2010 22nd International Conference on Tools with Artificial Intelligence

Automated Aortic and Mitral Valves Diseases Diagnosis from Heart Sound Signals Using Novel Ensemble Classification Techniques

Manolis Maragoudakis *Member, IEEE*, Euripides Loukis

***Abstract* - The development of 'intelligent' medical equipment, which can not only acquire various signals from the human body, but also process them and provide recommendations as to probable pathological conditions, will be highly beneficial for both the medical personnel and the patients. However, this necessitates the development and exploitation of advanced highly efficient classification techniques. In this direction this paper presents a novel ensemble classification technique, combining Random Forests with the 'Markov Blanket' notion, which is used for the automated diagnosis of aortic and mitral heart valves diseases from low-cost and easily acquired heart sound signals. It has been tested in a highly 'difficult' global and heterogeneous dataset of 198 heart sound signals, which been acquired from both healthy and pathological medical cases. The proposed ensemble classification technique exhibited a higher classification performance in comparison with the classical Random Forest algorithms, and also other widely used classification algorithms.**

I. Introduction

The development of 'intelligent' medical equipment, which can not only acquire various signals from the human body, but also process them and provide recommendations as to probable pathological conditions, will be highly beneficial for both medical personnel and patients. This capability will be particularly useful in the numerous small primary healthcare centers, which lack sufficient experienced medical personnel, but have however to examine at a first level big numbers of subjects, and identify the ones who need further and more complex examinations in bigger hospitals. The provision of such recommendations will improve the accuracy of these first diagnoses, reducing on one hand the numbers of subjects who proceed to further and more complex examinations without actually needing them, and on the other hand the numbers of subjects who are mistakenly diagnosed as healthy and do not proceed to the required additional examinations with negative consequences for their health. Therefore it will result in significant benefits for the small primary healthcare centers, the hospitals, the national health systems on one hand, and for the patients on the other. However, in order to develop these valuable recommendation capabilities it will be necessary to incorporate in this 'intelligent' medical equipment knowledge of experienced specialized doctors, and also to develop and use advanced highly efficient classification techniques for the optimal exploitation of this knowledge.

This paper contributes in this direction by presenting a novel ensemble classification technique, combining Random Forests with 'Markov Blanket', which performs diagnosis of aortic and mitral heart valves diseases from simple heart sound signals that can be easily acquired at a very low cost at any small primary healthcare center. It consists of three stages: initially it diagnoses the existence or not of murmurs, and in the former case it distinguishes between systolic and diastolic ones, and finally between aortic and mitral origin (in the case of systolic murmurs between aortic stenosis and mitral regurgitation, and in the case of diastolic murmurs between aortic regurgitation and mitral stenosis). For its training it requires a number of heart sound signals for each category (healthy, aortic stenosis, mitral regurgitation, aortic regurgitation and mitral stenosis), which have been correctly characterized (diagnosed) by experienced cardiologists, so it incorporates their specialized knowledge. This knowledge is efficiently transformed from its initial case-based form into the more structured form of an ensemble of decision trees, which can perform diagnosis of a new heart sound signal. The proposed ensemble classification technique has been tested in a highly 'difficult' global and heterogeneous dataset of 198 heart sound signals, which have been acquired from both healthy and pathological medical cases belonging to the above categories. Also, its classification performance is compared with the performance of the classical Random Forest algorithms, and also of other widely used classification algorithm, on the same data set.

The paper consists of six sections. In the following section II previous relevant research is briefly reviewed, while in section III the background is outlined (Random Forests and Markov Blanket). In section IV, the proposed novel 'Markov Blanket Random Forests' ensemble classification technique is presented, followed by the results of its application in the above dataset in section V. Finally section VI summarizes the conclusions.

II. Previous Research

There has been considerable previous research concerning the automated detection of various heart pathological conditions and diseases from heart sound signals. It can be broadly divided into two research streams: the first deals with the removal of noise, segmentation and preprocessing

Manolis Maragoudakis is with the Department of Information and Communication Systems Engineering, University of the Aegean. Phone: +30-22730-82261; fax: +30-22730-82009; e-mail: mmarag@aegean.gr.

Euripides Loukis is with the Department of Information and Communication Systems Engineering, University of the Aegean. Phone: +30-22730-82221; fax: +30-22730-82009; e-mail: eloukis@aegean.gr.

