Original Article

Parkinson's Disease Identification using Reduced Kernel Classification

Ahmed Ali Hussein Alemran

Department of Physiology, Faculty of Medicine, Misan University, Al-Amarah, Iraq

Abstract

Background: Artificial intelligence (AI) has begun to play an increasingly important role in the growing use of technology in several sectors. One of the areas where AI can greatly benefit is the medical sector. For instance, several hazardous diseases can be diagnosed early by applying neural networks, which are trained on an extensive array of populations. A good instance in which such applications are useful is Parkinson's disease, due to the evolving nature of its symptoms. **Materials and Methods:** The study discusses the extreme learning machine (ELM), which is a specific kind of neural network with a feed-forward having a single layer. The kernel version of ELM and the reduced kernel approach have been employed; in the reduced kernel approach, the ELM is trained on a segment of the population. **Results:** The reduced kernel technique has been used for ELM with the objective of detecting PD. The sampling factor's role and the kind of kernel were tested to examine their impact on the classifier's performance. It was discovered that the value of the sampling factor in the range from 0.7 to 0.9 provided the best precision for detection with a kernel of type RBF. Moreover, the study was evaluated against the other high-tech methods based on SVM classifier and it was found that RKELM achieves 100% precision compared to SVM precision of about 91%–92%. **Conclusion:** The ELM approach brings forth impressive results compared to the traditional support vector machine (SVM) approach. After applying the EVM approach on various kernels and with various figures of samples from the overall population, the kernel has displayed 100% accuracy.

Keywords: Artificial intelligence (AI), ELM, identification, KELM, MARK-ELM, Parkinson's disease

Introduction

Artificial intelligence (AI) is transforming several technical practices. The technical aspect of healthcare may gain from this AI revolution. Machine learning, a branch of AI, has the potential to offer very effective diagnostic tools to the medical field using data collected from subjects through prototypes processed on pivotal segments of the population. Advanced detection of Parkinson's disease (PD) would be a very promising application of AI.[1]

PD is one of the most widespread degenerative illnesses that affect a person's central nervous system. There's a marked prevalence of PD in several developing countries. The causes of PD are still unknown. Therefore, the earlier this ailment is diagnosed, the easier it becomes to deal with it. Symptoms of PD include vocal disorders, posture instability, slowness of movements, bodily tremors and rigidity, and asymmetry in hand coordination. Mainly, the vocal disorders are the earliest perceptible symptoms of PD and they are noticed in almost 90% of PD patients almost 5 years earlier than the medical diagnosis.

A considerable portion of this study discusses the literature related to ANNs and diagnosis of PD. A research conducted by Berus *et al*.[2] employed several models of artificial neural networks (ANNs) with feed-forwards. PD in several subjects was predicted by drawing 26 dissimilar vocal samples from each subject and studying the features. Further, the outcomes were

Access this article online	
Quick Response Code:	Website: https://journals.lww.com/mjby
	DOI: 10.4103/MJBL.MJBL_376_24

Address for correspondence: Dr. Ahmed Ali Hussein Alemran, Department of Physiology, Faculty of Medicine, Misan University, Al-Amarah, Iraq.
E-mail: ahmed.mcm@uomisan.edu.iq

Submission: 09-May-2024 **Accepted:** 06-Jun-2025 **Published:** 23-Jan-2026

How to cite this article: Alemran AAH. Parkinson's disease identification using reduced Kernel classification. Med J Babylon 2025;22:1522-7.

confirmed through the LOSO (Leave-One-Subject-Out) procedure. A few feature-selection methods derived from Kendall's correlation coefficient, Pearson's correlation coefficient, and self-organizing maps as well as principal component analysis, were used to better the algorithmic performance and reduce the dataset. The most poignant vocal samples were taken and using the feature-selection derived from Kendall's correlation coefficient, maximum test accuracy was achieved. The varied ANN method has been proven to be a high-quality classification approach for PD diagnosis as it does not depend on feature-selection methods that involve unprocessed data. Finally, after the fine-tuning of the neural network, a test accuracy of 86.47% was achieved. The study was conducted by Morisi *et al*.[3] encompassed diffusion tensor imaging, proton spectroscopy, and morphometric-volumetric information. The MR quantitative marker information which is supported by SVM systems was calculated using these. The goal was to recognize disorders related to PD. Feature-selection methods were also used to finalize the features for critical classification. A greater accuracy has been obtained using a graph-based method, which has been derived from quantitative markers to determine supplementary features.

