Univ Access Inf Soc (2014) 13:73–88
DOI 10.1007/s10209-013-0293-4

LONG PAPER

Heart sound screening in real-time assistive environments through MCMC Bayesian data mining

Manolis Maragoudakis · Euripides Loukis

Published online: 17 March 2013

Abstract Emerging pervasive assistive environment applications for remote home healthcare monitoring of the elderly, disabled and also patients with various chronic diseases generate massive amounts of sensor signal data, which are transmitted from numerous homes to local health centers or hospitals. While it is critical to process this data efficiently (in a fast and accurate manner) and cost-effectively, in a large-scale application of the above technologies, it is not possible to do so manually by specialized human resources. This paper proposes a methodology for automatic real-time screening of heart sound signals (one of the most widely acquired signals from the human body for diagnostic purposes) and identification of those that are abnormal and require some action to be taken, which can be applied to many other similar types of bio-signals generated in assistive environments. It is based on a novel Markov Chain Monte Carlo Bayesian Inference approach, which estimates conditional probability distributions in structures obtained from a Tree-Augmented Naïve Bayes algorithm. It has been applied and validated in a highly '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 methodology achieved high classification performance in this difficult screening problem. It performs higher than other widely used classifiers, showing great potential for contributing to a cost-effective large-scale application of ICT-based assistive environment technologies.

Keywords Bayesian inference · Markov Chain Monte Carlo · Tree-Augmented Naïve Bayes · Assistive environments · Heart sounds diagnosis

1 Introduction

There is growing interest worldwide in the development of ICT-based assistive environment technologies for home care of elderly and disabled people, and for patients with various chronic diseases, aiming to improve the quality of their life in a cost-effective manner. The aging population in many western countries, in combination with government austerity programmes including severe spending cuts, which reduce the available financial resources for elderly and disabled care, creates big social problems. The large-scale application of these ICT-based assistive environments can be an ideal solution for this problem, as they allow for low-cost and high-quality home care and close monitoring of vital health parameters for many subjects, reducing needs for hospitalization. Reviews of research conducted in this area and also descriptions of such environments are provided in [1] and also available in the website of the 'Ambient Assisted Living' (AAL) Program of the European Union (www.aal-europe.eu). Most of this research work, however, faces significant challenges when it comes to large-scale application for supporting large numbers of people. One such challenge involves the huge number of signals that can be generated at the houses of the elderly, chronically ill and disabled people who are monitored and supported, which are transmitted to health centers or hospitals in order to be monitored by medical personnel. In particular, these assistive environments

M. Maragoudakis (✉) · E. Loukis
Department of Information and Communication Systems Engineering, University of the Aegean, Samos, Greece
e-mail: mmarag@aegean.gr

E. Loukis
e-mail: eloukis@aegean.gr

include various types of sensors and devices, and each of them generates many bio-signals or other types of signals; these signals are 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 medical 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 the large-scale application of these ICT-based assistive environments and the quality of the services provided to the elderly, chronically ill and disabled; at the same time, if it is manually performed (without appropriate technological support and automation), it is going to be quite demanding for specialized human resources and therefore too costly, threatening the financial sustainability for large-scale application of these technologies. In order to perform this critical task in a cost-efficient way, it is important to develop tools and methods for automated screening of these signals in order to identify abnormal ones that require some medical action as soon as possible and produce notifications to authorized medical personnel. Such technological support can significantly reduce the needs for specialized human resources, and therefore the cost and at the same time improve the quality of the home care services offered to elderly, chronically ill and disabled people. In this sense, it can be critical for the financial sustainability and success of the large-scale application of these assistive technologies. Advanced data mining technologies can be quite useful in this direction. As mentioned in [1], the development of successful assistive environments for this purpose relies critically on three core technologies: sensor technologies, network technologies and data mining.

This paper contributes in this direction by focusing on the third of these technologies, proposing a methodology for automatic real-time screening of heart sound signals produced by home healthcare–assistive environments and identification of those that are abnormal and require some action to be taken, which can, however, be applied to many other similar types of bio-signals generated in 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. Heart signals are the most widely acquired signals from the human body for diagnostic purposes in both the 'traditional' medicine and in 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 at home. At the same time, the heart sound is highly sensitive to many important heart diseases. The development of digital electronic stethoscopes allows for easy acquisition of heart sounds at home as well as their digitization, storage and transmission to remote health centers or hospitals. These signals can then be visualized on a screen and processed in order to identify abnormal components (e.g., murmurs or additional heart sounds) indicating possible diseases. In such cases, appropriate medical action can then be taken, for example, 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, large-scale application of this approach for supporting large numbers of people will result in health centers or hospitals receiving numerous heart sound signals, which their medical staff will have to examine in order to diagnose any possible problems and prescribe appropriate actions. This, in turn, will necessitate more medical personnel and considerable financial resources, which would make this approach financially and operationally unsustainable. At the same time, the pool of skilled medical personnel for this particular task is very limited and the skills for heart auscultation are in shortage [2, 3]. Therefore, for the cost-effective and large-scale application of remote heart monitoring, it is important to develop mechanisms for automated screening of the incoming heart sound signals and identification of the ones having abnormal elements.

The proposed methodology has been applied and validated in a highly 'difficult' heterogeneous high-dimensional dataset of 198 heart sound instances with 100 input features, derived from both healthy medical cases and unhealthy ones having aortic stenosis, mitral regurgitation (with both these diseases resulting in systolic murmurs), aortic regurgitation or mitral stenosis (with both these diseases resulting in diastolic murmurs). It has achieved high classification performance, which exceeds the ones of other widely used classifiers.

In the following section (Sect. 2), previous relevant research is briefly reviewed. Then in Sect. 3, an overview of the proposed methodology is presented. In Sect. 4, the theoretical aspects of MCMC and Gibbs sampling, which are the main foundations of the proposed methodology are described, followed by a detailed description of the methodology in Sect. 5. The data used for the above-mentioned first application and validation of the proposed methodology and the preprocessing of them are described in Sect. 6, while the results of this application are presented in Sect. 7. Finally, the conclusions are summarized in Sect. 8.

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 diseases has been a strong motivation for this research. It can be broadly divided into two research streams. The first 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 this research stream is provided in [4]. The second research stream addresses the development of methods for the detection of heart pathological conditions and diseases from appropriately preprocessed heart sound signals.

This paper contributes to the second research stream, so the review of previous research will focus on it. Most of the studies in this area deal with the discrimination between normal and abnormal heart sound signals [5–9] or with the discrimination between innocent and pathological murmurs in children [10–15]. Some other studies are dealing with the detection of particular heart diseases from heart sound signals, such as coronary artery diseases [16–20] and heart valve diseases or murmurs [2, 21–29]. However, in general, they do not proceed to highly detailed diagnosis.

Most of the studies dealing with the diagnostic classification of the heart sound signals are based on neural networks of various types [5, 6, 8, 12, 14–19, 23, 26, 27]. There are only a few studies that examine the performance of other classification algorithms, such as discriminant functions [11, 25], decision trees [28, 29], Bayesian networks [7], Support Vector Machines [29] and Hidden Markov Models [2]. Therefore, the diagnostic potential of other classifiers besides neural networks have not been sufficiently explored, and further research is needed in this direction. It should also be noted that the risk that heart valve diseases pose for human life has motivated considerable research in computerized methods of their diagnosis, based on signals acquired from other more costly and sophisticated examinations, such as Doppler Heart Sound (DHS), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). A good review of them is provided in [29]. It should also be noted that these examinations require highly sophisticated and costly equipment which cannot be available for home-based monitoring. In this paper, the potential of Bayesian Networks for detailed diagnosis of heart pathological conditions and diseases from heart signals acquired through digital stethoscopes in the context of home healthcare ICT-based assistive environments is exploited, by overcoming some inherent limitations, as described in more detail in the following section.