1082-3409/10 $26.00 © 2010 IEEE
DOI 10.1109/ICTAI.2010.110

of heart sound signals, and calculation of appropriate features for their diagnostic exploitation, while the second deals with the detection of heart pathological conditions and diseases. The present paper belongs to the second research stream, so we are going to focus our review on it. Some of the studies of this stream are dealing with distinguishing between normal and abnormal (i.e. from subjects having a disease) heart sound signals [1-4], or with distinguishing between innocent and pathological murmurs in children [5-8]. Other studies are dealing with the more detailed detection from heart sound signals of particular heart diseases, such as coronary artery diseases [9,10] and heart valve diseases or murmurs [11-21]. With respect to the type of classification algorithms used it should be emphasized that in most of the studies of this research stream the diagnostic classification of the heart sound signals is based on neural networks of various types (e.g. back-propagation, radial basis function, probabilistic neural networks etc.) [1,2,4,5,7-12, 13, 16, 17]. However, there are only a few studies using other classification algorithms, such as discriminant functions [6,15], decision trees [19,20], Bayesian networks [3], Support Vector Machines [20] and Hidden Markov Models [21]. Therefore the diagnostic potential of other classifiers than the neural networks for the automated detection of heart pathological conditions and diseases from heart sound signals has not been sufficiently explored yet, so further research is required in this direction.

III. BACKGROUND

Recently, ensemble classification has gained much popularity within the Machine Learning and Data Mining community. The idea behind the aggregation of multiple single classifiers is based on the assumption that non-correlated classifiers have the potential to outperform the total prediction error when aggregated. The following example illustrates how an ensemble method can improve a classifier's performance. Suppose that a set of twenty-five binary classifiers is constructed, each of which predicts the class with an error rate of 0.35. As previously mentioned, an ensemble classifier performs the classification based on the majority vote of each base classifier. In the case that all base classifiers are identical, the error rate of the ensemble will remain 0.35, while, if the base classifiers are independent (their error is not correlated) then the ensemble will make false prediction if more than half of the base classifiers predict wrongly. From a mathematical perception the error rate of the ensemble is calculated using the following equation:

$$\varepsilon_{ensemble} = \sum_{13}^{25} \binom{25}{i} \varepsilon^i (1-\varepsilon)^{25-i} = 0.06\,, \qquad (1)$$

which is significantly lower than the error rate of the base classifiers.

The Random Forest (RF) algorithm is a popular classification technique, whose classifier is an ensemble of classification trees. It is considered particularly well suited to situations characterized by a large number of features, a circumstance that is becoming more prevalent as the ability to collect and store vast amounts of data becomes easier and increasingly common [22]. In such instances, the classical classification approaches tend to become overwhelmed by the number of features and fail, while RF continues to do well. For instance, with DNA microarray data, work by [23-25], shows that RF outperforms most of the other classification techniques. However, when, in addition to having a large number of features, the proportion of truly informative features is small, its performance tends to decline as well [26]. In this paper, a solution for this problem is proposed, which is founded on the notion of a feature selection and reasoning algorithm, based on the *Markov Blanket* of the class attribute. The identification of relevant variables is an essential component of the construction of efficient decision support and computer-assisted discovery models. Especially in the area of biomedicine the problem of variable selection is more pressing than ever, due to the recent emergence of extremely large datasets, sometimes involving tens to hundreds of thousands of variables.

A. Random Forests

A RF classifier consists of a number of decision trees, with each tree grown using some form of randomization. The leaf nodes of each tree are labelled by estimates of the posterior distribution over the data class labels. Each internal node contains a test that best splits the space of data to be classified. A new, unseen instance is classified by sending it down every tree and aggregating the reached leaf distributions. The process is described in Fig. 1. Each tree is grown as follows:

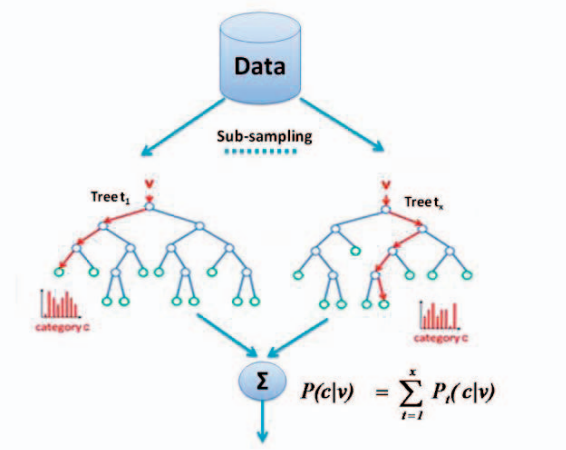

Fig. 1. Hierarchical Decomposition of a Random Forest

- If the number of cases in the training set is *N*, sample *N* cases at random but with replacement, from the original data. This sample will be the training set for growing the tree. This is actually the so-called bagging step [22] and the selected samples are called the in-bag cases; the rest are set aside as *out-of-bag cases*.
- If there are *F* input variables, a number $m<<F$ is specified such that at each node, m variables are

selected at random out of the F and the best split on these m is used to split the node. The value of m is held constant during the forest growing.

