Neurocomputing 184 (2016) 131–144

Contents lists available at ScienceDirect

Neurocomputing

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

# An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease

Hui-Ling Chen [a,f], Gang Wang [b], Chao Ma [c], Zhen-Nao Cai [d,a], Wen-Bin Liu [a], Su-Jing Wang [e,f,*]

[a] College of Physics and Electronic Information, Wenzhou University, Wenzhou, Zhejiang 325035, China
[b] College of Computer Science and Technology, Jilin University, Changchun 130012, China
[c] School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China
[d] School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[e] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[f] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, China

ARTICLE INFO

*Article history:*
Received 2 January 2015
Received in revised form
11 May 2015
Accepted 14 July 2015
Available online 14 December 2015

*Keywords:*
Kernel extreme learning machine
Feature selection
Medical diagnosis
Parkinson's disease

ABSTRACT

In this paper, we explore the potential of extreme learning machine (ELM) and kernel ELM (KELM) for early diagnosis of Parkinson's disease (PD). In the proposed method, the key parameters including the number of hidden neuron and type of activation function in ELM, and the constant parameter $C$ and kernel parameter $\gamma$ in KELM are investigated in detail. With the obtained optimal parameters, ELM and KELM manage to train the optimal predictive models for PD diagnosis. In order to further improve the performance of ELM and KELM models, feature selection techniques are implemented prior to the construction of the classification models. The effectiveness of the proposed method has been rigorously evaluated against the PD data set in terms of classification accuracy, sensitivity, specificity and the area under the ROC (receiver operating characteristic) curve (AUC). Compared to the existing methods in previous studies, the proposed method has achieved very promising classification accuracy via 10-fold cross-validation (CV) analysis, with the highest accuracy of 96.47% and average accuracy of 95.97% over 10 runs of 10-fold CV.



## 1. Introduction

Parkinson's disease (PD) is named after James Parkinson, who published the first paper describing this disease in 1817. Now PD has become the second most common degenerative disorders of the central nervous system after Alzheimer's disease [1]. It breaks out in large part of the world with fast rate, and the disease prevalence is expected to increase dramatically as the population ages [2], which might be particularly serious for developing countries such as China or India [3]. Till now, the cause of PD is still unknown, however, it is reported to be possible to alleviate symptoms significantly at the onset of the illness in the early stage [4]. Patients with PD are usually characterized by five symptoms including tremor, rigidity, bradykinesia or slowness of movement, hand asymmetry and posture instability [5,6]. Research has shown that approximately 90% of the patients with PD show vocal disorders [7]. It has also been proven that a vocal disorder may be one of the first symptoms to appear nearly 5 years before clinical diagnose [8]. The vocal impairment symptoms related with PD are known as dysphonia (inability to produce normal vocal sounds) and dysarthria (difficulty in pronouncing words) [9]. Little et al. [10] has made use of the dysphonic indicators in their study to help discriminate PD patients from healthy ones. In their study, Support Vector Machine (SVM) with Gaussian kernel functions in combination with the feature selection approach was taken to predict PD, the simulation results have demonstrated that the proposed method can discriminate PD patients from healthy ones with approximately 90% classification accuracy using only four dysphonic features. A more recent study [11] has lifted the accuracy to 93% by increasing the number of features, and the result has been further boosted up to 99% through the use of a group of feature selection algorithms.

Motivated by the pioneer work in [10], many researchers made use of a comprehensive machine learning techniques to handle the PD diagnosis problem. In [12], Das presented a comparative study of using artificial neural networks (ANNs), DMneural, Regression and Decision Tree for effective diagnosis of PD, the experimental results have shown that ANNs yield the best results with the overall classification score of 92.9%. In [13], AStröm and Koker proposed a parallel feed-forward neural network structure for

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2015.07.138
0925-2312/

diagnosis of PD, the highest classification accuracy of 91.20% was obtained. In [14], Sakar and Kursun used the mutual information based feature selection methods integrated with the SVM classifier for PD diagnosis, and the classification accuracy of 92.75% was achieved. In [15], Li et al. proposed a fuzzy-based non-linear transformation method in combination with the SVM classifier for prediction of PD, and the best classification accuracy of 93.47% was achieved. In [16], Shahbaba and Neal introduced a new nonlinear model based on Dirichlet process mixtures for classification of PD, the results have been compared with that of multinomial logit models, decision trees, and SVM, the best classification accuracy of 87.7% was obtained by the proposed approach. In [17], Psorakis et al. introduced novel convergence measures, sample selection strategies and model improvements for multiclass mRVMs, and finally, the improved mRVMs achieved the classification accuracy rate of 89.47% when applied to prediction of PD. In [18], Guo et al. combined genetic programming and the expectation maximization algorithm (GP-EM) to detect PD, and the best classification accuracy of 93.1% was obtained. In [19], Luukka employed the feature selection method based on fuzzy entropy measures together with the similarity classifier to predict PD, and mean classification accuracy of 85.03% with only two features was obtained. In [20], Ozcift and Gulten combined the correlation based feature selection (CFS) algorithm with the RF ensemble classifiers of 30 machine learning algorithms to identify PD, and the best classification accuracy of 87.13% was achieved by the proposed CFS-RF model. In [21], Spadoto et al. applied evolutionary-based techniques in combination with the Optimum-Path Forest (OPF) classifier to detect PD, and the best classification accuracy of 84.01% was achieved. In [22], Polat proposed to integrate the use of fuzzy c-means clustering-based feature weighting (FCMFW) with the k-NN classifier for the detection of PD, the classification accuracy of 97.93% was obtained. In [23], Chen et al. employed the Fuzzy k-nearest neighbor (FKNN) classifier in combination with the principle component analysis (PCA-FKNN) to diagnose PD, and the best classification accuracy of 96.07% was obtained by the proposed diagnosis system. In [24], Zuo et al. presented an effective and efficient diagnosis system based on particle swarm optimization enhanced FKNN for PD diagnosis, and the mean accuracy of 97.47% was reported. In [25], Hariharan et al. Developed a hybrid method by combining several feature pre-processing methods with classification techniques using least-square SVM, probabilistic neural network and general regression neural network, and the best classification accuracy of 100% was reported. In [26], Gök developed a discriminative model by using rotation-forest ensemble k-nearest neighbor classifier algorithm, and the diagnosis accuracy of 98.46% was achieved.

