Improving detection of apneic events by learning from examples and treatment of missing data

Hernández-Pereira ELENA [a,1], Álvarez-Estévez DIEGO [b] and
Moret-Bonillo VICENTE [a]

[a] *Department of Computer Science, Faculty of Informatics, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain*
[b] *Sleep Center, Medisch Centrum Haaglanden, Lijnbaan 32, 2512VA The Hague, The Netherlands*

Abstract. This paper presents a comparative study over the respiratory pattern classification task involving three missing data imputation techniques, and four different machine learning algorithms. The main goal was to find a classifier that achieves the best accuracy results using a scalable imputation method in comparison to the method used in a previous work of the authors. The results obtained show that the Self-organization maps imputation method allows any classifier to achieve improvements over the rest of the imputation methods, and that the Feedforward neural network classifier offers the best performance regardless the imputation method used.

Keywords. Medical Decision-Support Systems, Apneic events classification, Missing data imputation, Machine learning

Introduction

Medical decision-support systems (MDSS) have always played an important role in medical practice. The literature provides the major source of knowledge accompanied by local and practice-based evidence [1]. The knowledge of MDSSs exists in the form of guidelines. There are several areas in the process of developing a guideline-based decision-support system [2], and examples of this kind of systems [3,4,5]. An alternative to guideline-based approach is machine learning. Instead of the required process from guideline modeling to execution, these approaches gain knowledge automatically from clinical data and then use the knowledge to provide decision-support. The MDSS can help to physicians in the diagnosing of any disorder using clues obtained from signal or images taken from subject having the disorder. The objective of this work is in the field of the diagnosis of the sleep apnea-hypopnea syndrome (SAHS). In particular a machine learning MDSS is designed to distinguish sleep apneic events (apneas and hypopneas) from normal breathing.

---

[1]Corresponding Author: Departament of Computer Science, Faculty of Informatics, 15071, A Coruña, Spain; E-mail: elena.hernandez@udc.es

Sleep apnea is defined as a pause in breathing, or cessation of the airflow in the respiratory tracts, of at least 10 seconds in duration. The event is defined as a hypopnea when, rather than a complete cessation, a considerable reduction occurs in the airflow accompanied by a desaturation of oxygen levels in arterial blood. In addition, a micro-arousal happens during sleep that is related to the resolution of this apneic events. Since these micro-arousals happen at each event, the physiological structure of sleep becomes fragmented. The involuntary periodic repetition of these respiratory pauses constitutes one of the most frequent sleep disorders: the sleep apnea-hypopnea syndrome. The most effective method for the SAHS diagnosis is made on the basis of the analysis of a nocturnal polysomnogram, defined as a continuous and simultaneous recording during sleep of a set of variables including airflow in the upper air tracts, oxygen saturation in arterial blood (SaO2) and respiratory effort (both abdominal and thoracic). Following conventional clinical criteria, the apneic episodes are detected in the airflow signal, using the information derived from the electrophysiological and oxygen saturation signals as context for interpretation [6].

Diagnosis models in Sleep Apnea-Hipopnea Syndrome are usually constructed from records that include the polysomnogram information. However, clinical information databases commonly contain missing values or incomplete data that reduce the number of available cases for analysis or might distort the analysis by introducing a bias into the estimation and/or prediction process. In many cases using the simple and common strategy to deal with absent values continues to involve directly ignoring them. Such deletion can introduce substantial biases in the study, especially when missing data are not randomly distributed. Several strategies inspired in statistics and machine learning have been developed to address this problem.

In this work a machine learning MDSS for apneic event identification is presented. Three well-known methods, i.e. mean, hot-deck and Self-organization maps (SOM) are used to impute absent values in the data set and several linear and non linear models are applied to classify respiratory patterns as apneas, hypopneas or normal breathing. The paper is structured as follows: a review of works focused on the identification of apneic events is given in section 2; section 3 describes the materials and methods used in the research, section 4 presents the results obtained and finally, a discussion and the conclusions are presented in section 5.