3 Methodology overview

This section provides an overview of the proposed methodology, which includes the main challenges and limitations posed in using BN for heart sounds diagnosis, and in general for similar medical diagnosis problems, and also how they are addressed. Bayesian Networks (BN) [3, 30] are quite attractive for heart sounds screening and diagnosis, because they are flexible models for representing relationships among different interacting heart sound features that can be interpreted and visualized. Such relationships could be exploited both for diagnosing heart sounds and simultaneously for obtaining an insight into which are the input features that the classification process is based on. It is also important to note that the non-deterministic nature of BN enables them to handle data having high levels of noise generated due to biological or technical reasons.

In the formulation proposed, BN is used to represent two aspects of heart sound signals: a qualitative and a quantitative one. The qualitative structure depicts direct relationships between features and represents the relations' network as a Directed Acyclic Graph (DAG), where features are denoted as nodes and arcs represent probabilistic relationships among them. The quantitative structure describes such relationships as conditional probability distributions.

However, BN suffer from significant limitations when applied in the medical domain, particularly when prior knowledge is not available or difficult to be defined, and the available data are characterized as 'highly-dimensional' (i.e., having large numbers of features), with limited available training instances (as their collection is usually difficult and costly) in comparison with the number of users. For instance, in the application of the methodology proposed in this paper (presented in detail in Sect. 5) the dataset consists of 100 features and only 198 training instances.

Another important limitation of applying generic BN to the task at hand is that the features proposed have continuous ranges of values (which is quite usual in heart sound signals and in many other bio-signals). Despite the fact that alternative solutions exist for dealing with continuous values in BN, the majority of them focus on the use of discrete valued features, since in the former case (continuous range of values) there are significant topology restrictions and only the Gaussian distribution is supported. Nevertheless, discretization is not a preferred approach in such cases, due to loss of information it causes to the original data.

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 considered to be NP-hard [3], since, as the number of features grows, the number of candidate network structures increases super-exponentially to huge numbers. For example, a dataset of only 10 features would result in the evaluation of more than 15,000 possible network structures during the learning phase. Further, if the sample size is small compared to the number of features (something

quite usual in the medical domain, as mentioned above, for example, as it happens in the dataset used in the paper), then there is a plethora of sub-optimal models that can fit the data with equal likelihood [30]. Also, 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 $parents(X_i)$ refers to the set of parent nodes of X_i node in this network. This will necessitate huge amounts of calculations.

Finally, a fourth obstacle in BN is the lack of orientation toward the class feature (which should be the root node in the estimated network, as heart sounds classification is the main goal), which could pose significant problems to the classification process. BN are by principle designed to allow for reasoning under conditions of uncertainty. This does not necessarily mean that they are suitable for classification. Since the class node is treated in the same way as all other nodes, a BN does not have special knowledge on the class feature and the topology is not oriented to allow for reasoning over the class label, given evidence of the values of the other features; therefore, the class node will not be necessarily the root node in the estimated models.

To summarize the main challenges and limitations of using BN for heart sounds classification are as follows:

(a) BN cannot deal efficiently with high-dimensional datasets, especially when the available labeled dataset (i.e., training set) is limited.
(b) BN do not operate optimally when dealing with continuous variables.
(c) BN learning of structure and CPD is prone to errors and ambiguity when dealing with high-dimensional datasets and limited training samples and can necessitate large amounts of calculations.
(d) BN are not oriented toward classification, so the networks estimated do not necessarily have the class node as root (which is essential for classification).

This paper proposes a significant contribution in this area by presenting a BN analysis framework for identifying causal as well as independent relationships among features of heart sound signals, and in general, similar bio-signals generated in ICT-based assistive environments, which addresses the above challenges and limitations. In particular, similar to other machine-learning approaches, but unlike most BN methods, the framework proposed is handling features as continuous rather than discrete, addressing the above-mentioned challenge (b). Additionally, due to the high-dimensionality nature of the dataset, exact computation of the CPDs is infeasible and computationally costly. Hence, the joint distribution is approximated by stochastic simulation commonly referred to as '*sampling*.' Using Markov Chain Monte Carlo (MCMC), one can fit a distribution to the data that converges to the posterior distribution (i.e., the distribution of the class, treated as a random variable, conditional on the evidence obtained from the dataset) and retain the samples. MCMC can cope with domains where the state space is huge (i.e., large number of features) 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 [31]. In this the above-mentioned challenges, (a), (c) and (d) are addressed. In particular, a new approach is proposed to approximate the conditional probability distributions of complex BN using a MCMC algorithm; it is demonstrated that this allows for the creation of a robust system for a highly detailed diagnosis of heart sound signals. The present 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, which is oriented toward classification. This structure is called Tree-Augmented Naïve Bayes (TAN) and unlike general, unrestricted BNs, in TAN, the class node is the root node, that is, the parent of all other nodes, which can form a BN among them, addressing in this way the above-mentioned challenge (d). This type of structure is proven to be more efficient than BN for classification purposes [32], since in traditional BN, the class node is not considered as a special type of node, and it is treated as an ordinary one, so it may even not appear in the network resulting in a lack of classification capabilities. Then, it is used in order to perform highly detailed diagnosis of heart sound signals: (1) initially as healthy or unhealthy, (2) then, for the unhealthy, distinguish among those having systolic and having diastolic murmurs, and finally, (3) in both systolic and diastolic murmurs cases discriminating between aortic or mitral dysfunction. Also, some widely used alternative classifiers have been applied to the same data for comparison purposes.

4 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 [33]. Bayesian inference involves the mathematical integration of high-dimensional probability distributions. This process is analytically intractable; therefore, it is common to deploy Monte Carlo (MC) techniques. MC techniques are requiring sampling from the probability distributions which are to be integrated. However, in many cases, it is not possible to draw such samples directly from the distributions. MCMC methods provide a unified framework for coping with such issues. The way they operate is twofold: at first, a Markov Chain is generated that converges to the target probability distribution. Subsequently, the target sample values are obtained using Monte Carlo integration.

4.1 MCMC methods

A probability distribution is specified through a DAG G (a set of interconnected nodes, each of which corresponds to one of the features) and a set of conditional probability distributions (parameters), one for each feature X_i–node in G. A BN is actually a DAG G where the topology refers to its structure and the CPD is encoded as a table, named CPT (Conditional Probability Table). By definition, 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 efficient as it requires too much computation. Several methodologies for alleviating this problem have been proposed, such as the *K2 algorithm* [30] or the *Bayesian Scoring Method* [31]. The following Sect. 5 describes the suggested technique for obtaining graph structures more straightforwardly and thus constructing BNs that enable efficient classification process.

