Neural feature extraction for the analysis of Parkinsonian patient handwriting

Neural feature extraction for the analysis of Parkinsonian patient handwriting

Vincenzo Randazzo[1], Giansalvo Cirrincione[2,3], Annunziata Paviglianiti[1], Eros Pasero[1], Francesco Carlo Morabito[4]

[1] Politecnico di Torino, DET, Turin, Italy
{vincenzo.randazzo,annunziata.paviglianiti,
eros.pasero}@polito.it
[2] University of Picardie Jules Verne, Lab. LTI, Amiens, France
exin@u-picardie.fr
[3] University of South Pacific, SEP, Suva, Fiji Islands
{giansalvo.cirrincione}@usp.ac.fj
[4] Mediterranea University of Reggio Calabria, DICEAM, Reggio Calabria, Italy
{morabito}@unirc.it

Abstract. Parkinson's is a disease of the central nervous system characterized by neuronal necrosis. Patients at the time of diagnosis have already lost up to 70% of the neurons. It is essential to define early detection techniques to promptly intervene with appropriate therapy. Handwriting analysis has been proven as a reliable method for Parkinson's disease diagnose and monitoring. This paper presents an analysis of a Parkinson's disease handwriting dataset in which neural networks are used as a tool for analyzing the problem space. The goal is to check the validity of the selected features. For estimating the data intrinsic dimensionality, a preliminary analysis based on PCA is performed. Then, a comparative analysis about the classification performances of a multilayer perceptron (MLP) has been conducted in order to determine the discriminative capabilities of the input features. Finally, fifteen temporal features, capable of a more meaningful discrimination, have been extracted and the classification performances of the MLP trained on these new datasets have been compared with the previous ones for selecting the best features.

Keywords: Biplot, Feature Extraction, Handwriting, Intrinsic Dimensionality, MLP, Parkinson, PCA.

1 Introduction

Neurodegenerative diseases (NDD) [1] are a group of diseases of the central nervous system characterized by neuronal necrosis, which leads to an inevitable and irreversible damage of brain functions. The causes of the onset are still unclear [2]. For sure, several factors, such as genetic or environment, contribute to one another in giving rise to the pathology [3]. NDD follow a progressive course that is phenotypically highlighted when the anatomical brain damage is in an advanced stage: on average, the patient at

the time of diagnosis has already lost up to 70% of the neurons, thus reducing the possibility of therapeutic intervention effectively [4]. It is essential to define reliable early detection techniques to promptly intervene with appropriate therapy that can be more effective the more the neuronal destruction mechanism is in the early stages. The disabling forms arising from NDD, such as Alzheimer's, Parkinson's, Huntington's chorea and Amyotrophic Lateral Sclerosis (ALS), are characterized by the slow and progressive loss of one or more functions of the nervous system. Parkinson's disease (PD) [5], [6] is a degenerative disease of the central nervous system that affects muscle control, and therefore can influence movement, speech and posture. It is often characterized by muscle stiffness, tremor, slowing of physical movement, and in extreme cases, loss of physical movement. From a pathological point of view, it does not exist a reliable method for an objective and quantitative diagnosis of Parkinson's disease.

Human beings' skills are strongly related to their state of health; indeed, cognitive functions are closely linked to aging processes. Particularly, calligraphy and speech are motor control tasks performed by our brain, therefore the degradation of these abilities implies a neurological deterioration. Handwriting signals are useful for diagnostic and disease monitoring applications. Several tests [7], e.g. house drawing, can be performed in order to check the status of an NDD disease. One of the most effective studies for Parkinson's disease diagnosis concerns the analysis of patient calligraphy [8]. Indeed, it is usually characterized by the development of micrographia, which is a reduction in the size of the writing, and other deficits regarding geometry, kinematics, pressure patterns and air movement [9], [10].

Feature extraction and feature selection techniques have been used to process handwriting signals. A popular approach for PD detection from handwriting consists in extracting kinematic features, which can be either a single value or a sequence of values extracted through time [11]. On one side, feature transformation strategies, such as principal component analysis (PCA) [12] and independent component analysis [13], involve a transformation of the original inputs and produce a set of new variables. On the other hand, feature selection approaches reduce the dimensionality of the input data, removing the irrelevant features and retaining the original interpretations of inputs. A comparative analysis of these techniques and their application to handwriting of people affected from Parkinson's disease is presented in [14]. [15] proposes an experimental analysis of ANOVA ([16]), which is a technique used to determine whether differences in two or more datasets are statistically significant. [17] suggests another feature selection approach based on Support Vector Machine (SVM), with Radial Basis Function (RBF) as kernel [18], which is used as a classifier to predict class labels, in particular to discriminate the task samples into two classes (PD and healthy).