The study was conducted by Nilashi *et al*.[4] encompassed the benefits of employing an incremental computer learning-based approach, namely, the incremental SVM, to develop newer methods of predicting UPDRS. In addition to using the iterative, non-linear, partial least-squares formula, and self-organizing maps for reducing the size of the dataset, incremental SVM was also employed for Total-UPDRS and Motor-UPDRS predictions, related to task clustering. Shahbakhi et al.[5] used the genetic algorithm (GA) to select an optimized dataset. Healthy subjects were differentiated from the PD-afflicted ones using an SVM-based network. Vocal signals taken from 31 subjects, 23 of whom were PD-afflicted, and eight of whom were healthy, formed the research database. Each person was required to articulate the first letter of the alphabet, that is, "A," for 3 s. Total 22 linear and non-linear features were extracted systematically from these vocal data, with 14 taken from F0 (the pitch, that is, the fundamental frequency), jitter, shimmer, and the ratio of noise-to-harmonics. These are the key factors in classifying vocal data. Because the variations in these features are quite evident in PD-afflicted persons, optimized aspect sets were taken from among these individuals. Among the various types of optimized features, data classification methods were evaluated.

The study was conducted by Lahmiri and Shumel[6] focused on the evaluation of performances of eight various pattern ranking algorithmic approaches which are also called feature selection methods. These were combined with non-linear SVMs to differentiate the healthy individuals from the PD-afflicted ones. The Bayesian optimization system was used to optimize the criteria of the radial basis function kernel as related to the SVM classifier.

Another study was conducted by Lahmiri *et al*.[7] examined the outcomes of AI learning-based mechanisms for PD detection using dysphonia symptoms. Numerous machine learning-based procedures were employed and tested through a data of 22-voice disorder calculations, so as to differentiate healthy individuals from PD-afflicted ones. These machine learning-based techniques included linear discriminant analysis (LDA),[8] k-nearest-neighbors (k-NN), radial basis function neural network (RBFNN), naive Bayes (NB), regression trees (RT), SVMs, as well as the Mahalanobis distance differentiator. The outcomes of these methods were evaluated using a 10-fold cross-validation system. In yet another study undertaken by Bi *et al*.,[9] randomized SVM method based clusters were suggested for differentiating between healthy individuals and PD-afflicted patients. Twenty-five Alzheimer's disease patients and 35 HC patients were recruited from Disease Neuroimaging Initiative Database. A novel method was adopted by the researchers in Aich *et al*.[10] Various performance data based on different aspects including the original data were compared. Approaches based on the analysis of the principal constituent to reduce the number of features were also employed in selecting feature sets. Moreover, non-linear differentiation methods were employed for performance metric differentiations. In a study conducted by Er *et al*.,[11] several differentiation methods were juxtaposed for their effectiveness for PD detection. These methods included learning vector quantization, artificial immune systems, feed-forward, and a probability-based neural network algorithm. 197 PD vocal recordings obtained using 22-voice characteristic sets were examined in this study. A 10-fold cross-validation protocol was followed to ascertain algorithmic results in PD detection. It has been observed that superior performances in stratification accuracy are achieved using probability-based neural network algorithms. Likewise, the study results were compared to those from the previous studies which were conducted using the same PD database.

For the study, an RKELM (reduced kernel extreme learning mechanism) is created and trained for PD detection. Four kinds of kernels are examined and ten sampling factors are taken for each kind. The rest of the report is structured as follows. In section 2, we discuss the methodology. Section 3 presents the experimental results and the summary and conclusion are presented in section 4.

