Machine learning-based classification for diagnosis of neurodegenerative diseases

Eddy Sánchez-DelaCruz[1] and Pilar Pozos-Parra[2]

[1] Instituto Tecnológico Superior de Misantla, Veracruz, Mexico
esanchezd@itsm.edu.mx
[2] Instituto Tecnológico y de Estudios Superiores de Occidente, Tlaquepaque, Jalisco, Mexico
mariapozos@iteso.mx

Abstract. Many neurodegenerative diseases affect human gait. These pathologies may have similarities that make difficult their correct classification, so it is essential to distinguish them with a high degree of accuracy to prescribe appropriate treatment. In this study, we worked with gait biomarkers of a public dataset and then we implemented an ANN in order to obtain competitive results comparing to those from the literature. Our result shows that with only one machine learning algorithm, it is not possible to increase the percentages until an optimal classification in multiclass classification performed. Based on this, we propose the study of the multiclass classification of gait in the neurodegenerative diseases with four perspectives.

Keywords: Classification · neurodegenerative diseases · Artificial Neural Network.

1 Introduction

The neurodegenerative diseases cause, among others afflictions, human gait disorders due to impairment in the nervous system. In order to prescribe proper medications and physical therapy to patients with these diseases, it is essential to make a correct classification of these diseases as elaborated in [13,23,27]. There are online datasets that containing gait biomarkers of patients suffering these conditions [12,31,22]. Also, there are open source tools of machine learning for the implementation of classification algorithms that allow the use of these datasets. One of these, TensorFlow[3], is a highly promising tool developed by a research group from Google for implementing a variety of algorithms mainly Artificial Neural Networks (ANN) [1].

The impact of neurodegenerative diseases is drastic. Patients suffer additional problems such as psychological impairment, loss of quality life, labor incapacity, loss of social skills, and the detriment of their caregivers, besides the enormous economic costs in healthcare [24]. By common observation, neurodegenerative pathologies such as Parkinson disease (PD), Huntington disease (HD)

[3] https://www.tensorflow.org/

and Alzheimer disease (AD), cause similar movement disorders on human gait, however, there are subtle aspects that distinguish them, e.g. slow gait velocity (bradykinesia), short step and stride lengths, among other gait biomarkers [13,23]. In this regard, computational techniques such as ANN can be useful to detect the associated disease given a set of specific gait biomarkers as input.

The purpose of this research is to explore the machine learning approach using the ANN algorithm included in the TensorFlow library, in search of competitive percentages classifying neurodegenerative diseases. Therefore, we propose: *i)* a comparative results of ANN algorithm (using OVO and OVA approaches) with data from a public database, *ii)* a comparative state of the art results and *iii)* four prospects to study the multiclass classification.

The rest of the paper is organized as follows. In section 2 we describe previous work in the neurodegenerative diseases classification. Materials and methods are detailed in section 3. In section 4 we show results and conclusions. Finally, discussions and future work are shown in section 5.

2 Related work

All related works mentioned in this section experiments with a public dataset named Gait dynamics in Neurodegenerative Disease Database (GaitND-DB) [12]. Experiments involve classification of patients facing PD, HD, and Amyotrophic lateral sclerosis (ALS).

Yang *et al.*, worked with Support Vector Machine (SVM) obtaining 86.85% on accuracy in classification of all diseases [29].

Banaie *et al.*, proposed a Quadratic Bayesian classifier. Results showed a correct classification of 86.95% in total accuracy [4].

Iram *et al.*, divided into three stages the neurodegenerative diseases: retrogenesis, cognitive impairment, and gait disorder. Their work aimed to determine the grade of the disease, for which they use a Quadratic Bayes Normal Classifier. Results showed 65% of accuracy [16].

Dutta *et al.*, used an Elman's Recurrent Neural Network (ERNN), obtained an overall of classification of 87.1% [9].

Zeng and Wang, proposed a method employing deterministic learning theory. Results showed a good classification performance: 93.75% of accuracy [30].

Xia *et al.*, used the Leave-One-Out Cross-Validation (LOOCV) method and obtained 96.83% of accuracy [28].

