseases [19]. Due to diseases diagnosis importance to mankind, several studies have been conducted on developing methods for their classification (see Table 1). Although these techniques can be used to predict the PD through a set of real-world datasets, however the most methods developed by supervised prediction techniques in the previous researches do not support the incremental updates of the data for PD prediction. Furthermore, standard supervised techniques cannot be used for the incremental learning and therefore they require to recompute all the training data to build prediction models. As the computation time and accuracy are two important criteria for the assessment of the medical diagnosis systems, in this research a new method using Incremental Support Vector Regression (ISVR) [20,21], an effective regression machine learning technique, is proposed in order to improve the predictive accuracy and reduce the computation time of PD prediction. The ISVR have been successfully used to solve real-time prediction problems. To the best knowledge of the

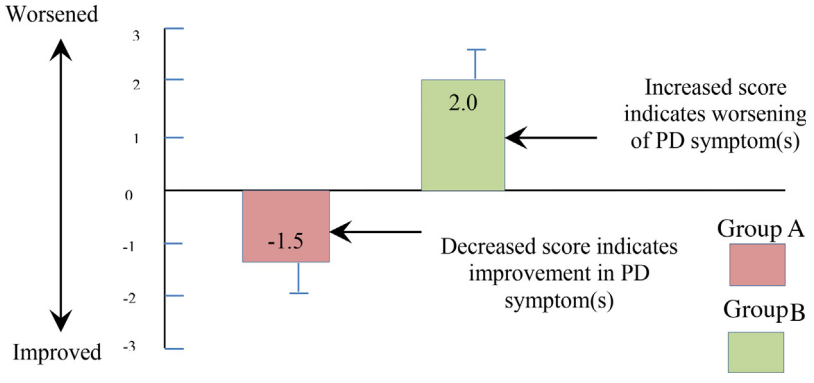


Fig. 1 – The effect of increasing and decreasing score on PD symptom(s).

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

Table 1 – Related work on PD diagnosis using machine learning techniques.

Disease	Author	Technique													
		SVM	KNN	NN	ANFIS	FL	KM	GP	EM	PCA	RF	LDA	DT	AR	PSO	NB
Parkinson	Guo et al. [48]							*	*							
	Das [24]			*												
	Bhattacharya and Bhatia [49]	*														
	Åström and Koker [50]					*										
	Li et al. [51]					*										
	Ozcift [52]	*														
	Polat [23]		*			*	*									
	Eskidere et al. [38]	*		*												
	Chen et al. [53]	*	*			*				*						
	Babu and Suresh [54]			*												
	Peterek et al. [55]										*					
	Hariharan et al. [56]			*					*	*		*				
	Froelich et al. [57]												*			
	Buza and Varga [58]			*												
	Al-Fatlawi et al. [59]			*												
	Jain and Shetty [60]		*	*										*		
	Behroozi and Sami [61]	*	*													*
	Avci and Dogantekin [62]							*								

Abbreviation used in this table: SVM: support vector machine, KNN: k-nearest neighbor, NN: neural network, ANFIS: adaptive neuro-fuzzy inference system, FL: fuzzy logic, KM: k-means, GP: genetic programming, EM: expectation maximization, PCA: principal component analysis, RF: Random Forest, LDA: linear discriminant analysis, DT: decision tree, AR: association rule, PSO: particle swarm optimization and NB: Naive Bayes.

authors, there is no implementation of a hybrid system using prediction (ISVR), clustering (SOM) and dimensionality reduction (NIPALS) in the context of PD diagnosis form the real-world datasets. Furthermore, since new information constantly is added to the medical datasets, it is important to consider the new data for constructed prediction models without recomputing all the training data. This will help to reduce the computation time in predicting the PD. Accordingly, the proposed method in the study at hand supports incremental updates by using ISVR to re-learn the PD data which can be more efficient in memory requirement. To evaluate the proposed method, we conduct several experiments with a real-world PD dataset. The dataset, Parkinson's telemonitoring dataset [22–24], is taken from Data Mining Repository of the University of California, Irvine (UCI) [25]. Overall, in comparison with the research efforts found in the literature, our work has the following contributions. In this research:

- a hybrid method is proposed for PD diagnosis using SOM, NIPALS, ISVR techniques. We believe that the combination of these techniques can accurately predict the UPDRS and at the same time improve the computation time of UPDRS prediction.
- SOM is used for the clustering of data in the PD dataset. SOM has been a robust clustering method in developing the disease diagnosis methods [26,27].
- NIPALS [28–30] is used for dimensionality reduction and dealing with the multi-collinearity problem in the dataset.
- ISVR technique is used for predicting the Total-UPDRS and Motor-UPDRS. Incremental machine learning techniques have proved to be effective when dealing with large datasets [31–33]. Cauwenberghs and Poggio [34] proposed incremental support vector machine for online learning when a new sample is added into the old samples. This idea has been extended to be used as an incremental leaning technique for regression problems. ISVR has been used for real-time prediction in several previous researches [35–37]. The results of using ISVR in the previous researches showed that it needs less memory capacity and less computation time compared with the non-incremental SVR. We believe that the use of ISVR for PD diagnosis will significantly reduce the computation time in relation to the non-incremental SVR.

The research is organized as follows: In Section 2 research method is presented. Section 3 presents the results of experimental analysis of SOM and NIPALS. Section 4 presents method evaluation. Section 5 presents ISVR results and method comparison. Discussion is presented in Section 6. Finally, conclusions and future work is provided in the Section 7.

2. Methodology

This paper aims to develop a hybrid intelligence method for PD diagnosis using a set of real data. The main aims of developing this method are to reduce the computation time and improve the prediction accuracy of PD diagnosis. Specifically, we take the advantages of ISVR to achieve the objectives of this research. We use ISVR to predict the Total-UPDRS and Motor-UPDRS. We also use non-linear iterative partial least squares (NIPALS) for data dimensionality reduction, self-organizing map (SOM) for clustering the data. We believe that the combination of these machine learning techniques can help to improve the prediction accuracy of previous methods and reduce the computation time of PD diagnosis systems. In Fig. 2, the machine learning techniques used in our method are presented. As can be seen from this figure, in the first step, data is clustered using SOM (1). In this step, the similar observations of patients are categorized in clusters generated by SOM. Then, we use NIPALS for data dimensionality reduction (2). Using

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

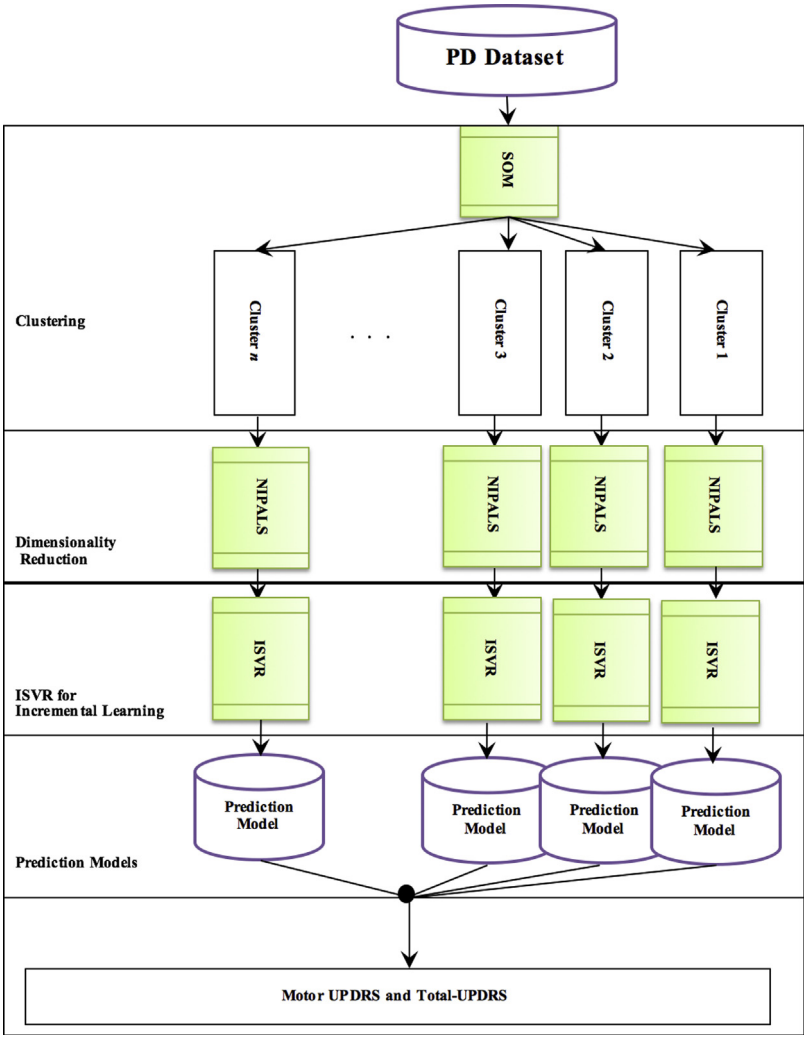


Fig. 2 – Research methodology.