Regardless of the structure-learning algorithm, given a structure G with nodes $X = \{X_1, X_2, \ldots, X_n\}$, the process of obtaining the CPD with sampling is described below. For reasons of comprehension suppose that G is referring to the example BN depicted in the following Fig. 1. Let us also assume that each node X_i is a binary node with values T or F. For each node X_i in G:

- Randomly select a state for all other nodes except for X_i.
  - *For example,* <?,T,T,F,F,T>
- Compute the probability distribution over the states of X_i, that is, $p(X_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

Note that since G is a Bayesian network, the above probability is simplified to include only the Markov Blanket of X_i [6], that is:

$$p(X_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = p(X_i|\text{parents}(X_i)) \prod_{j=1}^{k} (Y_j|\text{parents}(Y_j)),$$

where Y_j denotes the set of child nodes of X_i. For example:

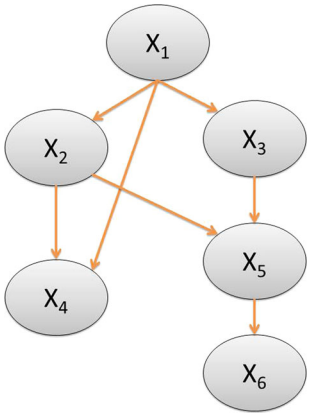


Fig. 1 An example BN consisting of 6 binary nodes with states True or False each

$$p(X_1 = T|X_2 = T, X_3 = T, X_4 = F, X_5 = F, X_6 = T) = p(X_1 = T)p(X_4 = F|X_2 = T)p(X_4 = F|X_1 = T) p(X_2 = T|X_1 = T)p(X_3 = T|X_1 = T)$$

and

$$p(X_1 = F|X_2 = T, X_3 = T, X_4 = F, X_5 = F, X_6 = T) = p(X_1 = F)p(X_4 = F|X_2 = T)p(X_4 = F|X_1 = F) p(X_2 = T|X_1 = F)p(X_3 = T|X_1 = F)$$

- From the probability distribution, randomly select a state of X_i to complete the sample vector.
  - For example suppose that value T is selected for node X_1.

Monte Carlo sampling requires drawing of n samples from the BN with each instance of feature states forming its value as explained above. For the purpose of this research, only continuous values are considered; therefore, the method of [34] is adopted and the samples are projected as a histogram. Afterward, the histogram is smoothed 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 dataset. 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 preferable for approximating the challenging high-dimensional distributions. The Gibbs sampler was chosen as an MCMC utilization method, because it is more suitable to DAG structures [33]. Furthermore, a Gibbs sampler can allow for convergence in reasonable computation time and its implementation code is widely available in the academic community (e.g., *WinBUGS* [35]).

4.2 MCMC and Gibbs sampling

Before describing the Gibbs sampler, a few introductory comments on Markov Chains are provided. Since Markov Chains by principle contain the concept of time, they used to be mostly associated with applications of data mining and pattern recognition which directly encompass this dimension, such as speech recognition and time series analysis [36]. However, Markov Chains could also be applied to BN search process, where each time step denotes a candidate network structure that is evaluated. In the approach presented, a Markov chain is designed where each state is a full joint instantiation of the distribution (i.e., values are assigned to all features of the network). Hence, a transition in time is a transfer from one joint instantiation to another. The target sampling distribution is the posterior joint distribution $P(x|e)$ where x is the class feature and e is the set of evidence features. It is typically the unknown that is to evaluate. Let X_t^i denote the value of a random variable X_i at time (or step) t, and let the state space refer to the range of possible X_i 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, that is:

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

In other words, for a random variable to be considered a Markov process, the only information about the past needed in order to predict the future is the current state of it. Any knowledge about the values of earlier states does not affect 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 matrix $P(j \rightarrow k)$, which is the probability that a process at state space s_j moves to state s_k in a single step, that is:

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

For reasons of readability, the notion of X_t^i into X_t shall be simplified 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 $\pi(t)$ denote the row vector of the state space probabilities at step t. The chain starts by specifying a starting vector $\pi(0)$. Often, all the elements of $\pi(0)$ 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, one can define the probability transition matrix P as the one whose element (i, j) denotes the $P(i \rightarrow j)$ transition kernel. The probability that the chain has state value s_i at time (or step) $t + 1$ is given by:

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

In other words, as the above equation implies, a Markov chain can reach a stationary (final) distribution π^*, regardless of the selection for the initial distribution parameters. In order to explain this more systematically, consider a random process in which the state S_0 is initialized according to an initial distribution p_0. On each time step t, with probability γ, the chain is 'stopped' and outputted the current state S_t. Moreover, with probability $1-\gamma$, a state transition step is taken and sample S_{t+1} according to the transition probabilities $p(S_{t+1}|S_t)$. Since the number of steps T is distributed according to a geometric distribution with parameter $(1 - \gamma)$, the random state that is generated by this process will also be distributed according to π.

A straightforward method of approaching this distribution includes sampling. While there are numerous sampling strategies, the Gibbs sampler [33] is well suited for DAGs, as shall be described in the next paragraphs.

4.3 Gibbs sampler

The main notion of this methodology is that only univariate conditional distributions are taken into account, that is, distributions where all of the random variables except for one are assigned fixed values. The reason for the above consideration lies to the fact that such conditional distributions are more straightforward to simulate than complex joint distributions and usually have simpler forms. To introduce the Gibbs sampler, consider a bivariate random variable (x, y) and suppose the computation of one or both probabilities, $p(x)$ and $p(y)$ is requested. 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 $p(x, y)$, for exsample, $p(x) = p(x, y)dy$. 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 $i \leq j \leq m < k$ 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 (1) 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 (2) 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 to simulate [34].

5 Methodology description

As mentioned above, one critical challenge is that in the occurrence of high-dimensional input vectors, the set of plausible network models is large; thus, a full comparison of all the posterior probabilities associated to the candidate models becomes infeasible. A solution to this can be grounded on the MCMC method and its variation, namely the Gibbs sampler. Note, however, that a direct application of the above algorithm for BN estimation within the heart sounds domain faces limitations, due to the high dimensionality of the data where the number of features is analogous to the number of available observations. 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. 2, can overcome this limitation. Initially inspired by the work of [34], which proposed the use of an initial set of 10–20 dissimilar but high-scoring BN [as regards to the probability of the network structure S given the input data D, $p(S|D)$], could be used for calculating the Bayesian posterior probability distribution of all features. The present approach is different than the previous one in two major points: the former deals with the fact that considering the top-k-ranked networks would result in obtaining very similar network structures and therefore, would result in having a set of distributions with limited variation. The reason is that when using traditional scoring algorithms, such as K2 [30] or Bayesian Scoring Method [31], each candidate network is produced from the previous, most-likely one by performing simple graph operation such as arc additions, removals or reversals. One possible solution would be to consider multiple and parallel search implementation, but this could create an extra computational overhead. The latter aspect that this framework differs is the orientation toward classification, which is not present in traditional BN approaches. The approach of [34] performs Bayesian inference on the class feature given the set of input variables having simulated generic network structures that do not consider the class node as a special one. The proposed suggestion focuses on creating simple and straightforward BN structures which are suitable for the classification process (since *classifying a heart sound is the final goal*). Such classification-oriented network structures are constructed using the Tree-Augmented Naïve Bayes (TAN) algorithm [32]. By definition, the TAN algorithm creates networks where the class node is a parent of all features nodes. The rest of the input features form a traditional BN among them in which each node has one parent at most, in order to retain the structure and the CPD simple. Furthermore, compared to the traditional BN-learning algorithms, the TAN methodology can produce networks approximately 50–100 times faster than the BN approach, depending on the number of input features and the number of states each feature has. Finally, TANs are considered more appropriate for classification than both BNs and the Naïve Bayes approach. This conclusion is attributed to the structural characteristics of the former, which considers the class node as a parent of all other nodes (as Naïve Bayes does) but also considers features as not to be conditional independent given the class node, a fact that is non-realistic in a plethora of domains.

