NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Khan, Maryam Mahsal; Chalup, Stephan K.; Mendes, Alexandre "Parkinson's disease data classification using evolvable wavelet neural networks". Originally published in Artificial Life and Computational Intelligence: Second Australasian Conference on Artificial Life and Computational Intelligence (presented in Lecture Notes in Computer Science, Vol. 9592) (Canberra, A.C.T. 2-5 September, 2016) p. 113-124 (2016)

Available from: http://dx.doi.org/10.1007/978-3-319-28270-1_10

The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-28270-1_10

Accessed from: <http://hdl.handle.net/1959.13/1324055>

Parkinson's Disease Data Classification Using Evolvable Wavelet Neural Networks

Maryam Mahsal Khan, Stephan K. Chalup, Alexandre Mendes

School of Electrical Engineering & Computer Science,
The University of Newcastle, Australia

{maryammahsal.khan@uon.edu.au, stephan.chalup@newcastle.edu.au, alexandre.mendes@newcastle.edu.au}

Abstract. Parkinson's Disease is the second most common neurological condition in Australia. This paper develops and compares a new type of Wavelet Neural Network that is evolved via Cartesian Genetic Programming for classifying Parkinson's Disease data based on speech signals. The classifier is trained using 10-fold and leave-one-subject-out cross validation testing strategies. The results indicate that the proposed algorithm can find high quality solutions and the associated features without requiring a separate feature pruning preprocessing step. The technique aims to become part of a future support tool for specialists in the early diagnosis of the disease reducing misdiagnosis and cost of treatment.

Keywords: Parkinson's Disease, Neuroevolution, Wavelet Neural Network, Cartesian Genetic Programming, Artificial Neural Network

1 Introduction

Parkinson's Disease (PD) is a central nervous system disorder that is chronic, progressive and incurable. The main symptoms of PD are rigidity, bradykinesia, tremor and postural instability. In addition to those, other symptoms include impaired speech; difficulty in chewing and swallowing; and urinary and constipation problems. The treatment of the disease is very costly, with the average lifetime financial cost for a patient living with PD for 12 years being \$144,000. For comparison purposes, that is similar to the lifetime cost of treatment for a cancer patient (\$165,000) [1].

Decision support systems play an important role in the field of health care. Given the huge volume of information present in medical data repositories, data mining solutions are frequently employed to extract relevant and significant information. The field of Artificial and Computational Intelligence (artificial neural networks, expert systems, fuzzy logic, evolutionary algorithms, genetic programming, etc) can thus provide a second opinion to the medical experts; reducing the long cycle it takes to diagnose the disease. There are several studies in the scientific literature specifically on PD diagnosis. It has been estimated that about 90% of individuals [2] with Parkinson's Disease have some kind of vocal disorder,

known as *dysphonia*. That high incidence is the main reason why this research focuses on biomedical voice measurements, or speech signals.

In this study we aim to improve our understanding on how the Wavelet Neural Network functions thus making it effective for data classification tasks. The neuroevolutionary algorithm has already been applied on classifying digital mammograms as benign or malignant [3]. The current research complements the authors' previous research by providing an extended study on the PD dataset. As a result we have more comprehensive performance results, a better explanation and understanding of how the method can be utilised in critical applications and how robust it is if the data set or application domain vary. For this particular case study, the long term aim is to obtain a non-invasive test that is highly sensitive to early signs of PD indicative features in speech signals. In the next section, we describe the features present in the PD dataset, and review some of the previous studies that have used this relatively challenging data. Then a novel neuroevolutionary algorithm, a Wavelet Neural Network evolved via Cartesian Genetic Programming (CGPWNN), is discussed and explained with examples. Section 3 presents the methodology; followed by an analysis of the results, in Section 4. Section 5 concludes with the main findings and an outlook to future work.

2 Background

2.1 Parkinson's Disease: Dataset and Features

The dataset used in this case study was acquired from an online machine learning database repository in the University of California at Irvine (UCI) [4–6]. It is a challenging dataset that has been investigated previously by several researchers [7–9]. The database is composed of 195 rows (samples), each with 22 different biomedical voice measurements. These voice measurements were taken from 31 individuals, where 23 had Parkinson's Disease (PD). Each patient had between 6 and 7 records in the data set, totalling 195 samples as mentioned before. The main objective is to classify healthy from unhealthy subjects, according to the status attribute (which is 0 for healthy and 1 for PD individuals). The attributes and their description are listed in Table 1 for reference.