NIPALS, the latent variables of principal component analysis (PCA) are obtained without correlation matrix diagonalization. In the last step, we use ISVR technique for the prediction of Total-UPDRS and Motor-UPDRS observed in the data (3).

We evaluate our method on a real-world PD dataset [22–24,38] which is available in the machine learning repository at UCI, University of California, Irvine. The dataset includes 5875 records (around 200 recordings per patient) from 28 men and 14 women. In this dataset, the Total-UPDRS includes the range of 0–176 (0 indicates healthy and 176 total disability) and the Motor-UPDRS, which refers to the motor section, is in the range of 0–108 (0 indicates healthy state and 108 severe motor impairment).

Table 2 presents the 16 features of dataset along with UPDRS scores [38]. From the correlation coefficients among the selected variables in Fig. 3, valuable information can be derived regarding the relationship between them. Overall, it can be seen that there is a strong correlation between the features and it may influence the accuracy result of disease prediction [39]. Accordingly, to reduce the multi-collinearity problem, we apply the NIPALS on the dataset before performing the prediction task.

2.1. *NIPALS*

PCA is a statistical technique for multivariate analysis and is used as a dimensionality reduction technique in data compression to retain the essential information which are easy to display (Moore, 1981). The method identifies patterns in data Q3 and represents the data in a way that highlights similarities and differences. The central idea is to reduce the number of dimensions of the data while retaining most of the data information. PCA has four goals. The first goal is to extract the most information from the data. The second goal is to compress the data by only keeping the most characterizing information. The third goal is to simplify the description of the data and the fourth goal is to enable analysis of the structure of the observations. The analysis enables conclusions to be drawn regarding the used variables and their relations. The analysis is performed through transforming the data to a new set of variables, called the principal components (PCs). The new variables PCs generate by PCA are uncorrelated and ordered so that the first few PCs retain most of the variations of the total dataset. The first PC describes the dimension in which the data has the biggest variation (variance) and the second component

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

Table 2 – Description of the features and UPDRS scores of the Parkinson's telemonitoring dataset [38].

Description	Label	Feature label	Min	Max	Mean	SD
Clinician's motor UPDRS score, linearly interpolated	Motor-UPDRS	Motor-UPDRS (baseline)	6	36	19.42	8.12
		Motor-UPDRS (after three months)	6	38	21.69	9.18
		Motor-UPDRS (after six months)	5	41	29.57	9.17
Clinician's total UPDRS score, linearly interpolated	Total-UPDRS	Total-UPDRS (baseline)	8	54	26.39	10.8
		Total-UPDRS (after three months)	7	55	29.36	11.82
		Total-UPDRS (after six months)	7	54	29.57	11.92
Several measures of variation in fundamental frequency	F1	MDVP:Jitter (%)	8E−4	0.1	0.006	0.006
	F2	MDVP:Jitter (Abs)	2E−6	4E−4	4E−5	3E−5
	F3	MDVP:Jitter:RAP	3E−4	0.057	0.003	0.003
	F4	MDVP:Jitter:PPQ5	4E−4	0.069	0.003	0.004
	F5	Jitter:DDP	10E−4	0.173	0.009	0.009
Several measures of variation in amplitude	F6	MDVP:Shimmer	0.003	0.269	0.034	0.026
	F7	MDVP:Shimmer (dB)	0.026	2.107	0.311	0.230
	F8	Shimmer:APQ3	0.002	0.163	0.017	0.013
	F9	Shimmer:APQ5	0.002	0.167	0.020	0.017
	F10	Shimmer:APQ11	0.003	0.276	0.028	0.020
	F11	Shimmer:DDA	0.005	0.488	0.052	0.040
Two measures of ratio of noise to tonal components in the voice	F12	NHR	3E−4	0.749	0.032	0.060
	F13	HNR	1.659	37.875	21.679	4.291
A nonlinear dynamical complexity measure	F14	RPDE	0.151	0.966	0.541	0.101
Signal fractal scaling exponent	F15	DFA	0.514	0.866	0.653	0.071
A nonlinear measure of fundamental frequency variation	F16	PPE	0.022	0.732	0.220	0.092

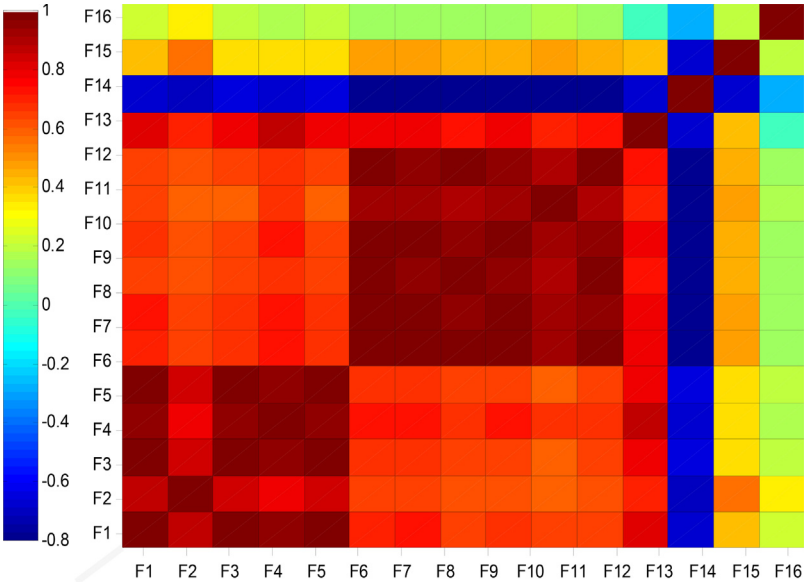


Fig. 3 – Correlation coefficients between different features of PD.

describes the dimension in which it has the second largest variation (variance). PCA is chosen for this study because the method exemplifies a category of analysis methods. If the data has linear relations and is correlated, as data often is in medical datasets, the method will give a compression that maintains a high amount of the information in the original dataset. The described solution saves a compact summary of the data, which is derived by applying ideas from statistics to enable an analysis while preserving its characteristics. The NIPALS algorithm is one of the many methods that exist for finding the eigenvectors [40]. Using NIPALS, the latent variables of PCA are obtained without correlation matrix diagonalization. NIPALS can handle missing values in the dataset. In NIPALS, scores, loadings and residuals can be retrieved directly and the algorithm is iterative. In the first iteration, the first PC is calculated, the second PC in the second iteration and so on. First the largest eigenvalue is found and then smaller and smaller ones. NIPALS algorithm is presented as follows:

Algorithm 1. NIPALS algorithm

Begin

➲ **Variables**

A: indicates a data matrix of the variable space.
Num: indicates the number of PCs. *Num*>=1 and *Num*<= *min*(*objects, variables*)
E(0): meancenter(*A*), the E-matrix for the zero-th PC (PC_0)
t: vector is set to a column in *A*. Will be the scores for PC_i
p: indicates the loadings for PC_i
threshold: (*t*= 0.00001, a low value) to do the convergence check towards zero.

