Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 181 (2021) 948–956

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2020

Features Selection Algorithms for Classification of Voice Signals

Letícia Silva[a,b], Bruno Bispo[c], João Paulo Teixeira[a,d] *

[a]*Instituto Politécnico de Bragança, Bragança (IPB), Portugal*
[b]*Federal Technological University of Paraná, Cornélio Procópio Campus, Brazil*
[c]*Department of Electrical and Electronic Engineering, Federal University of Santa Catarina, Florianópolis, Brazil.*
[d]*Research Centre in Digitalization and Intelligent Robotics (CEDRI), Applied Management Research Unit (UNIAG). IPB, Bragança, Portugal*

Abstract

In data mining problems, the high dimensionality of the input features can affect the performance of the process. In this way, the features selection methods appear as a solution to the problems encountered when analyzing databases with large dimensions. This article presents the implementation of the Pearson's linear correlation, ReliefF, Welch's t-test and multilinear regression based algorithms with forwards selection and backward elimination direction for the selection of acoustic features for the task of voice pathologies identification. The best set of selected features improved the accuracy and F1-score from 83% to 92% (9 points of percentage), using the ReliefF algorithm.

Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2020

Keywords: Pearson correlation; ReliefF; Welch's t-test; multilinear regression analysis; forward selection; backward elimination.

* Corresponding author. Tel.: +351 273 30 3129; fax: +351 273 30 3051.
E-mail address: joaopt@ipb.pt

Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2020
10.1016/j.procs.2021.01.251

1. Introduction

Since the early 1970s, feature selection has been the subject of constant research and development in the areas of data mining and reduction, machine learning and pattern recognition [1]. The selection of features is used to solve problems such as high dimensionality, overfitting risk and biased results, which are in sets of searches with excessive information due to the presence of many features [2, 3]. The learning algorithms are affected by redundancy and low relevance in input data [4, 5].

Therefore, the selection and removal of less relevant characteristics from the initial data set reduces the size of the data, computational expenditure and results in the improvement of the predictors' forecast accuracy [2]. The selection of features has three dimensions: the direction in which the search will be carried out, the strategy used in the search and the stopping criterion [6].

The search direction is related with the definition of the starting point. The forward selection is a linear incremental search strategy that selects individual available features, starting with the empty subset and adding one feature at a time [7, 8]. The backward elimination is the opposite of the forward selection approach, since it starts with the total set of features and removes one at a time. The bidirectional selection performs two parallel searches per iteration, one to add a feature and the other to exclude, and it can be advantageous when the number of features of the optimal subset is unknown. The random selection does not have a specific direction in which the search should take place. This approach is to prevent the search from being kept at a local minimum [1].

The search strategy is the second dimension, which can have local or global action. The exhaustive search analyzes all possible combinations of input features and selects the optimal set [7]. The non-deterministic search has in its search process the choice of the optimal subset of features in a random way [9]. The heuristic search is used in optimization problems where the search space is large. This type of algorithms implements a search that combines random evaluation of solutions across the search space with a mechanism to increase the focus of search in regions that lead to good solutions [7]. The sequential search selects one or more features, in a progressive and iterative process. It presents a complete and easy approach to implement [9].

The search stop is the third dimension and different criteria can be used. For example, stop removing or adding features when none of the alternatives improves the accuracy of the classification. Continue to review the feature subset while accuracy still reduced. Stop when the other end of the research space is reached and choose the best of these subsets. Stop when the selected subset of resources separates all classes perfectly. And finally, order the resources according to some importance score that uses a system resource to determine the breakpoint [6].

This paper describes the implementation of some algorithms applied to the selection of acoustic features for identification of pathologic voices using one multi-layer-perceptron (MPL) artificial neural network (ANN) for classification.

Next section presents the theoretical background for search strategy, followed by the description of the implemented algorithms based on selection criteria to sort features. Section 4 presents and discuss the results of the algorithms and discuss the improvement in the classification task, and last section finish with the conclusions.