Ren *et al.*, implemented Empirical Mode Decomposition method for decomposing the time series of gait rhythms into intrinsic mode functions. General values of AUC shows good performance in binary classification: 0.949, 0.900, and 0.934 for PD, HD, and ALS, respectively [26].

Bilgin, using Discrete Wavelet Transform (DWT) "bior2.6" decomposed the compound force signal for determination of features and applying Naïve Bayesian classifier achieved 90.93% of accuracy in binary classification [6].

3 Material and methods

3.1 Dataset

In this work, we use the GaitND-DB, a public dataset often used by researchers interested in gait analysis of patients with neurodegenerative diseases. GaitND-DB is a free access database with data of 15 patients with PD, 20 patients with HD, 13 with ALS and 16 healthy controls (Ctrl). Table 1 shows the patients age, and gender distribution.

Table 1. Age and gender distribution in GaitND-DB.

Case	**Age groups**						**Gender**		**Patients**
	18-29	30-39	40-49	50-59	60-69	70-79	F	M	**number**
Ctrl	6	3	2	2	2	1	14	2	16
PD			1	3	4	7	5	10	15
HD	1	5	7	4	1	2	14	6	20
ALS		2	2	2	5	2	3	10	13
								Total	**64**

For data acquisition of GaitND-DB, Hausdorff *et al.*, instructed subjects to walk at their usual pace along a 77 m long hallway for 5 min (300 s) [14]. The gait data were measured with Force-sensitive sensors placed inside each subject's shoes. Signals were recorded at 300-Hz sampling rate with a 12-bit resolution per sample. First 20 s of each record were excluded to minimize the start-up effects. Each record includes the following attributes [12]:

- Elapsed time.
- Left stride interval (seconds).
- Right stride interval (seconds).
- Left swing interval (seconds).
- Right swing interval (seconds).
- Left swing interval (% of stride).
- Right swing interval (% of stride).
- Left stance interval (seconds).
- Right stance interval (seconds).
- Left stance interval (% of stride).
- Right stance interval (% of stride).
- Double support interval (seconds).
- Double support interval (% of stride).

3.2 Algorithm

We use a multilayer perceptron ANN for classifying neurodegenerative diseases. The implementation was using the TensorFlow library as shown in Algorithm 1.

Algorithm 1 ANN algorithm [3,10]

1: Input: Entry of the training set $D = \{x_0, x_1, x_2, ..., x_k\}$
2: In the hidden layer, initialize weights and thresholds with Xavier method
3: Perform the sum of all weights $x_i = \sum_j p_{ij} x_j$
4: Apply the *Sigmoid* activation function $f(x) = \frac{1}{1+e^{-x}}$
5: Output: Activation of the output layer neurons $Y = y_0, y_1, y_2, ..., y_k$

Weights (for $\{x_1, x_2, \ldots, x_n\}$) are initialized using the Xavier method [11]: $Var(W) = \frac{2}{n_{in}+n_{out}}$. Where: $Var(W)$ is the variance of the weights for a layer, initialized with a normal distribution and n_{in}, n_{out} are the number of neurons in the parent and in the current layer.

3.3 Experimental design

Figure 1 shows the implemented method where the procedures included in this method are described as follows:

Attribute selection procedure: We use the raw data from the GaitND-DB (section A of Figure 1), with *(i)* all the 13 attributes and, *(ii)* based on classification criteria of Dutta *et al.* [9]. The best 5 attributes (in seconds), i.e., Double support interval, Right stance interval, Right stride interval, Left stance interval and Left stride interval were selected applying the Chi-Squared method (section B of Figure 1). The chi-squared method is one of the best methods measured in the selection [32,21,25,20,15].

Applying the stratified sampling, datasets were created using OVO and OVA approaches (section C of Figure 1), obtaining the following datasets respectively, OVO: Crtl vs. ALS, Crtl vs. HD, Crtl vs. PD, ALS vs. HD, ALS vs. PD and PD vs. HD; OVA: Crtl vs. {ALS,HD,PD}, ALS vs. {Crtl,HD,PD}, HD vs. {Crtl,ALS,PD} and PD vs. {Crtl,ALS,HD}. To construct the training and test sets the random criterion 2/3 and 1/3 was used respectively (section D of Figure 1), given that this criterion is used in the literature [17].