2.2 Literature Survey

The PD dataset used in this research was originally published by Little et al. [10] in 2009. Their study was divided into three main sections: 1) Feature acquisition and calculation; 2) Feature pruning, where 10 non-correlated features were identified; and 3) Use of the pruned feature set and its different combinations (via exhaustive search) on a Support Vector Machine with a Gaussian-based radial kernel under bootstrap re-sampling. The maximum average classification accuracy attained was 91.4% when using four features, namely HNR, RDPE, DFA and PPE; description of which can be found in the Table 1.

Table 1. Parkinson’s dataset features and their descriptions [10]

Features	Description
MDVP: $F_0(Hz)$	Average vocal fundamental frequency
MDVP: $F_{hi}(Hz)$	Maximum vocal fundamental frequency
MDVP: $F_{lo}(Hz)$	Minimum vocal fundamental frequency
MDVP: Jitter (%)	Several measures of variation in fundamental frequency
MDVP: Jitter (Abs)	
MDVP: RAP	
MDVP: PPQ	
Jitter: DDP	
MDVP: Shimmer	Several measures of variation in amplitude
MDVP: Shimmer (dB)	
Shimmer: APQ3	
Shimmer: APQ5	
MDVP: APQ	
Shimmer: DDA	Two nonlinear dynamical complexity measures
RPDE	
D2	Signal fractal scaling exponent
DFA	
Spread1	Three nonlinear measures of fundamental frequency variation
Spread2	
PPE	
NHR	Two measures of ratio of noise to tonal components in the voice
HNR	
Status	Healthy / Sick

Okan et al. [7] first pruned the dataset features by applying a mutual information measure, and ranked them based on maximum relevance minimum redundancy criteria ‘mRMR’. They came up with four key features, namely *Spread1*, *MDVP : F_0* , *Shimmer : APQ3* and *D2*, which were then used for building a classification model using an SVM. Using a leave-one-subject-out cross validation strategy they achieved an accuracy of 81.53%. The same strategy when applied to the original data of Little et al. [10] caused the classification accuracy to drop to 65.13%, while using 10-fold cross-validation with their features caused the accuracy to rise to 92.75%.

Akin [8] also pruned the feature set first, via a linear SVM feature selection strategy where 10 features (*spread1*, *MDVP : F_0* , *D2*, *spread2*, *MDVP : F_{hi}* , *MDVP : APQ*, *DFA*, *HNR*, *PPE* and *RPDE*) were selected. These features were then used to train six rotation forest (RF) ensemble classifiers created by using two ANN architectures (*MLP*, *RBF*), two lazy learners (*KSTAR*, *IBk* (*WEKA*¹ implementation of *KNN*)) and two decision trees (*LADTree*, *J48*). A 10-fold cross-validation strategy was adopted. Out of the six RF ensemble classifiers, the classification accuracy of *IBk* was the highest - 96.93%.

Caglar et al. [9] used an Adaptive Neuro Fuzzy Classifier with Linguistic Hedges (*ANFC-LH*) to first select the strong features from the dataset. The technique identified four features of interest: *Spread1*, *MDVP : F_{hi}* , *RPDE* and *D2*. Based on these features the dataset was split into equal parts: 50% for training and 50% for testing. As the dataset contained individuals with 6-7 records each, if the number of ‘true checks’ associated with an individual was

¹ <http://www.cs.waikato.ac.nz/ml/weka/>

high, then it was classified as healthy. Multilayer perceptrons (MLP), radial basis function neural networks (RBFNN) and ANFC-LH were used as classifiers for training. The author found ANFC-LH to yield the best result with 95.38% training and 94.72% testing accuracy.

Pei et al. [11] used a minimum distance classifier to assess genetic programming combined with expectation maximization algorithm. They used a 10-fold cross validation strategy for training and testing the classifiers using all features. The process yielded a training accuracy of 95.06% and a testing accuracy of 93.12%. Farhad et al. [12] used MLP and RBF neural networks from MATLAB's neural network toolbox to predict Parkinson's Disease. Samples were randomly split into 70% training and 30% testing, and all the features were used. He found that MLPs yielded the highest accuracy, at 93.22%.