2. Theoretical Framework - Feature Selection Criteria

The selection criteria present three different approaches: filter, wrapper and embedded methods [2]. In this way, the relationship between the classifier accuracy and the selected subset of features can be assessed independently or depending on the classifier [10]. The criteria for determining the relevant features used are based on the application of techniques such as Pearson's linear correlation, ReliefF, Welch's t-test and multilinear regression analysis.

2.1. Pearson's Linear Correlation Coefficient

One of the simplest filtering schemes is the evaluation of each feature individually, based on its correlation [1, 4]. A good subset of features must have low correlation between the features and a high correlation with the output. In other words, a feature is important if it is correlated with the predictive class; otherwise, it is irrelevant [10, 11].

Thus, Pearson's linear coefficient was used to determine the relevance of the attributes as in [12].

2.2. ReliefF

The ReliefF algorithm developed by Kononenko (1994) is one of the only filter approach algorithms capable of evaluating attribute dependencies [13-15]. It uses the concept of closest neighbors to obtain statistics of attributes that indirectly represent interactions [14].

The algorithm penalizes predictors that assign different values to neighbors of the same class and rewards predictors that assign different values to neighbors of different classes [15]. The weight of each resource reflects its ability to distinguish between class and can vary in the range of -1 to 1. A relevant attribute has positive values [10].

The ReliefF identifies two neighboring observations closest to the target; one with the same class, called the nearest hit occurrence H and the other with the opposite class, called the nearest error occurrence M. The last step of the cycle updates the weight of an attribute A in W, if it has a different value between target observation R_i and H or M observations. The function calculates the difference in the values of elements A between two instances I_1 and I_2 where $I_1 = R_i$ and I_2 is either H or M, when performing updates [14].

2.3. Welch's t-test

The Welch's t-test is a measure that evaluates the subsets of features according to the filter approach. The Welch's t-test is an alternative to the classic t-test [16]. It is used where is not assumed that the two data samples are from populations with equal variations, the statistic test under the null hypothesis has an approximate t-student distribution with a number of degrees of freedom given by the Satterthwaite approximation [16]. The Welch's t-test defines the t statistic as in Equation (1) [17]:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \tag{1}$$

where $\bar{x}$ and $\bar{y}$ are the sample means, S_x and S_y are the standard deviations of the sample, and n and m are the sample sizes. If it is assumed that the two data samples are from populations with equal variations, the test statistic under the null hypothesis has Student's t distribution with v degrees of freedom. The degrees of freedom are determined by Equation (2), where $v_1 = N_1 - 1$ and $v_2 = N_2 - 1$ are the degrees of freedom related to the first and second estimated variations [18], N_1 and N_2 are the length of each group, and s_1 and s_1 their standard deviation.

$$v = \frac{\left(\frac{s^2{}_1}{N_1} + \frac{s^2{}_2}{N_2}\right)}{\frac{s^4{}_1}{N^2{}_1 v_1} + \frac{s^4{}_2}{N^2{}_2 v_2}} \tag{2}$$

After calculating v and t, these values can be used with the t distribution to test the null hypothesis that the two means are equal [18].

2.4. Multilinear Regression Analysis

Stepwise regression is a systematic method for adding and removing features from a multilinear model based on its statistical significance in a regression. While a simple regression of two variables results in the equation that represents a line, a problem of three variables implies a plane, and a problem of k variables implies a k hyperplane. A multilinear model can be represented by Equation (3) [19]:

$$Y_c = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \tag{3}$$

where, a corresponds to the y-axis intercept, b_i corresponds to the weight of the i-th feature, k corresponds to the number of independent features, and Y is the output, or depend variable.

