Das, R.: A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with…

Article *in* Expert Systems with Applications · March 2010

DOI: 10.1016/j.eswa.2009.06.040 · Source: DBLP

1 author:

Resul Das
Firat University
6 PUBLICATIONS **367** CITATIONS

Expert Systems with Applications 37 (2010) 1568–1572

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A comparison of multiple classification methods for diagnosis of Parkinson disease

Resul Das

Department of Informatics, Firat University, 23119 Elazig, Turkey

ARTICLE INFO

Keywords:
Classification methods
Parkinson disease
Neural networks
Regression
Decision tree
SAS base software

ABSTRACT

In this paper, different types of classification methods are compared for effective diagnosis of Parkinson's diseases. The reliable diagnosis of Parkinson's disease is notoriously difficult to achieve with misdiagnosis reported to be as high as 25% of cases. The approaches described in this paper purpose to efficiently distinguish healthy individuals. Four independent classification schemas were applied and a comparative study was carried out. These are Neural Networks, DMneural, Regression and Decision Tree respectively. Various evaluation methods were employed for calculating the performance score of the classifiers. According to the application scores, neural networks classifier yields the best results. The overall classification score for neural network is 92.9%. Moreover, we compared our results with the result that was obtained by kernel support vector machines [Singh, N., Pillay, V., & Choonara, Y. E. (2007). Advances in the treatment of Parkinson's disease. *Progress in Neurobiology, 81*, 29–44]. To the best of our knowledge, our correct classification score is the highest so far.

1. Introduction

Parkinson's disease (PD) is a destroy neuropathological status, characterized by progressive neurodegeneration of dopamine (DA) neurons in the substantia nigra pars compacta (Manciocco et al., 2009). Nowadays, Parkinson's disease influences a large part of worldwide population. About 1% of the population over 55 years of age is affected by this disease (Betarbet, Sherer, & Greenamyre, 2002; Wooten, Currie, Bovbjerg, Lee, & Patrie, 2004). The PD represents the second most common neurodegenerative disorder after Alzheimer disease (Beal, 2001).

Recently, there is no medical treatment of PD, although medication is available offering significant alleviation of symptoms, especially at the early stages of the disease (Singh, Pillay, & Choonara, 2007). Many of people with Parkinson's disease will therefore be substantially dependent on clinical intervention. The requisite physical visits to the clinic for monitoring and treatment are difficult for many people with Parkinson's disease. Research has shown that the most important symptoms of PD are dysphonia (Little, McSharry, Hunter, Spielman, & Ramig, 2008) and gait variability (Cho, Chao, Lin, & Chen, 2009). In literature, it is claimed that approximately 90% of people with Parkinson's disease exhibit some form of vocal impairment (Ho, Iansek, Marigliani, Bradshaw, & Gates, 1998; Little et al., 2008). The people with Parkinson's typically display a constellation of vocal symptom that includes impairment in the normal production of vocal sounds, which is dysphonia.

The dysphonia is a broad term that refers to disorders of voice, and it includes any pathological or functional problem with one's voice (Ho et al., 1998). The voice will sound hoarse, strained or effortful. The voice may be difficult to understand because their voice sounds very quiet or is so distinctive that it distracts from the content of what the person is saying. Voice disorders may be due to a physical problem, such as vocal nodules or polyps, which are almost like a callous on the vocal cord, paralysis of the vocal cords because of strokes or after some surgeries, or contact ulcers on the vocal cords. These disorders may also be caused by misuse of the vocal instrument, such as using the voice too high or low in pitch, using the voice too softly or too loudly, or with insufficient breath support, often because of postural problems. Some dysphonias seem like a cross between misuse and something physiological.

Most of the current methods used for evaluating Parkinson's disease (PD) rely heavily on human expertise (Cho et al., 2009). In diagnosis of PD literature, there have been extensive studies of speech measurement for general voice disorders (Boyanov & Hadjitodorov, 1997; Godino-Llorente & Gomez-Vilda, 2004; Hadjitodorov, Boyanov, & Teston, 2000; Hansen, Gavidia-Ceballos, & Kaiser, 1998; Little, McSharry, Moroz, & Roberts, 2006; Little, McSharry, Roberts, Costello, & Moroz, 2007) and PD in particular (Cnockaert et al., 2008; Rahn, Chou, Jiang, & Zhang, 2007). In these studies, the speech sounds produced during standard speech tests are recorded using a microphone, and the recorded speech signals are subsequently analyzed using measurement methods (implemented in software algorithms) designed to detect certain properties of these signals.

