# MCMC Bayesian Inference for Heart Sounds Screening in Assistive Environments

Manolis Maragoudakis
University of Aegean
Department of Information and Communication Systems Engineering
+302273082261
mmarag@aegean.gr

Euripides Loukis
University of Aegean
Department of Information and Communication Systems Engineering
+302273082000
eloukis@aegean.gr

## ABSTRACT
The large scale application of ICT-based assistive environment technologies for the home care of elderly and disabled people is going to generate huge numbers of signals transmitted from homes to local health centers or hospitals in order to be monitored by medical personnel. This task is going to be of critical importance and at the same time - if manually performed - quite demanding for specialized human resources and costly. In order to perform it in a cost-efficient manner it is necessary to develop mechanisms and methods for automated screening of these signals in order to identify abnormal ones that require some action to be taken. This paper proposes a method for automatic screening of heart sound signals, which are the most widely acquired signals from the human body for diagnostic purposes in both the 'traditional' medicine and the emerging ICT-based assistive environments. It is based on a novel Markov Chain Monte Carlo (MCMC) Bayesian Inference approach, which estimates conditional probability distributions in structures obtained from a Tree-Augmented Naïve Bayes (TAN) algorithm. The proposed approach has been applied and validated in a difficult heterogeneous dataset of 198 heart sound signals, which comes from both healthy medical cases and unhealthy ones having Aortic Stenosis, Mitral Regurgitation, Aortic Regurgitation or Mitral Stenosis. The proposed approach achieved a good performance in this difficult screening problem, which is higher than other widely used alternative classifiers, showing great potential for contributing to a cost-effective large scale application of ICT-based assistive environment technologies.

## Categories and Subject Descriptors
G.3.3 [**Probability and Statistics**]: Probabilistic Algorithms

## General Terms
Algorithms, Measurement, Performance.

## Keywords
Bayesian Inference, Markov Chain Monte Carlo, Tree-Augmented Naïve Bayes, Assistive Environments, Heart Sounds Diagnosis.


PETRA'11, May 25 - 27, 2011, Crete, Greece.


## 1. INTRODUCTION
There is a growing interest worldwide in the development of ICT-based assistive environment technologies for home care of elderly and disabled people and improvement of their quality of life (e.g. see web site of the 'Ambient Assisted Living (AAL) Program of the European Union www.aal-europe.eu). However, their large scale application poses significant challenges. One of them is the huge numbers of signals that will be generated at the houses of the elderly and disabled people supported, which will be transmitted to health centers or hospitals in order to be monitored by medical personnel. These assistive environments will include various types of sensors and devices, and each of them will generate big numbers of bio-signals or other types of signals; these signals will be transmitted through the Internet or other networks (e.g. wireless) to the nearest local health centers or hospitals in order to be examined and monitored by specialized medical personnel, so that appropriate action can be taken whenever necessary (i.e. to send a nurse or doctor for home medical care, or to proceed to more sophisticated examinations). This task is going to be of critical importance for the success of these ICT-based assistive environments and the quality of the services provided to the elderly and disabled; at the same time, if it manually performed (without appropriate technological support and automation), it is going to be quite demanding for specialized human resources and too costly, threatening the financial sustainability of the large scale application of these technologies. In order to perform this critical task in a cost-efficient way it is important to develop mechanisms and methods for automated screening of these signals in order to identify abnormal ones that require some action. Such a technological support can significantly reduce the needs for specialized human resources, and therefore cost, and at the same time improve the quality of the home care services offered to elderly and disabled people. In this sense it can be critical for the financial sustainability and success of the large scale application of these technologies.

This paper contributes in this direction by proposing a method for automatic screening of heart sound signals, which are the most widely acquired signals from the human body for diagnostic purposes in both the 'traditional' medicine and the emerging ICT-based assistive environments. The heart sound auscultation is an operationally simple, low cost and non-invasive examination, which can be easily performed in the context of home care, and has high sensitivity to many important heart diseases. The development of digital electronic stethoscopes allows the easy acquisition of heart sound at home and then its

digitization, storage, transmission to remote systems of health centers or hospitals, where it can be presented on screen and processed in order to identify abnormal components (e.g. murmurs or additional heart sounds) indicating possible diseases; in such cases appropriate action can be taken, e.g. visit of a nurse or doctor at home, or a more sophisticated examination, such as Echocardiography or Medical Imaging (e.g. Ultrasound Imaging US, Computed Tomography CT, Magnetic Resonance Imaging MRI, etc.). However, the wide application of this approach will result in health centers or hospitals receiving numerous heart sound signals, which their medical staff will have to examine, diagnose possible problems and prescribe appropriate actions; this will necessitate more medical personnel and considerable financial resources. At the same time the pool of skilled medical personnel for this particular task, who have been trained in the era before echocardiography, continues to age, and the skills for heart auscultation is in shortage and in danger to disappear [27-28]. Therefore it will be of critical importance for the cost-effective large scale application and financial sustainability of this approach to develop mechanisms and methods for the automated screening of the incoming heart sound signals, and the identification of the ones having abnormal elements.