The multilinear regression method employs an embedded approach in the evaluation of features. It starts with an initial set of weights and then compares the explanatory power of increasingly larger and smaller models. At each stage, the p-value of an F-statistic is calculated to test models with and without a certain candidate feature. If the feature is not currently in the model, the null hypothesis is that the feature would have a zero coefficient if added. If there is sufficient evidence to reject the null hypothesis, the feature will be added. On the other hand, if a feature is currently in the model, the null hypothesis is that the term has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the feature will be removed [20].

3. Methodology

The acoustic features used in this work were extracted from the voice recordings of the SVD [21]. It contains 20 features, namely, the absolute jitter (Jittta), relative jitter (Jitt), absolute shimmer (ShimdB), relative shimmer (Shim), autocorrelation, harmonic to noise ratio *(*HNR), noise to harmonic ratio (NHR) [22] and 13 mel frequency cepstral coefficients (MFCC). Each parameter was determined from 3 vowels (/a/, /i/ and /u/) pronounced with 3 tones (high, normal and low), resulting in a total of 180 features [23]. This set of features was retrieved from the cured database of features [24, 25].

Each feature vector were pre-processed using a boxplot outlier's identification and filled with the feature limit value, and a z-score normalization [26]. The classification is performed by an ANN with MLP architecture. It has up to 180 input nodes (depending of the number of features) and one node at the output layer. The network output performs the binary classification with 0 (healthy) and 1 (pathologic) [27-28].

The architecture of the ANN were experimentally achieved by the comparison with different combinations of training function, activation functions and number of neurons in the hidden layer. The one that allowed for best classification accuracy was selected. The training function chosen was the back propagation of the conjugated gradient. The activation function in the hidden layer is the tangent sigmoidal transfer function and the linear transfer function at the output layer [27].

In this article, vocal fold paralysis was used as it is the pathology with the largest number of individuals in the database. The patients were subdivided into two classes, namely, healthy (194 subjects) and pathological (169 subjects), in a total of 363 subject. Classes can be considered approximately balanced.

The output was forced to be 0 or 1 in a post-processing procedure. The data set was divided into three subsets: training, validation and testing with 70%, 15%, 15%, respectively, according to Table 1.

Table 1 – Number of subjects of data used in the classification with MLP.

Classifier	Classes	Total	Training	Validation	Test
MLP	Control	194	136	29	29
	Paralysis	169	119	25	25
	Control x Paralysis	363	255	54	54

The algorithms to apply feature selection techniques described previously were implemented. These techniques will be characterized in terms of strategy, search direction, stop criteria and in assessing performance.

3.1. Pearson's Linear Correlation Coefficient Algorithms

Three algorithms were developed with different approaches (stop criterion) for the Pearson's linear correlation Coefficient (C1, C2, C3). The C1 algorithm is a progressive sequential selection method and uses linear correlation to determine the relevance of the parameters. For a parameter to be considered relevant, it is necessary that it be strongly correlated with the output. At the end, a set of parameters is returned that will be used at the entrance of the neural network. The search space starts empty ($i = 0$) and at each iteration a parameter is increased at the ANN input. The stopping criterion is reached when all features were included ($i = 180$). The performance of the model is obtained by calculating the accuracy. A new accuracy (in the test set) value is calculated and stored for each iteration. After completing the algorithm, the subset with the maximum accuracy in the test is selected.

The C2 algorithm is a sequential selection method, where the linear correlation determines the relevance of the parameters. Regarding the search direction, the algorithm has two variants: forward selection and backward elimination. The main difference in relation to the previous algorithm is that C2 has a different stopping criterion, since it is intended to select smaller representative subsets of parameters and reduce the search time. Regarding the stopping criterion, two variants were implemented: for $i = 3$ and for $i = 20$. The algorithm seeks maximum accuracy and its position. After identifying them, it continues looking for up to i more features, even without improvement in accuracy. The process ends when the maximum accuracy value along next i features (accmx) is less than the previous accuracy (accold).