Each tree is grown to the largest extent possible; therefore, no pruning procedures need to be applied. Upon completion of the tree construction step, the set of data are traversed down the tree, and proximity values are computed for each pair of cases.

As regards to the overall error rate of the Random Forests, this is affected by two different factors:

1. *Robustness* (strength) of each individual tree within the forest. Higher strength results in lower error rates.
2. Tree *inter-correlation*. Highly correlated trees result in high error rate.

The former, *individual performance*, is obtained by using robust tree classifiers and the latter, *inter-correlation*, is alleviated by randomly choosing cases on which to train each tree and by randomly choosing attributes at each node of each tree. Most classification algorithms tend to overfit when faced with datasets in which there are a large number features. However, a major advantage of RF is that they are able to keep the likelihood of overfitting low by using different subsets of the training data and different subsets of features for training the different base classifiers. Thus, only patterns truly existing in the data would be detected consistently by the majority of the base classifiers. As regards to the most common implementations of "injecting" randomness into the tree-learning phase, two approaches exist:

1) Random input forests

The simplest random forest with random features is formed by selecting at random, at each node, a small group of input variables to split on. Grow the tree using entropy as criterion for splitting to its maximum size and do not prune.

2) Random combination forests

This approach consists of defining more features by taking random linear combinations of a number of the input variables. That is, a feature is generated by specifying L, the number of variables to be combined. At a given node, L variables are randomly selected and added together with coefficients that are uniform random numbers on [-1,1]. F linear combinations are generated, and then a search is made over these for the best split.

B. Bayesian Networks and Markov Blanket

In order to better capture the significant properties of a Markov Blanket (MB), a brief introductory section of Bayesian networks is included. Bayesian networks graphically represent the joint probability distribution over a set of random variables. A Bayesian Network (BN) is composed of a qualitative portion (its structure) and a quantitative portion (its conditional probabilities). The structure BS is a directed acyclic graph where the nodes correspond to domain variables $x_1,\ldots,x_n$ and the arcs between nodes represent direct dependencies between the variables [27]. Likewise, the absence of an arc between two nodes x_i and x_j denotes that x_j is independent of x_i given its parent nodes. Following the notation of [28], the set of parent nodes of a node x_i in BS is denoted as π_i. The structure is annotated with a set of conditional probabilities, containing a term $P(x_i=X_i|\pi_i=\Pi_i)$ for each possible value X_i of x_i and each possible instantiation Π_i of π_i. In a BN, the structure BS encodes the Markov condition if each node x_i is probabilistically independent of all non-descendants given its parent nodes. From this condition, the so-called "chain rule" for BNs follows immediately: a BN can be factorized as a product, for all variables in the network, of their probabilities conditionally on their parents only, i.e.

$$p(x_1 = X_1, \ldots, x_n = X_n) = \prod_{i=1}^{n} p(x_i = X_i | \pi_i = \Pi_i) \qquad (1)$$

A Markov Blanket of a node x_i, denoted as MB(x_i), is a minimal attribute set, containing its immediate parent nodes, its child nodes and the immediate parent nodes of its child odes. Mathematically, the above statement is translated into:

$$\forall x_k \in \{x_1, \ldots, x_n\} \backslash MB(x_i) \cup \{x_i\}, x_i \vdots x_k | MB(x_i) \qquad (2)$$

where $\vdots$ denotes the conditional independence of x_i with x_k given MB(x_i).

Suppose B_i and B_j are two Bayesian networks that have the same probability distribution, then $MB_{Bi}(x_k) = MB_{Bj}(x_k)$ for any variable x_k. Certainly, MBs are not exclusive and may vary in size, but any given BN has a unique MB(x_i) for any x_i, which is the set of parents, children and parents of children of x_i. In Fig. 2, a BN is depicted along with the MB of a target node x, colored in blue. As regards to the dataset interpretation, feature x is independent of all other features given its MB(x)={U_i,U_j,Y_k,Y_l,Z_{km},Z_{ln}}.