Recently, different of novel measurement methods have been proposed to assess dysphonic symptoms. In generally, these

E-mail address: rdas@firat.edu.tr

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.06.040

methods are based on nonlinear dynamical systems theory (Kantz & Schreiber, 1999; Little et al., 2007). These measurements are activated by extensive modelling studies (Jiang & Zhang, 2002) and evidence (Hastie, Tibshirani, & Friedman, 2001) that vocal production is a highly nonlinear dynamical system, and that changes caused by impairments to the vocal organs, muscles and nerves will affect the dynamics of the whole system (Little et al., 2008). Consequently, these changes can be detected by nonlinear time series analysis tools (Little, McSharry, Moroz, & Roberts, 2006), such as correlation dimension and methods for characterizing pseudoperiodic time series (Zhang, Luo, & Small, 2006; Zhang & Small, 2006). In many studies, the randomness and noise are inherent to vocal production is used (Godino-Llorente & Gomez-Vilda, 2004) such as recurrence period density entropy (RPDE) and detrended fluctuation analysis (DFA). These methods have been applied to speech signals for showing the ability to detect general voice disorders (Godino-Llorente & Gomez-Vilda, 2004). Nonetheless, to gain as much reliability as possible, measurement methods should be chosen that are as robust as possible to such uncontrolled and in many cases, uncontrollable variations, which are absolute sound pressure level measurement requires costly calibration equipment and the requisite precision is often difficult to obtain. In (Little et al., 2008), it is introduced a new measure of dysphonia that we dub pitch period entropy (PPE), a robust measure sensitive to observed changes in speech specific to PD for controlling these uncontrollable variations. Methods from statistical learning theory are proposed such as linear discriminant analysis (LDA) and support vector machines (SVM) (Hastie et al., 2001).

In this study, various classifiers have been applied to recognize the PD by using SAS based software. SAS based software streamlines the entire data mining process from data access to model assessment. It supports all necessary tasks within a single, integrated solution while providing the flexibility for efficient collaborations. Four different classifiers are selected and implemented with SAS based software. These are DMNeural, Neural Network, Regression, and Decision tree respectively. Various evaluation methods employed for calculating the efficiency of the classifiers. According to the classification accuracy neural networks classifier yields 92.9% correct classification rate.

2. The application of classification methods using SAS based software

Classification is a form of predictive modelling. It is an important part of data mining because it defines groups within the population. There are many different classification methods. With so many different methods, comparing results to determine the best means of classification for a problem is critical.

It is the purpose of this paper to evaluate various classifiers for recognizing the PD. The selected classifiers, which are illustrated in Fig. 1, are implemented with the SAS base software 9.1.3 (Licence number: 291468) (Das, Turkoglu, & Sengur, 2009a; Das, Turkoglu, & Sengur, 2009b; Internet: http://support.sas.com/documentation/, Accessed 21.11.08).

SAS base software 9.1.3 includes two different programs. These programs are called SAS Enterprise Guide 4.3 and SAS Enterprise Miner 5.2. While SAS Enterprise Guide program 4.3 was used for data pre-processing, SAS Enterprise Miner 5.2 program was used to analyze and recognize the PD by combining several classification methods with model comparison node. As can be seen from the Fig. 1, the performed system is composed of 8 components. The brief description of each component is given in the following.

Parkinsons disease dataset holds the features that are used to characterize healthy persons and patients. As it was mentioned earlier, the database composed of 23 columns and 197 rows. The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. The original study published the feature extraction methods for general voice disorders.

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD. The data is in ASCII CSV format. The rows of the CSV file contain an instance corresponding to one voice recording. There are around six recordings per patient; the name of the patient is identified in the first column.

Variable selection component was used in reducing the number of inputs by setting the status of the input variables that are not related to the target as rejected. Although rejected variables were passed to subsequent nodes in the process flow, these variables were not used as model inputs by a successor modelling node. The purpose of the variable selection node is to perform a variable selection procedure in determining the best set of input variables for the predictive model from a pool of all possible input variables that best predicts the variability of the unique target variable from the training data set (Das et al., 2009a).