The C2 algorithm considers relevant feature when it has high correlation with the output. The performance of the model is obtained by calculating the accuracy. A new accuracy value is calculated and stored for each iteration (for the test set). After completing the algorithm, the subset with the maximum accuracy is selected.

The C3 algorithm is the sequential selection method that uses linear correlation to determine the relevance of the parameters. Unlike C2, the C3 algorithm considers a relevant attribute if it has a high correlation with the output and a low correlation between attributes (attribute-attribute). Regarding the search direction, it has two variants: forward selection and backward elimination. The main difference with respect to the C1 algorithm is that C3 has a different stopping criterion, since it is intended to select smaller representative subsets of features. Regarding the stopping criterion, similar to C2, two variants were implemented: with $i = 3$ and $i = 20$.

3.2. ReliefF Algorithm

The R1 algorithm is a progressive sequential selection method that uses ReliefF to determine the relevance of the parameters. For a parameter to be considered relevant, it must have large positive weight values. After applying the ReliefF algorithm, a set of parameters is returned, which will be used in the ANN input in a progressive sequencial method. The R1 algorithm works similarly to the C1 algorithm, differing only in the way it orders the features with interest.

3.3. Welch's t-test Algorithm

The WT1 algorithm is a progressive sequential selection method where the determination of the relevance of the parameters is based on Welch's t-test, classifying the best features based on their ability to separate the classes. A relevant feature is one that has high statistical significance according to Welch's t-test (a large positive value for the p-value variable). At the end, a set of features is returned and ordered, which will be used at the entrance of the ANN.

3.4. Multilinear Regression Analysis Algorithm

The MR1 algorithm is a progressive sequential selection method (similar to C1). The determination of the relevance of the features is based on a multilinear regression analysis. At each stage, the p-value of a statistic F test is determined to test the model's performance. For a feature to be relevant, it must have high statistical significance. The algorithm considers that the typical criterion for a parameter to enter the model has a *p-value* less than 0.05 and for a parameter to leave it has a *p-value* greater than 0.10. Subsequently, a set of features is returned and used at the entrance of the ANN.

The Algorithm 1 pseudocode describes the basic functioning of the C1, R1, WT1 and MR1 algorithms.

Algorithm 1 (C1, R1, WT1 and MR1)

```
1-  Input: initial set of parameters P = (180 parameters x 363 subjects);
2-  Determination of the importance of the parameters: according to the criterion of C1, R1, WT1 and MR1;
3-  Ordering: the parameters of P are ordered according to the importance of the parameter in descending order;
4-  Output: set of ordered parameters P0;
5-  Inicialization: n=180; i=0; acc=0; accF=[ ]; accmx = [ ];
6-  Sequential Selection: the elements of P1 are incremented
    While i < n
        Increments i
        P1 = P0 (1: i);
        Create the ANN with input P1
        Trains and determines the accuracy in Test set
```

```
            accF(i) = acc;
            i = i + 1;
        End While
            [accmx, pos] = max (accF);
            Features = 1: pos;
            acc = accmx;
7-      Finish.
```

4. Results and Discussion

To evaluate the results only subjects in the test set (not seen during training process) were used. Two measures are presented: the accuracy, given by the relation between correct decision by the total number of subject (decisions), and F1-score given by Eq. 4, where P is the precision and R the recall. Accuracy can be unreliable if the dataset were not balanced (this is not the case as shown in table 1), in this case The F1-score is more reliable.

$$F1 - score = 2 * \frac{P*R}{P+R} \qquad (4)$$

Several MLP ANNs configurations were tested and the best architecture was chosen based on the accuracy in training and validation sets. For C1 and C2 Algorithms, the tangent sigmoidal and linear activation functions were selected for hidden and output layers. The scaled conjugate gradient backpropagation training function were selected.

Table 2 presents the results for C1 and C2 algorithms for the best architecture. The [I, H, O] represents the number of nodes in the input, hidden and output layers of the ANN.