1. Background

Several attempts have been found in the literature for dealing with the identification of individual apneic episodes. In [7] an on-line signal classification method for the detection of the presence or absence of normal breathing is introduced. Four different artificial neural networks are presented for the recognition of three different patterns in the respiration signals (normal breathing, hypopnea, and apnea). Bystricky and Safer [8] combine neural networks with dynamic Markov models to assign each instant in the ECG signal recording to one of the following four states: "no apnoea", "onset of apnoea", "apnoea" and "end of apnoea". In this proposal, a neural network is employed to extract a set of morphological characteristics from the beats on the basis of the ECG signal. These characteristics constitute the input to a dynamic Markov model which only contemplates a sequence of transitions permitted between the four aforementioned states. Al-Ani et

al. [9] also use Markov models in the detection of apnoea episodes, in this case using respiratory flow, oesophageal pressure and gastric pressure signals. The last two measurements require invasive procedures and are not habitually recorded in the polysomnographic study of sleep. Tian and Liu [10] have used a time delay network to identify apneas on the basis of respiratory airflow and SpO2 signals. The neural network inputs are the area and the standard deviation of the respiratory airflow signal; the basal level and desaturation level of the SpO2 signal; and a correlation coefficient between the SpO2 and respiratory airflow signals. Fontenla-Romero et al. [11] propose an ad hoc technique for identifying apneas based on the respiratory airflow signal. They use a mobile window to calculate the absolute value of the difference between the instantaneous value of the respiratory airflow signal and its average value in the window. An adaptive threshold is then applied to the samples of the signal generated in the mobile window to determine whether they correspond with apnea or normal breathing. Polat et al. compare different classifier algorithms to detect the obstructive sleep apnea syndrome which is a particular type of SAHS. The classifier algorithms include C4.5 decision tree, artificial neural network, artificial immune recognition system, and adaptive neuro-fuzzy inference system. The clinical features used were arousals index, apnea-hypopnea index, SaO2 minimum value in stage of rapid eye movement, and percent sleep time in stage of SaO2 intervals bigger than 89%.

Yashar Maali and Adel Al-Jumaily [13] propose a genetic fuzzy approach for detecting apneic events by using airflow, thoracic and abdominal respiratory movement signals and oxygen desaturation as the inputs. In this approach fuzzy rules and weights are generated by genetic algorithm in three steps. In the first step, input biosignals are analyzed and events in each of them are determined. In the second step, the events are evaluated and reasoning units are constructed. In the third step, a genetic fuzzy approach is used for the final evaluating of reasoning units .

The system MIASOFT (Intelligent Monitoring of the Sleep Apnea-Hypopnea Syndrome), developed by the authors, is a comprehensive medical decision-support system for the diagnosis of SAHS [19]. MIASOFT is knowledge-based intensive, and it has been designed to allow explanative capabilities of their results. For that purpose, and with the aim to mimic human handling of generalization and reasoning procedures, MIASOFT has been implemented using a fuzzy logic inference engine to provide judgments on the basis of similarity and approximation. In MIASOFT, to walk-around the problem of missing values, the inference engine makes use of a chaining of different knowledge-bases to account for the situations where different missing attributes can be present [15,16]. Such a solution is far from being optimal and complicates the design when the number of features increases. The scope of this work is to develop a machine learning model that can learn from examples and effectively handle the occurrence of missing values. This approach represents a more straightforward and scalable solution than the one present in MIASOFT. However the question remains of whether such an approach can outperform the results of the first system, and thus the interest to include the MIASOFT system as an additional benchmark.

2. Material and Methods

The objective of this work is to obtain a machine learning model that achieves the best accuracy results in the respiratory pattern identification task. Besides, another goal is to

analyze the improvements in identification accuracy against MIASOFT system results when different algorithms are applied to impute missing data values.