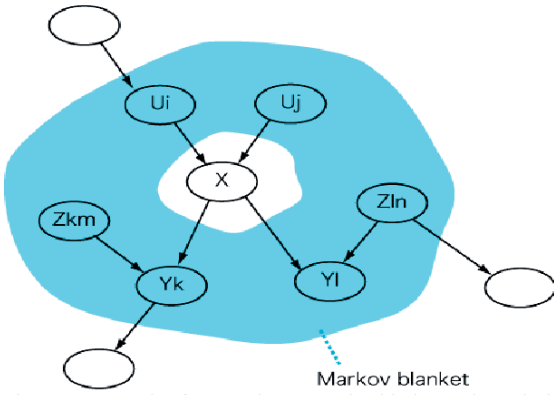

Fig. 2. An example of a Bayesian Network with the Markov Blanket of node x.

IV. Markov Blanket Random Forests

Based on the existing implementations of Random Forests and taking our initial concerns on feature relevance into consideration, we propose a novel algorithm for classification using RF. The algorithm is entitled "*Markov Blanket Random Forests-MBRF*", since the danger of selecting irrelevant and misleading features is remedied by using the Markov Blanket of the class node to provide the best splitting criteria for each tree. By selecting random

samples and obtaining the extracted MB of the target node, the probability of tree containing more informative features is increased. In case of high-dimensional datasets, the diversity of the ensemble is not compromised and is more robust that other, pre-filtering or weighting schemes. The algorithm is consisted of two distinct phases; the former regards the construction of the MB and latter is about constructing the trees. Its basic procedure can be sketched in the following steps:

MBRF (Data D, Features F, Target C)

1. Draw n_{tree} bootstrap samples from the original data D.
2. Build an unconstrained Bayesian network without learning the conditional probability table.
3. Obtain the MB of the class node C.
4. For each of the bootstrap samples, grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, use m_{try} of the Markov Blanket and choose the best split from among those variables.
5. Predict new data by aggregating the predictions of the n_{tree} trees (i.e., majority votes for classification, average for regression).

An important variation of the proposed algorithm in comparison with previous RF implementation is located on the second and third step respectively. In the following section, we provide the mathematical explanation of the aforementioned phases.

A. Determining the Bayesian network structure and the MB of the class node

Based on the framework of [28], the most probable BN structure is obtained by performing a hill-climb search over the space of candidate structures. The following equation along with Bayes' theorem provides a metric of relation r among two candidate network structures BS_i and BS_j respectively:

$$r = \frac{P(BS_i|D)}{P(BS_j|D)} = \frac{\frac{P(BS_i,D)}{P(D)}}{\frac{P(BS_j,D)}{P(D)}} \quad (3)$$

Therefore, the problem of calculating $P(BS|D)$ reduces to that of calculating $P(BS,D)$.

In order to estimate the above probability the following assumptions have to be made [29].

1. Variables are discrete and all are observed (i.e. there are no hidden variables).
2. Database records (cases) occur independently, given a belief network model.
3. There are no cases that have variables with missing values.
4. The density function $f(BS_i|BS_j)$ is uniform, i.e. the prior probabilities to place on a network structure BS is unconcerned.

According to the notation used so far, suppose F be a set of m discrete variables, where a variable x_i in F has r_i possible value assignments: $(v_{i1},...,v_{iri})$. Let D be a database of n cases, where each case contains a value assignment for each variable in F. Let BS denote a network structure containing just the variables in F. Each variable x_i in BS has a set of parents, represented as a list of variables π_i. Let w_{ij} denote the j^{th} unique instantiation of π_i relative to D. Suppose there are q_i such unique instantiations of π_i. Let N_{ijk} be defined as the number of cases in D in which variable x_i has the value v_{ik} and π_i is instantiated as w_{ij}. N_{ijk} is calculated as:

$$N_{ijk} = \prod_{i=1}^{n} p(x_i = X_i|\pi_i = \Pi_i) N_{ij} \sum_{k=1}^{r_i} N_{ij} \quad (4)$$

Then, given the assumptions outlined above,

$$p(BS, D) = \prod_{i=1}^{m} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}\,! \quad (5)$$

Equation 5 can be combined with Equation 4 to give a computable method of comparing the probabilities of two network structures, when given a database of cases for the variables in the structures. Since, by the third assumption listed above, the prior probabilities of all valid network structures are equal, $P(BS)$ is a constant. Therefore, to maximize $P(BS,D)$ just requires finding the set of parents for each node that maximizes the second inner product of Equation 5. The search strategy operates by initially assuming that a node has no parents, and then adding incrementally that parent whose addition most increases the probability of the resulting network. Parents are added greedily to a node until the addition of no one parent can increase the network structure probability. The function used in this procedure is taken from the second inner product of Equation 5:

$$gain(x_i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}\,! \quad (6)$$

In a single iteration, an arc is added to node i from the node z that maximizes $gain(x_i,\pi_i \cup \{z\})$. If $gain(x_i,\pi_i) > gain(x_i,\pi_i \cup \{z\})$ then no arc is added.