Materials and Methods

This section deals with the development of methodology for creating a differentiation of PD based on fast learning system. Sub-section 2.1 presents the fast learning network, and in sub-section 2.2, the algorithm of building feature prototypes related to the fast learning network through

PD data is explained. Finally, in sub-section 2.3, the evaluation criteria used to assess the methodology of this research are discussed.

Kernel concept

In mathematics, the kernel function is defined as a function between samples of input information explained over the space of characteristics and samples of input information explained over a particular feature.[12] Eqs. (1) and (2) depict the definition of Kernel and Eqs (3)–(6) depict famous kinds of kernels.

$$K : X \times X \to K \tag{1}$$

$$K\left(x_i, x_j\right) = k \tag{2}$$

Examples of kernels are:

- **Linear kernel**

$$K\left(x_i, x_j\right) = x_i^t . x_j \tag{3}$$

- **Radial basis function RBF**

$$K\left(x_i, x_j\right) = e^{-\gamma\left(X_i - X_j\right)^2}, \gamma > 0 \tag{4}$$

- **Polynomial kernel**

$$K\left(x_i, x_j\right) = \left(\gamma 2 X_i^T . X_j + r\right)^d, \gamma > 0 \tag{5}$$

- **Sigmoid kernel**

$$K\left(x_i, x_j\right) = \frac{e^{2 X_i^T . X_j} - 1}{e^{2 X_i^T . X_j} + 1} \tag{6}$$

Extreme learning machine

Initially, the ELM model was designed by Fijani *et al.*,[13] to serve as a hidden feed-forward separate layer NN (neural network) tool.[14] The standard ELM training algorithm consists of three steps:

- Random creation of input hidden layer weights
- Finding out the output matrix $\boldsymbol{H}$
- Finding the unspecified output weights $\boldsymbol{\beta}$ using equation of Moore–Penrose

This method intends to generate the unspecified node variables randomly and further find out output weights methodically. Alternatively, the traditional techniques for learning, for example, SVM or ANN model, require more human involvement and the procedure for tuning of the model. Superior generalization performance, more precise calculation, and faster learning speed are the most important ELM benefits in comparison to the conventional methods.[14,15]

To design the ELM model, let us consider N samples of training $\left(x_i, y_i\right)$ combined into an NN (neural network) design with H unspecified nodes. The x_i and y_i time-series are used to correspond to the input as well as target vectors, respectively, where

$$x_i = \left(x_{i1}, x_{i2}, \ldots, x_{in}\right) \in R^n \tag{7}$$

and

$$y_i = \left(y_{i1}, y_{i2}, \ldots, y_{im}\right) \in R^m \tag{8}$$

Hence, the ELM function will be[14,15]

$$\sum_{i=1}^{H} \beta_i f_i\left(x_j\right) = \sum_{i=1}^{H} \beta_i f\left(a_i . x_j + b_i\right), j = 1, \ldots, N, \tag{9}$$

where $f_i\left(x_j\right)$ is the function for activation, $a_i = [a_{i1}, a_{i2}, \ldots, a_{in}]^T$ represents the weight vector among the unspecified and input nodes, b_i is the threshold of the i^{th} concealed node, $\beta_i = \left[\beta_i, \beta_i, \ldots, \beta_i\right]^T$ is the output weights vector connected to the i^{th} concealed node with the output nodes.

Eq. (9) is as follows in the reduced form[13]:

$$\sum_{i=1}^{H} \beta_i f_i\left(x_j\right) = G\beta \tag{10}$$

$$G = \begin{bmatrix} f\left(a_1 . x_1 + b_1\right) & \cdots & f\left(a_H . x_1 + b_H\right) \\ \vdots & \ddots & \vdots \\ f\left(a_1 . x_N + b_1\right) & \cdots & f\left(a_H . x_N + b_H\right) \end{bmatrix}_{N \times H} \tag{11}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{H \times m} \quad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_L^T \end{bmatrix}_{N \times m} \tag{12}$$

where G is called hidden layer output matrix, β represents the output weight matrix, and T represents the transport operator.

The SLFN instance with L hidden neurons is displayed in Figure 1.