Chen et al. [13] used a fuzzy k-nearest neighbour (FKNN) approach on a feature reduced dataset based on PCA. He trained the FKNN using the 10-fold cross validation strategy and obtained an accuracy of 96.07%. While Polat [14] used a feature weighting method on the basis of fuzzy C-means clustering. He applied his approach on different benchmark datasets Parkinsons, Iris and Wine from the UCI machine learning repository. He found that weighting the features based on the C-means approach improves the performance of classification. He obtained 97.03% using a KNN classifier with a 50-50% training-testing partition.

Hariharan et al. [15] evaluated the dataset based on a three step process. First a Gaussian mixture model is used to weight the features, which is then ranked and reduced by using different feature selection/reduction strategies (PCA, LDA) and feature subset selection methods (SFS and SBS). The reduced feature space is then classified based on the 10-fold cross validation strategy where three classifiers were investigated LS-SVM, Probabilistic Neural Networks (PNN) and General Regression Neural Networks (GRNN). The author found that LDA and SBS feature reduction methods provided a feature space which was separable for LS-SVM, PNN and GNN classifiers. Their approach obtained an accuracy of 100%.

Spadoto et al. [16] used three evolutionary mechanisms namely Particle Swarm Optimization (PSO), Harmony Search (HS) and Gravitational Search Algorithm (GSA) to select features from the PD dataset. The reduced dataset is then classified via an Optimum-Path Forest (OPF) classifier. They obtained an average of 84.01% with an execution time of 0.16s on a hold-out dataset.

Most of the techniques applied on the Parkinson's dataset surround feature selection or weighting prior to classification, as shown in Table 3. Our intent is to implement an algorithm that does not require pre-processing of the features. The neuroevolution itself will select features that are significant and evolve to produce good quality solutions i.e. feature selection and classification all in one. There is no technique that uses neuroevolution to classify the dataset and thus our technique would be the first to be tested on this dataset. Even though neuroevolution has greater prospects in the reinforcement learning domain, where efficient intelligent controllers are designed, our research focus is to apply it on the supervised learning domain and check its applicability.

2.3 Wavelet Neural Network Evolved via CGP

Wavelet Neural Networks (WNN) represent a class of neural networks consisting of wavelets; i.e. they combine the theory of wavelets and neural networks. It has been found that WNN provide better function approximation ability than standard multilayer perceptrons (MLP) and radial basis function (RBF) neural networks. Wavelet Neural networks are generally feed-forward neural networks, with one hidden layer, whose activation functions are drawn from an orthonormal wavelet family. The common wavelet activation functions are Gaussian, Mexican hat, Morelet and Haar [17].

In this paper, a novel algorithm based on the concept of Cartesian Genetic Programming is used to evolve the wavelet neural network parameters. Cartesian Genetic Programming (CGP) is an evolutionary programming technique developed by Miller et al. [18] and has been particularly used for digital circuit optimization. The concept of CGP has also been used to train artificial neural networks; named Cartesian Genetic Programming Artificial Neural Network (CGPANN) and was proposed by Khan et al. [19]. The motivation behind using CGP for evolving parameters is that CGP doesn't bloat [20], are dominated by redundant genes that have a neutral effect [21–23] on the performance and most of the applications evolved via CGP are found to be generalized, robust and with overall good accuracy [24–27].

Fig. 1. Generalized structure of a CGPWNN genome

The main entity of WNN are the wavelons ' ψ '. Wavelons represent a wavelet function with a predefined scaling ' α ' and dilation ' β ' values. Given an input x_{ij} and wavelet function ψ_j with scaling α and dilation β , the input is transformed into $\psi_j((x_{ij} - \beta)/\alpha)$, where $x_{ij} = [1, \text{Total Inputs}]$, $\alpha = [-1, 1]$ and $\beta = [0, 1]$, $\psi_j = [1, \text{Total wavelet functions}]$. Input, switch, scale and dilate genes occur in pairs, so if there are 2 or more inputs than there would be 2 or more sets of input, switch, scale, dilate genes respectively. Each of the input in the wavelons are connected via a c_{ij} gene known as the switch gene where $c_{ij} = \{0, 1\}$. Each of the wavelons within a network have an associated weight represented as $w_{ij} \in [-1, 1]$. All these genes combine together to form a wavelon structure ' ω '.