Upon creation of the structure, the time-consuming process of calculating the conditional probabilities is not needed, since the proposed methodology is basically focused on identifying the MB of the class node. Therefore, learning of the conditional probability table on the created Bayesian network is skipped and the process of identifying the MB of the class node within the network can be straightforwardly implemented by considering the set of parents, children and parents of children nodes of the class node.

V. Data and Preprocessing

In order to investigate the usefulness and performance of Random Forests for the automatic identification of murmurs in heart sound signals, a global and representative heart sounds dataset has been created with heart sound signals from various different heart sound sources (educational audiocassettes, audio CDs, CD ROMs, files of existing heart sound databases, etc.), which are described in [8-10]. It includes heart sound signals acquired with various types of stethoscopes, sensors and filters, in various modes, subjects' positions and auscultation areas, from subjects of various ages, heart conditions and medical treatments. This dataset is more 'noisy' and therefore more 'difficult' for the classifiers, than the more 'homogeneous' ones used by most similar studies, but it enables a more realistic investigation of classifiers' performance in conditions better approximating the 'real-life' medical practice. For the purposes of the present study from this dataset we used 198 heart sound signals: 38 normal heart sounds, 41 heart sound signals with AS systolic murmur, 43 ones with MR systolic murmur, 38 ones with a AR diastolic murmur and 38 signals with a MS diastolic murmur. Each of these heart sound signals had been diagnosed by a specialized cardiologist and classified to one of the above four basic heart valve diseases. Initially a pre-processing of these heart sounds was performed, in order to remove noise and extract features from them, which consisted of three phases. In the first phase the segmentation of the heart sound signal is performed, i.e. the cardiac cycles in every signal are detected by locating the S1 and S2 peaks. In the second phase, for each of the heart sounds produced in the first phase were calculated the standard deviation of the duration of all the heart cycles it includes, the standard deviation of the S1 peak values of all heart cycles, the standard deviation of the S2 peak values of all heart cycles and the average heart rate; these values are the first four scalar features (F1-F4) of the feature vector of each heart sound signal. In the third phase, the rest of the features used for classification are extracted. For this purpose we calculated for each transformed heart sound signal two mean signals for each of the four structural components of the heart cycle, namely two signals for the S1, two for the systolic phase, two for the S2 and two for the diastolic phase. The first signal focused on the frequency characteristics and was calculated as the mean value of each component, after segmenting and extracting the heart cycle components, time warping them and aligning them. The second signal focused on the morphological time characteristics and was calculated as the mean value of the normalized average Shannon Energy Envelope of each component, after segmenting and extracting the heart cycles components, time warping them and aligning them. The second S1 mean signal is then divided into 8 equal parts, for each part the mean square value is calculated and the resulting 8 values are used as features (F5-F12). Similarly 24 scalar features for the systolic period (F13-F36), 8 scalar features for S2 (F37-F44) and 48 scalar features for the diastolic period (F45-F92) were calculated. Finally the systolic and diastolic phase components of the first mean signal were passed from four band-pass filters: a) a 50–250 Hz filter giving its low frequency content, b) a 100–300 Hz filter giving its medium frequency content, c) a 150–350 Hz filter giving its medium-high frequency content and d) a 200–400 Hz filter giving its high frequency content. For each of these 8 outputs, the total energy was calculated and was used as a feature in the heart sound vector (F93-F100). The above pre-processing produced for each heart sound signal a feature vector consisting of 100 components. These feature vectors of our heart sound signals were used for the identification of murmurs using Random Forests described in the next section.