It can be seen that C1 algorithm performed better than C2, with an accuracy and F1-score over 90% using 170 features. Regarding C2, it is visible a stop very early both in forward selection (few number of features) and backward elimination (large number of features), although *i* between 3 and 20 features.

In the analysis using the C3 algorithm with *i*=3, different correlation threshold between features (CTresh) are tested. It is the limit to consider a feature singular (can be selected), if it has a feature-feature correlation lower than CTresh. The results are shown in Table 3. The architecture of the ANN has 25 nodes in the hidden layer, one node at the output and the number of features in the input layer.

Regarding the C3 algorithm, the best results were obtained with *i*=3, forward selection and CTresh 0.7, which resulted in the selection of 7 features and reached 89.55% accuracy and 90,01% F1-score.

Table 2 – Comparison between C1 and C2 algorithms.

Algorithm	**C1**	**C2 (i=3)**		**C2 (i=20)**	
Direction	FS	FS	BE	FS	BE
[I, H, O]	[170,20,1]	[3,25,1]	[174,25,1]	[3,25,1]	[159,25,1]
# of Features	170	3	174	3	159
Accuracy	90.59	87.08	85.37	85.57	85.57
F1-score	90.64	87.02	85.75	86.95	86.95

FS – Forward Selection, BE – Backward Elimination.

Table 3 – Comparison between C3 algorithm with i=3 and i=20.

Algorithm	**C3 (i=3)**		**C3 (i=3)**		**C3 (i=3)**		**C3 (i=20)**		**C3 (i=20)**		**C3 (i=20)**	
CTresh	<0.7		<0.8		<0.9		<0.7		<0.8		<0.9	
Direction	FS	BE	FS	BE	FS	BE	FS	BE	FS	BE	FS	BE
# Features	7	177	8	7	177	8	19	165	6	19	165	6
Accuracy	89.55	87.36	85.57	89.55	87.36	85.57	88.49	87.04	88.27	88.49	87.04	88.27
F1-score	90.01	88.38	86.95	90.01	88.38	86.95	88.73	87.71	89.17	88.73	87.71	89.17

FS – Forward Selection, BE – Backward Elimination.

When comparing all selection algorithms based on correlation, it can be concluded that C1 presented higher accuracy in the classification. Therefore, the search method that considers the entire set of attributes proved to be more efficient. This result is more or less expected because it experiments the all combination of features organized by its correlation, meanwhile the other algorithms (C2 and C3) stop when no improvement appends between last 3 or 20 inserted/removed features. However, this algorithm (C1) can be very time consuming compared with C2 or C3.

Table 4 – Comparison between algorithms C1, R1, TW1 and RM1.

Algorithm	No Selection	C1	R1	WT1	MR1
# Features	180	170	30	92	80
Accuracy	83.10	90.59	92.21	90.75	90.79
F1-score	83.61	90.64	92.34	90.83	90.65

FS – Forward Selection, BE – Backward Elimination.

Finally, Table 4 presents the comparison between C1, R1, TW1 and RM1 algorithms with no selection method (all features). The all algorithms used the same ANN architecture with 20 nodes in the hidden layer, one node at the output and the input with the experimented number of features. They have the tangent sigmoidal transfer function in the hidden layer and linear function at the output layer, and were trained with scaled conjugate gradient backpropagation.

After comparing all the selection methods developed, the best results were obtained by the R1 algorithm, which selected through forward selection 30 features and obtained 92.21% accuracy and 92.34% F1-score.

Table 5 shows the set of selected features by the best algorithm (R1). The table shows the identification of the tones (high, low and normal) for the vowels (/a/, /i/ or /u/) and the features extracted (first column).

According to the list of features, in general the MFCCs were the most selected. Among the vowels, the most frequently selected vowel was /a/. Some parameters were not selected even once such as Jitt (relative jitter), HNR, NHR, MFFC2.

Table 5 – Comparison between the C1, R1, TW1 and RM1 algorithms.