The CGP genotype consists of nodes and output genes, whereas the CGPWNN genotype contains wavelons, output and bias ' θ ' genes. There are two

basic types of evolution strategies (μ, λ) -ES and $(\mu + \lambda)$ -ES [28]. μ represents the number of individuals in the parent population and λ refers to the number of offspring produced in a generation. In (μ, λ) -ES offspring replaces the parent as the fittest is selected from λ , while in $(\mu + \lambda)$ -ES the fittest is selected from both parents and offspring for the next generation. Cartesian Genetic programming uses the $(1 + \lambda)$ -ES strategy, a single parent is mutated based on a mutation rate ' τ ' to produce λ offsprings. The fittest of the genotypes becomes the parent and is forwarded to the next generation. α , β and w_{ij} are perturbed with in the range $[0.8v, 1.2v]$ where ' v ' is the parameter value. This captures the continuous aspect of the wavelet networks.

Figure 1 shows a CGPWNN genome structure. The genome consists of four main sections: (1) m-Wavelons ω_m , (2) a bias θ (3) n-outputs O_n and (4) a function gene O_{func} respectively labeled in Figure 1 for better understanding. Figure 2(a) is an example of 2×1 architecture CGPWNN genotype, where 2×1 corresponds to the number of rows and columns, respectively. As WNNs have one hidden layer, the number of columns is fixed to 1, while there are 2 wavelons ω_m along the row. The number of inputs to the network is 2, i.e. x_0 and x_1 , and the number of outputs is 1. The number of inputs to each wavelon structure is also taken as 2. Figure 2(b) shows the random assignment of values to each gene of the wavelons ω_m . Figure 2(c) is the phenotypic representation of the genome in Figure 2(a), along with its mathematical expression.

Fig. 2. (a) An example of a 2×1 CGPWNN genome with 2 inputs and 1 output. (b) Generalised representation of a wavelon ψ structure with random values assigned. (c) Phenotype of the example genome in (a) as it is a one output genome hence usage of a single wavelon only.

3 Experimental Setup

3.1 Training and Testing Sets

Two different training and testing sets were generated in order to investigate the performance of the algorithm.

- **Leave-one-subject-out strategy:** The dataset consists of 31 individuals or subjects where each individual can have 6 or 7 associated records. In the leave-one-subject-out crossvalidation approach not one record was left out for testing but one individual (i.e. all associated 6 or 7 records). The remaining 30 individuals (each with 6 or 7 records) were used for training. This approach is similar to [7] with the exception that no feature pruning was used as pre-processing step.
- **10-fold cross validation strategy:** The training set was composed of the records of 28 individuals while the testing set consisted of records of 3 individuals.

Due to the fact that each individual has between 6 and 7 records in the dataset, if more than half of an individual’s records are classified as PD, then the individual itself is classified as unhealthy (PD), otherwise not. This approach was adopted from [9]. The purpose of this approach is to avoid overfitting as the records of the same patient are potentially very similar.

3.2 Performance Measures

The classifiers are evaluated based on the following performance metrics:

1. **Training accuracy** (Tr_{Acc}): fraction of correctly classified training samples.
2. **Testing accuracy** (Te_{Acc}): fraction of correctly classified testing samples, also known as the classification accuracy. The higher the percentage, the better is the classifier performance.

3.3 CGPWNN Parameters

Random architectures of CGPWNN $[(5 \times 1), (10 \times 1), (15 \times 1)]$ were evolved at different parameter settings with a mutation rate of 0.01% and a (1+25)-ES as shown in Table 2. The wavelet functions used were Gaussian, Morelet, Mexican hat and Haar. Each network was evolved for 100,000 generations. The number of outputs O_p from the genotype were 5, 8, 10 and 15 respectively and the number of inputs to each neuron I_E was set to 6, 11 and 22 respectively. The genotypes were evolved based on two different strategies of training i.e. leave-one-subject-out and 10-fold cross validation strategies. Table 2 shows the result for CGPWNN, which are averaged for all the sets in a training and testing set. Fifty independent evolutionary runs were performed for each dataset within a training strategy. Table 2 shows results of CGPWNN in terms of training accuracy ‘ Tr_{Acc} ’, testing accuracy ‘ Te_{Acc} ’, the percentage of active neurons in the search space ‘ A_{Neu} ’ and the percentage of features selected ‘ F_{sel} ’.