In classification applications, attributes selected from initial dataset are given as input to the classification algorithms. According to [19] attributes that can better distinguish between classes (high-level attributes) are more important than the others in terms of performance. In the ReliefF algorithm [20], attributes are selected according to their suitability with target function; the principle is similar to the basic rules of k-NN algorithm. [21] proposes a simple and fast feature selection algorithm, Sequential forward selection (SFS), based on a greedy search algorithm. It extracts the subset of features by maximizing the efficiency of the feature subset.

2 Experimental Setup

The dataset has been built collecting data from 36 Parkinsonian subjects (18m and 18f, aged between 33 and 83 years old) and 10 healthy subjects (6m and 4f, aged between 49 and 67 years old) recruited at the Matarò Hospital in Barcelona. Every sick person was observed before and after the daily drug (L-dopa COMT catecolo–metal transferasi) administration. Unfortunately, we do not have access to patient clinical information such as Parkinson's disease rating scale part III, levodopa equivalent daily dose, etc. All the patients were right-handed. 22 of these had attended primary school (21 PD/1 Healthy), 17 secondary school (9 PD/8 Healthy), 6 University (5 PD/1 Healthy) and one had not attended any academic studies. Participants were individually tested in a laboratory free of auditory and visual disturbances.

At the beginning of the experiment, the study was explained to the participants and then they underwent a task concerning the writing of the sentence "*La casa de Barcelona es preciosa*" (in Spanish, the native language of the participants). Handwriting collection and analysis has been performed using a digitizing tablet with an ink pen. This approach has an advantage over the classic method based on handwriting and posterior scanning: the machine, actually, can record the pen pressure on the tablet and acquire the information even ''in the air'', that is, where there is no contact between the pen and the surface. The data acquisition was made by means of a tablet, specifically an Intuos Wacom digitizer, which acquired 100 samples per second (total number of samples amounts to around 244K). The acquired features are the same of [22]: X and Y pen positions (the spatial coordinates), Altitude (the angle between the pen and the tablet surface along the vertical), Azimuth (the horizontal angle between the pen and the tablet surface) and the pen pressure on the tablet surface.

3 The Proposed Approach

This paper presents an analysis of a Parkinson's disease handwriting dataset in which neural networks are used to describe the problem. The goal is not the classification in itself, but the validity of the corresponding selected features. Indeed, it is assumed that the best description of the phenomenon should correspond to the best possible classification. In this sense, it can be argued that neural networks are here used as a tool of exploratory data analysis.

This study requires a preliminary analysis (here based on a linear one), in order to have a first insight on the database and, particularly, on its intrinsic dimensionality.

4 Linear Analysis of the Dataset

The manifold of the proposed dataset has been deeply analyzed using the Principal Component Analysis (PCA) in order to understand its intrinsic dimensionality and select the best feature subset. The former has been studied using Pareto diagrams [23],

the latter with biplots [24]. The whole dataset (*H-Pre-Post*), i.e. healthy subjects together with sick patients before and after the drug treatment, has been projected using PCA. The corresponding Pareto diagram is shown in Fig. 1. It displays the explained data variance by the principal component (PC). The bars represent the associated singular values. Fig. 1 shows the importance of the first four components. They explain 88.47% and suggest the intrinsic dimensionality of the manifold is around five.

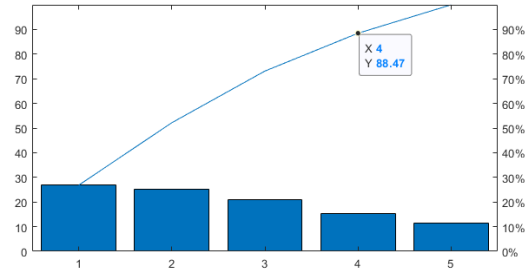

Fig. 1. Pareto diagram on whole dataset, *H-Pre-Post*.

4.1 Biplots

Additional information from the linear analysis of data can be retrieved from a biplot. It is a graphic representation which allows to display, at the same time, both samples and variables of a data matrix. By means of PCA, it is possible to show both the data projected into the principal component space together with the input variable directions. Fig. 2 shows the biplot computed on the whole dataset after being projected with PCA. Although data appear to be clustered along the third principal component, it is not clear which are the features that discriminate and explain the three clusters of data (healthy, pre-treatment, post-treatment).