From the samples drawn from a set of different TAN structures, the posteriors after convergence can be obtained, to then determine the probability estimates of the model in a straightforward manner. Despite the fact that 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 of these network structures are sampled to estimate the probability distribution accurately. The important element of this methodology is the use of fast-learned TAN structures and a rank ordering among them.

The following Fig. 2 presents a flowchart of the proposed methodology, showing its above-mentioned main components: TAN learning phase, Gibbs sampling phase and finally, convergence phase.

5.1 TAN phase

Based on the following process, a set of 10 TAN networks structures were obtained:

1. Built a Naïve Bayesian structure where the class node C is a parent to all feature nodes X_i and all feature nodes are not connected with each other.

For each pair of different features X_i and X_j, compute the conditional mutual information given the class $I(X_i; X_j|C)$, using the formula given below, i.e.:

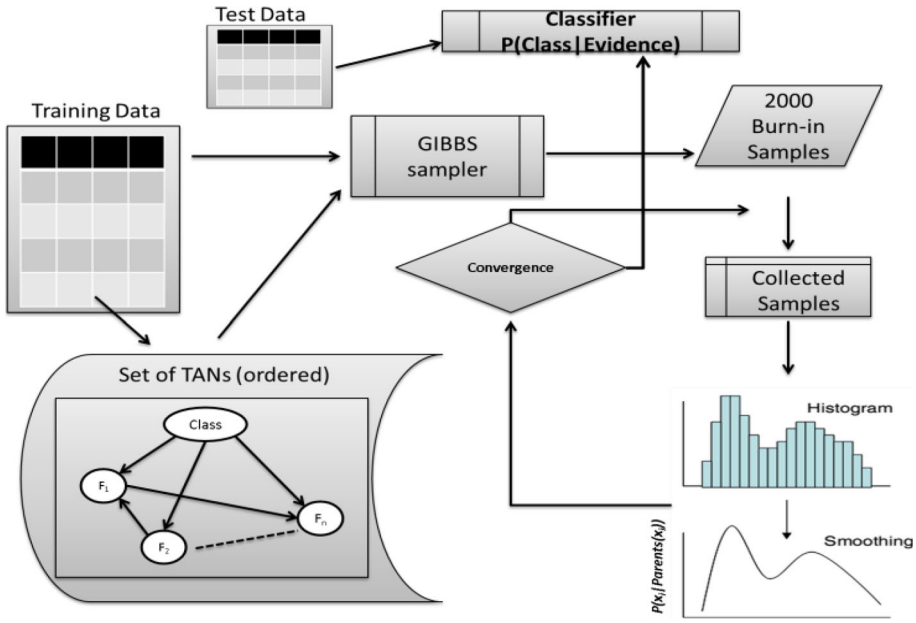


Fig. 2 The flowchart of the proposed methodology, showing its main components: TAN learning phase, Gibbs sampling phase and finally, convergence phase

$$I(X_i; X_j|C) = \sum_{X_i,X_j,C} \frac{p(X_i, X_j, C)\log\left(p(X_i, X_j|C)\right)}{p(X_i|C)p(X_j|C)}$$

Build a complete, undirected graph to connect all features and use $I(X_i; X_j|C)$ to weight all arcs.

2. Build a maximum-weighted spanning tree.
3. 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, a feature selection algorithm was applied based on SVM [37] and eliminated the features that scored below 0.1, thus achieving a mean value of 40–60 % reduction in the number of input features for the TAN learner. According to the authors of the aforementioned article, feature selection by SVM is more beneficial than other wrapper approaches such as information gain and odds ratio [38] when being applied in high-dimensional datasets. The different TAN structures were obtained by choosing different features as root, in the 5th step of the previously mentioned TAN algorithm. As mentioned above, an ordinary Gibbs sampler chooses features at random and then samples a new value from the estimated posterior of the neighboring variables. Friedman [4] 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. The score of each network S is calculated as the probability of S given dataset D, and $p(S|D)$ is given by the following formula [39]:

$$p(S|D) = \prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{k=1}^{r_i}N_{ijk}!$$

where n equals to the number of features; r_i denotes the number of values in the ith feature; q_i denotes the number of possible different value combinations the parent features can take; N_{ij} depicts the number of rows in data that have jth value combinations for parents of ith feature; and N_{ijk} corresponds to the number of rows that have kth value for the ith variable and which also have jth value combinations for parents of ith variable

Note that other graph-scoring metrics could be used as well, such as the *BIC-TAN* measure, proposed by [40]. The applied scoring metric was chosen because it is implemented in a variety of programming languages and is freely available.

5.2 Gibbs sampling phase

For the Gibbs sampling phase, uniform prior distributions for all the features in the domain need to be defined. Instead of applying random initial state of the network, a multivariate Dirichlet distribution was chosen, inspired by [33]. The initial distribution of the states of nodes in the network was assigned by using the density function. It was estimated after smoothening of the histogram of normalized feature data. Since all nodes have parent(s), sampling was made 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. 2 above. The probability density function $P(x)$ of a continuous feature was approximated by smoothening of the histogram.

5.3 Convergence phase

Convergence is the process of reaching a stationary probability distribution. The initial phase of the convergence is called the 'burn-in' phase. For the proposed approach, 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 ($X_1 = x_1$, $X_2 = x_2$, …, $X_n = x_n$), then a new value x_1 for feature X_1 can be sampled using $p(X_1 = x_1, |X_2 = x_2, \ldots, X_n = x_n)$. In similar manner, one can sample the remaining new values for features X_2, X_3 … X_n until a new state S_1, instantiated as: $\langle X_1 = x_1', X_2 = \text{x}_2, \ldots, X_n = x_{n\prime}\rangle$. In the following Fig. 3, a sample Markov Chain is depicted for a mockup TAN structure, with two features, each being binary. The chain represents four states for each instantiation of features X_1 and X_2.

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. Only the probability of the class c given evidence e (expressed as an input vector of the considered feature values) is computed, calculated as $p(c|e)$ and classify it to the most probable class.

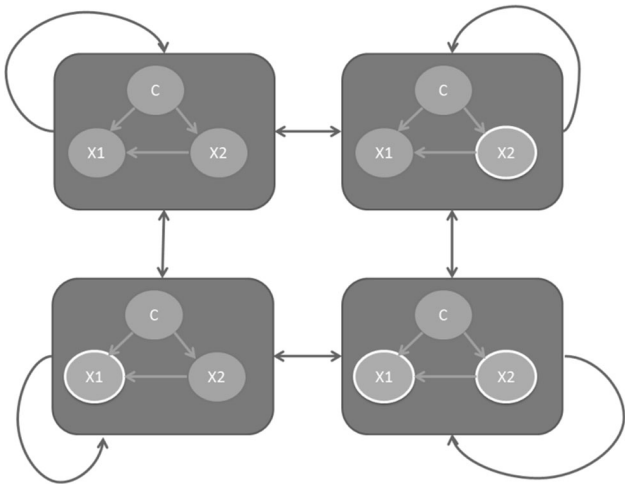


Fig. 3 An example Markov Chain for a mockup TAN structure—Each feature is instantiated to either true (*highlighted circle*) or false (*non-highlighted circle*) at each state