2.1. Data processing

Patient data (PSG recordings) were gathered from the Sleep Health Heart Study (SHHS) [14]. This prospective cohort study was originally implemented to study the consequences of obstructive sleep apnea and other sleep-disordered breathing on the development of cardiovascular diseases. The resulting database was then enabled to be used as a resource for subsequent studies. Further details about the design, motivation and methods of SHHS can be found in [14]. For the purpose of this work a sample of 95 recordings have been randomly selected from this database. Patient demographics from the resulting sample are shown in Table 1. Each recording contains expert consensus on the different events scored by clinicians during the manual offline analysis of the recordings. Annotations regarding the scoring of apneic events include hypopneas, obstructive apneas and central apneas for which onset and duration for each event are specified. These annotations will be used as the standard reference for the validation of our approach.

Number	Male	Age -mean(std)-	AHI -mean(std)-	BMI -mean(std)-
95	49	66.27 (10.02)	46.21 (27.63)	30.29 (6.03)

Table 1. Data set patient demographics.

For the construction of the data set, features are extracted from a subset of PSG signals that involve both respiratory and neurophysiological information. Specifically a total of 9 features are used which are described in Table 2. The process to automatically extract these features from the raw biomedical signals contained in the PSG is described elsewhere. For details the reader is referred to [15,16] in what concerns the extraction of features 1-8 and to [17] for the detection of EEG arousals (feature 9). Also in [15,16] an explanation can be found on how the individual features (1-8) that are extracted from each of the different PSG respiratory channels are then related in time to form what it has been called an apneic pattern (AP), that is, a set of features that together characterize a certain time interval of the PSG and point out to the possible occurrence of an apneic event [15,16]. On the other hand, for the association of an EEG arousal to the AP, the criterion described in [18] is used as reference. Specifically an EEG arousal (detected as described in [17]) is associated with an AP if the arousal begins less than 5 seconds after the end of the AP.

Following the previously described procedures a total of 39.539 patterns have been collected, each one with one possible output namely: (i) normal-respiration, (ii) hypopnea, or (iii) apnea. For the purposes of the validation obstructive apneas and central apneas are grouped together into a single category, namely apnea. The number of each class is 5.436 apnea patterns, 12.078 hypopnea patterns and 22.025 normal-respiration patterns. Occurrence of missing values in the data set is originated in the situations in which a certain feature cannot be evaluated in the context of the corresponding AP. Such a situation is actually common and may be caused by several reasons including presence of artifacts, inaccuracy of the detection algorithm or simply the current physiological condition (for example, a reduction in breathing may manifest differently across the individual respiratory channels). Characterization of the features and their related missingness can be found in Table 2.

Feature	Range	Mean/Mode	Type/scale	Missingness (%)
Desaturation	0-100	2.4673	Quantitative/ratio	2.77
Airflow reduction	0-100	49.7118	Quantitative/ratio	44.18
Abdominal respiration reduction	0-100	54.7780	Quantitative/ratio	33.80
Thoracic respiration reduction	0-100	55.0571	Quantitative/ratio	37.55
Desaturation	0-400	14.2083	Quantitative/seconds	2.77
Airflow reduction	0-400	21.3744	Quantitative/seconds	44.18
Abdominal respiration reduction	0-400	24.3508	Quantitative/seconds	33.80
Thoracic respiration reduction	0-400	25.0748	Quantitative/seconds	37.55
EEG arousal	0-1	0	Qualitative/nominal	0

Table 2. Feature characterization of the data set.

2.2. Data Imputation Methods

Imputation is the process used to determine and assign replacement values for missing data items [21]. Imputation methods are especially useful in situations where a complete data set is required for the analysis. A wide range of methods and tools for data imputation is available. Some methods try to make use of the available information, for example, listwise or casewise data deletion techniques (LD), based on the omission of all those records that contain a missing value for one or more variables. Other methods are proper imputation techniques as they compute appropriate values for replacing the missing data. So, according to their degree of complexity, we have implemented three of these methods: two statistical methods (mean and hot deck) and one machine learning based method (SOM).