Tuning procedure: In order to select layers, we run ten times the multilayer perceptron ANN algorithm for each layer followed by average of each layer. This was applied for each one approach: OVO (13 and 5 attributes) and OVA (13 and 5 attributes), (section E of Figure 1). The layers selected are mentioned in subsection 3.3: *Tuning procedure.*

We conducted test with a multilayer perceptron ANN algorithm of TensorFlow (section F of Figure 1), 30 times per selected layer (see subsection 3.3: *Tuning procedure*) and each one was run with different random weights.

Finally, results were analyzed (section G of Figure 1) along with drawing conclusions (section H of Figure 1).

Tuning procedure

- OVO. In this approach, we selected the first, second and third layers for 13 and 5 attributes. For each case: Crtl vs. ALS, Crtl vs. HD, Crtl vs. PD, ALS vs. HD, ALS vs. PD, PD vs. HD, we conducted 10 executions for layer 1, 10 for layer 2, and continuing like this 10 for layer 10. Then we obtained the average for each layer and analyzed results. This tuning was made for 13 and 5 attributes.
- OVA. In this approach, we selected the first, second, third and fourth layers for 13 and 5 attributes. For each case: Crtl vs. {ALS,HD,PD}, ALS vs. {Crtl,HD,PD}, HD vs. {Crtl,ALS,PD}, PD vs. {Crtl,ALS,HD}, we conducted 10 executions for layer 1, 10 for layer 2, and subsequently 10 for layer 10. Then we obtained and analyzed the average for each layer, so also results. This tuning was made for 13 and 5 attributes.

This criterion of layer selections was made based on stability, because they do not tend to over-training.

We conducted the experiments on an Alienware M17x laptop, with an Intel Core i7-2670QM CPU @2.20GHzx8, 8 GB in RAM, 500 GB in HDD, Ubuntu Linux 12.04 64-bit, Python 2.7.6 and TensorFlow v.0.5.0.

4 Results and Conclusions

4.1 Performance of OVO approach

In this approach, for 13 attributes we can see in Table 2 that the highest percentages were: 39% for Ctrl vs. ALS, 43% for Ctrl vs. HD, 38% for Ctrl vs. PD, 65% for ALS vs. HD, 61% for ALS vs. PD and 54% for PD vs. HD.

For 5 attributes, we can see in Table 2 that the highest percentages were: 39% for Ctrl vs. ALS, 44% for Ctrl vs. HD, 38% for Ctrl vs. PD, 65% for ALS vs. HD, 61% for ALS vs. PD and 55% for PD vs. HD.

Table 2. Results of OVO approach

OVO	**Ctrl vs. ALS**	**Ctrl vs. HD**	**Ctrl vs. PD**	**ALS vs. HD**	**ALS vs. PD**	**PD vs. HD**
13 attributes	39	43	38	65	61	54
5 attributes	39	44	38	65	61	55

To obtain the results shown in Table 2, first we performed 30 times per layer, then obtained the average of this performance.

4.2 Performance of OVA approach

In this approach, for 13 attributes we can see that the highest percentages were: 84% for Crtl vs. {ALS,HD,PD}, 64% for ALS vs. {Crtl,HD,PD}, 71% for HD vs. {Ctrl,ALS,PD} and 74% for PD vs. {Ctrl,ALS,HD}.

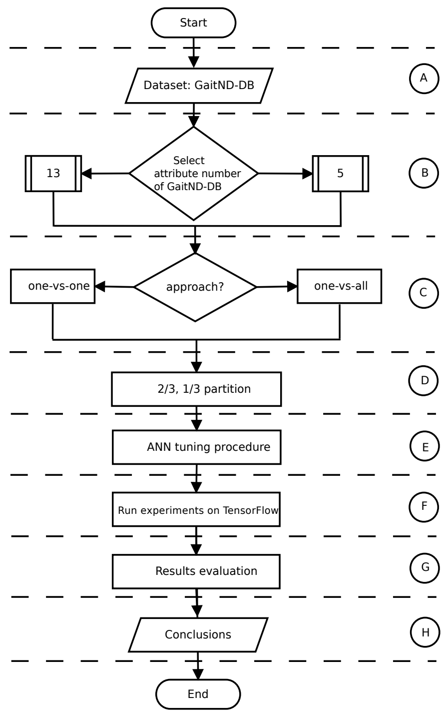

Fig. 1. Flowchart of experiment's method

For 5 attributes, we can see that the highest percentages were: 81% for Crtl vs. {ALS,HD,PD}, 74% for ALS vs. {Ctrl,HD,PD}, 71% for HD vs. {Ctrl,ALS,PD} and 76% for PD vs. {Ctrl,ALS,HD}.