Table 2. Performance of CGPWNN on leave-one-subject-out & 10-fold cross validation strategies. The best configurations are indicated in boldface.

Representation			Leave-one-subject-out				10-fold cross validation			
Structure	I_E	O_p	Tr_{Acc} (%)	$Te_{Acc}(\sigma)$ (%)	A_{Neu} (%)	F_{sel} (%)	Tr_{Acc} (%)	$Te_{Acc}(\sigma)$ (%)	A_{Neu} (%)	F_{sel} (%)
10×1	6		99.39	82.84(21.33)	30.93	26.27	99.32	88.73(15.26)	30.80	26.33
	11	5	98.48	82.00(18.84)	21.02	27.87	98.51	89.27(12.84)	21.44	28.35
	22		96.17	83.10(14.51)	7.72	30.33	95.94	89.53(11.01)	8.04	30.58
	6		99.98	83.03(19.68)	30.36	48.42	99.99	92.60(11.98)	30.64	47.68
	11	10	99.67	83.94(16.69)	26.97	56.03	99.61	92.73(11.04)	26.54	55.71
	22		96.87	82.90(15.14)	15.65	63.12	96.61	91.60(11.00)	15.74	63.93
15×1	6		99.94	85.29 (17.86)	37.85	43.12	99.91	91.93(12.65)	38.38	43.63
	11	8	99.60	82.39(18.88)	32.11	52.10	99.51	90.13(13.28)	31.83	51.85
	22		96.85	82.00(15.65)	19.50	59.55	96.54	90.00(11.95)	19.85	60.40
	6		100.00	84.39(20.50)	35.90	65.31	99.99	91.80(14.10)	36.55	65.78
	11	15	99.91	84.65(18.14)	31.29	74.75	99.91	92.93 (11.77)	32.33	75.46
	22		97.34	83.74(14.97)	24.95	87.72	96.97	91.27(11.49)	24.85	87.68

4 Results and Discussion

Table 2 shows the performance of CGPWNN on the PD dataset classification. CGPWNN obtained a maximum classification accuracy of 85.29% in the leave-one-subject-out strategy and 92.93% when using the 10-fold cross validation strategy respectively. From Table 2 it is observed that with increasing the number of inputs to each wavelon ω_i the training accuracy reduces; no clear trend is observed on the testing accuracy. While with increasing output O_p and constant input I_E the training accuracy increases; also no such trend is observed in the testing accuracy. In the former case, the reduction in the training accuracy is associated with the property of the wavelet neural network i.e. WNN performs better in application of small input dimensions [17]. While in the later case, output gene has a predominant role in improving the accuracy. Output represents either a wavelon(s) or an input multiplied by a weight value or the sum of both. Thus increasing the number of links an output layer can have with the previous layers (input/hidden) consequently improves accuracy.

The percentage of active neurons/wavelons indicates the complexity of the problem investigated. It is observed in Table 2 that the active percentage of neurons/wavelons is quite similar for each structure in both of the training/testing datasets. This indicates that any training /testing split of a dataset would not affect the active percentage of neurons/wavelons in the cartesian space. The percentage of the active neurons is found to be dependent on the number of inputs to each wavelon; where a structure with minimum inputs I_E have higher percentage of active neurons and lower percentage of features selected. This indicates

Fig. 3. Histogram of the average active wavelons for all the genotypes in the simulated 12 structures for both leave-one-individual-out & 10-fold cross validation strategies.

that the neuroevolutionary algorithm is able to find a solution with minimum features that are transformed by the wavelons in the structure. We can also observe a clear ascending trend in the percentage of features selected when the number of inputs to a wavelon I_E and the output O_p is increased. This clearly indicates that the feature percentage contribution increases two-fold when both the number of links and the wavelon’s input connectivity are increased.

Figure 3 shows the histogram of the number of wavelet activation functions utilized by the best solutions for all the simulated structures. It was found that for both the training strategies the behaviour of selection was the same. Mostly Gaussian and leastly Haar activation function were utilized. Future research shall also be focused on observing the behaviour of the genotypes when using a single activation function.