- Mean/mode imputation. It's a method where any missing value of a quantitative variable is replaced by the mean of the observed values for that variable. If the variable is qualitative, the missing values are replaced by the mode.
- Hot-deck imputation. Given an incomplete pattern, this method estimates missing values from similar but complete records of the same data set. The similarity criterion used is the heterogeneous Euclidean-overlap metric (HEOM) [22], which uses the so-called overlap metric for categorical attributes and a normalized city-block distance for linear numeric quantitative attributes. The overlap metric is a normalized Hamming distance given as the percentage of coordinates that differ. The HEOM distance is intended to remove the effects of the arbitrary ordering of categorical values, and it constitutes an overly simplistic approach to handling these kinds of attributes.

Consider that a patient case is represented by an n-dimensional input vector, $x = [x_1, x_2, ..., x_n]^T$, and that m is a vector of binary variables such that $m_j = 1$ if x_j is unknown and $m_j = 0$ if x_j is present. Given a pair of patient cases, represented by x_a and x_b, the HEOM distance between them is:

$$d(x_a, x_b) = \sqrt{\sum_{j=1}^{n} d_j(x_{aj}, x_{bj})^2} \qquad (1)$$

where $d_j(x_{aj}, x_{bj})$ is the distance between x_a and x_b on its jth attribute:

$$d_j(x_{aj}, x_{bj}) = \begin{cases} 1 & \text{if } (1 - m_{aj})(1 - m_{bj}) = 0 \\ d_O(x_{aj}, x_{bj}) & \text{if } x_j \text{ is a categorical attribute} \\ d_N(x_{aj}, x_{bj}) & \text{if } x_j \text{ is a quantitative attribute} \end{cases} \tag{2}$$

Unknown data are handled by returning a distance value of 1 (i.e., maximal distance) if either of the input values is unknown. The overlap distance function d_O assigns a value of 0 if the discrete attributes are the same; otherwise, the value is 1. The range normalized difference distance function d_N is given by:

$$d_N(x_{aj}, x_{bj}) = \frac{\mid x_{aj} - x_{bj} \mid}{max(x_j) - min(x_j)} \tag{3}$$

where $max(x_j)$ and $min(x_j)$ are the maximum and minimum values, respectively, observed in the training set for the numerical attribute x_j; thus, the normalization attempts to scale the attribute down to the point where differences are almost always less than one, and the resulting distance matrix is set to range between 0 and 1. The difference $\mid x_{aj} - x_{bj} \mid$ is the city-block distance [20].

- Self-organization maps (SOM). A SOM is a neural network model made out of a set of nodes that are organized on a 2D grid and fully connected to the input layer. Each node has a specific topological position in the grid, as well as a vector of weights of the same dimension used for the input vectors [23]. After the SOM model has been trained, it can be used to estimate missing values. When an incomplete observation is presented to the SOM, the missing input variables are ignored during the selection of the best matching unit (BMU). The incomplete data are imputed by the feature values of the BMU in the missing dimensions [20], as following:

  1. Presentation of an incomplete observation in the input layer.
  2. Selection of the BMU by minimizing the distance between the observation and nodes. Missing components are excluded from the distance calculation.
  3. The replacement value for a missing item in the input vector is the value for that item in the corresponding BMU.

  The SOM imputation approach is implemented using the SOM toolbox [24].

2.3. Classification Methods

In this section, we provide an overview of the methods used in the research for respiratory pattern classification: apnea, hypopnea or normal breathing. Several approaches were considered, two linear models – linear discriminant analysis and a proximal support vector machine–, and two non linear ones – a multilayer feedforward neural network and a classification tree–.

- Linear discriminant analysis
  The linear discriminant analysis is a classification method originally developed by R. A. Fisher [25]. It is simple, mathematically robust and often produces models whose accuracy is as good as more complex methods. It consists of searching, some linear combinations of selected variables, which provide the best separation between the considered classes. These different combinations are called discriminant functions. It assumes that different classes generate data based on different Gaussian distributions [26].