➲ **Algorithm**

Step 1: Set *E*(0) equal to mean centered *A*
Step 2: Set *t* equal to an acceptable attribute vector of *A* (a non-zero vector)
Step 3: For each *i* in range(1, *Num)*:
S1: Project *A* onto *t* to find the corresponding loading *p*
$p = (E_{(i-1)}^T t)/t^T t$
S2: Normalize loading vector *p* to length 1
$p = p * (p^T p)^{-0.5}$
S3: Project *A* onto *p* to find corresponding score vector *t*
$t = (E_{(i-1)} p)/p^T p$
S4: Check for convergence. If difference between eigenvalues $\tau_{new} = t^T t$ and *t* old (from last round) is larger than *threshold**τ_{new} return to Step S1
S5: Remove the estimated PC component from $E_{(i-1)}$
$E(i) = E(i\text{-}1)\text{-}(tp^T)$
S6: Collect scores and loadings: current *t* and *p* (and current *E*(*i*) if *E*-matrices are needed)

End.

2.2. SOM

SOM has been proved useful as data clustering and projection tools [41,42]. It uses a small set of well-organized neurons to represent a data set and creates a neighborhood preserving mapping from a high-dimensional data space onto a low-dimensional grid of neurons. These neurons are located on a low-dimensional, usually 2-dimensional, grid which provides a convenient surface to project and summarize data. In fact, SOM creates a neighborhood preserving mapping from a high-dimensional data space onto a low-dimensional grid of neurons.

The SOM establishes a projection through an iterative learning procedure. It behaves like a flexible net that folds onto the 'cloud' formed by the input data during the learning. In each learning step, the data item is chosen randomly, and the distances between it and all the weight vectors of the output neurons are calculated using a certain distance metric (e.g., the Euclidean distance).

For clustering task, the SOM has been the most commonly used network as after the training phase in batch mode the cluster can be effectively visualized. Basically, the SOM network includes one layer, with neurons organized in a grid of rows and columns (see Fig. 4). In addition, the main goal of the SOM's learning in clustering task is the satisfactory preservation of topology from the data space onto the grid of neurons. When using a SOM for data clustering, each neuron is regarded as a subcluster center.

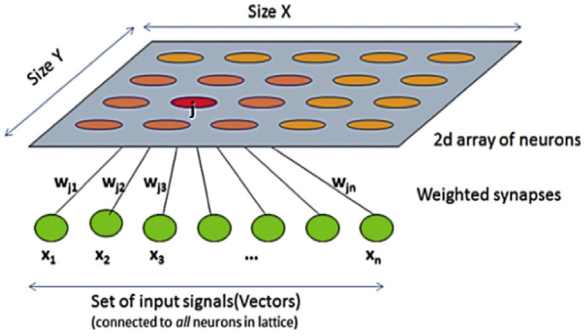


Fig. 4 – A SOM of size X × Y.

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

2.3. Incremental SVR

wAs a discriminative classifier, SVM, also known as support vector network, is becoming increasingly popular [43]. For two class classification problems, SVM needs to separate the data points and identify them in positive or negative classes. The aim is to maximize the margin around the separating hyper-plane that separates the data points. The mathematical background of SVM for solving the problem is as following:

$$\begin{aligned} &\min_{\vec{w},b} \quad \frac{1}{2}\|w\|, \quad s.t \quad \min_{\vec{w},b} \quad \frac{1}{2}\|w\|, \quad s.t \\ &y_i = +1 \Rightarrow \vec{w}\cdot\vec{x}_i + b \geq +1 \\ &y_i = -1 \Rightarrow \vec{w}\cdot\vec{x}_i - b \leq -1 \\ &s.t \quad y_i(\vec{w}\cdot\vec{x}_i + b) \geq 1, \quad \forall i \end{aligned} \tag{1}$$

The scalar b is also termed the bias.

In SVM, the identification and separation tasks for each data point x_i is based on y_i, which indicate the classes and can have a binary value. The solution hyper-plane is presents as follow:

$$u = \vec{w}\cdot\vec{x} + b \tag{2}$$

Theory of Lagrange is used as a standard method to solve the above problem. Using this theory, the above problem is converted to a dual Lagrangian problem. The dual problem can been defined as following:

$$\begin{aligned} &\min_{\alpha} \quad \Psi(\vec{\alpha}) = \min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j (\vec{x_i}\cdot\vec{x_j})\alpha_i\alpha_j - \sum_{i=1}^{N}\alpha_i \\ &\sum_{i=1}^{N}\alpha_i y_i = 0 \\ &\alpha_i \geq 0, \quad \forall i \end{aligned} \tag{3}$$

The variables α_i are the Lagrangian multipliers for corresponding data point x_i.

SVR is an extension of the support vector classifier which estimates the continuous function of certain training data sets. SVR as a nonparametric technique also relies on kernel functions. Suppose we have the training dataset with l samples,

$$\mathbf{X} = \{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_l, y_l)\} \quad x_i \in R^n, \quad i = 1, \ldots, l \tag{4}$$

SVR seeks an optimal function $f(x) = (wx) + b$ where, w is the weight vector and $b \in R$ is the threshold value. The aim of using SVR is to minimize the expected risk of prediction. By introducing the slack variable ξ_i and ξ_i^*, the prediction problem can be formulated as:

$$\begin{aligned} \min W = &\frac{1}{2}\sum_{i,j}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i . x_j) \\ &+ \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l} y_i(\alpha_i^* + \alpha_i) \\ s.t. \quad &\begin{cases} \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, 2, \ldots, l. \end{cases} \end{aligned} \tag{5}$$

After the introduction of kernel function $K(x_i, x_j)$, the regressive function becomes:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{6}$$

where α_i^* and α_i are Lagrangian multipliers.

Cauwenberghs and Poggio [34] proposed incremental SVM by analyzing the changes of the Karush–Kuhn–Tucker (KKT) conditions for online learning when a new (incremental) sample was added into the old samples. This idea can also be used as the incremental SVR. This study applies ε-SVR and develop the regression models for the PD prediction.

3. Results and discussion

The results of experimental analysis of the proposed method for the prediction of PD progression are explained in this section. Here, the results of applying all incorporated methods in the proposed system are discussed.

3.1. Step 1: Clustering using SOM

As we discussed in the methodology section, SOM technique for clustering is applied on the PD dataset. We selected different clustering size for SOM. For SOM, SOM 2×2 (4 clusters), SOM 2×3 (6 clusters), SOM 3×3 (9 clusters) and SOM 3×4 (12 clusters) were considered. Fig. 5 visualizes the generated clusters for SOM 2×2 (4 clusters) and SOM 2×3 (6 clusters) on PD dataset. In addition, all cluster centroids are presented in the tables of Appendix A. In Tables 3–6, the number of instances in PD dataset along with the map quality are presented in each cluster of SOM 2×2, SOM 2×3, SOM 3×3 and SOM 3×4. As the map quality of SOM 3×3 was higher than other SOM Map qualities, we selected 9 clusters for further experimental analysis.

3.2. Step 2: Dimensionality reduction using NIPALS

After a number of different SOM clusters are developed, we applied NIPALS on the clusters for data dimensionality reduction. According to Algorithm 1, we performed PCA with fully informative analyses using the NIPALS and accordingly computed PCs until above than 99% of variation was explained (Wold et al., 1987). In Table 7, it can be seen that for Cluster 1 Q4 ten PCs provide the 99.24% information of original data.