Data partition component was used to randomly partition the input data into train and validation data sets. Partitioning provides mutually exclusive data sets. Two or more mutually exclusive data sets share no observations with each other. Partitioning the input data reduces the computation time of preliminary modelling runs.

Regression node is one of the few modelling nodes that have a variable selection routine within the node along with various modelling selection criteria. The purpose of the regression node in enterprise miner is to perform either linear regression modelling or least-squares modelling. The node will allow you to perform forward, backward or stepwise regression. After variable selection routine terminates, the model that optimizes the modelling selection criteria from the validation data set is selected as the final model.

DMNeural node is to perform both principal component analysis and DMneural network training. Principal component modelling is based on constructing an independent linear of input variables in which the coefficients capture the maximum amount of variability in the data. DMneural network training is an additive nonlinear model that uses a set of principal components that are placed into separate buckets to predict either a binary-valued or interval valued target variable.

Neural network node was used to classify the feature space. Three independent neural networks models were used to construct this component. There are many types of neural networks architectures; however, multi-layer feed-forward neural network is the most widely used for prediction. A multi-layer feed-forward neural network typically has an input layer, an output layer, and one or more hidden layers. In multi-layer feed-forward networks, neurons are arranged in layers and there is a connection among the neurons of other layers. The inputs are applied to the input layer the output layer contributes to the output directly. Other layers between input and output layers are called hidden layers. Inputs are propagated in gradually modified form in the forward direction, finally reaching the output layer. The back propagation learning algorithm has been used in the feed-forward, single hidden layer neural network. The variants of the algorithm used in the study are the Levenberg–Marquardt (LM), scaled conjugate gradient (SCG) and Pola–Ribiere conjugate gradient (CGP) algorithms. A tangent sigmoid transfer function has been used for both the hidden layer

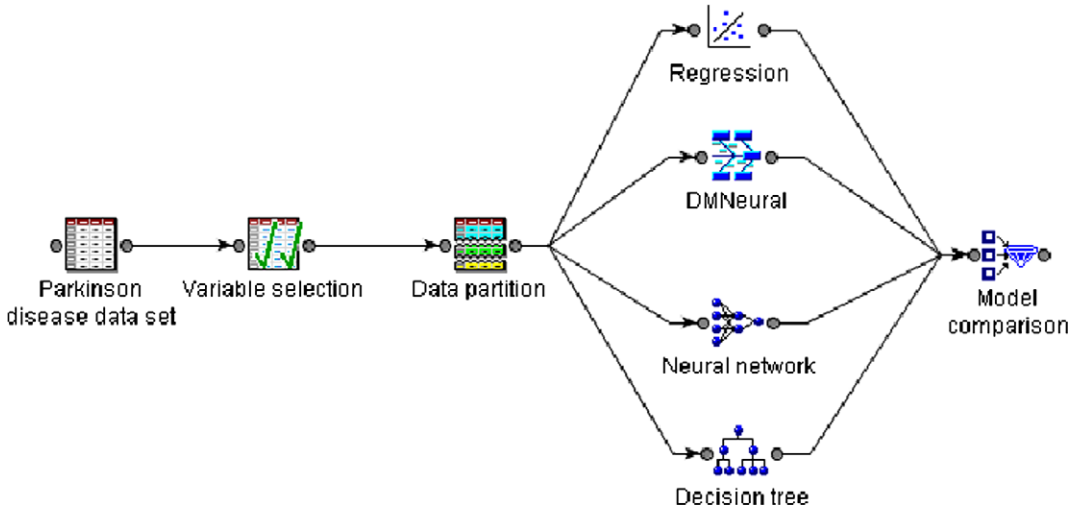

Fig. 1. The applied methods for PD recognition.

and the output layer. The initial weights were chosen randomly (Das et al., 2009b).

Decision tree node provides a completely different approach to classification. A decision tree develops a series of if-then rules. Each rule assigns an observation to one segment of the tree, at which point another if-then rule is applied. The initial segment, containing the entire data set, is the root node for the decision tree. Unlike neural networks and regression, decision trees do not work with interval data. Decision trees work with nominal outcomes that have more than two possible results and with ordinal outcome variables.