- Proximal Support Vector Machine (pSVM)
  The proximal Support Vector Machine [29] is a method that classifies points assigning them to the closest of two parallel planes (in input or feature space) that are pushed as far apart as possible. The difference with a SVM is that this one classifies points by assigning them to one of two disjoint half-spaces. The pSVM leads to an extremely fast and simple algorithm by generating a linear or nonlinear classifier that merely requires the solution of a single system of linear equations.
- Multilayer Feedforward Neural Network
  The multilayer feedforward neural network as it is one of most commonly used neural network classification algorithms [27]. The architecture used for the classifier consisted of a two layer feed-forward neural network: one hidden and one output layer. It has been demonstrated that, with an appropriate number of hidden neurons, one hidden layer is enough to model any continuos function [28]. The optimal number of hidden neurons for this problem was empirically obtained.
- Classification Trees
  Classification trees are used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on one or more predictor variables. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels [30]. Each internal (non-leaf) node of the tree is labeled with an input feature. The arcs coming from a node labeled with a feature are labeled with each of the possible values of the feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner. The recursion is completed when the subset at a node has all the same value of the target variable, or when splitting no longer adds value to the predictions. This process of top-down induction of decision trees is by far the most common strategy for learning decision trees from data [31].

2.4. Performance measures

After the classifiers were trained, the performance of the system is evaluated in terms of the following measures:

- The classification accuracy, computed as the percentage of correctly classified positive and negative instances.
- The sensitivity quantifies the ability to correctly identify positive instances. It is the proportion of true positives that are correctly identified.
- The specificity quantifies the ability to correctly identify negative instances. It is the proportion of true negatives that are correctly identified.

3. Results

In this section the results obtained after applying missing data imputation techniques and several classifiers, are shown and compared in terms of the three effectiveness measures described in section 2.4. To compare and study the convenience of imputing data, the reference model was first estimated by simply removing missing values from the original

data set; this process is usually described as listwise or case deletion (LD). Then, the methods described in section 2.2 were applied to input absent values, and the classification methods (section 2.3) were used to predict the respiratory patterns.

Table 3 shows the accuracy measures obtained by the selected models over a 10-fold cross validation for the respiratory pattern classification. These results are yield against the standard reference, i.e. the medical expert scores.

	LD	Mean	Hot-deck	SOM
LDA	75.85	76.06	74.13	76.30
pSVM	71.84	74.25	72.85	74.77
FNN	**80.28**	**78.61**	**79.50**	**81.15**
Class. Tree	72.66	76.65	74.68	75.74

Table 3. Respiratory pattern classification results. Mean test set accuracy (%) of a 10-fold cv. Best values marked in bold font

Tables 4 to 6 show the sensitivity and specificity obtained for each of the respiratory pattern.

	Sensitivity				Specificity			
	LD	Mean	Hot-deck	SOM	LD	Mean	Hot-deck	SOM
LDA	92.71	**85.82**	**84.60**	**87.36**	88.38	95.53	93.91	93.80
pSVM	**93.81**	83.48	80.45	82.01	84.50	95.83	94.40	94.19
FNN	87.30	73.03	70.51	80.54	**93.15**	**97.67**	**97.12**	**97.02**
Class. Tree	80.65	74.94	72.83	73.25	91.18	96.68	96.23	96.40

Table 4. Apnea classification results. Mean test set sensitivity and specificity (%) of a 10-fold cv. Best classifier values marked in bold font

	Sensitivity				Specificity			
	LD	Mean	Hot-deck	SOM	LD	Mean	Hot-deck	SOM
LDA	69.37	36.78	42.39	45.88	**82.54**	95.42	91.39	92.26
pSVM	71.37	29.28	31.39	38.62	73.59	**96.50**	**95.00**	**93.60**
FNN	**81.52**	57.24	**64.32**	**65.09**	79.71	91.38	88.93	90.11
Class. Tree	72.42	**60.86**	59.04	59.99	73.86	85.62	84.10	84.91