4. ISVR method evaluation

The prediction models constructed by ISVR were trained under a 4 GHz processor PC and Microsoft Windows 7 running MATLAB 7.10 (R2010a). For evaluating the ISVR, we initially considered 20% of data in each cluster for test set, 20% for initial data clusters and 60% for incremental set (see Fig. 6). The increment ratio is considered 5% and 10% of incremental set and added to training set and calculated computation time. Specifically, we consider 12 and 6 measurement points and add 5% (for computation time) and 10% (for accuracy) of data to

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

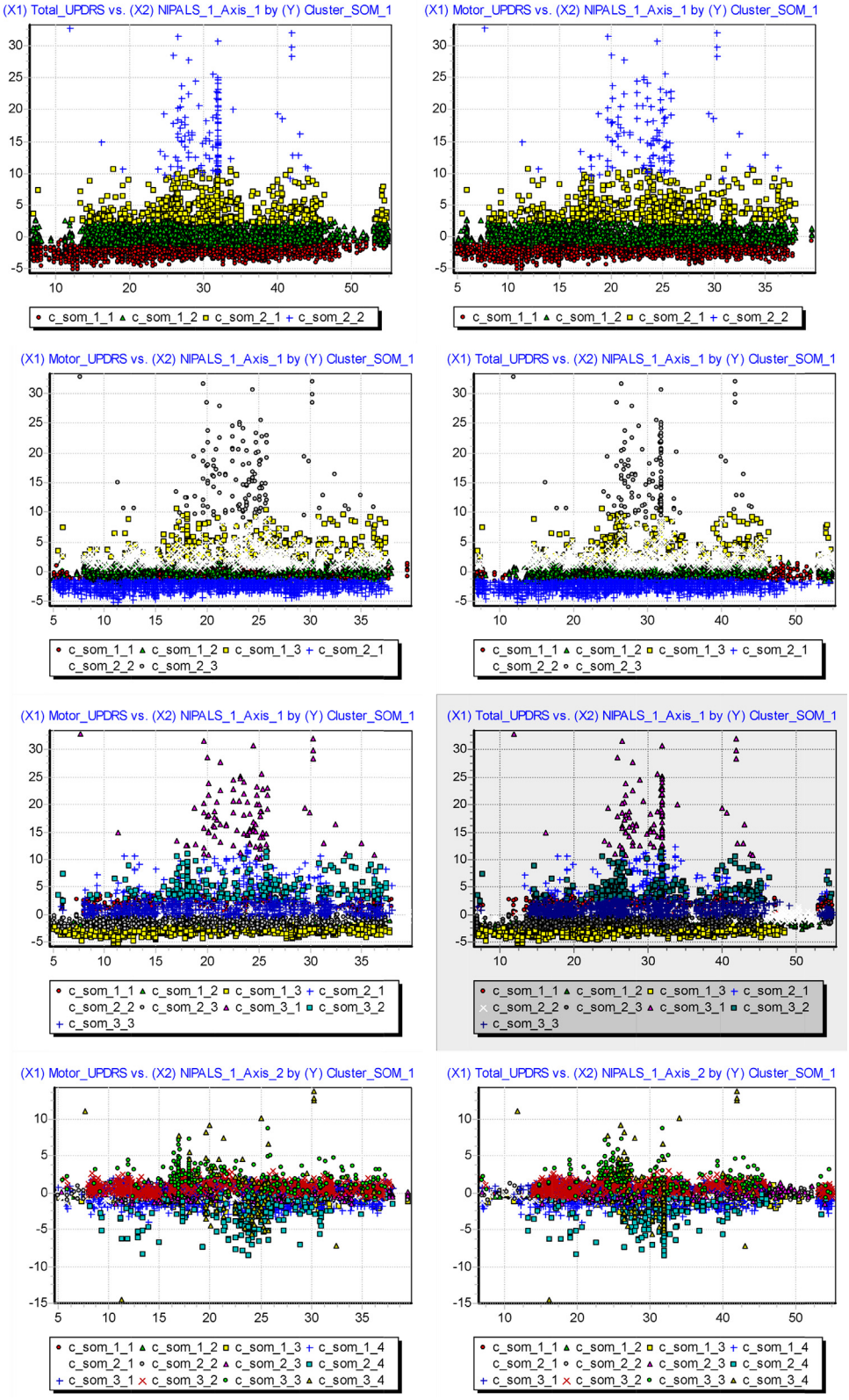


Fig. 5 – SOM 2 × 2 (4 clusters), SOM 2 × 3 (6 clusters), SOM 3 × 3 (9 clusters) and SOM 3 × 4 (12 clusters).

the initial clusters in each measurement point. In that direction, for different measurement points, the average computation time and accuracy were calculated for all clusters. The increment ratio is considered 5% and 10% of incremental set and added to training set and calculated computation time. Specifically, we consider 12 and 6 measurement points and add 5% (for computation time) and 10% (for accuracy) of data to the initial clusters in each measure-

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

Table 3 – MAP topology for SOM 2 × 2.

Map quality = 0.7943	1	2
1	2866	2386
2	508	115

Table 4 – MAP topology for SOM 2 × 3.

Map quality = 0.7817	1	2	3
1	1498	1400	364
2	1887	599	127

Table 5 – MAP topology for SOM 3 × 3.

Map quality = 0.8814	1	2	3
1	675	1037	978
2	177	928	1009
3	98	246	727

Table 6 – MAP topology for SOM 3 × 4.

Map Quality = 0.7618	1	2	3	4
1	604	730	670	521
2	676	581	337	166
3	628	597	263	102

Table 7 – NIPALS result of Cluster 1.

PCs	Eigenvalue	Explained (%)	Cumulutive (%)
PC1	8.088228	50.55	50.55
PC2	3.456244	21.60	72.15
PC3	1.285030	8.03	80.18
PC4	1.115777	6.97	87.16
PC5	0.585436	3.66	90.82
PC6	0.462398	2.89	93.71
PC7	0.332143	2.08	95.78
PC8	0.280873	1.76	97.54
PC9	0.169520	1.06	98.60
PC10	0.103169	0.64	99.24

ment point. In that direction, for different measurement points, the average computation time and accuracy were calculated for all clusters. Note that we applied SVR with RBF kernel on experimental dataset clustered by SOM algorithm.

To evaluate the prediction models of ISVR technique, we used mean absolute error (MAE), root mean square error (RMSE) and R^2, which are respectively presented in Eqs. (7)–(9):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|(\hat{y}_i - y_i)\right| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{8}$$

$$R2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\sum(yi - \hat{y}i)2}{\sum(yi - \bar{y}i)2} \tag{9}$$

where n, y, $\hat{y}$, $\bar{y}$ are respectively the entire number of sample in each cluster of the dataset, the actual value, the predicted value by the method and the average of $[y_1, y_2, \ldots, y_n]$.

In order to experimentally show the effectiveness of ISVR, we divide the data in the clusters into two categories. The first category is considered as initial clustering and the second one is considered for incremental phase that is incrementally added to the initial clusters data. The aim is to calculate the prediction time and accuracy of method after adding the second category data incrementally. We perform this procedure on all clusters and present the average computation time and accuracy.

5. Results and comparisons of methods

This section provides the experimental results of PD prediction with non-incremental and incremental SVR on PD dataset. In addition, one of the main objectives of this research is also to make a consistent comparison between the prediction accuracy achieved by our method and the prediction accuracy achieved by other corresponding results of several prediction techniques. Hence, in this section, comparison experiments with other methods are performed on the same dataset.

As we have selected RBF kernel for ISVR, there are two parameters, C and γ, which are unknown and we need to set a best value for them. Hence, some kind of model selection methods are required to find an optimal value for C and γ. The aim of this task is to find good parameters for RBF kernel so that SVR can provide good prediction models and accurately predict the unknown classes in testing data. To do so, we used k-fold cross-validation ($k = 10$) [44] with 10 trials as a statistical model selection method. By trying several values for the parameters C and γ, we then set the value of penalty parameter C and γ in RBF kernel equal to the optimal one determined via 10-fold cross-validation. The method of choosing C and γ was exhaustive search as it is the most popular method in determining the SVR parameters. Specifically, we tried exponentially growing sequences of C form 2^{-15} to 2^{10} and γ from 2^{-10} to 2^{9} to find the optimal values. After testing there ranges, we found that the best (C, γ) is $(2^3, 2^{-2})$.