Model comparison node uses the expected profits and actual profits to produce standard charts and tables that describe the usefulness of the model that was used to create the scored data set (Internet: http://support.sas.com/documentation/, Accessed 21.11.08). Model comparison node allows comparing models more directly. We were used to compare four models. We were connected the models to the Model Comparison node, as shown in output Fig. 1. Visual representations to compare models are provided with the receiver operating curve (ROC) and the cumulative lift function. In addition, fit statistics allow making a more direct comparison between models.

3. Application results

The implementation schema is illustrated in Fig. 1. The input dataset was randomly partitioned into train and test dataset. 65% of the input dataset was used for training and the rest of the dataset was used for testing. The adjustable parameters of each classifier were tuned. For neural networks classifier, the following adjustments were carried out: The backpropagation learning algorithm has been used in the feed-forward, single hidden layer neural network. The algorithm used in the study is the Levenberg–Marquardt (LM) algorithms. A tangent sigmoid transfer function has been used for both the hidden layer and the output layer. We used 10 neurons in the hidden layer. The initial weights were chosen randomly. In regression node, logistic regression was used. Moreover, default values were chosen for both DMneural and Decision tree nodes.

Various performance evaluation tests were used for calculating the score of the applied classifiers. Table 1 gives the classification accuracies for each classifier. Neural networks classifier obtained higher classification score. 100% correct classification rate was obtained at the training stage of the neural networks classifier. Second best score was obtained by regression node. 88.6% accuracy was conducted for regression node. On the other hand regression node's training performance was the worse than other classifiers. DMneural node and Decision tree node produced the same correct classification score. But Decision tree produced much more correct classification rate at the training stage.

Table 1
Model classification rates.

Method	Classification training (%)	Classification testing (%)
Neural network	100	92.9
DMNeural	89.6	84.3
Regression	89	88.6
Decision tree	93.6	84.3

The ROC maps the sensitivity of the test against 1-specificity and indicates the overall accuracy of the model. Sensitivity is the proportion of target values that are predicted as a value of 1 and are actually equal to 1. Specificity is the proportion of observations that are predicted as a value of 0 and are actually equal to 0. The values 1 and 0 can represent the levels of a target variable with only two possible outcomes. The result for the training set appears the most accurate, with the area under the curve nearly equal to 1. Unfortunately, the ROC for the testing set shows less accuracy (Cerrito, 2006).

Besides correct classification rates, ROC representations of each classifier were given in Fig. 2. Each colour[1] represents different classifiers. As can be seen from the ROC curves of training stage, the neural networks classifier obtained the highest correct classification score. Because the area under ROC curves of neural network classifier was the biggest. Moreover, the best result was obtained by neural networks at the test stage.

Cumulative lift is another performance indication graph that SAS based software represents us. As shown in Fig. 3, the cumulative lifts appear similar when comparing the four different classification models. For the test set, the DMNeural node provides the better model; for the testing sets, neural network model is the better predictive model. The cumulative lift reduces quickly, indicating a poor model fit. In the testing set, the neural network result has a highest cumulative lift, indicating a best predictive model.

[1] For interpretation of colour in Fig. 2, the reader is referred to the web version of this article.