As described in more detail in the following section 2, various classification algorithms have been successfully used for the detection of various heart pathological conditions and diseases from heart sound signals. Bayesian Networks (BN) [1-2] are quite attractive for this purpose, due to the fact that they are interpretable, flexible models for representing relationships between interacting heart sound features. Such relationships could be exploited in terms of both diagnosing heart sounds and simultaneously obtaining an insight on which input features really contribute to the classification process. In particular, Bayesian networks consist of two parts: a qualitative and a quantitative one. At the qualitative level, the structure of the network (in a form of a Directed Acyclic Graph-DAG, where features are denoted as nodes and arcs represent probabilistic relationships among them) depicts direct relationships between features. At the quantitative level, such relationships are described as conditional probability distributions. Also, the non-deterministic nature of Bayesian networks enables them to handle better data having high levels of noise generated due to biological or technical reasons.

Nevertheless, conducting Bayesian network analysis in the medical domain, poses a series of important problems. The most important of them is the usually high dimensionality of the datasets in this domain in comparison with the number of the available training instances (e.g. our dataset – described in more detail in section 5 - consists of 100 features and only 198 training instances). Another significant problem is that most Bayesian network learning approaches are suitable for discrete domains, with only a few solutions for continuous ones; discretization is not a good solution in such cases due to loss of information it causes. Also, BN learning consists of two separate processes, executed in a serial manner: the former is called structure learning and the latter is called parameter estimation. Structure learning is NP-hard [1], since as the number of features grows the number of candidate network structures increases super-exponentially to huge numbers, e.g. if we have only 10 features the learning algorithm needs to evaluate more than 15000 possible network structures. Further, if the sample size is small compared to the number of features (something quite usual in the medical domain, as mentioned above, like in our dataset) there is a plethora of sub-optimal models that can fit the data with equal likelihood [2]. Upon evaluating the most probable network structure estimation of parameters (i.e. Conditional Probability Distributions-CPDs) of each BN is carried out. Estimating CPDs involves the calculation of $p(X_i|parents(X_i))$ for each of the features $X_i$ where the term $parents(X_i)$ refers to the set of parent nodes of $X_i$ in the given network.

In this paper it is our goal to cope with the aforementioned issues and present a Bayesian Networks (BN) analysis framework for identifying causal (and independence) relationships between features of heart sound signals and perform highly detailed diagnosis of them: i) initially as healthy and unhealthy, ii) then the latter as having systolic or diastolic murmurs, and finally iii) in both cases discriminating between aortic or mitral dysfunction. Similar to other machine learning approaches, but unlike most BN methods, we are handling features as continuous rather than discrete. Additionally, due to the high dimensionality nature of our dataset, exact computation of the CPDs is infeasible and computationally costly. Hence, the joint distribution can only be approximated by stochastic simulation commonly named as *"sampling"*. Using Markov Chain Monte Carlo (MCMC) we can fit a distribution to the data that converges to the posterior and retain the samples. MCMC can cope with domains where the state space is huge with large number of samples needed to approximate the probabilities reasonably well, by selecting each sample using the previous sample resulting in the well-known Monte Carlo Markov Chain (MCMC) methods and its variants [3]. In particular, we propose a new approach to approximate the conditional probability distributions of complex BN using a MCMC algorithm; we demonstrate that this allows us to create a robust system for a highly detailed diagnosis of heart sound signals. Our work is principally based upon a novel idea, in which the CPD computation is based on the ordered ranking of a structure similar to traditional BNs, but is oriented towards classification.

This structure is called Tree-Augmented Naïve Bayes (TAN) and unlike general, unrestricted BNs, TAN considers the class node to be the parent of all other nodes which can form a BN among them. This type of structure is proven to be more efficient than BN for classification purposes [4], since in traditional BN the class node is not considered as a special type of node, and it is treated as ordinary one, so it may not appear in the network resulting in poor classification performance. The proposed methodology has been applied and validated in a difficult heterogeneous high-dimensional dataset of 198 heart sound instances with 100 input features, which comes from both healthy medical cases and unhealthy ones having Aortic Stenosis, Mitral Regurgitation (both these diseases result in systolic murmurs), Aortic Regurgitation or Mitral Stenosis (both these diseases result in diastolic murmurs). Also, some widely used alternative classifiers have been applied to the same data for comparison purposes.

In the following section 2 previous relevant research is briefly reviewed, while in section 3 the theoretical aspects of MCMC and Gibbs sampling are presented. The proposed methodology based on MCMC Bayesian analytics is described in section 4. The data we used for the abovementioned first application and validation of the proposed methodology and the preprocessing of them are described in section 5. Finally in section 6, the results of this application are presented.