6 Data and preprocessing

The dataset used for a first application and validation of the proposed methodology consisted of 198 heart sound signals, which have been acquired from both healthy medical cases and pathological ones having one of the following four frequent and severe heart valve diseases: aortic stenosis (AS), mitral regurgitation (MR), aortic regurgitation (AR) or mitral stenosis (MS). A heart sound signal from a healthy medical case has the form shown in the upper part of Fig. 4. It consists of four main components: (a) the first heart sound (S1), which is generated by the nearly simultaneous closure of the mitral and the tricuspid valve after the return of blood from the body and the lungs; (b) it is followed by the systolic phase; and then (c) the second heart sound (S2), which is generated by the nearly simultaneous closure of the aortic and the pulmonic valve as the blood is pushed to the body and the lungs; and finally, (d) the diastolic phase [4, 41]. Most heart diseases generate additional components in the heart sound, from which they can be diagnosed. The valve diseases correspond to pathological functioning of one of the four valves of the heart (aortic, mitral, tricuspid, pulmonic), due to either stenosis (resulting in blood flow reduction) or regurgitation (problematic closure leading to back flow of blood). Valve diseases generate additional noise components, referred to as murmurs, in the systolic phase (systolic murmurs) or/and the diastolic phase (diastolic murmurs), having the form shown in the lower part of Fig. 4. With respect to the above-mentioned four valve diseases dealt with in this paper, AS and MR generate systolic murmurs, while AR and MS generate diastolic murmurs [4, 22, 41].

In particular, 38 of the heart sound signals of the dataset were healthy, while the remaining 160 were pathological: 41 from patients with AS, 43 from patients with MR, 38 from patients with AR and 38 from patients with MS.

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 patient's 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., these factors are generating additional components), making the detection of various heart diseases and pathological conditions from

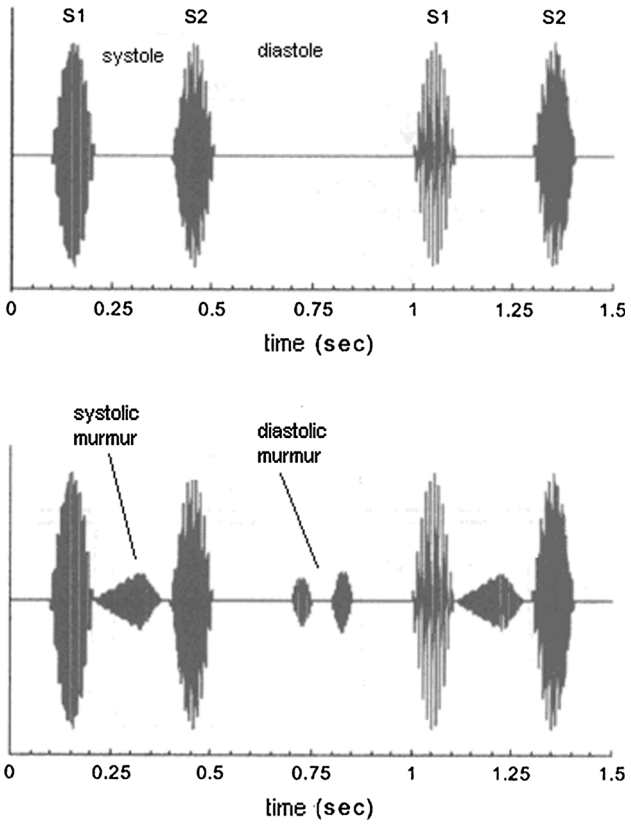


Fig. 4 Heart sound signals from a healthy heart (*upper* part); Heart sound signals from a pathologic heart generating systolic and diastolic murmurs (*lower* part)

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 this research more realistic, it was decided to create a 'global' and representative dataset (and not use one of the publicly available datasets), including 'heterogeneous' heart sounds recorded with various different acquisition methods and 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. For creating this dataset, heart sounds from several different heterogeneous sources were combined: educational audio cassettes, audio CDs, CD ROMs, files of existing heart sound databases, etc., which had been recorded with various different acquisition methods and values of the above factors, and then had been diagnosed by experienced cardiologists and classified to one of the above five heart health conditions.

Initially, a preprocessing of these heart sounds was performed, in order to remove noise and extract features from them. The preprocessing method is described in detail in [4, 28]. 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, for each heart sound signal, two mean signals were calculated 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 preprocessing 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.

7 Experimental results

The experimental part of the study is organized as follows:

A. Initially, a broad classification of heart sounds as normal–healthy (NRM) or sick–unhealthy (SCK) was performed.

B. The inferred instances that were predicted as SCK were further classified as having systolic (STL) or diastolic (DTL) murmur.
C. Finally, for each of the aforementioned two classes formed in B, 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 also compared against other well-known alternative classification algorithms that have previously been referred to as having provided 'state-of-the-art' results in the heart sounds diagnosis domain. In particular, the performance-proposed methodology was evaluated against Naïve Bayes, Decision Trees, Neural Networks (with Radial Basis Functions) and k-Nearest Neighbor ($k = 3$) for the same 'difficult' dataset, using for all a 10-fold cross-validation approach. This approach is regarded by the relevant literature as the most appropriate for the evaluation of classification performance. For instance, Kohavi [42] compared numerous approaches for evaluating classification performance, both cross-validation and bootstrap (sample with replacement) ones, and concluded that 10-fold cross-validation is the best approach, as it tends to provide less biased accuracy estimations. The *RapidMiner*® data mining suite [43] was used for implementing both the proposed methodology and the alternative classification algorithms.

7.1 Discrimination between healthy and unhealthy signals

From a medical expert's perspective, the accuracy of discrimination between healthy–normal and unhealthy–sick is of major importance. Due to the significance of this 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:

Predicted class	Actual class	
	Sick	Normal
Sick	TP	FP
Normal	FN	TN

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

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

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

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

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

The anticipated classification outcome is the one that does not erroneously classifies as healthy a patient with a heart disease. Therefore, focus is particularly placed on the performance of the proposed methodology in situations where sick patients were identified as healthy (FPR), and this can be plotted against TPR in order to produce the widely used ROC (Receiver Operating Characteristic) curve, which is shown in Fig. 5. This curve is a very good visualization of classification performance. Robust classifiers are expressed by ROC curves which retain high values of TPR for most of the horizontal axis area (FPR). The ROC curve is calculated according to the following process:

I. The test instances are sorted according to the probability of belonging to the healthy class, in increasing order.
II. Select each instance, respectively, starting from the instance with the lowest rank:
A. Assign a label to it and to all other instances that are ranked above it to the healthy class.
B. Calculate the TP, FP, TN, FN and FNR, TRP metrics as mentioned above.
C. Plot the FNR, TPR values
III. Repeat 2 until the top-ranked instance is selected.