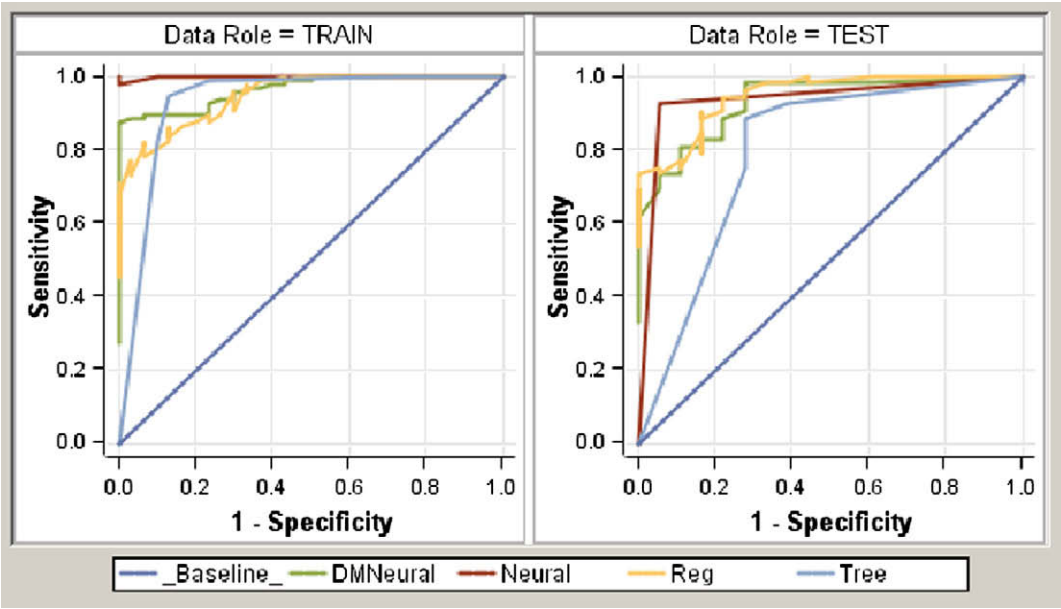

Fig. 2. ROCs for the classification models.

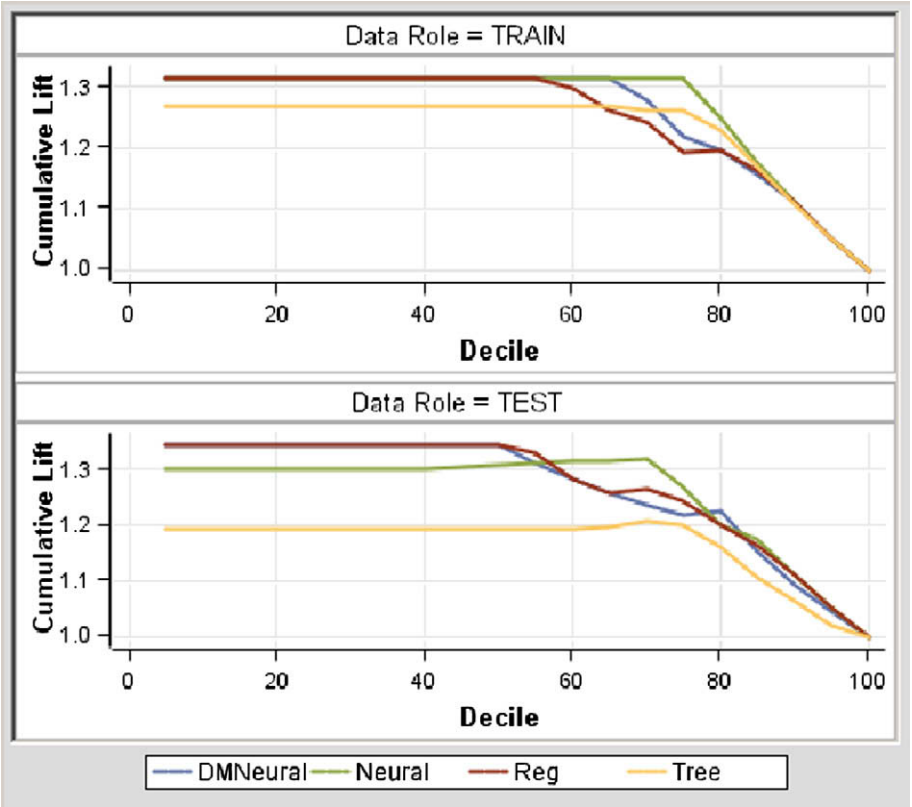

Fig. 3. Cumulative lift curves from model comparison.

We also compare our application results with the result that was presented in reference Singh et al. (2007). In reference, Little et al. proposed a kernel support vector machines for recognition of PD. It is reported that by using kernel support vector machines 91.4% correct classification rate was obtained. Thus, the classifier result that was produced by SAS based software is the best score to our knowledges.

4. Conclusions

So far, several studies have been reported focusing on PD diagnosis. In these studies different methods were applied to the given problems. In this study, various classification methods were used for diagnosing of the Parkinson's disease by using SAS base software. Implementations were carried out on the PD dataset to diag-

nose Parkinson's disease in a fully automatic manner. Four independent classification models were used. These are DMNeural, Neural Network, Regression, and Decision tree respectively. Various evaluation schemas were employed for calculating the performance score of the classifiers. Moreover a comparative study was aimed indeed. The neural network classifier yielded the best score. The experimental results gained 92.9% classification accuracy for neural networks. Another comparative study was also conducted. We also compared our score with the score that was obtained with kernel support vector machines.