## 2. PREVIOUS RESEARCH

Considerable previous research has been conducted on the automated detection of various heart pathological conditions and diseases from heart sound signals. The wide availability of these signals and their high sensitivity to most heart problems has been a strong motivation for this research. It can be broadly divided into two research streams. The first of them deals with the development of methods for the preprocessing of heart sound signals (e.g. removal of noise, segmentation of heart cycles, partitioning of each heart cycle into S1, systolic phase, S2 and diastolic phase, etc.); a good review of them is provided in [5]. The second research stream aims at the development of methods for the detection of heart pathological conditions and diseases from appropriately preprocessed heart sound signals.

Since this paper contributes to the second research stream, we are going to focus our review on it. Most of the studies of this stream are dealing with the discrimination between normal and abnormal heart sound signals [6-10], or with the discrimination between innocent and pathological murmurs in children [11-16]. Some other studies are dealing with the detection of particular heart diseases from heart sound signals, such as coronary artery diseases [17-21] and heart valve diseases or murmurs [22-31].

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 [6,7, 9, 13, 15-20, 24, 27, 28]. There are only a few studies using other classification algorithms, such as discriminant functions [12,26], decision trees [29,30], Bayesian networks [8], Support Vector Machines [30] and Hidden Markov Models [31]. 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, so further research is required in this direction.

It should also be noted that the risk that heart valve diseases pose for human life has motivated considerable research on its computerized diagnosis from other more costly signals, such as Doppler Heart Sound (DHS), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI); a good review of them is provided in [30]. However, these signals require highly sophisticated and costly equipment which cannot be available in home care context.

## 3. MCMC BAYESIAN INFERENCE

In this section the theoretical principles of MCMC sampling are outlined, with a focus on Gibbs sampling, a variation of MCMC more suitable for DAG structures [32].

### 3.1 MCMC Methods

A probability distribution is specified through a DAG *G* (the BN structure – a set of interconnected nodes, each of which corresponds to one of the features) and a collection of conditional probability distribution (parameters) for each feature $X_i$ in *G*. Every node is conditionally independent of all other nodes given the set of its parents. The CPD of a BN is encompassing the probabilities of observing all values of feature (node) $X_i$ given the values of its parent nodes. Large network models will introduce more parameters, so exact computation will be infeasible and thus approximation of the CPD is achieved through sampling techniques. The structure of G is essential for sampling and can be obtained by applying a greedy search over the entire space of all possible structures. However, the number of possible DAG structures increases super-exponentially as the number of features grows, so greedy search on the space of all possible structures is not realistic as it requires too much computation. There have been several methodologies for alleviating this problem, such as the K2 algorithm [2] or the Bayesian Scoring Method [3]. In the next section 4 we shall present our approach for obtaining graph structures more straightforwardly and producing BN that favor the classification process.

Regardless of the structure learning algorithm, given a structure *G* with nodes $X=\{X_1,X_2,...,X_n\}$, the process of obtaining the CPD with sampling is described below: For each node $X_i$ in the network:

- Randomly select a state for all other nodes except for $X_i$.
- Compute the probability distribution over the states of $X_i$, i.e. $p(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$.

  Note that since G is a Bayesian network the above probability is simplified to include only the Markov Blanket of $X_i$ [3], *i.e.:* $p(X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)=p(Xi|parents(X_i)\prod_{j=1}^{k}(Y_j|parents(Y_j))$,

  where $Y_j$ denotes the set of child nodes of $X_i$.
- From the probability distribution, randomly select a state of $X_i$ to complete the sample vector.

Monte Carlo based sampling requires drawing of *n* samples from the BN with each vector of feature states forming its value as explained above. For our research, we only consider continuous values, therefore, we adopt the method of [33] and project the samples as a histogram and afterwards we smooth the histogram to obtain the probability density function of the features of the dataset. In most approaches, the selection of a state of the features is performed using the distribution that best resembles the available data set. This approach is however not suitable for large feature sets, such as the task at hand, because they tend to be slow and cannot converge to the actual posterior distribution. Therefore, a Markov Chain Monte Carlo (MCMC) approach is more suitable in such cases for approximating the challenging high dimensional distributions. The Gibbs sampler was chosen as an MCMC alternative because it is more appropriate for DAG structures [32]. Furthermore, a Gibbs sampler can allow for convergence in reasonable computation time and its implementation code is widely available (e.g. WinBUGS [34]).