Figure 4 shows the frequency of selection of the 22 features of CGPWNN genotypes for both training strategies. The histogram represents the sum of features used for all the genotypes within each structure thus providing an average of 50×12 evolutionary runs; where 50 are the number of genotypes and 12 are the simulated structures. We can clearly see that out of 22 features, 4 of them were predominantly used in both the training strategies. They were PPE, Spread1, Spread2 and D2. The selected features are found to be a subset of features in most of the feature selection algorithms used by researchers in literature [10, 8]. This indicates that the feature selection process is an independent process and the presence of any features in the feature set doesn’t affect the performance of the classifier.

Table 3 shows the comparison of the neuroevolutionary algorithm with other techniques using 10-fold cross validation and leave-one-subject-out strategy as training sets. We can clearly see that CGPWNN performed competitively even though no feature pre-processing step was adopted. This indicates its potential to classify vocal features for PD diagnosis.

5 Conclusion

In this research, a novel neuroevolutionary algorithm - WNN evolved via CGP (CGPWNN) was tested to verify its effectiveness towards an accurate classifica-

Fig. 4. Histogram of the average selected features for all the genotypes in the simulated 12 structures for both leave-one-individual-out & 10-fold cross validation strategies.

Table 3. Comparison of techniques using leave-one-subject-out & 10-fold cross validation strategies in Parkinson’s Disease classification

Algorithm	Training acc. (%)	Testing acc. (%)	Reference
Leave-one-subject-out			
mRMR+SVM	—	81.53 ± 2.17	[7]
CGPWNN	99.94	85.29	
10-fold cross validation			
Preselection filter+exhaustive search+SVM	—	91.4 ± 4.4	[10]
mRMR+SVM	—	92.75 ± 1.21	[7]
SVM+MLP-ensemble	—	90.8	[8]
SVM+RBF-ensemble	—	88.71	[8]
SVM+LADTree-ensemble	—	92.82	[8]
SVM+J48-ensemble	—	92.3	[8]
SVM+KSTAR-ensemble	—	96.41	[8]
SVM+IBk-ensemble	—	96.93	[8]
GP-EM	95.06	93.12	[11]
PCA+FKNN	—	96.07	[13]
GMM+LDA+LS-SVM or PNN or GRNN	—	100	[15]
CGPWNN	99.90	92.93	

tion system for PD diagnosis. The dataset contained features based on biomedical voice measurements of both PD patients and healthy individuals.

It was found that CGPWNNs are successful in locating high-quality solutions; and are not dependent on a ‘*feature selection process*’ as a pre-processing step before training occurs. CGPWNN obtained an accuracy of 92.93% on a 10-fold cross validation dataset, performing competitively with other techniques presented in literature. Therefore, vocal features have a clear prospect in the diagnosis of PD.

CGPWNN selected four dominant features from the feature set more frequently than the rest, which illustrates their importance in the classification process. The four features identified were: Spread1, Spread2, D2 and PPE which are found to be subsets of features selected in other feature selective algorithms reported in literature.

6 Acknowledgments

We acknowledge Max Little, from the University of Oxford, UK, who created the database in collaboration with the National Center for Voice and Speech, Denver, Colorado, USA, who recorded the speech signals.