The output weights are computed with the method of least-squares. The outcome can be stated as[16]:

$$\beta = H^* Y \tag{13}$$

where H^* is the Moore–Penrose basic inverse solution for Y, the hidden layer output matrix, and can be obtained from the singular value decomposition (SVD).[13] As recommended by Cao and Lin,[14] the entire technique for

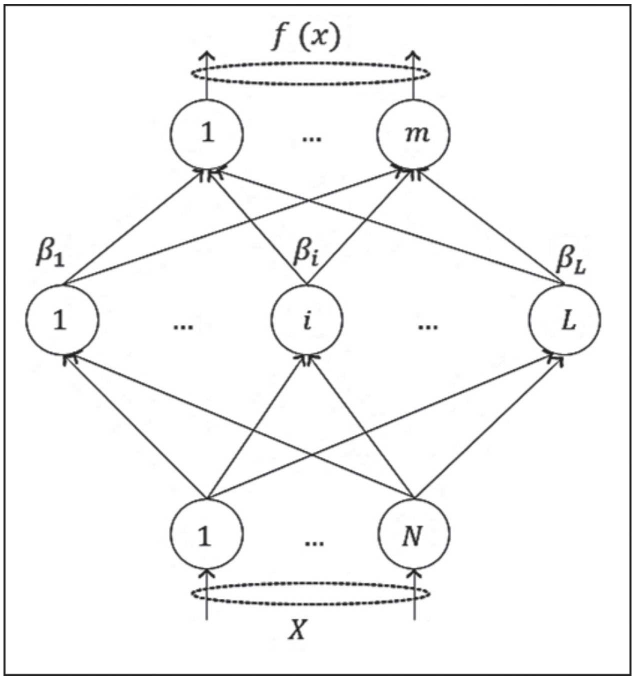

Figure 1: Single hidden layer feed-forward neural network in extreme learning machine

learning and technique for cross-validation is combined in the training phase so as to solve the overtraining problem and enhance the predictive stability of the ELM model.

Kernel extreme learning machine

A novel alternative of ELM is the kernel ELM.[17] The ELM training equation is

$$\beta = H^T \left(\frac{1}{\lambda} + HH^T \right)^{-1} T \tag{14}$$

In the kernel ELM, the $H^T H$ part is changed and substituted by kernel

$$f_L(\vec{x}) = K\left(x, x^{'}\right)\left(\frac{1}{\lambda} + K\left(x, x^{'}\right)\right)^{-1} T \tag{15}$$

RK – extreme learning machine

The idea of classification based on Kernel is useful in separating highly non-linear classes by converting the feature space into high order space (as per the kernel). Here, the separation is carried out and later it returns again to the initial space. Nonetheless, computation of the kernel cannot be done on the complete training data when it increases because of the complexity of the calculation. The kernels serve as high order mathematical functions. It has caused some experts to compute the kernel on a portion of the training data on the basis of the rate of sampling [Table 1].

Table 1: Depicts the process of reduced kernel extreme learning machine

Input
TrainingData
TestingData
samplingFactor
KenrelTYpe
Output
TrainedModel
TrainingAccuracy
TestingAccuracy
Start
determine the number of samples for calculating the reduced kernel based on the sampling factor
Select the samples from the training data with the constraint of N samples from each class at least
calculate the kernel matrix based on the selected trained samples
build the classifier
calculate the kernel matrix between the remaining training samples and selected training
calculate the kernel matrix between the testing samples and the selected training data
predict based on the classifier and kernel matrix for remaining training samples
predict based on the classifier and kernel matrix for testing samples
evaluate the training and testing accuracy
End

Table 2: Binary confusion matrix for Parkinson's Disease

	Prediction of Parkinson is positive	Prediction of Parkinson is negative
Parkinson is positive	TP	FN
Parkinson is negative	FP	TN

Evaluation measures

The measures of assessment that will be employed for comparison of our approaches' performance are provided in the equations. These measures are computed on the basis of the binary confusion matrix elements. This matrix is shown in Table 2.

