Classification of Parkinson's disease patients based on spectrogram using local binary pattern descriptors

Journal of Physics: Conference Series **2153** (2022) 012014 doi:10.1088/1742-6596/2153/1/012014

Classification of Parkinson's disease patients based on spectrogram using local binary pattern descriptors

E Gelvez-Almeida[1,2], A Vásquez-Coronel[1], R Guatelli[3], V Aubin[3], and M Mora[1]

[1] Laboratorio de Investigaciones Tecnologicas en Reconocimiento de Patrones, Universidad Católica del Maule, Talca, Chile
[2] Facultad de Ciencias Básicas y Biomédicas, Universidad Simón Bolívar, San José de Cúcuta, Colombia
[3] Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza, Buenos Aires, Argentina

E-mail: e.gelvez@unisimonbolivar.edu.co

Abstract. Extreme learning machine is an algorithm that has shown a good performance facing classification and regression problems. It has gained great acceptance by the scientific community due to the simplicity of the model and its sola great generalization capacity. This work proposes the use of extreme learning machine neural networks to carry out the classification between Parkinson's disease patients and healthy individuals. The descriptor used corresponds to the feature vector generated applying the local binary Pattern algorithm to the grayscale spectrograms. The spectrograms are obtained from the audio signal samples from the considered repository. Experiments are conducted with single hidden layer and multilayer extreme learning machine networks comparing the results of each structure. Results show that hierarchical extreme learning machine with three hidden layers has a better general performance over multilayer extreme learning machine networks and a single hidden layer extreme learning machine. The rate of success obtained is within the ranges presented in the literature. However, the hierarchical network training time is considerably faster compared to multilayer networks of three or two hidden layers.

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease. It involves slow and irreversible progression, the first symptoms becoming evident after several years of suffering from the disease. Motor symptoms, such as tremors, are the most popular ones. However, other symptoms can be, in many cases, incapacitating [1]. These are cognitive disorders, limb rigidity, gait and balance problems, and slowness. PD symptoms appear early as speech and voice disorders.

Several studies have proposed the processing of voice signals to obtain acoustic parameters as an objective and non-invasive method for PD detection [2]. Sakar B E, *et al.* [3] found that sustained vowels carry enough PD-discriminative information using machine learning tools. In Wodzinski M [4], a previously trained ResNet architecture is used for PD detection using ImageNet and singular value decomposition (SVD) databases. For classification, audio spectrograms of vowels with sustained phonation are used from the PC-GITA database [5]. The

validation set has shown an accuracy rate higher than 90%. In Zahid, *et al.* [6], deep learning methodologies are combined with machine learning methodologies to classify Parkinson's disease patients. The multilayer perceptron applied to monologue recordings showed the highest accuracy of 99.7%. In Trinh, *et al.* [7], the detection of voice pathologies in patients who suffer from Parkinson's or Alzheimer's disease is proposed. A convolutional neural network (CNN) architecture is used with a classification accuracy higher than 95%.

The aforementioned works use machine learning methods known for their higher levels of fit to classification issues, reporting accuracy rates between 80% and 90%. It is well known that said methodologies experience a high computational cost in training algorithms and the estimation of the hyperparameters of the model. extreme learning machine (ELM) [8] has gained great acceptance from the scientific community because of the simplicity of the models and their generalization capacity. They began to be massively used to address complex problems of pattern recognition due to the low computational cost of their training algorithm.

In this paper, the use of ELM neural networks to address the classification issue of Parkinson's disease patients and healthy individuals is proposed. Experiments are conducted with single and multilayer ELM networks and results are compared. Grayscale spectrograms obtained from the voice signals from the repository presented in Giuliano M, *et al.* [9] are used. Descriptors for those images are generated using the local binary pattern (LBP) algorithm. Said algorithm highlights textures creating a feature vector. Multilayer extreme learning machine (ML-ELM) and hierarchical extreme learning machine (H-ELM) have shown a better performance addressing the classification problem of Parkinson's disease than single-layer ELM networks, with reduced training times.

2. Background

In this section, a brief description is presented of the spectrogram images, the LBP algorithm and the ELM variants used in this paper.

2.1. Spectrogram

A spectrogram is the representation of a signal in time. The result is a three-dimensional graphic of the spectrum of a signal formed by time, frequency, and amplitude of energy distribution. The amplitude of energy distribution is represented by color intensity. That is why it can be represented in a two-dimensional image, despite it having three dimensions. Speech representations through spectrograms have proven to be stable and robust, even with high noise levels [10, 11].