### 3.2 MCMC and Gibbs Sampling

Before familiarizing with the Gibbs sampler, a few introductory comments on Markov Chains are in order. Let $X_t^i$ denote the value of a random variable $X_i$ at time *t*, and let the state space refer to the range of possible *X* values. This random variable is a Markov process if the transition probabilities between different values in the state space depend only on the random variable's current state, i.e.:

$$p(X_{t+1}^i = s_j | X_0^i = s_l, \ldots, X_t^i = s_k) = p(X_{t+1}^i = s_j | X_t^i = s_k)$$

In other words, for a Markov random variable the only information about the past needed to predict the future is the current state of it. Knowledge of the values of earlier states does not change the transition probability. A Markov chain refers to a sequence of random variables generated by a Markov process. A particular chain is defined most critically by its transition kernel

P(j→k), which is the probability that a process at state space $s_j$ moves to state $s_k$ in a single step, i.e.:

$$P(j \to k) = p(X_{t+1}^i = s_k | X_t^i = s_j)$$

For reasons of readability, we shall simplify the notion of $X_t^i$ into $X_t$ to denote that a random variable X takes a specific value at time t. Let $\pi_j(t) = p(X_t = s_j)$ denote the probability that the chain is in state j at time t, and let *π(t)* denote the row vector of the state space probabilities at step t. We start the chain by specifying a starting vector *π(0)*. Often, all the elements of π(0) are zero except for a single element of 1, corresponding to the process starting in that particular state. As the chain progresses, the probability values get spread out over the possible state space. Using matrix notation, we can define the probability transition matrix P as the one whose element *(i,j)* denotes the *P(i→j)* transition kernel. The probability that the chain has state value $s_i$ at time (or step) t +1 is given by:

$$\pi(t+1) = \pi(t)P = (\pi(t-1)P)P = \cdots = \pi(0)P^{t+1}$$

In simple words, as the above equation implies, a Markov chain may reach a stationary distribution $\pi^*$, regardless of the selection for the initial distribution parameters. A straightforward method of approaching this distribution includes sampling. While there are numerous sampling strategies, the Gibbs sampler [32] is well-suited for DAGs, as we shall describe in the next paragraphs.

The key to the Gibbs sampler is that one only considers univariate conditional distributions, i.e. distributions where all of the random variables except for one are assigned fixed values. Such conditional distributions are far easier to simulate than complex joint distributions and usually have simple forms. To introduce the Gibbs sampler, consider a bivariate random variable *(x, y)* and suppose we request the computation of one or both probabilities, *p(x)* and *p(y)*. The idea behind the sampler is that it is far easier to consider a sequence of conditional distributions, p(x|y) and p(y|x), than it is to obtain the probability by integration of the joint density *p(x,y)*, e.g. $p(x) = \int p(x,y)dy$. The sampler starts with some initial value $y_0$ for *y* and obtains $x_0$ by generating a random variable from the conditional distribution $p(x|y = y_0)$. Then, the sampler uses $x_0$ to generate a new value of $y_1$, drawing from the conditional distribution based on the value of $x_0$, $p(y|x = x_0)$ and so forth. It proceeds as follows:

$$x_i \sim p(x|y = y_{i-1})$$
$$y_i \sim p(y|x = x_i)$$

Repeating this process *k* times, generates a Gibbs sequence of length *k*, where a subset of points $(x_j, y_j)$ for $i \le j \le m < k$ are taken as the simulated draws from the full joint distribution.

To obtain the desired total of *m* sample points (here each "point" on the sampler is a vector of the two parameters), one samples the chain (i) after a sufficient burn-in process (i.e. a number of initial samples to be removed due to removal of the bad effects of the initial sampling values) and (ii) at set time points (say every *n* samples) following the burn-in. The Gibbs sequence converges to a stationary distribution that is independent of the starting values, and by the principle of MCMC, this stationary distribution is the target distribution we are trying to simulate [33].

# 4. METHODOLOGY

Direct application of the aforementioned Gibbs sampling for BN estimation within the heart sounds domain is somewhat limited, due to the high dimensional data where the number of features is comparable to the number of available samples. This implies that the variance in the values taken by each variable is high and this phenomenon may prohibit producing independent uniform samples. The suggested novel MCMC sampling framework, shown in Fig. 1, can overcome this limitation. Initially inspired by the work of [33], which states that an initial set of 10-20 dissimilar but high scoring networks (as regards to the probability of the network structure given the input data, *p(S|D)*) could be used for calculating the Bayesian posterior probability distribution of all features. Clearly, we could not simply take the top-*k* networks which achieve high probability from a distinct learning procedure, because all of these networks would be very similar in structure. Therefore, the proposed idea of [33] is valid in theory but lacks practicality. Our suggestion focuses on creating simple and straightforward BN structures which are suitable for the classification process (since *classifying a heart sound is our final goal*). Such structures could be obtained from the TAN algorithm [4]. The TAN algorithm creates networks where the class node is a parent of all features nodes. Features form a simplified Bayesian network amongst them in which each node has one parent at most, in order to retain the structure and the CPD simple. The learning phase of the TAN algorithm will be explained below and, as we shall see, TANs are very fast learners. Compared to the traditional BN learning algorithms, the TAN structure is obtained 50-100 times faster than the BN approach, depending on the number of input features and the number of states each feature has. Moreover, TANs are considered better classifiers than BNs, a fact that is attributed to the structural characteristics of the former, which consider the class node as a parent of all other nodes.