Table 5. Hypopnea classification results. Mean test set sensitivity and specificity (%) of a 10-fold cv. Best classifier values marked in bold font

	Sensitivity				Specificity			
	LD	Mean	Hot-deck	SOM	LD	Mean	Hot-deck	SOM
LDA	66.35	95.19	88.94	90.26	91.09	61.84	66.95	70.70
pSVM	41.74	**96.64**	**93.71**	**92.81**	**95.36**	55.48	57.45	64.39
FNN	**67.56**	91.70	90.04	90.10	94.39	69.75	**76.70**	**78.74**
Class. Tree	61.84	85.73	83.71	84.98	90.30	**76.31**	75.09	75.90

Table 6. Normal breathing classification results. Mean test set sensitivity and specificity (%) of a 10-fold cv. Best classifier values marked in bold font

The listwise deletion method (LD) is improved by all the imputation methods for the pSVM and the Classification Tree. In the other two classifiers, it doesn't result a good

method too. The mean method offer better results than the hot-deck except for the FNN classifier. In this case the FNN is the best over the two imputation methods but this result is achieved by the classifier not the imputation method. The reason for the mean method to be slightly better than the hot-deck method could be because using the mean/mode value for replacing missing values is more appropriate for the input variables than the HEOM distance. This distance is obtained taking into account all the variables of the example and it seems that not all of them are equally related. Finally, the SOM method is the best one over all the classifiers.

If we analyze these results by the classifier point of view, the FNN outperforms all the other classifiers. These results were obtained with a 9-40-3 model. Several tests were made over the FNN architecture. For the mean imputation method, the best results were obtained with an 9-100-3 FNN. but the improvement over the 9-40-3 model was very small. Among the linear models tested (LDA and pSVM), the LDA performs better using any imputation method. Over the non-linear models, the FNN obtain the best results.

Besides validation against the standard reference comprising expert annotations, results from the presented approach are compared against the performance achieved by the expert system MIASOFT, previously developed by the authors. The accuracy results obtained by MIASOFT were 78.67% and the sensitivity and specificity results, for each respiratory pattern are shown in Table 7.

	Apnea	Hypopnea	Normal respiration
sensitivity	81.22	64.65	85.73
Specificity	96.41	86.68	79.72

Table 7. Respiratory pattern classification results for MIASOFT. Mean test set sensitivity and specificity (%) of a 10-fold cv.

The results obtained by the different classifiers with the four imputation methods used, are not better than the MIASOFT results in terms of accuracy except for the FNN classifier. For this classifier, any of the imputation methods except the mean, outperforms the MIASOFT results.

Analyzing sensitivity and specificity using the mean of the two measures (which is equivalent to the area under the ROC curve with one operation point), against MIASOFT results and over the three respiratory patterns, the following can be stated. For the apnea pattern (Table 4), the lineal classifiers perform better than MIASOFT no matter what imputation method used. The FNN classifier only offers good results if the LD imputation method is used. For the hypopnea pattern (Table 5), the FNN outperforms the MIASOFT results except for the mean imputation method. Finally, for the normal breathing pattern (Table 6) none of the classifiers with any of the imputation methods improve the MIASOFT values.

4. Conclusions

This paper presents a comparative study over the respiratory pattern classification task involving three missing data imputation techniques, and four different machine learning algorithms. The main goal was to find a classifier that achieves the best accuracy results using a scalable imputation method in comparison to the method used by MIASOFT. As we pointed out, in contrast to the data-driven approach followed in this work, MIASOFT

is more knowledge-based intensive, and it has been designed to allow explanative capabilities of their results. But for the respiratory pattern classification task the developed approach seems to be slightly better.

The imputation techniques include two statistical methods, mean and hot-deck, and one machine learning method, SOM. These techniques were compared with the listwise deletion method to show the danger of eliminating records with mission values from the original data set. Such deletion can introduce substantial biases in the study. Once the unknown data were imputed, a classification model was created comparing two linear models, LDA and pSVM, and two non linear ones, FNN and a Classification tree. The results obtained show that the SOM imputation method allows any classifier to achieve improvements over the rest of the imputation methods. Besides, the FNN classifier offers the best performance regardless the imputation method used. It seems that lineal classification methods are not appropriate for the respiratory pattern classification and a deeper work must be done over the FNN architectures and over the SOM imputation method configuration.