2.2. Local binary patterns

LBP is an operator for image texture analysis which has been used in several applications, including visual inspection, image retrieval, remote sensing, biomedical image analysis, face image analysis, motion analysis, environment modeling, and outdoor scene analysis libre [12]. The operator consists in labelling the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. This operator traces the whole image and is mathematically represented as shown in Equation (1).

$$LBP_{(gp_x,gp_y)} \sum_{p=0}^{P-1} S(gp - gc) \times 2^P. \tag{1}$$

In Equation (1), gc is the intensity value of the central pixel, gp is the intensity of the neighboring pixel with index p and P is the number of sampling points on a circle of radius R

(circular neighborhood) [12]. Function S can be expressed as shown in Equation (2).

$$S_x = \begin{cases} 1 & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases}, \tag{2}$$

This operator highlights textures from the original image creating a new one. The histogram is extracted from this new image, obtaining a feature vector as a result. This descriptor is used in the experiments to observe the classification behavior using the ELM, ML-ELM and H-ELM algorithms.

2.3. Extreme learning machine

Being $\aleph = \{(\boldsymbol{x}_i, \boldsymbol{t}_i) \in \mathbb{R}^n \times \mathbb{R}^m | i = 1, \ldots, N\}$ a random training set, $g(s) : \mathbb{R}^n \to \mathbb{R}^m$ an activation function, and L the number of neurons in the hidden layer with $L < N$ [13], the SLFN training algorithm is given by Equation (3).

$$\sum_{i=1}^{L} \boldsymbol{\beta}_i g(\boldsymbol{w}_i \boldsymbol{x}_j + b_i) = \boldsymbol{t}_j, \qquad j = 1, \ldots, N. \tag{3}$$

In Equation (3), $\boldsymbol{\beta}_i$ are the weights of the output layer, $\boldsymbol{w}_i$ are the weights of the hidden layer and b_i, the i-th bias of the hidden layer [8, 13]. $\boldsymbol{w}_i$ weights and b_i biases of the hidden layer are pseudo-randomly assigned. Equation (3) can be written in the matrix form $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{T}$, where, the $\boldsymbol{H}$ matrix in Equation (4) is called the matrix of the output hidden layer of the neural network [14]. Last, the $\boldsymbol{\beta}$ weights of the output layer are calculated using Equation (5).

$$\boldsymbol{H} = \begin{bmatrix} g(\boldsymbol{w}_1 \boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{w}_L \boldsymbol{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\boldsymbol{w}_1 \boldsymbol{x}_N + b_1) & \cdots & g(\boldsymbol{w}_L \boldsymbol{x}_N + b_L) \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}, \quad T = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{bmatrix}. \tag{4}$$

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger} \boldsymbol{T}. \tag{5}$$

The $\boldsymbol{H}^{\dagger}$ matrix in Equation (5) is the Moore–Penrose generalized inverse of the $\boldsymbol{H}$ matrix [15].

2.4. Extreme learning machine auto encoder

Extreme learning machine auto encoder (ELM-AE) is an ELM modification to perform unsupervised learning. In this case, input data $\boldsymbol{x}$ is used as output data, so $\boldsymbol{t} = \boldsymbol{x}$. Weights and biases of the hidden layer are chosen to be randomly orthogonal in order to obtain a better performance [16]. In this sense, random orthogonal weights and biases are presented through Equation (6).

$$\boldsymbol{h}_i = g(\boldsymbol{w}_i \boldsymbol{x}_j + b_i), \quad \boldsymbol{w}^T \boldsymbol{w} = \boldsymbol{I}, \quad \boldsymbol{b}^T \boldsymbol{b} = 1. \tag{6}$$

In Equation (6), $\boldsymbol{w}$ and $\boldsymbol{b}$ are the weights and biases of the hidden layer, respectively [16]. Last, weights of the output layer are determined analytically with Equation (7).

$$\boldsymbol{\beta} = \left(\frac{\boldsymbol{I}}{C} + \boldsymbol{H}^T \boldsymbol{H} \right)^{-1} \boldsymbol{H}^T \boldsymbol{X}. \tag{7}$$

In Equation (7), $\boldsymbol{H}$ is the hidden layer output matrix, $\boldsymbol{X}$ is ELM-AE input and output data and C is a regularization parameter [16].