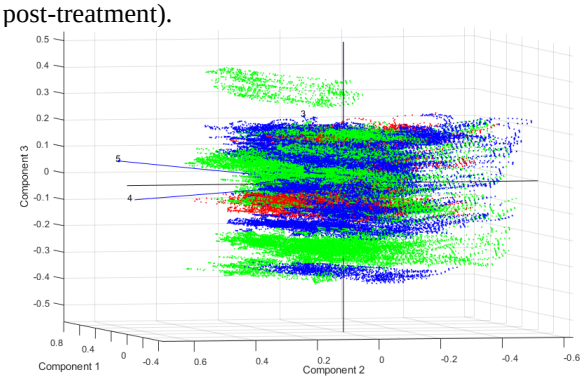

Fig. 2. Biplot on *H-Pre-Post*: healthy (red), pre-treatment (green), post-treatment (blue).

In order to determine which is this subset of features, three new datasets have been created:

1. *H-Pre*: Healthy and pre-treatment subjects.
2. *H-Post*: Healthy and post-treatment subjects.
3. *Pre-Post*: Pre-treatment and post-treatment subjects.

The former, *H-Pre*, has been analyzed in Fig. 3 (left). It can be noticed that the first two input variables (blue directions 1 and 2 in the figure) are nearly parallel to the first two axis, PC1 and PC2, while the rest is explained by the last principal component, PC3. This behavior can be clearly understood by looking at Fig. 3 (right), which is a zoom near the origin. Here, it is evident that the first two components of the PCA projection represent the first two input variables (X and Y pen positions); in fact, it is possible to directly read the original subject handwriting "*La casa de Barcelona es preciosa*". Although the direction of maximum variance, i.e. PC1, obviously follows the X component (writing from left to right), the most significative feature is the Y pen position; indeed, as shown in Fig. 3, this direction clearly discriminates between the healthy and the pre-treatment clusters.

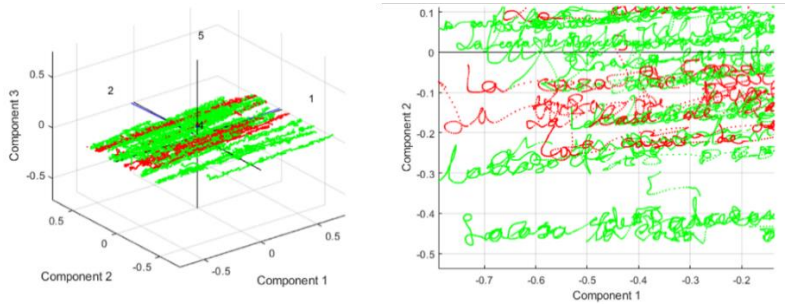

Fig. 3. Biplot on *H-Pre*: healthy (red), pre-treatment (green): whole (left), zoom (right).

Fig. 4 (left) shows the biplot for the second subset *H-Post*. The first two input features behave as in the previous case, while, in this case, the remaining three are, also, meaningful for distinguishing between the clusters. Indeed, Fig. 4 (right), which is the Z-view of the same biplot, shows that the clusters are linearly separated along PC3.

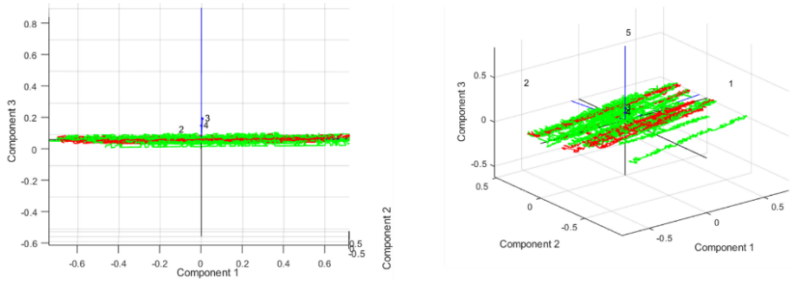

Fig. 4. Biplot on H-Post: healthy (red), post-treatment (green): whole (left), Z-view (right).