VI. Experimental Results

In this section, we present the results of the application of the above proposed ensemble classification technique in a dataset consisted of 198 heart sound signals, which have been acquired from both healthy and pathological medical cases, having one of the following four most frequent heart valve diseases: aortic stenosis, mitral regurgitation, aortic regurgitation or mitral stenosis. It should be noted that the acquired heart sounds are affected significantly by several factors related to the acquisition method, such as the type of stethoscope used, the type of sensor that the stethoscope has (e.g. microphone, piezoelectric film, etc.), the stethoscope use mode (e.g. bell, diaphragm, extended), the filtering applied to the heart sound signals (e.g. anti-tremor filter, respiratory sound reduction filter, etc.), the way the stethoscope is pressed on the patients skin (firmly or loosely), the patient's position (e.g. supine position, standing, squatting), the auscultation areas (i.e. apex, lower left sternal border, pulmonic area, aortic area), the medicines that the patient is taking, etc. These factors cannot be controlled in the everyday medical practice, and this adds high levels of noise to the acquired heart noise signals (i.e. generates additional components), making the detection of various heart diseases and pathological conditions from these heart sound signals even more difficult. Therefore an effective system for the diagnosis of heart diseases from heart sounds should cope with this problem. So, in order to make our research more realistic, we decided the above dataset, which we used both for constructing the classifiers and for testing them, to be a 'global' and representative one, consisting of 'heterogeneous' heart sounds recorded with different acquisition methods and different values of the above factors. Such a dataset is much more 'difficult' for the classifiers than a 'homogeneous' one (in which all heart sound have been recorded using the same acquisition method and values of the above factors), however it enables a more realistic investigation of classifiers' construction and performance. Using the 10-fold cross validation method on the aforementioned dataset, the classification performance of the proposed Markov Blanket Random Forests is compared against the classification performance of traditional Random Input and Random Combination Forests, C4.5 Decision Trees, Naïve Bayes, Radial Basis Function Neural Networks, K-Nearest Neighbor (KNN) and Support Vector

Machines (SVM). For each heart sound of this dataset a preprocessing was performed resulting in a feature vector consisting of 100 components. The preprocessing method is described in [19].

We have categorized the experimental results as follows:

(a) Initially, the classifier distinguishes normal (NRM) from sick (SCK) records.

(b) Those instances that belong to the unhealthy class are further classified according to whether they belong to the systolic (STL) or diastolic (DTL) class.

(c) Finally, for each of the aforementioned classes, the classifier identifies between either Aortic Stenosis (AS) or Mitral Regugritation (MR) cases and between either Aortic Regugritation (AR) or Mitral Stenosis (MS).

A. Distinguishing between healthy and unhealthy examples

From a medical expert's perspective, the accuracy of a predictive model in this situation is of major importance, since a misclassification of a sick case as healthy could be fatal for a patient. Thus, the desired classification performance is as high as possible. In the Table 1, results of MBRF are tabulated against all other methodologies described earlier. RF are generally performing better than other approaches, while MBRF portray better performance than the other two RF implementations in a scale of 4%-5.3%. The other classifiers depict rather poor performances.

Table 1. Comparison of selected classifiers for Healthy-Unhealthy data

Methodology	%Accuracy (Healthy-Unhealthy)
Markov Blanket Random Forests (MBRF)	90.22%
Random Input Forests (RIF)	86.15%
Random Combination Forests (RCF)	84.04%
C4.5 Decision Trees (C45)	75,38%
Naïve Bayesian Classifier (NB)	72,31%
Radial Basis Functions (RBF)	77,45%
K-Nearest Neighbor (KNN)	81,53%
Support Vector Machines (SVM)	78.46%

B. Distinguishing between systolic and diastolic murmurs

Since a patient is found to suffer from heart disease, a distinction with regards to the nature of the heart pulse is of great importance. Thus, we evaluate the set of machine learning algorithms, including the MBRF, in an attempt to classify heart murmurs as systolic or diastolic. Table 2 summarizes the percentage of correctly classified instances. The difference of MBRF from all other RF implementations is between 5.5% and 6.3% while the difference from other algorithms surpasses 27% in certain cases. This significant variation could be attributed to the elimination of noisy features from the Random Trees, due to the presence of the Markov Blanket of the class, at each tree learning step.

Table 2. Comparison of selected classifiers for Systolic-Diastolic data

Methodology	% Accuracy (S-D)
Markov Blanket Random Forests (MBRF)	98.67%
Random Input Forests (RIF)	93.12%
Random Combination Forests (RCF)	92.32%
C4.5 Decision Trees (C45)	87.53%
Naïve Bayesian Classifier (NB)	71.87%
Radial Basis Functions (RBF)	80.76%
K-Nearest Neighbor (KNN)	79.37%
Support Vector Machines (SVM)	83.74%

C. Distinguishing between AR-MS and AS-MR diseases

The final round of experimental evaluations focuses on classifying the types of heart diseases, according to the type of heart pulse. Note that when the heart pulse is diastolic, the patient could either suffer from either aortic regugritation (AR) or mitral stenosis (MS). Accordingly, when a patient's heart pulse is systolic, the disease is either aortic stenosis (AS) or mitral regugritation (MR).