References

Beal, M. F. (2001). Experimental models of Parkinson's disease. *Nature Reviews Neuroscience, 2*, 325–334.

Betarbet, R., Sherer, T. B., & Greenamyre, J. T. (2002). Animal models of Parkinson's disease. *Bioessays, 24*, 308–318.

Boyanov, B., & Hadjitodorov, S. (1997). Acoustic analysis of pathological voices. *IEEE Engineering in Medicine and Biology Magazine, 16*, 4–82.

Cerrito, P. B. (2006). *Introduction to data mining using SAS enterprise miner*. Cary, NC, USA: SAS press, SAS Institute Inc..

Cho, C.-W., Chao, W.-H., Lin, S.-H., & Chen, Y.-Y. (2009). A vision-based analysis system for gait recognition in patients with Parkinson's disease. *Expert Systems with Applications, 36*(3), 7033–7039.

Cnockaert, L., Schoentgen, J., Auzou, P., Ozsancak, C., Defebvre, L., & Grenez, F. (2008). Low-frequency vocal modulations in vowels produced by Parkinsonian subjects. *Speech Communication, 50*, 288–300.

Das, R., Turkoglu, I., & Sengur, A. (2009a). Diagnosis of valvular heart disease through neural networks ensembles. *Computer Methods and Programs in Biomedicine, 93*(2), 185–191.

Das, R., Turkoglu, I., & Sengur, A. (2009b). Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications, 36*(4), 7675–7680.

Godino-Llorente, J. I., & Gomez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering, 51*, 380–384.

Hadjitodorov, S., Boyanov, B., & Teston, B. (2000). Laryngeal pathology detection by means of class-specific neural maps. *IEEE Transactions on Information Technology in Biomedicine, 4*, 68–73.

Hansen, J. H. L., Gavidia-Ceballos, L., & Kaiser, J. F. (1998). A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment. *IEEE Transactions on Biomedical Engineering, 45*, 300–313.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: data mining, inference, and prediction: With 200 full-color illustrations*. New York: Springer.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology, 11*, 131–137.

Jiang, J. J., & Zhang, Y. (2002). Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds. *Journal of the Acoustical Society of America, 112*, 2127–2133.

Kantz, H., & Schreiber, T. (1999). *Nonlinear time series analysis* (New ed.). Cambridge, New York: Cambridge University Press.

Little, M. A. (2007). *Biomechanically informed nonlinear speech signal processing*. D.Phil. Thesis, University of Oxford, Oxford.

Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*.

Little, M., McSharry, P., Moroz, I., & Roberts, S. (2006). Testing the assumptions of linear prediction analysis in normal vowels. *Journal of the Acoustical Society of America, 119*, 549–558.

Little, M., McSharry, P., Moroz, I., & Roberts, S. (2006). Nonlinear biophysically-informed speech pathology detection. In *Proceedings of the ICASSP 2006*. New York: IEEE Publishers.

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*.

Manciocco, A., Chiarotti, F., Vitale, A., Calamandrei, G., Laviola, G., & Alleva, E. (2009). The application of Russell and Burch 3R principle in rodent models of Neurodegenerative disease: The case of Parkinson's disease. *Neuroscience and Biobehavioral Reviews, 33*, 18–32.

Rahn, D. A., Chou, M., Jiang, J. J., & Zhang, Y. (2007). Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis. *Journal of Voice, 21*, 64–71.

Singh, N., Pillay, V., & Choonara, Y. E. (2007). Advances in the treatment of Parkinson's disease. *Progress in Neurobiology, 81*, 29–44.

Wooten, G. F., Currie, L. J., Bovbjerg, V. E., Lee, J. K., & Patrie, J. (2004). Are men at greater risk for Parkinson's disease than women? *Journal of Neurology Neurosurgery and Psychiatry, 75*, 637–639.

Zhang, J., Luo, X., & Small, M. (2006). Detecting chaos in pseudoperiodic time series without embedding. *Physical Review E, 73*.

Zhang, J., & Small, M. (2006). Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters, 96*.