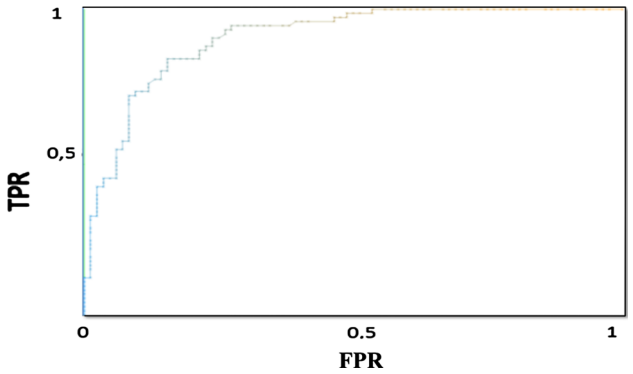


Fig. 5 ROC curve for the Healthy/Unhealthy experiment

The above ROC curve shows that the method presented is capable of classifying more that 87 % of the instances correctly. Moreover, the results of the proposed methodology are presented, as well as the above-mentioned alternative classifiers utilizing the *F-measure*, that is, the harmonic mean of precision P and recall R:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad F = \frac{2\mathrm{PR}}{\mathrm{P} + \mathrm{R}}$$

Additionally, two metrics that originate from the domain of statistical analysis of classification models were incorporated, in order to further evaluate classification performances:

7.1.1 Root mean squared error (RMSE)

The RMSE E of an algorithm is calculated using the equation:

$$E = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(P_j - T_j)^2},$$

where P_j is the value that algorithm forecasted for the sample j (from a set of examples) and T_j is the value of the 'actual class' for the j-th example. For an ideal classification, $P_j = T_j$ and $E = 0$. So, the *RMSE* indicator varies from 0 to infinity, with 0 to correspond to the ideal classification.

7.1.2 Root relative squared error (RRSE)

The RRSE E of an algorithm is calculated by the equation:

$$E = \frac{\sqrt{\sum_{j=1}^{n}(P_j - T_j)^2}}{\sqrt{\sum_{j=1}^{n}\left(T_j - \left(\frac{1}{n}\sum_{j=1}^{n}T_j\right)\right)^2}},$$

where again P_j is the value that algorithm forecasted for the sample j and Tj is the value of the 'actual class' for the j-th example. For an ideal classification, $P_j = T_j$ and $E = 0$. So, the RRSE indicator varies from 0 to infinity, with 0 to correspond to the ideal classification.

The results of the first experimental phase (i.e., classification of heart signals as healthy–normal or unhealthy–sick) measured in F-measure ratio are shown in Fig. 6.

The results illustrate that the proposed MCMC methodology shows a good classification performance (85, 65 %) for this heterogeneous and 'difficult' dataset in this first phase of diagnosis, and performs better than all other alternative methodologies, providing a gain between 3 and 15 % in certain cases. Moreover, Fig. 7 presents the RMSE and RRSE metrics for each classification algorithm. Again, MCMC Bayesian inference projects a substantially lower error rate for both RRSE and RMSE metrics, which indicates more robust classification.

7.2 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 presented in Fig. 8, while the statistical error metrics are presented in Fig. 9.

It should be remarked that the 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. Additionally, in the 'Appendix,' the best scoring TAN structure for this more detailed classification is available 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 subset of them (20) on which the classification is based. This reduction on one hand caused significant improvement of the MCMC step in terms of computational complexity. On the other hand, it shows to the medical experts in a visualized and easy-to-understand manner on which characteristics of the heart signals the classification has been based on, which is highly beneficial and enhances the acceptability of these tools.

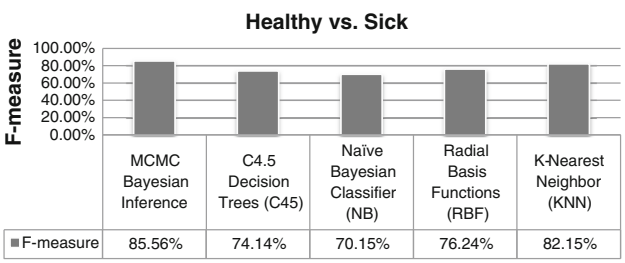
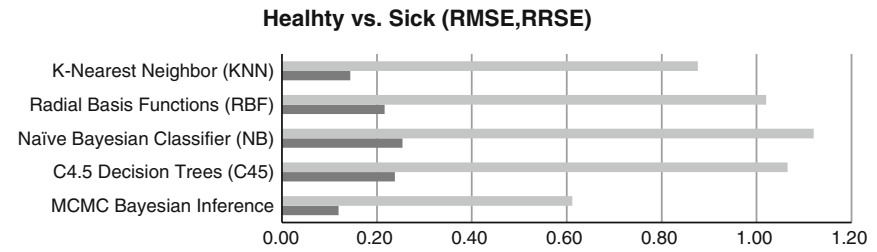


Fig. 6 Performance of the proposed MCMC methodology and the alternatives for the discrimination between Healthy and Sick heart signals



	MCMC Bayesian Inference	C4.5 Decision Trees (C45)	Naïve Bayesian Classifier (NB)	Radial Basis Functions (RBF)	K-Nearest Neighbor (KNN)
RRSE	0.612	1.065	1.120	1.020	0.876
RMSE	0.119	0.238	0.254	0.216	0.144

Fig. 7 Error rates for the proposed MCMC methodology and the alternatives measured in RMSE and RRSE for the Healthy–Unhealthy discrimination

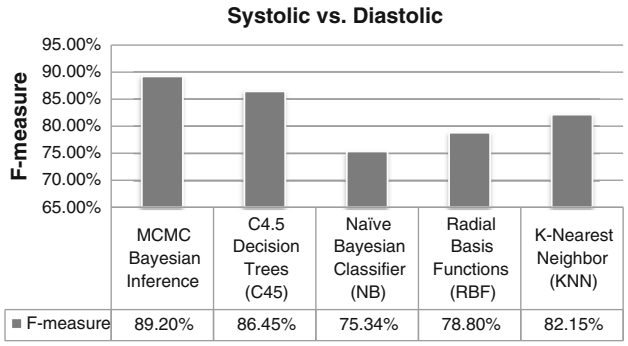


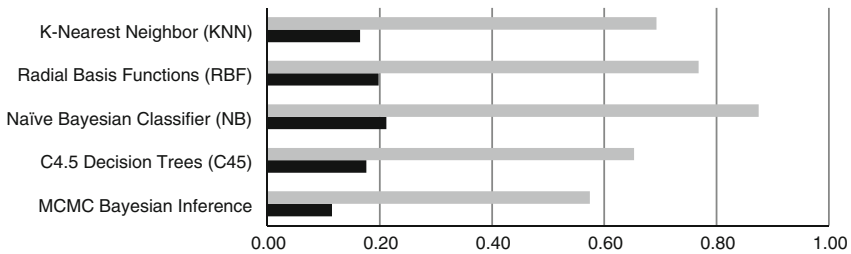
Fig. 8 Performance of the proposed MCMC methodology and the alternatives for discrimination between Systolic and Diastolic murmurs

7.3 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 as mentioned, 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 Figs. 10 and 11. MCMC is again the most efficient approach and outperforms all other alternative approaches. As regard to the latter case, the results are shown in Figs. 12 and 13, respectively. It is observed that MCMC still exhibits the highest classification performance, which outperforms the other alternative approaches by a varying percentage of 0.5–15 %.