The biplot for the last subset, *Pre-Post*, is shown in Fig. 5. As in the previous cases, the first feature (X pen position), is able to fully explain the clusters while along the second one (Y pen position) is possible to discriminate between the clusters. The main difference with the previous cases is that their directions are slightly rotated with regard

to the first two PCs; it can derive from the absence of the healthy cluster. The remaining features are quite useless because the manifold is nearly a hyperplane.

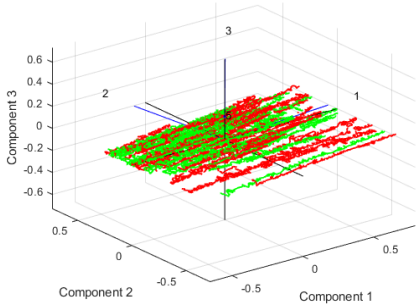

Fig. 5. Biplot on Pre-Post: pre-treatment (red), post-treatment (green).

In conclusion, it can be stated that the selected features represent only approximately the data manifold. The first two PCs roughly coincide with the X and Y pen positions, which is obvious because most variance in writing is in these two directions. Hence, the most meaningful information should stem from the other three components, which, as seen in Fig. 2 and Fig. 4, do not discriminate well enough. It can be argued that Y pen position ability to discriminate among clusters is related to vertical micrographia and with the activation of interphalangeal and metacarpophalangeal joints. This idea is worth to be further deepened and it will be explored in a future work.

5 Neural Classification

A comparative analysis about the classification performances of a multilayer perceptron (MLP) has been conducted in order to determine the discriminative capabilities of the input features. The MLP has been chosen because it is well-suited for pattern recognition [23]. At this purpose, it has a single hidden layer, composed of twenty neurons, and output units equipped with the soft-max activation function [23]. Because of the use of the cross-entropy error function, they yield the probability of membership for the following classes: healthy, pre-treatment, post-treatment. The input layer is mapped one-to-one to the input features; hence, it is always composed of five neurons. The MLP has been trained, by using the Scaled Conjugated Gradient technique [23], both on the whole dataset (three-neurons output layer) and on three subsets (two-neurons output layer) defined in the previous section; then, for each of these training sets, fifteen statistical features, based on the temporal behavior, have been extracted and fed to other MLPs to check their classification performances. Due to the absence of clinical information, in all the experiments, labels (healthy, pre-treatment, post-treatment) were used to split the input dataset into training, validation and test subsets such that their distribution over the labels (healthy, pre-treatment, post-treatment) was always balanced.

5.1 Raw features

The first experiment deals with data drawn directly from *H-Pre-Post*. Each record has been labelled according to the cluster it belongs: healthy, pre-treatment, post-treatment. The resulting set is a matrix made of five columns and as many rows as the number of samples (~ 244K). 70% of this set, i.e. the training set, has been fed to the MLP. The overall accuracy is 77.9%.

The second experiment deals with data drawn from the *H-Pre* subset. Only two labels have been used: healthy and pre-treatment. The input matrix has about 134K samples; as before, 70% is used for training and the rest is divided in equal parts between test and validation sets. An overall accuracy of 95.9% is reached. This classification is very accurate, which is obvious because healthy and sick patients have a significantly different motor control and, so, handwriting.

The experimental setup for the MLP trained on the *H-Post* dataset (~129K examples) is the same as before: two output classes (healthy and post) and 15% of input data used, respectively, for testing and validating. Compared to the previous experiment, the overall test performance decrease to 95.0%. However, this is not a negative result; indeed, it suggests that, after drug treatment, some patients have recovered enough to be confused with the healthy ones.

The last experiment regards the MLP trained on the *Pre-Post* subset. The dataset is made of around 224K records. Two class labels have been chosen: pre-treatment and post-treatment. The classification is worsened with regard to the previous methods (83.2%). Obviously, this is the most difficult pair of classes to be discriminated. All the patients are sick; as a consequence, their handwritings have similar characteristics. Unfortunately, Parkinson's disease treatments are not very effective yet, so, even after drug administration, improvements are quite limited especially when the pathology is, already, in an advanced stage. Another possible way of interpreting it, is that, maybe, patients are in early stages of PD, therefore the effect of levodopa is not so significant.

Resuming, it can be observed that the healthy state is the easiest to classify, because it is based on very peculiar values of the features. It can be used as a basis for determining if the post-treatment state tends to an improvement for the patient, in the sense that data post drug administration yield values of the features closer to the healthy state ones.