In Fig. 7a and b, the prediction accuracy of ISVR measured by R^2 and MAE in each cluster for Total-UPDRS and Motor UPDRS. From all plots in Fig. 7a and b, we can see that the ISVR has provided good prediction accuracies for each cluster measured by MAE. In addition, the results showed that the average prediction accuracy of all clusters measured by MAE for the Total-UPDRS is MAE = 0.4656. In addition, for the Motor-UPDRS the average prediction accuracy of all clusters measured by MAE is MAE = 0.4967. These indicate that the estimated Total-UPDRS and Motor-UPDRS are respectively 0.4656 points (out of 176) and 4967 points (out of 108). We also compare the results of our study with other prediction machine learning techniques, neural network (NN), adaptive neuro-fuzzy inference systems (ANFIS), multiple linear regression (MLR) and SVR. We apply these techniques on the same dataset. The results are provided in Table 8. The comparisons are made on the average prediction accuracy of the Total-UPDRS and Motor-UPDRS. It can be seen that the

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

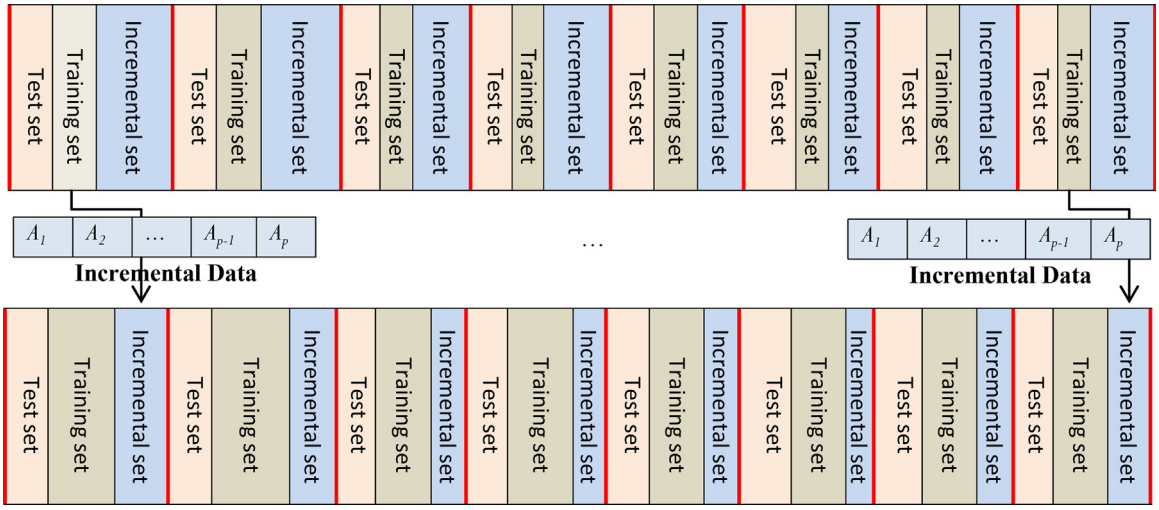


Fig. 6 – The procedure of ISVR method evaluation.

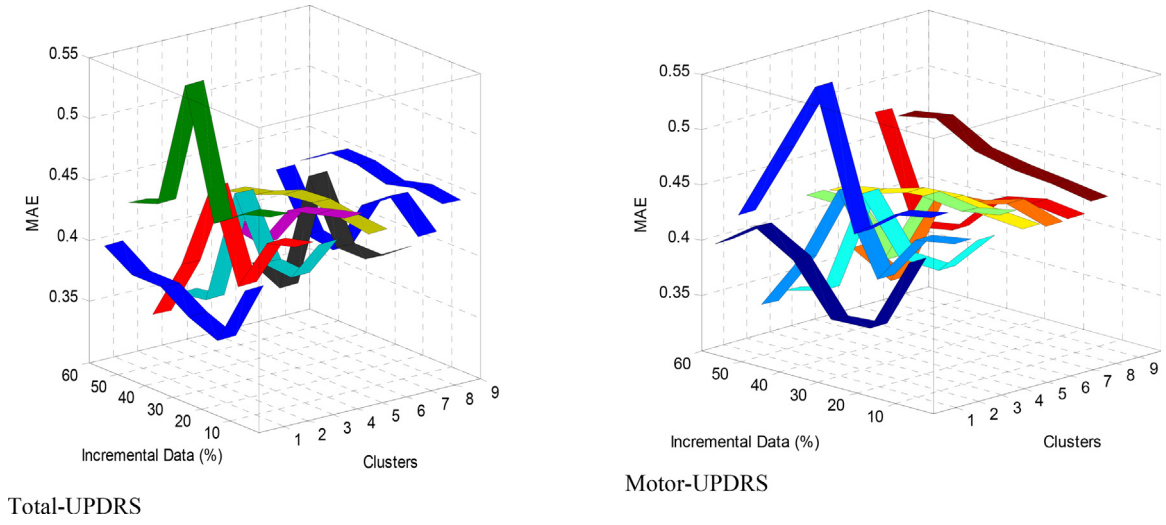


Fig. 7 – ISVR evalution for PD predicion accuracy.

method which uses SOM, NIPALS and ISVR (SOM-NIPALS-ISVR) is more accurate compared with the other methods which use solely prediction techniques.

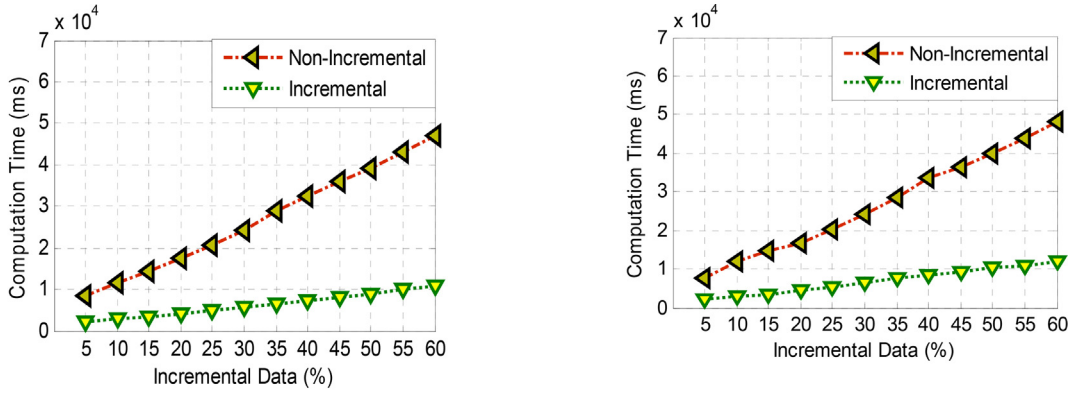
In order to compare the proposed method with previous work, we also evaluate the proposed method on the PD dataset using an additional metric. We present the results in Table 8. In this table, we report RMSE values as used in Zhao et al. [22]. In Zhao et al. [22], they developed a method, HSLSSVR, based on SVR and evaluated it on the PD dataset. In addition, they reported the RMSE results for Motor-UPDRS and Total-UPDRS. From Table 8, we can see that their method has also provided low prediction errors for Motor-UPDRS and Total-UPDRS. The prediction errors are about 0.8158 and 0.8004 for Motor-UPDRS and Total-UPDRS, respectively. However, compared with the SOM-NIPALS-ISVR method, we can see that the proposed method produce lower errors than the HSLSSVR method, leading to a lower RMSE values.

To experimentally show the effectiveness of ISVR for the computation time, we conduct several experiments by ISVR on the PD dataset and compared with the non-incremental SVR. Fig. 8a and b presents the computation time results of our analysis for the non-incremental SVR and the proposed method in the incremental situation. The computation time is plotted as a function of the incremental data percentage. From all curves in Fig. 8a and b, we can see that the incremental methods have significantly reduced the computation time in relation to the non-incremental one. In addition, as the figures show, non-incremental method performs poor with respect to the computation time for the Total-UPDRS and Motor-UPDRS prediction. From the curves as shown in the figure, it can be also observed that by increasing the number of incremental data, the computation time is slightly raised for ISVR compared with the non-incremental SVR. A possible explanation could be that, since the non-incremental method

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

Table 8 – Results of MAE and R^2 for all prediction methods.