The measures can be found in the given equations

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{16}$$

$$\text{recall} = sensitivity = TP / \left(TP + FN\right) \tag{17}$$

$$specificity = TN / \left(FP + TN\right) \tag{18}$$

$$precision = \frac{TP}{TP + FP} \tag{19}$$

Table 3: Performance measures of various values of sampling factor and different kinds of Kernels

Kernel	Sampling factor	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
Linear	Accuracy	0.76	0.70	0.94	0.70	0.88	0.85	0.85	0.82	0.91	0.91
Linear	Recall	0.76	0.70	0.94	0.70	0.88	0.85	0.85	0.82	0.91	0.91
Linear	Precision	0.76	0.70	0.94	0.70	0.88	0.85	0.85	0.82	0.91	0.91
Linear	F_measure	0.76	0.70	0.94	0.70	0.88	0.85	0.85	0.82	0.91	0.91
Linear	G mean	0.57	0.49	0.88	0.49	0.77	0.72	0.72	0.67	0.83	0.83
Poly	Accuracy	0.76	0.82	0.73	0.88	0.91	0.97	0.85	0.76	0.94	0.85
Poly	Recall	0.76	0.82	0.73	0.88	0.91	0.97	0.85	0.76	0.94	0.85
Poly	Precision	0.76	0.82	0.73	0.88	0.91	0.97	0.85	0.76	0.94	0.85
Poly	F_measure	0.76	0.82	0.73	0.88	0.91	0.97	0.85	0.76	0.94	0.85
Poly	G mean	0.57	0.67	0.53	0.77	0.83	0.94	0.72	0.57	0.88	0.72
Sig	Accuracy	0.76	0.70	0.79	0.82	0.79	0.79	0.70	0.73	0.76	0.94
Sig	Recall	0.76	0.70	0.79	0.82	0.79	0.79	0.70	0.73	0.76	0.94
Sig	Precision	0.76	0.70	0.79	0.82	0.79	0.79	0.70	0.73	0.76	0.94
Sig	F_measure	0.76	0.70	0.79	0.82	0.79	0.79	0.70	0.73	0.76	0.94
Sig	G mean	0.57	0.49	0.62	0.67	0.62	0.62	0.49	0.53	0.57	0.88
RBF	Accuracy	0.79	0.85	0.94	0.88	0.91	0.88	1.00	1.00	1.00	0.94
RBF	Recall	0.79	0.85	0.94	0.88	0.91	0.88	1.00	1.00	1.00	0.94
RBF	Precision	0.79	0.85	0.94	0.88	0.91	0.88	1.00	1.00	1.00	0.94
RBF	F_measure	0.79	0.85	0.94	0.88	0.91	0.88	1.00	1.00	1.00	0.94
RBF	G mean	0.62	0.72	0.88	0.77	0.83	0.77	1.00	1.00	1.00	0.88

Table 4: Comparison with other approaches

Approach	Author	Accuracy
SVM + BO	Citation	91.82%
SVM	Citation	0.92
RKELM	Current	100%

$$F - measure = \frac{2 \times precision \times Recall}{Precision + Recal} \quad (20)$$

$$G - mean = \sqrt{sensitivity \times precision} \quad (21)$$

Experimental work

This portion contains the experimental assessment of fast learning machine for detecting PD. This technique has been experimented on the dataset given in 3.1 sub-section and the measures of assessment were revealed, with their comparison made as illustrated above.

Dataset

The test is carried out on the dataset of PD taken from the repository of UCI Machine Learning.[18] This dataset was developed by Max Little of Oxford University, in association with the NCVS (National Centre for Voice and Speech), Denver, Colorado, which made a recording of the signals of speech. The initial review published the methods of feature extraction for common voice disorders. This dataset contains a series of biomedical measurements of voice from 31 people, out of which 23 have PD. Every table column is a specific measure of voice, and every row signifies one out of 195 recordings of voice from these people ("name" column). The primary objective of this data is to distinguish people with PD from healthy people, as per the "status" column, that is, set to 1 for PD and 0 for healthy.

Results and Discussion

The ELM technique of reduced kernel has been examined for the four kinds of kernels: linear, RBF, sigmoid, and polynomial. The corresponding precision, recall, and accuracy were produced. Every measure was created on a distinct sample factor from 0.1 to 0.95. The results are depicted in Table 3.