From the samples drawn from the TAN structures, we can obtain the posteriors after convergence, and then determine the state sequence and probability estimates of the model in a straightforward manner. Although the inferred high scoring TAN structures are disjoint (i.e. cannot be combined into one network structure), they can all be combined independently to the underlying probability distribution. Hence, all these network structures are sampled to estimate the probability distribution accurately. The important element of our methodology is the use of fast-learned TAN structures and a rank ordering amongst them.

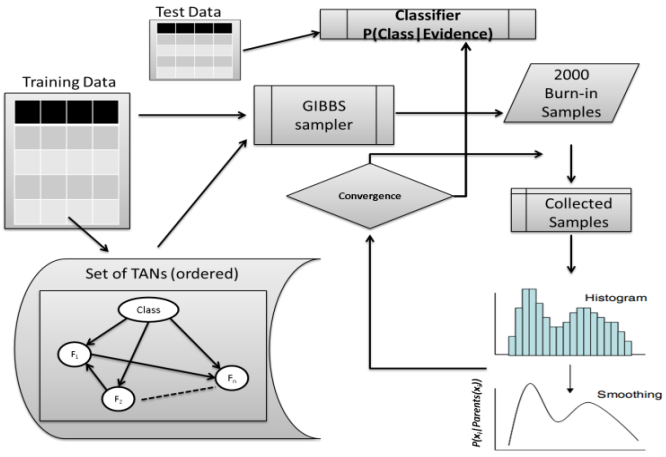


*Fig 1.The flowchart of the proposed methodology.*

As we can see in the methodology flowchart of Fig. 1, the main components of the proposed methodology are the TAN learning

phase, the Gibbs sampling phase and finally, the convergence phase. With regard to the former, a set of 10 TAN network structures were extracted using the following steps:

- Built a naïve Bayesian structure where the class node C is a parent to all feature nodes $F_i$.
- For each pair of features $F_i$, $F_j$, compute the conditional mutual information given the class, i.e.:

$$I(F_i; F_j|C) = \sum_{F_i,F_j,C} p(F_i, F_j, C) log \frac{p(F_i, F_j|C)}{p(F_i|C)p(F_j|C)}$$

- Build a complete undirected graph to connect features and use $I(F_i; F_j|C)$ to weight all arcs.
- Build a maximum weighted spanning tree.
- Transform the resulting undirected tree to a directed one by choosing a root feature and setting the direction of all edges to be outward from it.

For maximizing the performance of TAN, we applied a feature selection algorithm based on SVM [35] and eliminated the features that scored below 0.1, thus achieving a 20% reduction in the number of input features for TAN. By changing the root feature we managed to produce 10 different TAN structures. As mentioned before, an ordinary Gibbs sampler chooses features at random and then samples a new value from the estimated posterior of the neighboring variables. Friedman [1] argued that sampling from the space of total orders on variables rather than directly sampling DAGs was more efficient than application of ordinary MCMC directly in random manner. Since the Gibbs sampler also samples the new value of a feature based on the parent variables, an ordering of the rank of the TANs, based on their scores was applied. As regards to the Gibbs sampling phase, uniform prior distributions for all the features in the domain needed to be defined. Instead of applying a random initial state of the network, a multivariate Dirichlet distribution was chosen, inspired by [32]. This distribution is assigned to both the initial state distribution and also to the state transition distribution of the Markov chain (note that each state in our experiments represents the previous sample drawn). The initial distribution of the variables in the network was assigned using the density function. It was estimated after smoothening of the histogram of normalized feature data. Since all nodes have parent(s) we sampled from the conditional distribution of their TAN. Similarly, *n* independent samples were drawn from the target distribution *P(x).* The samples collected were plotted using a histogram with *n* bins as depicted in Fig. 1 above. The probability density function *P(x)* of a continuous feature was approximated by smoothening of the histogram.

Finally, as regards to the convergence phase, multiple TAN structures were fed to a parallelized series of Markov Chains, in order to obtain a large number of samples from the entire input space of the domain. Recall that each Markov Chain connects states of the network instantiation and sampling process. In other words, if $S_0$ represents the first instantiation of features $(X_1=x_1, X_2=x_2,\ldots,X_n=x_n)$ then we can sample a new value $x_1'$ for feature $X_1$ using $p(X_{1=x1'}|X_2 = x_2,\ldots,X_n = x_n)$. In similar manner, we can sample the remaining new values for features $X_2$, $X_3$ …$X_n$ until we have a new state $S_1$, instantiated as: $X_1 = x_1'$, $X_2 = x_{2'}$, . . . ,$X_n = x_{n'}$. Throughout the process of multiple chain runs, samples are exchanged between the chains and the overall samples of a number of variables in the top of the specified order are monitored. When the sample values do not exceed a variation threshold (manually defined to 0.01) after a large number of iterations, convergence is assumed.