5.2 Temporal features

The data manifold analysis in Sec. 4 and the previous subsection (5.1) have proven that the initial set of features was not able to distinguish properly the three clusters of subjects. Therefore, a new set of features, capable of a more meaningful discrimination have been proposed. The idea is to exploit their temporal content; fifteen temporal features have been extracted from each record of the four previous datasets (*H-Pre-Post*, *H-Pre, H-Post, Pre-Post*). The selected features are the following: mean, max value, root mean square (RMS), square root mean (SRM), standard deviation, variance, shape factor (with RMS), shape factor (with SRM), crest factor, latitude factor, impulse fac-

tor, skewness, kurtosis, normalized 5th central moment, normalized 6th central moment. Then, the comparative analysis about the classification performances of the multilayer perceptron has been repeated for each of the four new datasets: *H-Pre-PostT, H-PreT, H-PostT and Pre-PostT.*

For all the following experiments the chosen MLP has a single hidden layer, composed of forty neurons and an input layer of fifteen units. The rest of the setup is the same as the previous section.

The first experiment deals with data drawn directly from *H-Pre-PostT*. Each record has been labelled according to the cluster it belongs: healthy, pre-treatment, post-treatment. The resulting set is a matrix made of five columns and as many rows as the number of samples (~ 244K). 70% of this set, i.e. the training set, has been fed to the MLP. The overall accuracy is 99.3%, that is an 27% increase.

The second experiment deals with data drawn from the *H-PreT* subset. As before, only two labels have been used: healthy and pre-treatment. The input matrix has the same size of the raw case (*H-Pre*); again, 70% of data are used for training and the rest is divided in equal parts between test and validation sets. Despite this classification is more accurate (99.2%) than its corresponding raw case, the overall accuracy is not significantly improved (3%). The considerations done for *H-Pre* also hold for this experiment.

In the third experiment, the MLP has been trained using the *H-PostT* dataset (~129K examples). The experimental setup is the same as before: two output classes (healthy and post) and 15% of input data used, respectively, for testing and validating. The overall test reaches its maximum (100%) with an increase of 5.3%. It is worth of notice that, in this case, the network does not confuse patients who have recovered with the healthy ones. It may suggest that even if the handwritings are closer to normality, the temporal features are now able to discriminate from the healthy case.

The validity of the proposed approach is proved by the last experiment, which regards the MLP trained on the *Pre-PostT* subset. Two class labels have been chosen: pre-treatment and post-treatment. A dataset hard to cluster (83.2% of accuracy) like *Pre-Post*, is now perfectly learnt (100% of accuracy) by the classifier, with an increase of performance of more than the 20%.

TABLE I. MLP PERFORMANCE AND CLASSIFICATION

	# Epochs	***Final Error***	***% Training***	***% Test***
H-Pre-Post	990	0.18	77.8	77.9
H-Pre-PostT	1000	0.01	99.3	99.3
H-Pre	629	0.57	96.0	95.9
H-PreT	831	0.013	99.3	99.2
H-Post	497	0.07	94.8	95
H-PostT	1000	0.0008	100	100
Pre-Post	972	0.175	83.5	83.2
Pre-PostT	1000	0.0004	100	100

Some resuming considerations (see Table I) can be added. Considering that the input layer requires fewer units in case of raw features and the neural network is fully

connected, the use of temporal features requires more *epochs* for training. However, the final training error is several orders of magnitude smaller than in the raw case. This observation is enforced by the classification rates and proves that the temporal model represents better the database (the cross-entropy error yields the correlation between data and model). Hence, the temporal features describe better the phenomenon. This approach justifies the medical consideration of the importance of the temporal behavior in the handwriting.

6 Conclusions

Parkinson's disease is hard to diagnose timely. Indeed, when symptoms are evident, 70% of neurons are already compromised. Techniques for early detection are essential to intervene with appropriate therapies. Handwriting analysis has proved to be a reliable tool for Parkinson's disease diagnose. Starting from a Parkinson's disease database collected at Matarò Hospital in Barcelona, multiple features sets have been extracted and compared in order to select the best feature subset. The PCA-based analysis has shown that the dataset lays on a five-dimensional manifold and the raw features have been studied. Then, a comparative analysis based on an MLP has proved temporal features to be both a reliable model of the Parkinson's disease dataset and more effective in discriminating the different sub-clusters, with an upper bound performance of 100% and a final training error of 0.0004.

Future works will deal with a more accurate analysis of the post-treatment cluster, in order to assess the response of the patient. Also, a non-linear study can be performed for determining the shape of the cluster manifolds in a more accurate way.