Method	Measure	MAE	RMSE	R^2
MLR	Motor-UPDRS	0.997	2.4142	0.697
	Total-UPDRS	0.987	2.3911	0.709
NN	Motor-UPDRS	0.977	2.3836	0.719
	Total-UPDRS	0.951	2.3135	0.734
ANFIS	Motor-UPDRS	0.771	1.7047	0.785
	Total-UPDRS	0.743	1.6062	0.798
SVR	Motor-UPDRS	0.721	1.4942	0.814
	Total-UPDRS	0.689	1.4526	0.819
HSLSSVR [22]	Motor-UPDRS	–	0.8158	–
	Total-UPDRS	–	0.8004	–
SOM-NIPALS-ISVR	**Motor-UPDRS**	**0.4656**	**0.6268**	**0.885**
	Total-UPDRS	**0.4967**	**0.7097**	**0.868**



(a) Motor-UPDRS **(b) Total-UPDRS**

Fig. 8 – ISVR evalution for computation time.

cannot learn in the incremental situation, it requires to recompute all the training data to build the prediction models. In the other word, those medical records in the experimental dataset, which have been incrementally added, need to be retrained along with the previous data in each cluster through non-incremental prediction method. However, the method that use ISVR reduce computation time results as it needs to train only the data which has been added incrementally. Overall, the results showed that the main practical advantage of using ISVR as a prediction method is a great saving in the computation time.

6. Discussion

The main findings of this study are that the prediction method is integrated with dimensionality reduction and clustering techniques improved the accuracy prediction of PD and reduced the computation time. The superiority of the present method can be explained by the fact that our model support incremental updates of the data. In addition, the proposed method in this study supports incremental updates and re-learning of data which is more efficient in memory requirement. It is worth noting that our proposed method achieved the best performance on the PD dataset. The accuracies of prediction measured by MAE and R^2 for the Total-UPDRS were respectively MAE = 0.4656 and R^2 = 0.885, and for Motor-UPDRS were respectively MAE = 0.4967 and R^2 = 0.868. Beside the improvement in prediction accuracy, our incremental method proved to have a good performance in reducing the computation time in relation to the non-incremental one. A further advantage of the proposed method was that our method is efficient in memory requirement and can be effectively implemented for large datasets. Moreover, in the medical domains where the new data arrive at a high rate, incremental updates of our method allows to constantly improve the prediction models without computationally costly retraining the whole dataset. Overall, the prediction accuracy and computation time yielded by the proposed method compared well with the other corresponding results of several prediction techniques. In summary, the findings of our experiments on a public PD dataset showed the effectiveness of incorporating the clustering and dimensionality reduction techniques in improving the prediction accuracy and reducing the computation time of PD diagnosis.

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

7. Conclusion and future work

In this paper, a hybrid method was proposed for the UPDRS (Total-UPDRS and Motor-UPDRS) prediction using machine learning techniques. We used ISVR to predict the Total-UPDRS and Motor-UPDRS. SOM and NIPALS were respectively used for clustering and data dimensionality reduction. To evaluate the proposed method, several experimental analyses were conducted on a real-world PD dataset taken from UCI. The results indicated that the method which combines SOM, NIPALS, and ISVR techniques is effective in predicting the Total-UPDRS and Motor-UPDRS. The results also indicated that the proposed method can significantly reduce the prediction computation time of Total-UPDRS and Motor-UPDRS.

The method proposed in this study has been evaluated by a public datasets from UCI which have input and output parameters for PD diagnosis. In addition, compared to the big healthcare data, the nature of the data in these datasets is not complex. In addition, in case of big healthcare data which can be complex datasets with unique characteristics, the future studies need to consider this issue in the development of new methods in order to overcome the challenges of data processing time and take advantage of big data. Furthermore, as big healthcare data include multi-spectral, heterogeneous, imprecise and incomplete observations (e.g., diagnosis) which are derived from different sources, therefore new methods are needed and relying solely on conventional machine learning techniques may include some shortcomings in predicting the disease.

There is still plenty of work in conducting researches on clustering, dimensionality reduction and prediction techniques for the Total-UPDRS and Motor-UPDRS prediction in order to exploit all their potential and usefulness. In our future study, we plan to evaluate the proposed method on additional PD datasets and in particular on large datasets which includes other attributes for PD diagnosis to show the effectiveness of the method for computation time of large data. In addition, our future work will investigate that how the proposed method can be extended to be applicable to the other types of datasets in medical domain.

Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this article.

Q5 Uncited references

[45–47].

Appendix A

Table A1 – Cluster centroids for SOM 3 × 4.

Attribute	Cluster n1	Cluster n2	Cluster n3	Cluster n4	Cluster n5	Cluster n6	Cluster n7	Cluster n8	Cluster n9	Cluster n10	Cluster n11	Cluster n12
MDVP:jitter (%)	0.002548	0.003553	0.004871	0.006210	0.003879	0.004970	0.007336	0.010027	0.005359	0.008027	0.015244	0.036359
MDVP:jitter (Abs)	0.000014	0.000023	0.000033	0.000047	0.000024	0.000038	0.000059	0.000077	0.000042	0.000064	0.000123	0.000190
MDVP:jitter:RAP	0.001187	0.001592	0.002358	0.003033	0.001878	0.002194	0.003513	0.005022	0.002496	0.003842	0.008119	0.019137
MDVP:jitter:PPQ5	0.001288	0.001744	0.002491	0.003295	0.002017	0.002451	0.003739	0.005560	0.002719	0.004212	0.007778	0.025255
Jitter:DDP	0.003562	0.004776	0.007075	0.009100	0.005635	0.006581	0.010538	0.015067	0.007487	0.011525	0.024359	0.057413
MDVP:Shimmer	0.013690	0.019663	0.036142	0.056636	0.020703	0.025776	0.039267	0.103107	0.021242	0.035087	0.052077	0.159890
MDVP:Shimmer (dB)	0.127288	0.182762	0.330494	0.509910	0.187259	0.238475	0.365068	0.935319	0.192854	0.320765	0.492221	1.411794
Shimmer:APQ3	0.006729	0.009550	0.018976	0.030354	0.010507	0.012487	0.020144	0.051984	0.010075	0.017352	0.026255	0.077042
Shimmer:APQ5	0.007678	0.011101	0.021483	0.034528	0.012121	0.014834	0.023120	0.062885	0.012091	0.020612	0.029594	0.102926
Shimmer:APQ11	0.010823	0.016360	0.029133	0.044666	0.016763	0.021974	0.031499	0.083849	0.018378	0.030214	0.041363	0.108722
Shimmer:DDA	0.020188	0.028650	0.056929	0.091061	0.031520	0.037460	0.060432	0.155952	0.030226	0.052056	0.078764	0.231126
NHR	0.007695	0.013630	0.021172	0.032878	0.011670	0.023749	0.044982	0.117535	0.016738	0.027432	0.074372	0.392025
HNR	27.820805	24.690945	21.675654	18.546440	24.242070	21.786630	18.753881	13.656404	22.217949	19.581333	16.780973	6.514500
RPDE	0.397571	0.520151	0.511494	0.580710	0.438961	0.603123	0.636380	0.674960	0.577159	0.563672	0.650592	0.709122
DFA	0.583230	0.584296	0.627634	0.678195	0.673877	0.609387	0.623405	0.658057	0.725306	0.746597	0.724785	0.631107
PPE	0.114257	0.152192	0.198093	0.246234	0.161298	0.213722	0.280895	0.328034	0.217204	0.291059	0.408071	0.481860

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

Table A2 – Cluster centroids for SOM 2 × 2.