Upon convergence on the stationary distribution, the process of classification of a previously unseen example is straightforward. We only compute the probability of the class *c* given evidence *e* (expressed as an input vector of 100 feature values), noted as *p(c|e)*and classify it to the most probable class.

## 5. DATA & PREPROCESSING

Our dataset consisted of 198 heart sound signals, which have been acquired from both healthy and pathological medical cases having one of the following four frequent and severe heart valve diseases: Aortic Stenosis, Mitral Regurgitation (both these diseases resulting in systolic murmurs), Aortic Regurgitation or Mitral Stenosis (both these diseases resulting in diastolic murmurs). In particular, 38 of these heart sound signals were healthy, while the remaining were unhealthy: 41 ones with AS systolic murmur, 43 with MR systolic murmur, 38 with a AR diastolic murmur and 38 with a MS diastolic murmur. All these signals had been diagnosed by experienced cardiologists and classified to one of the above five heart health conditions. It should be pointed out that the heart sounds acquired using a stethoscope are influenced considerably by numerous factors related to the acquisition process, 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. A big problem is that these factors cannot be controlled in the everyday medical practice, adding high levels of noise to the acquired heart noise signals (i.e. generating 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 the high level of noise that this problem generates. So in order to make our research more realistic, we decided to make the above dataset, ‘global’ and representative, including ‘heterogeneous’ heart sounds recorded with different acquisition methods and different values of the above factors. Such a dataset is much more ‘difficult’ to cope with 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 the performance of the proposed methodology.

Initially a pre-processing of these heart sounds was performed, in order to remove noise and extract features from them. It consisted of three phases. In the first phase the segmentation of the heart sound signal was performed; in each signal the cardiac cycles were detected by locating the S1 and S2 peaks. In the second phase, for each of the segmented 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 were the first four features (F1-F4) of the feature vector of each heart sound signal. In the third phase, the rest of the features

used for classification were extracted. For this purpose we calculated for each 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. In particular the first of these signals 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 was then divided into 8 equal parts, for each part the mean square value was calculated and the resulting 8 values were used as features (F5-F12). Similarly 24 features for the systolic period (F13-F36), 8 features for S2 (F37-F44) and 48 features for the diastolic period (F45-F92) were calculated (the number of features per component was decided taking into account the time duration of each: for longer components more features 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 providing its low frequency content, b) a 100–300 Hz filter providing its medium frequency content, c) a 150–350 Hz filter providing its medium-high frequency content and d) a 200–400 Hz filter providing 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 198 feature vectors were used for the validation of the proposed methodology presented in the following section.

## 6. RESULTS

We have organized the experimental part of our work as follows:

A. Initially, discrimination between normal (NRM) from sick (SCK) heart sound was performed.
B. Those instances that belong to the sick class were further classified as having systolic (STL) or diastolic (DTL) murmur.
C. Finally, for each of the aforementioned classes, there was further classification into two sub-classes corresponding to aortic or mitral origin of murmurs: the heart sound signals classified as having systolic murmur (STL) were further classified as Aortic Stenosis (AS) or Mitral Regurgitation (MR) cases; similarly, the ones classified as having diastolic murmur were further classified as Aortic Regurgitation (AR) or Mitral Stenosis (MS) cases.

Results were compared against other well-known classification algorithms that have previously been referred to as having provided "state-of-the-art" results in the heart disease domain. In particular, we evaluated the proposed methodology against Naïve Bayes, Decision Trees, Neural Networks (with Radial Basis Functions) and k-Nearest Neighbor (k=3), using a 10-fold cross validation approach. The *RapidMiner* data mining suite was used for the evaluations [35].

### A. Discriminating between healthy and unhealthy signals

From a medical expert's perspective, the accuracy of a diagnosis is of major importance, since a misclassification of a sick case as healthy could have severe consequences for a patient. Due to the significance of the decision, the following cases have to be distinguished:

(a) The classification result is **sick** and the patient was **actually sick**. In such a case, classification is correct and these cases are labeled as **True Positives (TP)**.
(b) The classification result is **normal** and the patient is **actually healthy**. Similarly, such a classification is correct and these cases are labeled as **True Negatives (TN)**.
(c) The classification result is **sick** and the patient is **actually healthy**. In such an erroneous case the classification is incorrect, and these cases are labeled as **False Positives (FP)**.
(d) The classification result is **normal** and the patient is **actually sick**. Similarly in such a case the classification is incorrect, and these cases are labeled as **False Negatives (FN)**.