Table 3. Results of OVA approach

OVA	Crtl vs. {ALS,HD,PD}	ALS vs. {Ctrl,HD,PD}	HD vs .{Ctrl,ALS,PD}	PD vs. {Ctrl,ALS,HD}
13 attributes	84	64	71	74
5 attributes	81	74	71	76

To obtain the final results shown in Table 3, first we performed 30 times per layer, then we obtained the average of this performance.

4.3 Analysis of results for OVO and OVA approaches

In OVO approach, for both 13 and 5 attributes, we can note that the percentages for all classes (Crtl vs. ALS, Crtl vs. HD, Crtl vs. PD, ALS vs. HD, ALS vs. PD, PD vs. HD) are less than or equal to 65% (Figure 2), which means a low degree of reliability in binary classification.

Fig. 2. OVO approach: graphical of higher percentages with 13 and 5 attributes.

In OVA approach, we note that for 5 attributes the percentages for four cases (Crtl vs. {ALS,HD,PD}, ALS vs. {Ctrl,HD,PD}, HD vs. {Ctrl,ALS,PD}, PD vs. {Ctrl,ALS,HD}) are greater than or equal to 71% and the highest percentage is 84 for 13 attributes (Figure 3), which means an acceptable degree of reliability in multiclass classification.

4.4 Comparison of results

Our highest result was in OVA approach (multiclass classification) with an acceptable percentage of 84% for 13 attributes, which shows the good performance of the multilayer perceptron algorithm in classification of a group of neurodegenerative diseases. We must emphasize that the purpose of this research was to perform machine learning with ANN algorithm in search of competitive percentages, using OVO and OVA approaches, and comparing with state of the art results. With this method we showed that, for the classification of neurodegenerative diseases, it is not possible to increase the accuracy percentages using one machine learning algorithm until we reach an optimal classification (i.e., approximately 100%) in multiclass classification, as we can see in Table 4. Therefore, we suggest some recommendations in section 5: *Discussions and Future work*.

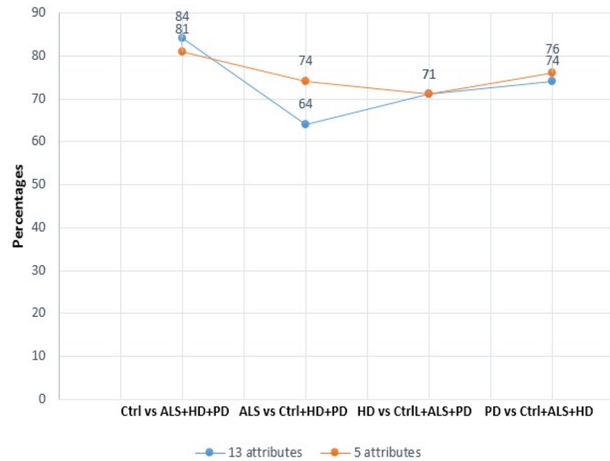

Fig. 3. OVA approach: graphical of higher percentages with 13 and 5 attributes.

Table 4. Comparison of multiclass classification reported in the state of the art using the GaitND-DB

Year	Study	Classifier	Higher accuracy
2009	Yang *et al.* [29]	SVM	86.85
2011	Banaie *et al.* [4]	Quadratic Bayes	86.95
2012	Iram *et al.* [16]	Quadratic Bayes Normal Classifier	65.00
2013	Dutta *et al.* [9]	ERNN	87.10
2015	Zeng and Wang [30]	Deterministic learning theory	93.75
2015	Xia *et al.* [28]	LOOCV	96.83
2017	Ren *et al.* [26]	Empirical Mode Decomposition method	AUC of 0.949
2017	Bilgin [6]	DWT and Naïve Bayesian Classifier	90.93
2018	Our result	ANN	84.00