8 Conclusions

The research presented in this paper focuses on a highly important part of an ICT-based assistive environment for the remote home healthcare monitoring of the elderly, the disabled and also patients with various chronic diseases: a component for on-line screening of the numerous signals



	MCMC Bayesian Inference	C4.5 Decision Trees (C45)	Naïve Bayesian Classifier (NB)	Radial Basis Functions (RBF)	K-Nearest Neighbor (KNN)
RRSE	0.574	0.653	0.875	0.768	0.693
RMSE	0.115	0.176	0.212	0.198	0.165

Fig. 9 Error rates for the proposed MCMC methodology and the alternatives measured in RMSE and RRSE for the Systolic–Diastolic murmurs discrimination

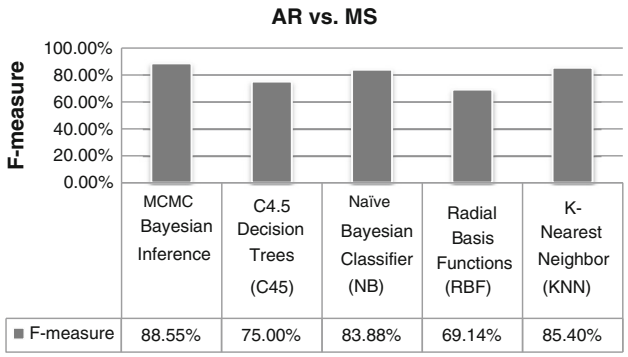
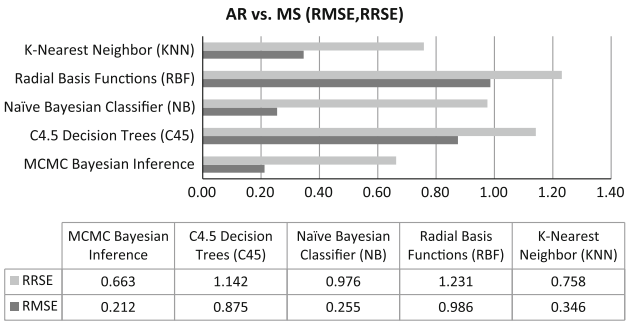


Fig. 10 Performance of the proposed MCMC methodology and alternatives for discrimination between AR–MS for Diastolic murmurs



	MCMC Bayesian Inference	C4.5 Decision Trees (C45)	Naïve Bayesian Classifier (NB)	Radial Basis Functions (RBF)	K-Nearest Neighbor (KNN)
RRSE	0.663	1.142	0.976	1.231	0.758
RMSE	0.212	0.875	0.255	0.986	0.346

Fig. 11 Error rates for the proposed MCMC methodology and alternatives measured in RMSE and RRSE, for the AR–MS Diastolic experiment

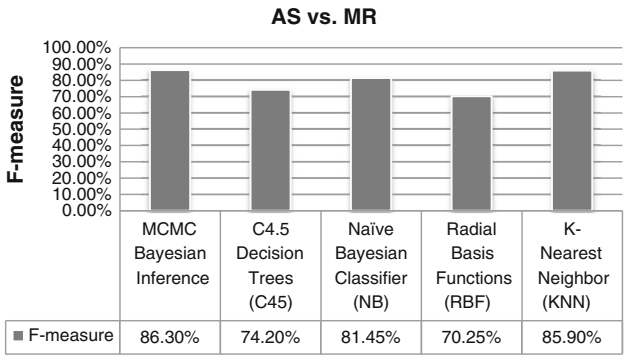
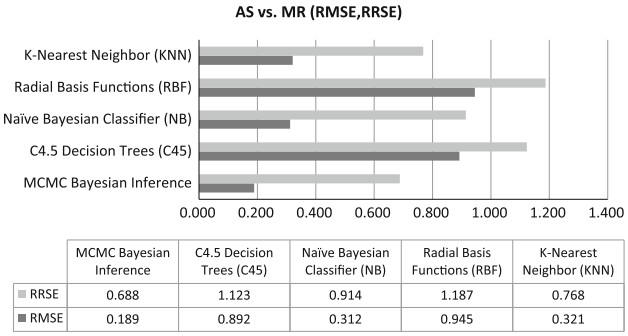


Fig. 12 Performance of the proposed MCMC methodology and alternatives for discrimination between AS–MR for Systolic murmurs



	MCMC Bayesian Inference	C4.5 Decision Trees (C45)	Naïve Bayesian Classifier (NB)	Radial Basis Functions (RBF)	K-Nearest Neighbor (KNN)
RRSE	0.688	1.123	0.914	1.187	0.768
RMSE	0.189	0.892	0.312	0.945	0.321

Fig. 13 Error rates for the proposed MCMC methodology and alternatives measured in RMSE and RRSE, for the AS–MR Systolic experiment

generated in the houses of the supported people in order to identify abnormal ones that require some medical action and produce notifications to authorized medical personnel. In order to enable cost-effective and sustainable large-scale remote health monitoring for the elderly, chronically ill and disabled at home, it is critical to develop capabilities of automating the screening of the numerous signals generated and identify abnormal ones, reducing the required human effort and associated costs. This paper addresses this need. It proposes a methodology for the automatic screening of heart sound signals acquired in home care context, which can, however, be applied to many other similar types of bio-signals generated in 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 methodology can handle datasets characterized by numerous continuous input features and limited training data; it addresses the inherent limitations and challenges of using BN for such medical diagnosis problems. It enabled a highly detailed diagnosis of heart sound signals including (1) classification as healthy or unhealthy, (2) classification of unhealthy ones as having systolic or diastolic murmurs, and then, (3) classification of both groups as being of aortic or mitral origin. The findings of this first application showed a good performance in a highly heterogeneous and difficult dataset, which is higher than the most widely used alternative classifiers. The proposed methodology can be very well incorporated in ICT-based home-assistive environments: the electronic stethoscope can be a small wearable device, which is wirelessly linked to a home station that can transmit the acquired heart sound signals to the nearest health center or hospital through fixed line or mobile phone, in order to be processed there. It is argued that if in the context of home-assistive environments some of the heart sound acquisition factors are controlled (e.g., type of stethoscope/sensor, stethoscope use mode and filtering, patient's position and auscultation area) and more training data is available, higher performance levels can then be achieved. The proposed approach can be used for other similar types of signals generated in home-assistive environment contexts, in order to identify pathological situations that need medical action. Further research is required in this direction.

Appendix

See Fig. 14.