Nevertheless, the results obtained in terms of accuracy are not as good as expected. The improvements over MIASOFT results are limited so a deeper study might be done. It would be desirable to study the relationship between the input variables used in this work by means of the use of feature selection methods. Although these methods are commonly applied in data sets with a big amount of variables, they offer potencial benefits as reducing training and utilization times and defying the curse of dimensionality to improve prediction performance. Besides, after applying feature selection, more complex missing data imputation methods could be used.

We conclude that machine learning techniques may be a better approach to imputing missing values, as they led to improvements in prediction accuracy, and in the SAHS diagnosis field these techniques offer better performance in classification tasks. Imputation techniques depend on the available data and the prediction method used; thus, the results obtained might not generalize to different data sets.

References

[1] Sim I., Gorman P., Greenes R.A., Haynes R.B., Kaplan B., Lehmann H. and Tang P.C. *Clinical decision support systems for the practice of evidence-based medicine*. Journal of the American Medical Informatics Association, 8(6), 2001,527-534.

[2] de Clercq P.A., Blom J.A., Korsten H.H. and Hasman A. *Approaches for creating computer-interpretable guidelines that facilitate decision support*. Artificial Intelligence in Medicine, 31(1), 2004, 1-27.

[3] Paul D. Clayton, T. Allan Pryor, Ove B. Wigertz and George Hripcsak. *Issues and Structures for Sharing Medical Knowledge Among Decision-Making Systems: The 1989 Arden Homestead Retreat*. Proc. Annu. Symp. Comput. Appl. Med. Care, 1989, 116–121.

[4] L. Ohno-Machado, J.H. Gennari, S.N. Murphy, N.L. Jain, S.W. Tu, D.E. Oliver, E. Pattison-Gordon, R.A. Greenes, E.H. Shortliffe and G.O. Barnett. *The guideline interchange format: a model for representing guidelines*. Journal of the American Medical Informatics Association, 5(4), 1998, 357-372.

[5] John Fox, Nicky Johns and Ali Rahmanzadeh. *Disseminating medical knowledge: the PROforma approach*. Artificial Intelligence in Medicine, 14(1-2), 1998, 157–182.

[6] R.B. Berry, R. Brooks, C.E. Gamaldo, S.M. Harding, C.L. Marcus, B.V. Vaughn and M.M. Tangredi. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0*. Darien, Illinois: American Academy of Sleep Medicine, 2012.

[7] P. V'arady, T. Micsik, S. Benedek and Z. Benyó. *A novel method for the detection of apnea and hypopnea events in respiration signals*. IEEE Trans. Biomed. Eng., 49(9), 2002, 936-942.

[8] W. Bystricky and A. Safer, *Identification of individual sleep apnea events from the ECG using neural networks and a dynamic markovian state model*. Computers in Cardiology, 2004, pp. 297–308.

[9] T. Al-Ani, Y. Hamam, R. Fodil, F. Lofaso and D. Isabey, *Using hidden markov models for sleep disordered breathing identification*. Simulation Modelling Practice and Theory, 2004, 12(2), 117–128.

[10] J.Y. Tian and J.Q. Liu, *Apnea detection based on time delay neural network*. 27th IEEE EMB Conference, 2005, 2571–2574.

[11] O. Fontenla-Romero, B. Guijarro-Berdiñas, A. Alonso-Betanzos and V. Moret-Bonillo, *A new method for sleep apnea classification using wavelets and feedforward neural networks*. Artificial Intelligence in Medicine, 2005, 34, 65–76.

[12] K. Polat, S. Yosunkaya and S. Gne. *Comparison of Different Classifier Algorithms on the Automated Detection of Obstructive Sleep Apnea Syndrome*. J. Med. Syst. 2008, 32(3), 243-250.