The table shows that the factor of sampling contributes significantly in the RKELM performance. The sampling factor's lower values might not lead to better performance because of under-fitting, whereas the sampling factor's higher value might also not lead to better performance because of under-fitting. The most favorable values range from 0.7 to 0.9. One more thing to note is that the kind of kernel is very significant to compute the classifier's performance. For RBF kernel, the performance achieved has been the best with 100% accuracy. Compared to the high-tech SVM method, RKELM is superior in general. The accuracy-based performance is given in Table 4 for all of them.

Conclusion

In this research, the reduced kernel technique has been used for ELM for the objective of detecting PD. The sampling factor's role and the kind of kernel were tested to examine their impact on the classifier's performance. It was discovered that the value of the sampling factor in the range from 0.7 to 0.9 provided the best precision for detection with a kernel of type RBF. Moreover, study was evaluated against the other high-tech methods based on SVM classifier and it was found that RKELM achieves 100% precision compared to SVM precision of about 91%–92%. Later, we can use classification based on ensemble for detecting PD which is based on more complex and larger datasets.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Vashistha R, Yadav D, Chhabra D, Shukla P. Chapter 5: Artificial intelligence integration for neurodegenerative disorders. In: Kobeissy F, Alawieh A, Zaraket FA, Wang K, editors. Leveraging Biomedical and Healthcare Data. Academic Press, 2019, p. 77-89.
2. Berus L, Klancnik S, Brezocnik M, Ficko M. Classifying Parkinson's disease based on acoustic measures using artificial neural networks. Sensors (Basel, Switzerland) 2018;19:16.
3. Morisi R, Neil D, Gnecco G, Lanconelli N, Testa C, Evangelisti S, *et al*. Parkinsonism and related disorders multi-class Parkinsonian disorders classification with quantitative MR markers and graph-based features using support vector machines. Parkinsonism Relat Disord 2018;47:64-70.
4. Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L, Farahmand M. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybernet Biomed Eng 2018;38:1-15.
5. Shahbakhi M, Far DT, Tahami E. Speech analysis for diagnosis of Parkinson's disease using genetic algorithm and support vector machine. J Biomed Sci Eng 2014;7:147-56.
6. Lahmiri S, Shmuel A. Biomedical signal processing and control detection of Parkinson' s disease based on voice patterns ranking and optimized support vector machine. Biomed Signal Proc Control 2019;49:427-33.
7. Lahmiri S, Dawson DA, Shmuel A. Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. Biomed Eng Lett 2017;8:29-39.
8. Alemran A, Rahmatullah BB, Hadi A. Systematic review on ear identification. Int J Eng Technol 2018;7:251-9.
9. Bi XA, Shu Q, Sun Q, Xu Q. Random support vector machine cluster analysis of resting-state fMRI in Alzheimer's disease. PLoS One 2018;13:1-17.
10. Aich S, Hui KL, Al-absi AA, Sain M. A nonlinear decision tree based classification approach to predict the Parkinson's disease using different feature sets of voice data. International Conference on Advanced Communications Technology (ICACT) 2018:638-42.
11. Er O, Cetin O, Bascil MS, Temurtas F. A comparative study on Parkinson's disease diagnosis using neural networks and artificial immune system. J Med Imaging Health Informat 2016;6:264-8.
12. Fossaceca JM. Application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection. ProQuest Dissertations and Theses 2015;42:171.
13. Fijani E, Barzegar R, Deo R, Tziritis E, Konstantinos S. Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. Sci Total Environ 2019;648:839-53.
14. Cao J, Lin Z. Extreme learning machines on high dimensional and large data applications: A survey. Math Probl Eng 2015;2015:1-13.
15. Huang G, Song S, You K. Trends in extreme learning machines: A review. Neural Netw 2015;61:32-48.
16. Huang G, Wang DH, Lan Y. Extreme learning machines: A survey. Int J Mach Learn Cybern 2011;2:107-22.
17. Lahmiri S, Shmuel A. Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. Biomed Signal Proc Control 2019;49:427-33.
18. Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online 2007;6:1-19.