Attribute	Cluster n1	Cluster n2	Cluster n3	Cluster n4
MDVP:Jitter (%)	0.003710	0.006607	0.011445	0.034299
MDVP:Jitter (Abs)	0.000024	0.000051	0.000089	0.000187
MDVP:Jitter:RAP	0.001704	0.003177	0.005906	0.018132
MDVP:Jitter:PPQ5	0.001861	0.003417	0.006043	0.023435
Jitter:DDP	0.005112	0.009532	0.017718	0.054395
MDVP:Shimmer	0.020047	0.036678	0.072918	0.156053
MDVP:Shimmer (dB)	0.184518	0.335133	0.668642	1.380548
Shimmer:APQ3	0.009900	0.018725	0.037574	0.075214
Shimmer:APQ5	0.011477	0.021746	0.043557	0.099502
Shimmer:APQ11	0.016541	0.030229	0.058256	0.107157
Shimmer:DDA	0.029701	0.056175	0.112722	0.225643
NHR	0.013396	0.028237	0.078296	0.375337
HNR	24.662103	20.047328	15.851876	6.954522
RPDE	0.488765	0.574114	0.647187	0.710837
DFA	0.620282	0.685647	0.691527	0.633103
PPE	0.158660	0.253220	0.348722	0.469848

Table A3 – Cluster centroids for SOM 2 × 3.

Attribute	Cluster n1	Cluster n2	Cluster n3	Cluster n4	Cluster n5	Cluster n6
MDVP:Jitter (%)	0.005052	0.006492	0.014377	0.003259	0.006823	0.031704
MDVP:Jitter (Abs)	0.000037	0.000051	0.000116	0.000020	0.000052	0.000171
MDVP:Jitter:RAP	0.002344	0.003095	0.007562	0.001503	0.003345	0.016641
MDVP:Jitter:PPQ5	0.002534	0.003383	0.007408	0.001639	0.003615	0.021765
Jitter:DDP	0.007032	0.009285	0.022686	0.004508	0.010035	0.049924
MDVP:Shimmer	0.031544	0.029892	0.053606	0.017353	0.064719	0.156146
MDVP:Shimmer (dB)	0.290352	0.271998	0.502610	0.159696	0.586224	1.383465
Shimmer:APQ3	0.016102	0.014780	0.026891	0.008528	0.034280	0.075303
Shimmer:APQ5	0.018526	0.017573	0.030666	0.009844	0.039323	0.100001
Shimmer:APQ11	0.025828	0.025440	0.043198	0.014144	0.051233	0.110575
Shimmer:DDA	0.048305	0.044339	0.080672	0.025584	0.102839	0.225909
NHR	0.024127	0.022727	0.070945	0.010825	0.048138	0.359523
HNR	21.561025	20.729431	16.928992	25.710736	17.470629	7.119654
RPDE	0.559647	0.560117	0.645553	0.460055	0.609518	0.712065
DFA	0.614691	0.730310	0.723034	0.611569	0.662601	0.633288
PPE	0.208337	0.251415	0.396719	0.140476	0.264315	0.458323

Table A4 – Cluster centroids for SOM 3 × 3.

Attribute	Cluster n1	Cluster n2	Cluster n3	Cluster n4	Cluster n5	Cluster n6	Cluster n7	Cluster n8	Cluster n9
MDVP:Jitter (%)	0.005928	0.003946	0.002783	0.009407	0.005500	0.004920	0.036554	0.015996	0.008372
MDVP:Jitter (Abs)	0.000044	0.000026	0.000015	0.000072	0.000041	0.000036	0.000188	0.000128	0.000068
MDVP:Jitter:RAP	0.002889	0.001804	0.001302	0.004681	0.002553	0.002307	0.019200	0.008603	0.004033
MDVP:Jitter:PPQ5	0.003144	0.001964	0.001406	0.005178	0.002755	0.002516	0.025694	0.008177	0.004368
Jitter:DDP	0.008668	0.005411	0.003907	0.014043	0.007659	0.006921	0.057601	0.025811	0.012098
MDVP:Shimmer	0.053156	0.022640	0.015149	0.099845	0.032746	0.021991	0.163097	0.056563	0.035262
MDVP:Shimmer (dB)	0.480327	0.209555	0.139865	0.904446	0.301553	0.199302	1.440388	0.532110	0.323922
Shimmer:APQ3	0.028441	0.011232	0.007477	0.050617	0.016671	0.010700	0.078376	0.028741	0.017406
Shimmer:APQ5	0.032297	0.012995	0.008560	0.061245	0.019228	0.012714	0.105172	0.031989	0.020647
Shimmer:APQ11	0.042088	0.018803	0.012064	0.081717	0.026757	0.018659	0.110868	0.043820	0.030231
Shimmer:DDA	0.085324	0.033697	0.022431	0.151852	0.050012	0.032099	0.235128	0.086224	0.052218
NHR	0.031190	0.015883	0.008473	0.108055	0.027809	0.015138	0.403527	0.082533	0.031415
HNR	19.010993	23.716084	27.027103	13.937294	20.861878	22.788262	6.234663	16.396561	19.317602
RPDE	0.570578	0.520331	0.415842	0.669819	0.584631	0.528570	0.711761	0.651953	0.584843
DFA	0.667061	0.597457	0.598792	0.658594	0.615492	0.715435	0.627223	0.719864	0.734744
PPE	0.240024	0.167722	0.121616	0.317708	0.223787	0.202764	0.479027	0.412509	0.300251