[13] Yashar Maali and Adel Al-Jumaily. *Genetic Fuzzy Approach based Sleep Apnea/Hypopnea Detection*. International Journal of Machine Learning and Computing, 2012, 2(5), 685–688.

[14] SF. Quan, BV. Howard, C. Iber, JP. Kiley, FJ. Nieto, GT. O'Connor, DM. Rapoport, S. Redline, J. Robbins, JM. Samet and PW. Wahl. *The Sleep Heart Health Study: design, rationale, and methods*. Sleep, 1997, 20(12), 1077-1085.

[15] D. Álvarez-Estévez, and V. Moret-Bonillo. *Fuzzy Reasoning Used to Detect Apneic Events in the Sleep Apnea-Hypopnea Syndorme*. Expert Systems with Applications, 2009, 36(4), 7778-7785.

[16] V. Moret-Bonillo, D. Alvarez-Estevez, A. Fernández-Leal, E. Hernández-Pereira, *Intelligent approach for analysis of respiratory signals and oxygen saturation in the sleep apnea-hypopnea syndrome*. The Open Medical Informatics Journal, 2014, 8, 1-19.

[17] D. Álvarez-Estévez, N. Sánchez-Maroño, A. Alonso-Betanzos and V. Moret-Bonillo. *Reducing dimensionality in a database of sleep EEG arousals*. Expert Systems with Applications, 2011, 38(6), 7746-7754.

[18] Sleep Health Heart Study SRC, *Sleep Health Heart Study. Reading center manual of operations*. Case Western Reserve University, Tech Report VMLA-039-02, 2002.

[19] D. Alvarez-Estevez, *Intelligent diagnosis of the Sleep Apnea-Hypopnea Syndrome: a comprehensive approach through an intelligent system to support medical decision*. PhD. Thesis, Department of Computer Science, University of A Coruna, Spain, 2012.

[20] Nuria Ribelles, Miguel Martin and Leonardo Franco. *Missing data imputation using statistical and machine learning methods in a real breast cancer problem*. Artificial Intelligence in Medicine, 50, 2 (October 2010), 105-115.

[21] Roderick J. Little and Donald B. Rubin. *Statistical Analysis with Missing Data (2nd edn)*. John Wiley & Sons, New York, 2002. ISBN: 0-471-18386-5.

[22] D. Randall Wilson and Tony R. Martinez. *Improved heterogeneous distance functions*. Journal of Artificial Intelligence Research, 6, 1 (January 1997), 1-34.

[23] T. Kohonen. *Self-organizing maps 3rd ed.*. Berlin: Springer Series in Information Sciences, Springer-Verlag, 2001.

[24] J. Vesanto et al. *SOM Toolbox for Matlab 5*. Technical Report A57. SOM Toolbox Team, Helsinki University of Technology. in http://www.cis.hut.fi/projects/somtoolbox/.

[25] R. A. Fisher. *The Use of Multiple Measurements in Taxonomic Problems*. Annals of Eugenics, 1936, 7, 179188.

[26] M. S. Srivastava and E. M. Carter, *Applied multivariate statistics*, North Holland Amsterdam, 1983.

[27] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1995.

[28] K. Hornik, M. Stinchcombe, H. White, *Multilayer feedforward networks are universal approximators*, Neural Networks 2 (5) (1989) 359–366.

[29] G. Fung, O. Mangasarian, *Proximal support vector machine classifiers*, in: F. Provost, R. Srikant (eds.), Proceedings KDD-2001: Knowledge Discovery and Data Mining, August 26-29, 2001, San Francisco, CA, Asscociation for Computing Machinery, New York, 2001, pp. 77–86, ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-02.ps.

[30] L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone. *Classification and Regression Trees*. Chapman & Hall, New York.

[31] J. R. Quinlan. *Induction of Decision Trees*. Machine Learning 1, 1986, 81-106, Kluwer Academic Publishers.