As regards to the former case, results are tabulated in Table 3. MBRF is again the most efficient classifier and outperforms all other approaches. The classification accuracy reaches 92.45% while other RF variations present lower results by a factor of 1.67% and 2.8%.

Table 3. Comparison of selected classifiers for AR-MS Diastolic data

Methodology	% Accuracy (D, AR-MS)
Markov Blanket Random Forests (MBRF)	92.45%
Random Input Forests (RIF)	90.78%
Random Combination Forests (RCF)	89.65%
C4.5 Decision Trees (C45)	75%
Naïve Bayesian Classifier (NB)	84.21%
Radial Basis Functions (RBF)	67.10%
K-Nearest Neighbor (KNN)	86.84%
Support Vector Machines (SVM)	89.47%

As regards to the latter case (Table 4), for the first time KNN and MBRF are portraying similar accuracy. This is acceptable as KNN have previously referred as good classifiers for the AR-MS problem. Nevertheless, MBRF are still better classifiers than other RF implementations.

Table 4. Comparison of selected classifiers for AS-MR Systolic data

Methodology	% Accurcay (S,AS-RM)
Markov Blanket Random Forests (MBRF)	90.34%
Random Input Forests (RIF)	86.90%
Random Combination Forests (RCF)	85.33%
C4.5 Decision Trees (C45)	86.90%
Naïve Bayesian Classifier (NB)	62.10%
Radial Basis Functions (RBF)	80.95%
K-Nearest Neighbor (KNN)	90.47%
Support Vector Machines (SVM)	79.76%

VII. Conclusion

A novel ensemble classification technique was presented, extending the traditional RF implementations with 'Markov Blanket', which has been used for the automated diagnosis of aortic and mitral heart valves diseases from low cost and easily acquired heart sound signals.

A first application in a global and heterogeneous dataset of 198 heart sound signals, which have been acquired from both healthy and pathological medical cases having one of the most frequent heart valve diseases (aortic stenosis, mitral regurgitation, aortic regurgitation or mitral stenosis) gave encouraging results, taking into account the high levels of noise that such a heterogeneity in acquisition method and conditions adds, and the difficulties this creates for diagnosis. Furthermore, the proposed ensemble classification technique exhibited a higher classification performance than the traditional RF algorithms, and also other, widely used classifiers. This can be attributed to the incorporation of the Markov-Blanket, thus suppressing poorly-correlated trees.

The above preliminary results provide some first evidence of the potential of RF in general, and the proposed extension of them based on 'Markov Blankets' in particular, for automated medical diagnosis, and for incorporation in 'intelligent' medical equipment, which can not only acquire various signals from the human body, but also process them and provide recommendations as to probable pathological conditions. Further research is required for validation of this potential in other medical problems using various types of signals from the human body and improvement of the proposed technique.

References

[1] Cathers I.: Neural network assisted cardiac auscultation. Artificial Intelligence in Medicine, February 1995, Volume 7, Issue 1, pp. 53-66.

[2] Wu C. H., Lo C. W., Wang J. F.: Computer-aided analysis of classification of heart sounds based on neural networks and time analysis. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signals Processing (ICASSP) 1995, 5, pp. 4355-3458.

[3] Wu C. H.: On the analysis and classification of heart sounds based on segmental Bayesian networks and time analysis. Journal of the Chinese Institute of Electrical Engineering - Transactions of the Chinese Institute of Engineers, Series E November 1997, 4(4), pp. 343-350.

[4] Leung T. S., White P. R., Collis W. B., Brown E., Salmon A. P.: Classification of heart sounds using time-frequency method and artificial neural networks. In Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2000, 2:988-991.

[5] Nakamitsu T., Shino H., Kotani T., Yana K., Harada K., Sudoh J., Harasawa E., Itoh H.: Detection and classification of systolic murmur using a neural network. In Proceeding of the 15th IEEE Southern Biomedical Engineering Conference March 1996, pp. 365-366.

[6] Leung T. S., White P. R., Collis W. B., Brown E., Salmon A. P.: Analysing paediatric heart murmurs with discriminant analysis. In Proceedings of the 19th Annual conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong 1998, pp. 1628-1631.

[7] DeGroff C., Bhatikar S., Mahajan R.: A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics. Artificial Intelligence in Medicine 2005, 33, pp. 251-260.