5 Discussions and Future work

For a long time, machine learning has made notable progress in the data classification process, mostly due to the efforts of the scientific community to develop better algorithms, whether binary or multiclass approaches. In the medical area, the gait classification of neurodegenerative diseases has been addressed by machine learning, reaching good percentages in different controlled experiments [29,4,16,9,30,28], with values ranging from 65% to 96.83% accuracy. Specifically in this research, we obtained about 84% of correct classification. In this sense, we showed that with the traditional classification methods the limit has been reached, concerning percentages of correct classification using one machine learning algorithm. With the aim to improve results, we propose the study of multiclass classification in the gait of neurodegenerative diseases with four prospects as follows:

1. To implement an alternative data preprocessing technique, for example, creating new features.
2. To combine algorithms for classification (ensemble methods).
3. To use *Deep Learning* approach in GaitND-DB, since this technique has led to good percentages of prediction in different areas, such as sentimental analysis [2], business intelligence [8], astronomy (classification of supernovas) [7], robotic grasps detection [19], Person's exposure time to ultraviolet rays prediction [5] and other fields [1]. According to LeCun *et al.*: Deep learning "allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction" [18].
4. To create new or experiment with available datasets containing gait data of patients with neurodegenerative diseases.

We hypothesize that points 1–3 above may allow to diminish the error margin, but the corresponding experiments must be conducted.

Acknowledgment

We would like to thank Oscar Chávez-Bosquez for the technical support conducting the tests and Juana Canul-Reich for the technical advises.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, M., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2016), software available from tensorflow.org
2. Araque, O., Corcuera-Platas, I., Sánchez-Rada, J.F., Iglesias, C.A.: Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications **77**, 236–246 (2017). https://doi.org/https://doi.org/10.1016/j.eswa.2017.02.002
3. Baldridge, J.: Simple code for trying out tensorflow with simulated datasets. GitHub Inc. (2015), ANN algorithm available at `https://github.com/jasonbaldridge/try-tf/blob/master/hidden.py`
4. Banaie, M., Pooyan, M., Mikaili, M.: Introduction and application of an automatic gait recognition method to diagnose movement disorders that arose of similar causes. Expert Systems with Applications **38**(6), 7359–7363 (Jun 2011). https://doi.org/10.1016/j.eswa.2010.12.091, `http://dx.doi.org/10.1016/j.eswa.2010.12.091`
5. Barrera, J.F.P., Hurtado, D.A., Moreno, R.J.: Prediction system of erythemas for phototypes i and ii, using deep-learning. Vitae, ISSN 0121-4004 / ISSNe 2145-2660 **22**(3), 188–196 (2016)

6. Bilgin, S.: The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects. Biomedical Signal Processing and Control **31**, 288–294 (2017)
7. Charnock, T., Moss, A.: Deep recurrent neural networks for supernovae classification. The Astrophysical Journal Letters **837**(2), L28 (2017)
8. Chong, E., Han, C., Park, F.C.: Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications **83**, 187 – 205 (2017). https://doi.org/https://doi.org/10.1016/j.eswa.2017.04.030
9. Dutta, S., Chatterjee, A., Munshi, S.: Hybrid correlation-neural network synergy for gait signal classification. In: Advances in Heuristic Signal Processing and Applications, pp. 263–285. Springer (2013)
10. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256 (2010)
11. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: International conference on artificial intelligence and statistics. pp. 249–256 (2010)
12. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000), circulation Electronic Pages: `http://circ.ahajournals.org/content/101/23/e215.full.pdf+html?sid=9dfce19a-f77d-4d0b-96d6-9c9e819ea980`. Database: `http://www.physionet.org/physiobank/database/gaitndd/`
13. Hass, C.J., Malczak, P., Nocera, J., Stegemoller, E.L., Shukala, A.: Quantitative normative gait data in a large cohort of ambulatory persons with parkinson's disease. PloS one **7**(8), e42337 (2012)
14. Hausdorff, J.M., Ladin, Z., Wei, J.Y.: Footswitch system for measurement of the temporal parameters of gait. Journal of biomechanics **28**(3), 347–351 (1995)
15. Hernández-Torruco, J., Canul-Reich, J., Frausto-Solís, J., Méndez-Castillo, J.J.: Feature selection for better identification of subtypes of guillain-barré syndrome. Computational and mathematical methods in medicine **2014** (2014)
16. Iram, S., Al-Jumeily, D., Fergus, P., Randles, M., Hussain, A.: Computational Data Analysis for Movement Signals Based on Statistical Pattern Recognition Techniques for Neurodegenerative Diseases. In: Proceedings of the 13th Annual Post Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting (PGNet2012). pp. 1–6 (2012)
17. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI. vol. 14, pp. 1137–1145. Stanford, CA (1995)
18. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
19. Lenz, I., Lee, H., Saxena, A.: Deep learning for detecting robotic grasps. The International Journal of Robotics Research **34**(4-5), 705–724 (2015)
20. Liu, H., Li, J., Wong, L.: A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. Genome informatics **13**, 51–60 (2002)
21. Liu, Y., Schumann, M.: Data mining feature selection for credit scoring models. Journal of the Operational Research Society **56**(9), 1099–1108 (2005)