2.5. Multilayer extreme learning machine

ML-ELM is a multilayer neural network based on ELM-AE similar to deep networks and with better performance [16]. ML-ELM hidden layer weights are initialized with ELM-AE, without requiring fine tuning. In this case, $\boldsymbol{H}$ output matrix values of each hidden layer are obtained through Equation (8).

$$\boldsymbol{H}^k = g((\boldsymbol{\beta}^k)^T \boldsymbol{H}^{k-1}). \tag{8}$$

The $\boldsymbol{H}^k$ in Equation (8) is the k-th hidden layer output matrix [16]. Weights of the first hidden layer are the $\boldsymbol{\beta}^1$ results of the first ELM-AE that has samples $\boldsymbol{X}$ as input data. The weights that connect the different hidden layers are the $\boldsymbol{\beta}^{i+1}$ results of the ELM-AE that has the output results of the previous ML-ELM hidden layer as input [16]. Last, weights of the ML-ELM output layer are determined analytically using the Moore–Penrose generalized inverse [15].

2.6. Hierarchical extreme learning machine

Faced with the need to address application issues of greater complexity among the scientific community, an H-ELM, non-iterative multilayer perceptron model is proposed in en [17]. H-ELM training is divided into an unsupervised feature learning stage and a final supervised stage. In the first stage, an ELM-based sparse autoencoder (ELM-SAE) is introduced in the deep layers for high level feature extraction of input data. In ELM-SAE, weights and biases of the output layer are pseudo-randomly assigned, whereas the output layer weights matrix is subject to an optimization problem, with L1 norm regularization term, as shown in Equation (9).

$$\min_{\boldsymbol{\beta}} : F(\boldsymbol{\beta}) = \|\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{X}\|^2 + \lambda\|\boldsymbol{\beta}\|_1. \tag{9}$$

In Equation (9), λ is a regularization parameter and $\|\boldsymbol{\beta}\|_1$ represents the L_1 norm [17]. The fast iterative shrinkage-thresholding algorithm (FISTA) [18] is adopted to solve the convex scheme of Equation (9) and to further improve the training speed. Equation (10) establishes the connection between ELM-SAE and the H-ELM deep layers.

$$\boldsymbol{H}^{(k)} = g(\boldsymbol{H}^{(k-1)}\boldsymbol{\beta}^{(k)}). \tag{10}$$

In Equation (10), $\boldsymbol{H}^{(k)}$ is the k-th output layer matrix, $\boldsymbol{H}^{(k-1)}$ the $(k-1)$-th output layer matrix, $\boldsymbol{\beta}^{(k)}$ the k-th ELM-SAE output layer weights, and $g(\boldsymbol{H}^{(k-1)}\boldsymbol{\beta}^{(k)})$ is an activation function [17]. In the supervised stage, the final decision of the classifier is the consequence of the ELM method, which takes as input data the last hidden layer output matrix and the labels as output.

3. Materials and method

In order to conduct the experiments, a machine with central processing unit (CPU) intel(R) Core(TM) i7-8550U 1.80GHz-1.99GHz was used, with 4 cores and 8 logic processors, apart from 12 Gb RAM memory. Codes were made in MATLAB R2017b. Datasets and methods used in this study are presented in the following section.

3.1. Database description

A database whit grayscale spectrograms generated from audio signals from Parkinson's disease patients and healthy individuals were used for the experiments [9]. The database is composed of spectrograms images from 58 Parkinson's disease patients and 77 healthy individuals. Patients' ages vary from 38 to 79 years, and they have been diagnosed with Parkinson's for six years, on average.

3.2. Methodology of the experiments
Experiments are divided into two phases: a pre-processing phase of the database with the LBP algorithm and a classification phase with the ELM algorithm and the variants presented in this paper. The two phases conducted in this work are presented below.

3.2.1. Pre-processing phase In this phase, the LBP algorithm is applied to the spectrogram images to extract their features. In order to do so, the MATLAB `extractLBPFeature` function is used. The result obtained by this algorithm generated a feature vector of 59 values. A label is assigned to each value, 0 for healthy individuals and 1 for Parkinson's disease patients. After the features are extracted from the spectrograms, the database is distributed as shown in Table 1.

The training, validation and test sets are randomly selected to ensure confidence in the results and prevent its unintentional manipulation.

Table 1. Distribution of the database of spectrograms from Parkinson's disease patients and healthy subjects.