[8] De Vos J. P., Blanckenberg M. M. Automated pediatric cardiac auscultation. IEEE Transactions on Biomedical Engineering February 2007, 54(2), pp. 244-252.

[9] Xuesong Y., Qiang C. Yuquan C.: Noninvasive detection of coronary artery disease based on heart sounds. In Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 29 Oct. – 1 Nov.1998, 3, pp. 1546-1548.

[10] Tateishi O.: Clinical significance of the acoustic detection of coronary artery stenosis. Journal of Cardiology November 2001, 38(5), pp. 255-262.

[11] Nygaard H., Thuesen L., Hasenkam J. M., Pedersen E. M., Paulsen P. K.: Assessing the severity of aortic valve stenosis by spectral analysis of cardiac murmurs (spectral vibrocardiography). Part I: Technical aspects. Journal of Heart Valve Disease July 1993, 2(4), pp. 454-67.

[12] Hebden J. E, Torry J. N: Identification of Aortic Stenosis and Mitral Regurgitation by Heart Sound Analysis. Computers in Cardiology 1997, 24, pp. 109-112.

[13] Brusco M., Nazeran H.: Development of an Intelligent PDA-based Wearable Digital Phonocardiograph. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference 2005, 4, pp. 3506-3509.

[14] Herold J., Schroeder R., Nasticzky F., Baier V., Mix A., Huebner T., Voss A.: Diagnosing aortic valve stenosis by correlation analysis of wavelet filtered heart sounds. Medical and Biological Engineering and Computing 2005, 43, pp. 451-456.

[15] Voss A., Mix A., Huebner T.: Diagnosing Aortic Valve Stenosis by Parameter Extraction of Heart Sound Signals. Annals of Biomedical Engineering September 2005, 33(9), pp. 1167-1174.

[16] Higuchi K., Sato K., Makuuchi H., Furuse A., Takamoto S., Takeda H. Automated diagnosis of heart disease in patients with heart murmurs: application of a neural network technique. Journal of Medical Engineering and Technology 2006 March-April, 30(2), pp. 61-68.

[17] Ahlstrom C., Hult P., Rask P., Karlsson J. E., Nylander E., Dahlstrom U., Ask P. Feature extraction for systolic heart murmur classification. Annals of Biomedical Engineering 2006 November, 34(11), pp. 1666-77.

[18] Pavlopoulos S., Stasis A., Loukis E.: A decision tree – based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds. BioMedical Engineering OnLine, June 2004.

[19] Stasis, A., Loukis, E., Pavlopoulos, S., Koutsouris, D.: A multiple decision trees architecture for medical diagnosis: The differentiation of opening snap, second heart sound split and third heart sound. Computational Management Science, Springer Verlag, Autumn 2004, pp. 245-274.

[20] Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A., Support Vectors Machine based Identification of Heart Valve Diseases Using Heart Sounds. Computer Methods and Programs in Biomedicine, Volume 95, Issue 1, July 2009, pp. 47-61 .

[21] Chauhan, S., Wang, P., Lim, C. S., Anantharaman, V.: A Computer Aided MFCC based HMM system for automatic auscultation. Computers in Biology and Medicine 2008, Volume 38(2), pp. 221-233.

[22] Breiman, L.: Random forests. Machine Learning Journal, 45(1), 2001, pp.5-32.

[23] Díaz-Uriarte,R. and de Andrés,S.A: Gene selection and classification of microarray data using random forest. BMC Bioinformatics, 7,3. 2006.

[24] Dudoit, S: Comparison of discrimination methods for the classification of tumors using gene expression data. J. Am. Stat. Assoc., 97, 2002, pp.77–87.

[25] Lee, J.W: An extensive evaluation of recent classification tools applied to microarray data. Comput. Stat. Data Anal., 48, 2005, pp. 869–885.

[26] Freund, Y., Shapire, R. E.: Experiments with a new boosting algorithm. In Lorenza Saitta (Editor), Machine Learning: Proceedings of the Thirteenth International Conference (ICML96). Morgan Kaufmann, 1996.

[27] Heckerman, D., Geiger, D., Chickering, D. M.: Learning Bayesian networks: the Combination of knowledge and statistical data. Machine Learning, 20, 1995, pp. 197–243.

[28] Cooper, G., Herskovits, E.: A Bayesian Method for the Induction of Probabilistic Networks from Data, Machine Learning, 9, 1992, pp. 309-347.

[29] Pearl, J. Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, 1988.