22. Makihara, Y., Mannami, H., Tsuji, A., Hossain, M.A., Sugiura, K., Mori, A., Yagi, Y.: The ou-isir gait database comprising the treadmill dataset. IPSJ Transactions on Computer Vision and Applications **4**(0), 53–62 (2012), `http://www.am.sanken.osaka-u.ac.jp/BiometricDB/index.html`
23. Pérez-Sanpablo, A.I., González-Mendoza, A., Quiñones-Uriostegui, I., Rodríguez-Reyes, G., Núñez-Carrera, L., Hernández-Arenas, C., Boll-Woehrlen, M.C., Montero, A.A.: Evidence-based design and development of a vr-based treadmill system for gait research and rehabilitation of patients with parkinson's disease. Revista de investigación clínica; organo del Hospital de Enfermedades de la Nutrición **66**, s39–s47 (2014)
24. de Pipaón, I.S., Larumbe, R.: Programa de enfermedades neurodegenerativas. In: Anales del Sistema Sanitario de Navarra. vol. 24, pp. 49–76 (2001)
25. Pitt, E., Nayak, R.: The use of various data mining and feature selection methods in the analysis of a population survey dataset. In: Proceedings of the 2nd international workshop on Integrating artificial intelligence and data mining-Volume 84. pp. 83–93. Australian Computer Society, Inc. (2007)
26. Ren, P., Tang, S., Fang, F., Luo, L., Xu, L., Bringas-Vega, M.L., Yao, D., Kendrick, K.M., Valdes-Sosa, P.A.: Gait rhythm fluctuation analysis for neurodegenerative diseases by empirical mode decomposition. IEEE Transactions on Biomedical Engineering **64**(1), 52–60 (2017)
27. Sánchez-DelaCruz, E., Acosta-Escalante, F., Boll-Woehrlen, C., Álvarez-Rodríguez, F.J., Hernández-Nolasco, A., Wister, M.A., Pancardo, P.: Categorización de enfermedades neurodegenerativas a partir de biomarcadores de la marcha. Komputer Sapiens **2**, 16–19 (2015)
28. Xia, Y., Gao, Q., Ye, Q.: Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: experiments with statistical features and different classification models. Biomedical Signal Processing and Control **18**, 254–262 (2015)
29. Yang, M., Zheng, H., Wang, H., McClean, S.: Feature Selection and Construction for the Discrimination of Neurodegenerative Diseases Based on Gait Analysis. In: International Conference on Digital Object Identifier. pp. 1–7 (2009)
30. Zeng, W., Wang, C.: Classification of neurodegenerative diseases using gait dynamics via deterministic learning. Information Sciences **317**, 246–258 (2015)
31. Zheng, S., Zhang, J., Huang, K., He, R., Tan, T.: Robust view transformation model for gait recognition. In: Image Processing (ICIP), 2011 18th IEEE International Conference on. pp. 2073–2076. IEEE (2011), `http://www.cbsr.ia.ac.cn/users/szheng/?page_id=71`
32. Zheng, Z., Wu, X., Srihari, R.: Feature selection for text categorization on imbalanced data. ACM Sigkdd Explorations Newsletter **6**(1), 80–89 (2004)