References

1. Parkinson's Australia, Living with Parkinson's disease update - October 2011; De-loitte Access Economics Pty Ltd. <http://www.parkinsonsnsw.org.au/>, accessed August 2014.
2. R. Das. A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, 37(2):1568–1572, 2010.
3. M.M Khan, S.K. Chalup, and A. Mendes. Evolving wavelet neural networks for breast cancer classification. In *12th Australasian Data Mining Conference*, Brisbane, Australia, 27th-28th November 2014.
4. M. Lichman. UCI machine learning repository. <http://archive.ics.uci.edu/ml>.
5. UCI machine learning repository - Parkinsons data set. <https://archive.ics.uci.edu/ml/datasets/Parkinsons.html>, accessed May 2014.
6. M.A. Little, P.E. McSharry, Roberts S.J., D.A.E. Costello, and I.M. Moroz. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering Online*, 6(23), 2007.
7. C.O. Sakar and O. Kursun. Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of Medical Systems*, 34(4):591–599, 2010.
8. A. Ozcift. Svm feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. *Journal of Medical Systems*, 36:2141–2147, 2012.
9. M.F. Caglar, B. Cetisli, and I.B. Toprak. Automatic recognition of Parkinson's disease from sustained phonation tests using ann and adaptive neuro-fuzzy classifier. *Journal of Engineering Science and Design*, 2:59–64, 2010.
10. M.A. Little, P.E. McSharry, E.J. Hunter, J.L. Spielman, and L.O. Ramig. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4):1015–1022, 2009.
11. P.F. Guo, P. Bhattacharya, and N. Kharma. Advances in detecting Parkinson's disease. In *Lecture Notes in Computer Science (LNCS)*, volume 6165, pages 306–314. Springer, 2010.
12. F.S. Gharehchopogh and P. Mohammadi. A case study of Parkinson's disease diagnosis using artificial neural networks. *International Journal of Computer Applications*, 73:1–6, July 2013.
13. Hui-Ling Chen, Chang-Cheng Huang, Xin-Gang Yu, Xin Xu, Xin Sun, Gang Wang, and Su-Jing Wang. An efficient diagnosis system for detection of parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst. Appl.*, 40(1):263–271, January 2013.
14. Kemal Polat. Classification of parkinson's disease using feature weighting method on the basis of fuzzy c-means clustering. *International Journal of Systems Science*, 43(4):597–609, 2012.
15. M. Hariharan, Kemal Polat, and R. Sindhu. A new hybrid intelligent system for accurate detection of parkinson's disease. *Computer Methods and Programs in Biomedicine*, 113(3):904 – 913, 2014.

16. A.A. Spadoto, R.C. Guido, F.L. Carnevali, A.F. Pagnin, A.X. Falcao, and J.P. Papa. Improving parkinson's disease identification through evolutionary-based feature selection. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 7857–7860, Aug 2011.
17. A.K. Alexandridis and A.D. Zaprakis. Wavelet neural networks: A practical guide. *Neural Networks*, 42:1–27, June 2013.
18. J.F. Miller and P. Thomson. Cartesian genetic programming. In *European Conference on Genetic Programming, Lecture Notes in Computer Science (LNCS)*, volume 1802, pages 121–132. Springer-Verlag, 2000.
19. M.M. Khan, G.M. Khan, A.M. Ahmad, and J.F. Miller. Fast learning neural networks using cartesian genetic programming. *Neurocomputing*, 121:274–289, 2013.
20. Julian Miller. What bloat? cartesian genetic programming on boolean problems. In *Genetic and Evolutionary computation conference - Late Breaking Papers*, pages 295–302, 2001.
21. V.K. Vassilev and J.F. Miller. The advantages of landscape neutrality in digital circuit evolution. In *International Conference on Evolvable Systems, Lecture Notes in Computer Science (LNCS)*, volume 1801, pages 252–263. Springer, 2000.
22. T. Yu and J.F. Miller. Neutrality and the evolvability of boolean function landscape. In *European Conference on Genetic Programming, Lecture Notes in Computer Science (LNCS)*, volume 2038, pages 204–217. Springer, 2001.
23. T. Yu and J.F. Miller. Finding needles in haystacks is not hard with neutrality. In *European Conference on Genetic Programming, Lecture Notes in Computer Science (LNCS)*, volume 2278, pages 13–25. Springer, 2002.
24. M.M. Khan, G.M. Khan, and J.F. Miller. Efficient representation of recurrent neural networks for markovian/non-markovian non-linear control problems. In *International Conference on System Design and Applications (ISDA)*, pages 615–620, 2010.
25. A. Walker and J. F. Miller. Solving real-valued optimisation problems using cartesian genetic programming. In *Proceedings of the Genetic and Evolutionary Computation Conference GECCO*, 2007.
26. J. A. Walker and J. F. Miller. Predicting prime numbers using cartesian genetic programming. In *Proceedings of 10th European Conference on Genetic Programming EuroGP*, volume LNCS 4445, pages 205–216. Springer, 2007.
27. J. A. Walker and J. F. Miller. Changing the genospace: Solving ga problems using cartesian genetic programming. In *Proceedings of 10th European Conference on Genetic Programming EuroGP*, volume LNCS 4445, pages 261–270. Springer, 2007.
28. H.G. Beyer and H.P. Schwefel. Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1):3–52, May 2002.