Vowels	/a/	/i/	/u/
Jitt			
Jitta		H	
Shim			H
ShimdB	L		
Autocor.	H		H
HNR			
NHR			
MFCC1	N		N
MFCC2			
MFCC3		N	N
MFCC4		L	
MFCC5	L		
MFCC6		L	N
MFCC7		H	H
MFCC8			N
MFCC9	L, N	N	
MFCC10	L, N	H, N	L
MFCC11	N	L	
MFCC12	H, L	L	
MFCC13			N

Tones: L - Low, N - Normal, H - High.

5. Conclusions

The paper presents the implementation of several feature selection algorithms. Forward selection and backward elimination of features were applied. Three methods to stop inserting/removing features were tested. Four methods to sort features by its importance were used. The algorithms were tested with a set of 180 acoustic features for the classification between healthy/pathologic voices using an ANN for classification.

The most efficient search algorithm was the one that considers all parameters in an exhaustive search space, against the other search methods with different stopping criteria. When comparing all the algorithms developed with respect to the accuracy and F1-score in the test set, it can be highlighted that the ReliefF algorithm presented the best performance with a final accuracy of 92.21% and 92.34% F1-score. There was an increase of 9 percentage points in accuracy and 8 percentage points in F1-score, respectively, in relation to the method without parameter selection. Thirty parameters were selected through this algorithm.

However, it must be considered that the search space is relatively small, that is, 180 input features. If the number of input features were much higher, the processing time required to select the parameters would be too long. Therefore, in this case, the C3 algorithm with i=3 and attribute-attribute correlation (CTresh) less than 0.7 are indicated (accuracy = 89.55%, F1-score = 90.01%). This algorithm improved in 7 and 6 percentage points in accuracy and F1-score, respectively, against the method without features selection. Seven parameters were selected through progressive selection.

As a final conclusion, the R1 and C3 method (i = 3) with forward selection and attribute-attribute correlation less than 0.7 are recommended to select the parameters of the data sets for the recognition of vocal pathology.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UIDB/05757/2020.

References

[1] Lee, H. D. (2005), “Seleção de atributos importantes para a extração de conhecimento de bases de dados”. Universidade de São Paulo.

[2] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. In L. P. Kaelbling (Ed.), Journal of Machine Learning Research - JMLR (Vol. 3).

[3] Rodrigues, P. M.; Teixeira, João Paulo. (2010) “Classification of Electroencephalogram Signals Using Artificial Neural Networks”. Proceedings of 3rd International Conference on BioMedical Engineering and Informatics (BMEI’10).

[4] Teixeira, J. P. & Freitas D. (2003). "Segmental Durations Predicted With a Neural Network", Proceedings of Eurospeech’03 – International Conference on Spoken Language Processing, Geneva. Pages 169-172.

[5] Fernandes, C. A. F. S. (2017), “Algoritmo do Tipo Filter-Wrapper de Seleção de Features para Utilização na Seleção de Genes “. Dissertação, Universidade de Coimbra.

[6] Rodrigues, P. M. & Teixeira, J. P., (2013). “Alzheimer’s Disease Recognition with Artificial Neural Networks” - chapter 7 (pag. 102-119) of the book “Information Systems and Technologies for Enhancing Health and Social Care”, by Ricardo Martinho, Rui Rijo, Maria Manuela Cunha and João Varajão. IGI Global. DOI: 10.4018/978-1-4666-3667-5.

[7] May, R., Dandy, G., & Maier, H. (2011), “Review of input variable selection methods for artificial neural networks”. University of Adelaide.

[8] Doak, J. (1992), “An evaluation of feature selection methods and their application to computer security permalink”. Comp. Science, 1–90.

[9] Kumar, V., & Minz, S. (2014), “Feature selection: a literature review. International Journal of Industrial and Systems Engineering”, 4(3), 211–229.