The following table (known as *confusion matrix*) summarizes the above descriptions:

| ***Confusion Matrix*** | | | |
|---|---|---|---|
| | | ***Actual Class*** | |
| | | ***Sick*** | ***Normal*** |
| ***Predicted class*** | ***Sick*** | TP | FP |
| | ***Normal*** | FN | TN |

Since identification only of the percentage of the correctly identified instances (TP+TN) is not indicative, two additional metrics of the robustness of the classification process are required:

(a) True Positive Rate (TPR): the percentage of sick instances correctly classified as sick:

$$TPR = \frac{TP}{TP + FN}$$

(b) False Negative Rate (FNR): the percentage of sick instances incorrectly classified as healthy:

$$FNR = \frac{FN}{TP + FN}$$

The anticipated classification outcome is the one that does not erroneously considers as healthy a patient with a heart disease. Therefore, we are particularly focusing on the performance of the proposed methodology in situations where sick patients were identified as healthy (FPR) and plot this against TPR in order to produce the ROC (Receiver Operating Characteristic) curve. This curve is a graphical depicter of the classification performance. Robust classifiers are expressed by ROC curves which retain high values of TPR for most of the horizontal axis area (FPR).

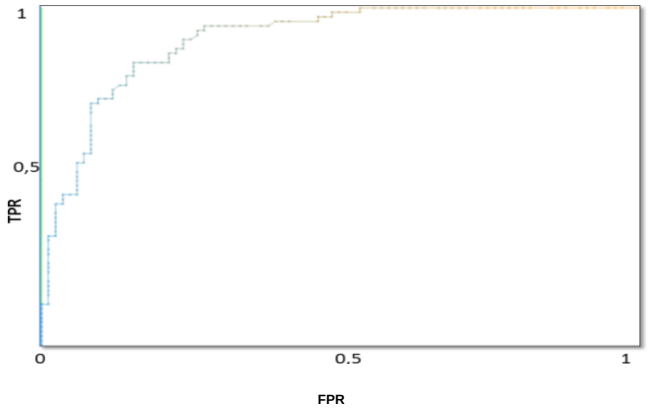


*Fig 2.ROC curve for the Healthy/Unhealthy dataset.*

Furthermore, apart from the above ROC curve, which illustrates that our method is capable of classifying more that 87% of the instances correctly, we also present the results of each benchmark algorithm against the proposed methodology, utilizing the *F-measure*, i.e. the harmonic mean of precision P and recall R:

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN}$$

***Table 1. Performance of MCMC methodology and alternatives for discrimination between Healthy and Unhealthy***

| Algorithm | %F-measure |
|---|---|
| MCMC Bayesian Inference | 85.56% |
| C4.5 Decision Trees (C45) | 74,14% |
| Naïve Bayesian Classifier (NB) | 70,15 % |
| Radial Basis Functions (RBF) | 76,24% |
| K-Nearest Neighbor (KNN) | 82,15% |

The results of Table1 show that MCMC performs better in this first level of discrimination than all other alternative methodologies, providing a gain between 3%-15% in certain cases for this highly heterogeneous and 'difficult' dataset.

### B. Discriminating between systolic and diastolic murmurs

Proceeding to a more detailed classification the healthy heart sounds were next classified as having systolic or diastolic murmurs, and the results (F-measure) are shown in Table 2.

***Table 2. Performance of MCMC methodology and alternatives for discrimination between Systolic and Diastolic Murmurs***

| Algorithm | %F-measure |
|---|---|
| MCMC Bayesian Inference | 89.20% |
| C4.5 Decision Trees (C45) | 86,45% |
| Naïve Bayesian Classifier (NB) | 75,34 % |
| Radial Basis Functions (RBF) | 78,80% |
| K-Nearest Neighbor (KNN) | 82,15% |

We remark that MCMC in this more detailed classification as well exhibits a better performance than all the other alternative methodologies. This could be attributed to the elimination of non-informative features from the TAN step of the proposed methodology and due to the convergence attribute of the MCMC process. Additional to the aforementioned table, in the Appendix we have included a figure which shows the best scoring TAN structure as obtained from the MCMC approach. From the plethora of initial features, the weighting of features using a SVM classifier has resulted in producing only a small set of them (18) that are found to influence the class attribute. This reduction caused significant improvement of the MCMC step in terms of computational complexity.

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

The final round of experimental evaluations proceeds to even higher diagnostic detail and focuses on identifying the exact heart disease (problem of aortic or mitral valve). Note that when the heart murmur is diastolic, the patient could suffer from either Aortic Regurgitation (AR) or Mitral Stenosis (MS); when the heart pulse is systolic, the disease can be either Aortic Stenosis (AS) or Mitral Regurgitation (MR). For the former case results are tabulated in Table 3. MCMC is again the most efficient approach and outperforms all other alternative approaches.