Acknowledgment

A special thanks to Prof. Marcos Faundez-Zanuy of the Escola Universitària Politècnica de Mataró and Prof. Anna Esposito of Università degli Studi della Campania for providing the original dataset.

References

[1] M. Heemels, «Neurodegenerative diseases,» Nature, vol. 539, n. 7628, 2016.

[2] A. Gitler, P. Dhillon e J. Shorter, «Neurodegenerative disease: models, mechanisms, and a new hope,» Disease Models & Mechanism, pp. 499-502, 2017.

[3] J. Cambier, M. Masson e H. Cambier, Neurologia, Elsevier.

[4] B. Rizzi, P. Mori, A. Scaglioni, A. Mazzucchi e M. Rossi, La malattia di Parkinson Guida per pazienti e familiari, Fondazione Don Gnocchi, 2013.

[5] J. Parkinson, «An Essay on the Shaking Palsy,» *The Journal of Neuropsychiatry and Clinical Neurosciences.*

[6] R. Pahwa, K. Lyons e W. Koller, Handbook of Parkinson's Disease: Neurological Disease & Therapy, New York: Marcel Dekker Inc, 2003.

[7] A. Larner, «Addenbrooke's cognitive examination-revised (ACE - R) in day today clinical practice,» *Age Ageing*, vol. 36, n. 6, 2007.

[8] M. Lepelley, F. Thullier, B. Bolmont e F. Lestienne, «Age-related differences in sensorimotor representation of space in drawing by hand,» *Clinical Neurophysiology*, 2010.

[9] S. Pinto e J. Velay, «Handwriting as a marker for PD progression: a shift in paradigm,» *Neurodegenerative Disease Management*, vol. 5, n. 5, 2015.

[10] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh e I. Schlesinger, «Handwriting as an objective tool for Parkinson's disease diagnosis,» *Journal of Neurology*, vol. 260, n. 9, 2013.

[11] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal e M. Faundez-Zanuy, «A new modality for quantitative evaluation of Parkinson's disease: In-air movement,» in *13th IEEE International Conference on BioInformatics and BioEngineering*, Chania, 2013.

[12] H. Hotelling, «Analysis of a complex of statistical variables into principal components,» *Journal of Educational Psychology*, vol. 24, 1933.

[13] W. Lu e J. C. Rajapakse, «Approach and applications of constrained ICA,» *IEEE Transaction on Neural Network*, vol. 16, n. 1, 2005.

[14] P. Drotár, J. Mekyska, Z. Smékal, I. Rektorová, L. Masarová e M. Faundez-Zanuy, «Contribution of Different Handwriting Modalities to Differential Diagnosis of Parkinson's Disease,» in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, Turin, 2015.

[15] Z. A. Bakar, D. I. Ispawi, N. F. Ibrahim e N. M. Tahir, «Classification of Parkinson's disease based on Multilayer Perceptrons (MLPs) Neural Network and ANOVA as a feature extraction,» in *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, Melaka.

[16] L. Zoubek, «Introduction to Educational Data Mining Using MATLAB».

[17] C. Taleb, M. Khachab, C. Mokbel e L. Likforman-Sulem, «Feature selection for an improved Parkinson's disease identification based on handwriting,» in *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, Nancy, 2017.

[18] V. Vapnik, Statistical Learning Theory, London: John Willey & Sons, 1998.

[19] M. Peker, A. Arslan, V. Şen, F. V. Çelebi e A. But, «A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+RF),» in *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, Madrid, 2015.

[20] I. Kononenko, «Estimating attributes: Analysis and extensions of RELIEF,» in *European Conference on Machine Learning*, Catania, 1994.

[21] T. Liogienė e G. Tamulevičius, «SFS feature selection technique for multistage emotion recognition,» in *3rd IEEE Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, Riga, 2015.

[22] M. Faundez-Zanuy, A. Hussain, J. Mekyska, E. Sesa-Nogueras, E. Monte-Moreno, A. Esposito, M. Chetouani, J. Garre-Olmo, A. Abel, Z. Smekal e K. Lopez-de-Ipiña, «Biometric Applications Related to Human Beings: There Is Life beyond Security,» *Cognitive Computation*, vol. 5, n. 1, 2013.

[23] C. M. Bishop, «Neural networks for pattern recognition,» *Oxford university press* , 1995.

[24] J. Gower, S. Lubbe e N. Le Roux, Understanding Biplots., Wiley, 2010.