Patient	Training (60%)	Validation (20%)	Test (20%)	Total
Sick	34	12	12	58
Healthy	47	15	15	77

3.2.2. Classification phase The database described in the previous section is used for this phase; for the purpose of obtaining the optimum values of hyperparameters, different iterations are conducted in the experiments, varying the amount of neurons in the hidden layer from 100 to 5000 and the regularization parameter from 10^{-20} and 10^{20}. Moreover, the internal structure of the network varies from one hidden layer (standard ELM) to three hidden layers (ML-ELM and H-ELM). The sigmoid function of Equation (11) is used as an activation function for all cases.

$$g(w_i x_j + b_i) = \frac{1}{1 + e^{-(w_i x_j + b_i)}}. \tag{11}$$

In Equation (11), w_i and b_i are the weights and biases of the hidden layer, respectively, and x_j the input data to the network [8,13]. Once the best results are obtained, a test set is carried out to verify the robustness of the network.

4. Results
To obtain hyperparameters presented in Table 2, tests were carried out with different combinations of the number of neurons in the hidden layer and the regularization parameters. Regularization hyperparameters improve accuracy, especially in those situations when data is imbalanced. The results of these experiments have shown that for this particular dataset, regularization parameter C does not improve results for the single-hidden-layer ELM network. That is why said parameter is not used.

Table 2. Hyperparameters used with best results in ELM and its variants.

Model	Hidden layer	Neurons	Regularization
H-ELM	3	$L_1 = L_2 = 500$, $L_3 = 1900$	$C = 10^{-4}$, $\lambda = 10^{-3}$
ML-ELM	3	$L_1 = L_2 = 1100$, $L_3 = 2700$	$C_1 = C_2 = 10^5$, $C_3 = 10^3$
ML-ELM	2	$L_1 = 1500, L_2 = 1200$	$C_1 = 10^5$, $C_2 = 10^6$, $C_3 = 10^2$
ELM	1	$L = 2100$	Not used

Once the hyperparameters are obtained, 10 iterations are made over the same dataset, registering accuracy average and standard deviation. The standard deviation among the results obtained varies from 0 to 0.009, which proves that the results are quite stable. Table 3 shows the comparison in terms of the accuracy and training speed between the standard ELM, ML-ELM, and H-ELM networks whit two and three hidden layers. The results presented were obtained using regularization parameters of Equation (7) and Equation (9).

As it can be seen in Table 3, multilayer networks (ML-ELM and H-ELM) have better validation and test results than a single-layer ELM network. For multilayer networks, the accuracy is 92.59% for validation and 81.48% for test. Said results are higher than the 88.89% and 77.78% obtained in validation and test, respectively, for single-layer ELM networks. Results for ML-ELM and H-ELM networks are equal for the training, validation, and test sets. However, training of H-ELM networks is performed 10.2 and 5.5 times faster in relation to three and two hidden layers of ML-ELM networks, respectively. It represents a significantly high and favorable reduction in training time which improves with databases at a larger scale.

Table 3. Accuracy in the classification of Parkinson's patients using standard ELM, ML-ELM, and H-ELM.

Model	Hidden layers	Training		Validation		Test	
		Accuracy	Time (s)	Accuracy	Time (s)	Accuracy	Time (s)
H-ELM	3	0.9012	0.2438	0.9259	0.0083	0.8148	0.0090
ML-ELM	3	0.9012	2.4849	0.9259	0.0055	0.8148	0.0060
ML-ELM	2	0.9012	1.3315	0.9259	0.0043	0.8148	0.0051
ELM	1	0.9136	0.0641	0.8889	0.0172	0.7778	0.0156

5. Conclusions

Standard extreme learning machine, multilayer extreme learning machine and hierarchical extreme learning machine to address the classification problem are used; said classification consists in identifying Parkinson's disease patients using a database composed of 135 spectrograms. The spectrograms are processed by the local binary pattern operator to extract features from the image to train the network; the dataset has been divided into three subsets, 60% for training (81 samples), 20% for test (27 samples), and 20% for validation (27 samples).

Results show that multilayer networks have a better performance than single-layer extreme learning machine networks. Multilayer extreme learning machine and hierarchical extreme learning machine networks present higher accuracy rates for the validation and test sets when compared to single-layer extreme learning machine networks. Although multilayer extreme learning machine and hierarchical extreme learning machine networks have the same accuracy rates for the three sets, training time with hierarchical extreme learning machine is 10.2 and 5.5 times faster than for three and two hidden layers multilayer extreme learning machine networks, respectively. Accuracy with multilayer extreme hierarchical extreme learning machine networks is 90.12% for the training set, 92.59% for the validation set, and 81.48% for the test set.