[10] Kohavi, R., & John, H. G. (1997), “Artificial intelligence wrappers for feature subset selection”. Artificial Intelligence, 97(1–2), 273–324.

[11] Hall, M. A. (1999), “Correlation-based Feature Selection for Machine Learning”. University of Waikato.

[12] Gibbons, J. D., & Chakraborti, S. (2003), “Nonparametric Statistical Inference”. In Fundamentals of Biostatistics (4 ed.).

[13] Cavique, L., Mendes, A. B., Funk, M., & Santos, J. M. A. (2013), “A feature selection approach in the study of azorean proverbs”. In Exploring Innovative and Successful Applications of Soft Computing (pp. 38–58).

[14] Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018), “Relief-based feature selection: Introduction and review”. Journal of Biomedical Informatics,Volume 85, Pages 189-203.

[15] Kononenko, I., Šimec, E., & Robnik-Šikonja, M. (1997), “Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF”. Applied Intelligence, 7(1), 39–55.

[16] Welch, A. B. L. (1947), “The Generalization of Student ’s’ Problem when Several Different Population Variances are Involved. Biometrika”. 34(12), 28–35.

[17] Williams, R. B. (1984), “Introduction to Statistics”. Macmillan International Higher Education.

[18] Gliozzo, J. (2016), “Network-based methods for outcome prediction in the sample space.” . Retrieved from http://arxiv.org/abs/1702.01268.

[19] Song & Song, X., & Lu, H. (2017), "Multilinear regression for embedded feature selection with application to fMRI analysis". 31st AAAI Conference on Artificial Intelligence, AAAI 2017, (2), 2562–2568.

[20] Draper, N.R. and Smith, H. (1998), “Applied Regression Analysis”. 3th Edition, Wiley, New York.

[21] Barry, W. J., & Pützer, M. (2018), “Saarbrücken Voice Database”. Retrieved from http://www.stimmdatenbank.coli.uni-saarland.de

[22] Fernandes, J., Teixeira, F., Guedes, V. Junior, A. & Teixeira, J. P. (2018). “Harmonic to Noise Ratio Measurement - Selection of Window and Length”, Procedia Computer Science - Elsevier. Volume 138, Pages 280-285. https://doi.org/10.1016/j.procs.2018.10.040

[23] Teixeira, J. P., Fernandes, J., Teixeira, F. & Fernandes, P. O. (2018). “Acoustic Analysis of Chronic Laryngitis – Statistical Analysis of Sustained Speech Parameters”, In proceedings of 11th International Joint Conference on Biomedical Engineering Systems and Technologies, pp 168-175.

[24] Fernandes, J., Teixeira, F., Fernandes, P. & Teixeira, J. P. (2018), "Cured Database of Sustained Speech Parameters for Chronic Laryngitis Pathology", In Proceedings of 31st International Business Information Management (IBIMA) Conference. Milan.

[25] Fernandes, J., Cena, L., Teixeira, F., Guedes, V., Santos, J., Teixeira, J. P, (2019), "Parameters for Vocal Acoustic Analysis - Cured Database", Procedia Computer Science – Elsevier.

[26] Silva, L.V., Hermsdorf, J., Guedes, V., Teixeira, F., Fernandes, J., Bispo, B., Teixeira, J. P. (2019), "Outliers Treatment to Improve the Recognition of Voice Pathologies". Procedia Computer Science, Volume 164, 2019, Pages 678-685.

[27] Teixeira, F. and Teixeira, J. (2020). Deep-learning in Identification of Vocal Pathologies. In Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4 BIOSIGNALS: pages 288-295. DOI: 10.5220/0009148802880295

[28] Teixeira, F., Fernandes, J., Guedes, V. Junior, A. & Teixeira, J. P. (2018) "Classification of Control/Pathologic Subjects with Support Vector Machines", Procedia Computer Science - Elsevier. Volume 138, Pages 272-279. https://doi.org/10.1016/j.procs.2018.10.039