***Table 3. Performance of MCMC methodology and alternatives for discrimination between AR-MS for Systolic Murmurs***

| Algorithm | %F-measure (Healthy-Unhealthy) |
|---|---|
| MCMC Bayesian Inference | 88,55% |
| C4.5 Decision Trees (C45) | 75,00% |
| Naïve Bayesian Classifier (NB) | 83,88 % |
| Radial Basis Functions (RBF) | 69,14% |
| K-Nearest Neighbor (KNN) | 85,40% |

As regards to the latter case the results are shown in Table 4. We remark that MCMC still exhibits the highest classification performance, which outperforms the other alternative approaches by a varying percentage of 0.5%-15%.

***Table 4. Performance of MCMC methodology and alternatives for discrimination between AS-MR for Diastolic Murmurs***

| Algorithm | %F-measure |
|---|---|
| MCMC Bayesian Inference | 86,30% |
| C4.5 Decision Trees (C45) | 74,20% |
| Naïve Bayesian Classifier (NB) | 81,45 % |
| Radial Basis Functions (RBF) | 70,25% |
| K-Nearest Neighbor (KNN) | 85,90% |

## 7. CONCLUSIONS

For the cost-efficient large scale application of ICT-based assistive environments for the home care of elderly and disabled people it is of critical importance to develop capabilities for automated first screening of signals generated in subjects' homes and transmitted to local health centers or hospitals, and identifying abnormal ones that require action to be taken. The present paper contributes to addressing this need. It proposes a methodology for the automatic screening at various levels of detail of heart sound signals acquired in home care context (classification as healthy or unhealthy, with systolic or diastolic murmurs, and then of aortic or mitral origin). It is based on a novel MCMC Bayesian approach, which can handle datasets characterized by numerous input features and limited training data. It has been concluded that the proposed methodology shows a good performance in a highly heterogeneous and difficult dataset, which is higher than the most widely used alternative methodologies. We believe that if we can control some of the heart sound acquisition factors (e.g. type of stethoscope/sensor, stethoscope use mode and filtering), and have more training data, even higher performance can be achieved.

## 8. REFERENCES


[1] Friedman, N., Koller, D.: Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. Machine Learning 50, 95–126 (2003)

[2] Heckerman, D.: A Tutorial on Learning with Bayesian Networks. In: Jordan, M. (ed.) Learning in Graphical Models. MIT Press, Cambridge (1999)

[3] Pearl, J.: Causality: Models, Reasoning and Inference. Cambridge University Press, Cambridge (2000)

[4] Pearl, J. Probabilistic reasoning in Intelligent Systems: networks for plausible inference. Morgan Kaufmann (1988)

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] White PR, Collis WB, Salmon AP: Analysing heart murmurs using time-frequency methods. In Proceeding of the IEEE-SP International Symposium Time-Frequency and Time-Scale Analysis June 1996, pp. 385-388.

[11] 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.

[12] Leung T. S., White P. R., Collis W. B., Brown E., Salmon A. P.: Analysingpaediatric 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.

[13] Leung T. S., White P. R., Collis W. B., Brown E., Salmon A. P.: Characterisation of paediatric heart murmurs using self-organising map. In Proceedings of the Joint Meeting of the BMES & IEEE Engineering in Medicine and Biology Society, 1999: 926.

[14] Noponen A. L., Lukkarinen S., Angerla A., Sikio K., Sepponen R.: How to recognize the innocent vibratory murmur. Computers in Cardiology 2000, pp. 561-564.

[15] 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.

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

[17] Akay M.: Noninvasive diagnosis of coronary artery disease using a neural network algorithm. Biological Cybernetics 1992, 67, pp. 361-367.

[18] Akay Y., Akay M., Welkowitz W., Kostis J.: Noninvasive detection of coronary artery disease. IEEE Engineering in Medicine and Biology 1994, 13(5), pp. 761-764.

[19] Bahadirlar Y., OzcanGulcur H., Aytekin S., Gulmez U.: Acoustical detection of coronary artery disease. In Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society November 1994, 2, pp. 1278-1279.

[20] 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.

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

[22] 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.

[23] 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.

[24] 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. 3506-9.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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, pp. 47-61. (2009)

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

[32] Ram, R., Chetty, M.: Constraint Minimization for Efficient Modeling of Gene Regulatory Network. In: Chetty, M., Ngom, A., Ahmad, S. (eds.) PRIB 2008. LNCS (LNBI), vol. 5265, pp. 201–213. Springer, Heidelberg (2008)

[33] Liu, J.S.: Monte Carlo Strategies in Scientific Computing. Springer, Heidelberg (2001)

[34] Lunn, A, Thomas, G, Best, H, Spiegelhalter, D.WinBUGS – A Bayesian modeling framework: Concepts, structure, and extensibility, Statistics and Computing 10, 325–337 (2000)

[35] Rapid-Miner, A software for Data Mining Tasks, www.rapid-i.com

# APPENDIX

The best scoring TAN structure of the Systolic-Diastolic dataset.