The hierarchical extreme learning machine structure which shows the best performance is the one formed by three hidden layers. The hyperparameters used $L_1 = L_2 = 1100$, $L_3 = 2700$, $C = 10^{-4}$, and $\lambda = 10^{-3}$, where L_1, L_2, and L_3 is the number of neurons in the first, second and third hidden layers, respectively, C and λ, the regularization parameter associated to the Moore–Penrose generalized inverse and fast iterative shrinkage-thresholding algorithm algorithm, respectively. The results show that the local binary pattern operator is a good alternative to generate spectrogram image descriptors to classify through hierarchical extreme learning machine networks. In future works, it is intended to try local binary pattern descriptors with other databases of larger-scale images using different extreme learning machine variants to compare accuracy and training times.

Acknowledgments
This work was funded by the National Agency for Research and Development (ANID), Scholarship Program, Doctorado Becas Chile, 2020-21201000. The first author of the article also thanks for the licenses for doctoral studies, the Fund for Teacher and Professional Development of the Universidad Simón Bolívar, Colombia.

References
[1] Logemann J A, Fisher H B, Boshes B, Blonsky E R 1978 Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients *Journal of Speech and hearing Disorders* **43(1)** 47
[2] Arora S, Baghai-Ravary L, Tsanas A 2019 Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice *The Journal of the Acoustical Society of America* **145(5)** 2871
[3] Sakar B E, Isenkul M E, Sakar C O, Sertbas A, Gurgen F, Delil S, Apaydin H, Kursun O 2013 Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings *IEEE Journal of Biomedical and Health Informatics* **17(4)** 828
[4] Wodzinski M, Skalski A, Hemmerling D, Orozco-Arroyave J R, Nöth E 2019 Deep learning approach to Parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Berlin: IEEE) p 717
[5] Orozco-Arroyave J R, Arias-Londoño J D, Vargas-Bonilla J F, Gonzalez-Rativa M C, Nöth E 2014 New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (Reykjavik: European Language Resources Association) p 342
[6] Zahid L, Maqsood M, Durrani M Y, Bakhtyar M, Baber J, Jamal H, Mehmood I, Song O Y 2020 A spectrogram-based deep feature assisted computer-aided diagnostic system for Parkinson's disease *IEEE Access* **8** 35482
[7] Trinh N, Darragh O 2019 Pathological speech classification using a convolutional neural network *Irish Machine Vision and Image Processing (IMVIP 2019)* (Dublin: Technological University Dublin) p 72
[8] Huang G B, Zhu Q Y, Siew C K 2004 Extreme learning machine: a new learning scheme of feedforward neural networks *International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)* (Budapest: IEEE) p 985
[9] Giuliano M, Perez S N, Maldonado M, Bondar P, Linari D, Adamec D A, Debas M I, Morales C M, León L, Yaco A Y, Birelli J B, Martínez R M, Lacaze M L, Gurlekian J A 2021 Construction of a Parkinson's voice database *International Conference on Industrial Engineering and Operations Management* (Sao Paulo: IEOM Society International) p 940
[10] Greenberg S, Kingsbury B E 1997 The modulation spectrogram: in pursuit of an invariant representation of speech *International Conference on Acoustics, Speech, and Signal Processing* (Munich: IEEE) p 1647
[11] Kingsbury B E, Morgan N, Greenberg S 1998 Robust speech recognition using the modulation spectrogram *Speech Communication* **24** 117
[12] Pietikäinen M 2005 Image analysis with local binary patterns *Scandinavian Conference on Image Analysis (SCIA 2005)* (Berlin: Springer) p 115
[13] Huang G B, Zhu Q Y, Siew C K 2006 Extreme learning machine: theory and applications *Neurocomputing* **70(1-3)** 489
[14] Ding S, Zhao H, Zhang Y, Xu X, Nie R 2015 Extreme learning machine: algorithm, theory and applications *Artificial Intelligence Review* **44(1)** 103
[15] Barata J C A, Hussein M S 2012 The Moore-Penrose pseudoinverse: a tutorial review of the theory *Brazilian Journal of Physics* **42(1-2)** 146
[16] Chamara L, Zhou H, Huang G B, Vong C M 2013 Representational learning with extreme learning machine for big data *IEEE Intelligent Systems* **28(6)** 31
[17] Tang J, Deng C, Huang G B 2015 Extreme learning machine for multilayer perceptron *IEEE Transactions on Neural Networks and Learning Systems* **27(4)** 809
[18] Beck A, Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM Journal on Imaging Sciences* **2(1)** 183