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

REFERENCES

[1] Postuma RB, Montplaisir J. Predicting Parkinson's disease – why, when, and how? Parkinsonism Relat Disord 2009;15: S105–9.
[2] Armañanzas R, Bielza C, Chaudhuri KR, Martinez-Martin P, Larrañaga P. Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach. Artif Intell Med 2013;58(3):195–202.
[3] Dunnewold RJW, Jacobi CE, Van Hilten JJ. Quantitative assessment of bradykinesia in patients with Parkinson's disease. J Neurosci Methods 1997;74(1):107–12.
[4] Nutt JG, Wooten GF. Diagnosis and initial management of Parkinson's disease. N Engl J Med 2005;353(10).
[5] Priyadarshi A, Khuder SA, Schaub EA, Priyadarshi SS. Environmental risk factors and Parkinson's disease: a metaanalysis. Environ Res 2001;86(2):122–7. 1021–7.
[6] Renfroe JB, Bradley MM, Okun MS, Bowers D. Motivational engagement in Parkinson's disease: preparation for motivated action. Int J Psychophysiol 2016;99:24–32.
[7] Daneault JF, Carignan B, Sadikot AF, Duval C. Are quantitative and clinical measures of bradykinesia related in advanced Parkinson's disease? J Neurosci Methods 2013;219(2):220–3.
[8] Olanow CW, Watts RL, Koller WC. An algorithm (decision tree) for the management of Parkinson's disease (2001): treatment guidelines. Neurology 2001;56(Suppl. 5):S1–88.
[9] Rascol O, Goetz C, Koller W, Poewe W, Sampaio C. Treatment interventions for Parkinson's disease: an evidence based assessment. Lancet 2002;359(9317): 1589–98.
[10] Martignoni E, Franchignoni F, Pasetti C, Ferriero G, Picco D. Psychometric properties of the unified Parkinson's disease rating scale and of the short Parkinson's evaluation scale. Neurol Sci 2003;24(3):190–1.
[11] Van Hilten JJ, Van Der Zwan AD, Zwinderman AH, Roos RAC. Rating impairment and disability in Parkinson's disease: evaluation of the Unified Parkinson's Disease Rating Scale. Mov Disord 1994;9(1):84–8.
[12] Farnikova K, Krobot A, Kanovsky P. Musculoskeletal problems as an initial manifestation of Parkinson's disease: a retrospective study. J Neurol Sci 2012;319(1): 102–4.
[13] Tsanas A, Little M, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng 2010;57 (4):884–93.
[14] Romenets SR, Gagnon JF, Latreille V, Panniset M, Chouinard S, Montplaisir J, et al. Rapid eye movement sleep behavior disorder and subtypes of Parkinson's disease. Mov Disord 2012;27(8):996–1003.
[15] Loukas C, Brown P. A PC-based system for predicting movement from deep brain signals in Parkinson's disease. Comput Methods Programs Biomed 2012;107(1):36–44.
[16] Gulbinat W. What is the role of WHO as an intergovernmental organisation. The coordination of telematics in healthcare. Geneva, Switzerland: World Health Organisation; 1997. http://www.hon.ch/libraray/papers/gulbinat.html.
[17] Han J, Kamber M. Data mining: concepts and techniques. 2nd ed. San Francisco, CA, USA: Morgan Kaufmann; 2001.
[18] Huang G, Song S, Gupta JN, Wu C. Semi-supervised and unsupervised extreme learning machines. IEEE Trans Cybernet 2014;44(12):2405–17.
[19] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001;23(1):89–109.
[20] Gu B, Sheng VS, Wang Z, Ho D, Osman S, Li S. Incremental learning for ν-support vector regression. Neural Netw 2015;67:140–50.
[21] Laskov P, Gehl C, Krüger S, Müller KR. Incremental support vector learning: analysis, implementation and applications. J Mach Learn Res 2006;7:1909–36.
[22] Zhao YP, Li B, Li YB, Wang KK. Householder transformation based sparse least squares support vector regression. Neurocomputing 2015;161:243–53.
[23] Polat K. Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering. Int J Syst Sci 2012;43(4):597–609.
[24] Das R. A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Syst Appl 2010;37 (2):1568–72.
[25] Newman DJ, Hettich S, Blake CL, Merz CJ. UCI repository of machine learning databases; 1998.
[26] Zhang K, Chai Y, Yang SX. Self-organizing feature map for cluster analysis in multi-disease diagnosis. Expert Syst Appl 2010;37(9):6359–67.
[27] Ortiz A, Górriz JM, Ramírez J, Salas-Gonzalez D, Llamas-Elvira JM. Two fully-unsupervised methods for MR brain image segmentation using SOM-based strategies. Appl Soft Comput 2013;13(5):2668–82.
[28] Miyashita Y, Itozawa T, Katsumi H, Sasaki SI. Comments on the NIPALS algorithm. J Chemom 1990;4(1):97–100.
[29] Gonzalez M, Goracci L, Cruciani G, Poggesi I. Some considerations on the predictions of pharmacokinetic alterations in subjects with liver disease. Expert Opin Drug Metab Toxicol 2014;10(10):1397–408.
[30] Muratore M. Raman spectroscopy and partial least squares analysis in discrimination of peripheral cells affected by Huntington's disease. Anal Chim Acta 2013;793:1–10.
[31] LeCun YA, Bottou L, Orr GB, Müller KR. Efficient backprop. Neural networks: tricks of the trade; 2012. pp. 9–48.
[32] Bordes A, Ertekin S, Weston J, Bottou L. Fast kernel classifiers with online and active learning. J Mach Learn Res 2005;6(September):1579–619.
[33] Tsang IW, Kwok JT, Cheung PM. Core vector machines: fast SVM training on very large data sets. J Mach Learn Res 2005;6(April):363–92.
[34] Cauwenberghs G, Poggio T. Incremental and decremental support vector machine learning. NIPS, vol. 13; 2000.
[35] Su H, Zhang L, Yu S. Short-term traffic flow prediction based on incremental support vector regression. Third International Conference on Natural Computation, 2007, ICNC 2007, vol. 1. IEEE; 2007. p. 640–5.
[36] Yan J, Tian C, Wang Y, Huang J. Online incremental regression for electricity price prediction. 2012 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI). IEEE; 2012. p. 31–5.
[37] Zhang J, Sato T, Iai S. Support vector regression for on-line health monitoring of large-scale structures. Struct Saf 2006;28(4):392–406.
[38] Eskidere Ö, Ertaş F, Hanilçi C. A comparison of regression methods for remote tracking of Parkinson's disease progression. Expert Syst Appl 2012;39(5):5523–8.
[39] Nilashi M, Ibrahim O, Ahani A. Accuracy Improvement for Predicting Parkinson's Disease Progression. Sci Rep 2016;6.
[40] Wold H. Causal flows with latent variables: partings of the ways in the light of NIPALS modelling. Eur Econ Rev 1974;5 (1):67–86.
[41] Vesanto J, Alhoniemi E. Clustering of the self-organizing map. IEEE Trans Neural Netw 2000;11(3):586–600.
[42] Kohonen T, Oja E, Simula O, Visa A, Kangas J. Engineering applications of the self-organizing map. Proc IEEE 1996;84 (10):1358–84.

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002

[43] Amari SI, Wu S. Improving support vector machine classifiers by modifying kernel functions. Neural Netw 1999;12(6):783–9.
[44] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI 1995;14 (2):1137–45.
[45] Loh WY. Classification and regression trees. Wiley Interdisc Rev: Data Mining Knowl Discov 2011;1(1):14–23.
[46] Pelleg D, Moore AW. X-means: extending K-means with efficient estimation of the number of clusters. ICML. 2000. pp. 727–34.
[47] Quinlan JR. Simplifying decision trees. Int J Man-Mach Stud 1987;27(3):221–34.
[48] Guo PF, Bhattacharya P, Kharma N. Advances in detecting Parkinson's disease. Medical biometrics. Berlin Heidelberg: Springer; 2010. p. 306–14.
[49] Bhattacharya I, Bhatia MPS. SVM classification to distinguish Parkinson disease patients. Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India. ACM; 2010. p. 14.
[50] Åström F, Koker R. A parallel neural network approach to prediction of Parkinson's Disease. Expert Syst Appl 2011;38 (10):12470–4.
[51] Li DC, Liu CW, Hu SC. A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. Artif Intell Med 2011;52(1):45–52.
[52] Ozcift A. SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. J Med Syst 2012;36(4):2141–7.
[53] Chen HL, Huang CC, Yu XG, Xu X, Sun X, Wang G, et al. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Syst Appl 2013;40(1):263–71.
[54] Babu GS, Suresh S. Parkinson's disease prediction using gene expression – A projection based learning meta-cognitive neural classifier approach. Expert Syst Appl 2013;40(5):1519–29.
[55] Peterek T, Dohnalek P, Gajdos P, Smondrk M. Performance evaluation of Random Forest regression model in tracking Parkinson's disease progress. 2013 13th International Conference on Hybrid Intelligent Systems (HIS); 2013. pp. 83–7.
[56] Hariharan M, Polat K, Sindhu R. A new hybrid intelligent system for accurate detection of Parkinson's disease. Comput Methods Programs Biomed 2014;113(3):904–13.
[57] Froelich W, Wrobel K, Porwik P. Diagnosis of Parkinson's disease using speech samples and threshold-based classification. J Med Imaging Health Inform 2015;5(6):1358–63.
[58] Buza K, Varga NÁ. ParkinsoNET: estimation of UPDRS score using hubness-aware feedforward neural networks. Appl Artif Intell 2016;30(6):541–55.
[59] Al-Fatlawi AH, Jabardi MH, Ling SH. Efficient diagnosis system for Parkinson's disease using deep belief network. 2016 IEEE Congress on Evolutionary Computation (CEC). IEEE; 2016. p. 1324–30.
[60] Jain S, Shetty S. Improving accuracy in noninvasive telemonitoring of progression of Parkinson'S Disease using two-step predictive model. 2016 Third International Conference on Electrical, Electronics, Computer Engineering and their Applications (EECEA). IEEE; 2016. p. 104–9.
[61] Behroozi M, Sami A. A multiple-classifier framework for Parkinson's disease detection based on various vocal tests. Int J Telemed Appl 2016.
[62] Avci D, Dogantekin A. An expert diagnosis system for Parkinson disease based on genetic algorithm-wavelet kernel-extreme learning machine. Parkinson's Dis 2016.

Please cite this article in press as: Nilashi M, et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybern Biomed Eng (2017), http://dx.doi.org/10.1016/j.bbe.2017.09.002