From the above works, we can see that ANNs and SVM have gained much more popularity due to their mature theory background as well as the satisfactory classification performance. The main advantages of ANNs are their outstanding capability of capturing the nonlinearity relationship between the input and output existed in the data. However, it should be noted that the traditional gradient descent based training algorithm such as back propagation method may be easily trapped in the local minima as well as leaving many network parameters to be specified. Recently, Huang et al. proposed a new learning algorithm, extreme learning machine (ELM) [27], for a single hidden layer feed-forward neural networks (SLFNs). ELM chooses input weights and hidden biases randomly, and the output weights are analytically determined by using Moore–Penrose (MP) generalized inverse. However, one drawback of ELM is that the randomly assigned weights can produce a large variation in the classification accuracy in different trials. In order to solve this problem, more recently Huang et al. [28] proposed the kernelized version of ELM (KELM), which requiring no randomness in assigning connection weights between input and hidden layers. Compared with SVM, KELM can achieve comparative or better performance with much easier implementation and faster training speed in many classification or regression tasks [28–30].

Motivated by the excellent performance achieved by the ELM or KELM classifier on the disease diagnosis problems such as thyroid disease diagnosis [31], erythemato-squamous diseases diagnosis [32] and paraquat-poisoned patients diagnosis [33], in this study, an attempt was made to explore the potential of ELM and KELM in constructing an automatic diagnostic system for diagnosis of PD. Previous studies [10,14,15,19,23] on PD diagnosis have proven that using dimension reduction before conducting the classification task can improve the diagnosis accuracy. Here, an attempt is made to diagnose PD by using the ELM and KELM classifiers in combination with the feature selection methods. Four common feature selection techniques including maximum relevance minimum redundancy (mRMR), Information Gain (IG), Relief and *t*-test are employed for pre-processing before the classification models are constructed. The effectiveness of the proposed hybrid method is examined in terms of the classification accuracy, sensitivity, specificity and AUC on the PD data set taken from UCI machine learning repository. Promisingly, as can be seen that the developed method for this dataset in which a more reliable result is found (96.47% highest accuracy) over 10 runs of 10-fold cross validation (CV).

In summary, the main contributions of this paper can be summarized as follows: (1) The potential of ELM and KELM are explored in constructing an automatic diagnostic system for diagnosis of PD; (2) The detailed investigation on the impact of feature selection to the classification performance of PD diagnosis and interesting discovery are presented; (3) The most relevant measurement has been identified with the aid of the feature selection method.

The remainder of this paper is organized as follows. Section 2 offers brief background knowledge on ELM and KELM. In Section 3 the detailed implementation of the proposed method is presented. Section 4 describes the experimental design. The experimental results and discussions of the proposed approach are presented in Section 5. Finally, conclusions and recommendations for future work are summarized in Section 6.

## 2. Background materials

### 2.1. ELM and KELM

This section gives a brief description of ELM. For more details, one can refer to [27,34]. Given a training set $\aleph = \{(x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, 2, \ldots, N\}$, where $x_i$ is the $n \times 1$ input feature vector and $t_i$ is a $m \times 1$ target vector. The standard SLFNs which has an activation function $g(x)$, and the number of hidden neurons $\tilde{N}$ can be mathematically modeled as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, 2, \ldots, N \tag{1}$$

where $w_i$ is the weight vector between the $i$th neuron in the hidden layer and the input layer, $b_i$ means the bias of the $i$th neuron in the hidden layer; $\beta_i$ is the weight vector between the $i$th hidden neuron and the output layer; and $o_j$ is the target vector of the $j$th input data. Here, $w_i \cdot x_j$ denotes the inner product of $w_i$ and $x_j$.

If SLFNs can approximate these $N$ samples with zero error, we will have $\sum_{j=1}^{N} ||o_j - t_j|| = 0$, i.e., there exists $\beta_i$, $w_i$, $b_i$ such that $\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j, \ j = 1, 2, \ldots, N$. The above equation can be