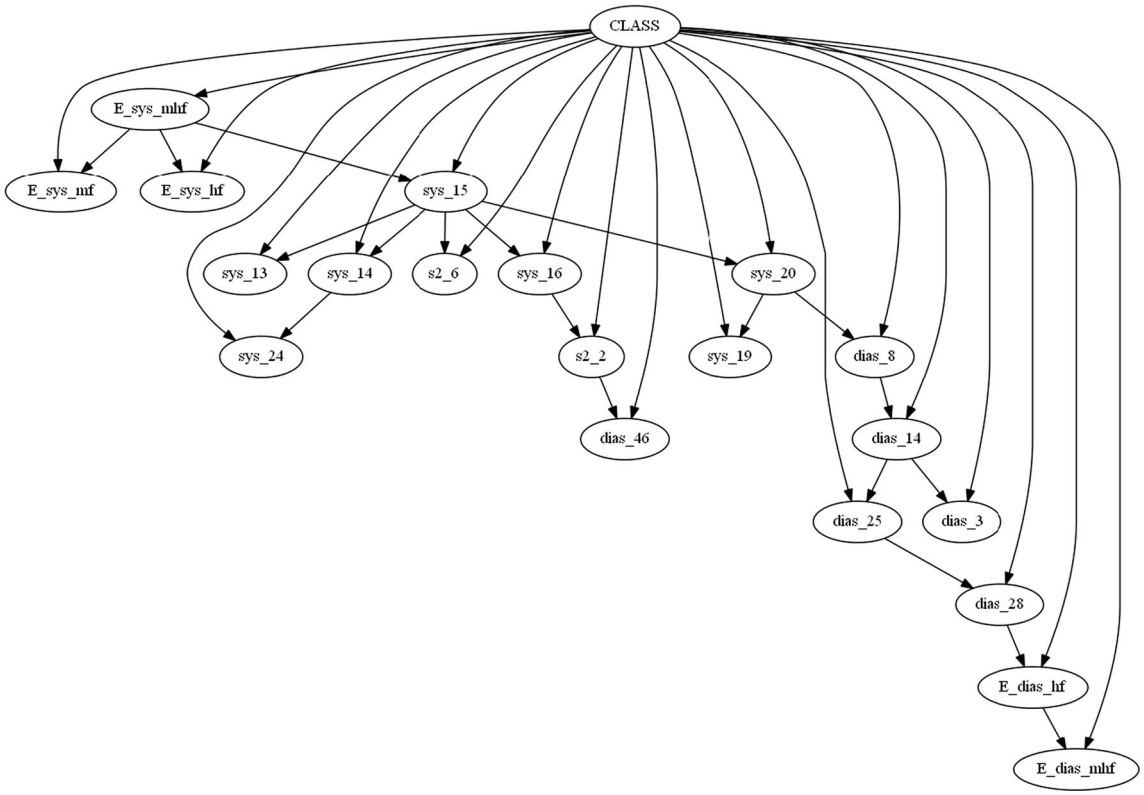


Fig. 14 The best scoring TAN structure for the Systolic–Diastolic discrimination upon performing SVM feature selection

References

1. Vassis, D., Belsis, P., Skourlas, C., Pantziou, G.: Providing advanced remote medical treatment services through pervasive environments. Pers. Ubiquit. Comput. **14**, 563–573 (2010)
2. Chauhan, S., Wang, P., Lim, C.S., Anantharaman, V.: A Computer Aided MFCC based HMM system for automatic auscultation. Comput. Biol. Med. **38**(2), 221–233 (2008)
3. Friedman, N., Koller, D.: Being Bayesian about network structure: a Bayesian approach to structure discovery in Bayesian networks. Mach. Learn. **50**, 95–126 (2003)
4. 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, pp. 245–274. Springer Verlag, Autumn (2004)
5. Cathers, I.: Neural network assisted cardiac auscultation. Artif. Intell. Med. **7**(1), 53–66 (1995)
6. 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 Signals Process. (ICASSP) **5**, 3455–3458 (1995)
7. Wu, C.H.: On the analysis and classification of heart sounds based on segmental Bayesian networks and time analysis. J. Chin. Inst. Electr. Eng. Transact. Chin. Inst. Eng. Ser. E **4**(4), 343–350 (1997)
8. 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 2: 988–991 (2000)
9. White, P.R., Collis, W.B., Salmon, A.P.: Analysing heart murmurs using time-frequency methods. In: Proceeding of the IEEE-SP International Symposium Time-Frequency and Time-Scale Analysis, pp. 385–388. (1996)
10. 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, pp. 365–366. (1996)
11. 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, pp. 1628–1631. (1998)
12. 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, pp. 926. (1999)
13. Noponen, A.L., Lukkarinen, S., Angerla, A., Sikio, K., Sepponen, R.: How to recognize the innocent vibratory murmur. Comput. Cardiol. **25**, 561–564 (2000)

14. DeGroff, C., Bhatikar, S., Mahajan, R.: A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics. Artif. Intell. Med. **33**, 251–260 (2005)
15. De Vos, J.P., Blanckenberg, M.: Automated pediatric cardiac auscultation. IEEE Trans. Biomed. Eng. **54**(2), 244–252 (2007)
16. Akay, M.: Noninvasive diagnosis of coronary artery disease using a neural network algorithm. Biol. Cybern. **67**, 361–367 (1992)
17. Akay, Y., Akay, M., Welkowitz, W., Kostis, J.: Noninvasive detection of coronary artery disease. IEEE Eng. Med. Biol. **13**(5), 761–764 (1994)
18. 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, 2:1278–1279. (1994)
19. 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, 3, pp. 1546–1548. 29 Oct–1 Nov (1998)
20. Tateishi, O.: Clinical significance of the acoustic detection of coronary artery stenosis. J. Cardiol. **38**(5), 255–262 (2001)
21. 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. J. Heart Valve Dis. **2**(4), 454–467 (1993)
22. Hebden, J.E., Torry, J.N.: Identification of aortic stenosis and mitral regurgitation by heart sound analysis. Comput. Cardiol. **24**, 109–112 (1997)
23. 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, 4:3506–3509. (2005)
24. 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. Med. Biol. Eng. Comput. **43**, 451–456 (2005)
25. Voss, A., Mix, A., Huebner, T.: Diagnosing aortic valve stenosis by parameter extraction of heart sound signals. Ann. Biomed. Eng. **33**(9), 1167–1174 (2005)
26. 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. J. Med. Eng. Technol. **30**(2), 61–68 (2006)
27. Ahlstrom, C., Hult, P., Rask, P., Karlsson, J.E., Nylander, E., Dahlstrom, U., Ask, P.: Feature extraction for systolic heart murmur classification. Ann. Biomed. Eng. **34**(11), 1666–1677 (2006)
28. Pavlopoulos S., Stasis A., Loukis E.: A decision tree—based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds. BioMed. Eng. Online, 2004
29. Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A.: Support vectors machine based identification of heart valve diseases using heart sounds. Comput. Methods Progr. Biomed. **95**(1), 47–61 (2009)
30. Heckerman, D.: A Tutorial on Learning with Bayesian Networks. In: Jordan, M. (ed.) Learning in Graphical Models. MIT Press, Cambridge (1999)
31. Pearl, J.: Causality: Models, Reasoning and Inference. Cambridge University Press, Cambridge (2000)
32. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks for Plausible Inference. Morgan Kaufmann, San Mateo (1988)
33. 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)
34. Liu, J.S.: Monte Carlo Strategies in Scientific Computing. Springer, Heidelberg (2001)
35. Lunn, A., Thomas, G., Best, H., Spiegelhalter, D.: WinBUGS—A Bayesian modeling framework: concepts, structure, and extensibility. Stat. Comput. **10**, 325–337 (2000)
36. Nummelin, E.: General irreducible Markov chains and non-negative operators. Cambridge University Press, 1984, 2004
37. Bi, J., Bennett, K., Embrechts, M., Breneman, C., Song, M.: Dimensionality Reduction via Sparse Support Vector Machines. J. Mach. Learn. Res. **3**, 1229–1243 (2003)
38. Brank, J., Grobelnik, M., Milić-Frayling, N., Mladenić, D.: Feature selection using support vector machines. Proceedings of the 3rd International Conference on Data Mining Methods and Databases for Engineering, Finance, and Other Fields, Bologna, Italy, 2002
39. Cooper, G., Hersovits, E.: A Bayesian method for the induction of probabilistic networks from data. Mach. Learn. **9**, 309–347 (1992)
40. Chickering, D., Geiger, D., Heckerman, D.: Learning Bayesian networks: Search methods and experimental results. In Proceedings of Fifth Conference on Artificial Intelligence and Statistics, pp. 112–128. (1995)
41. Ericson, B.: Heart Sounds and Murmurs: A Practical Guide, Inc. Mosby-Year Book, St. Louis (1997)
42. Kohavi R: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of International Joint Conference on AI, pp. 1137–1145. (1995)
43. Rapid-Miner, A.: Software for Data Mining Tasks, www.rapid-i.com