Research

A decision tree – based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds

Sotiris A Pavlopoulos*[1], Antonis CH Stasis[1] and Euripides N Loukis[2]

Address: [1]National Technical University of Athens, School of Electrical and Computer Engineering, Biomedical Engineering Laboratory, Iroon Polytexniou Zografou 15773, Athens, Greece and [2]University of Aegean, Dept. of Information and Communication Systems Engineering, Karlovassi 83200, Samos, Greece

Email: Sotiris A Pavlopoulos* - spav@biomed.ntua.gr; Antonis CH Stasis - astasis@biomed.ntua.gr; Euripides N Loukis - eloukis@aegean.gr

* Corresponding author

Abstract

Background: New technologies like echocardiography, color Doppler, CT, and MRI provide more direct and accurate evidence of heart disease than heart auscultation. However, these modalities are costly, large in size and operationally complex and therefore are not suitable for use in rural areas, in homecare and generally in primary healthcare set-ups. Furthermore the majority of internal medicine and cardiology training programs underestimate the value of cardiac auscultation and junior clinicians are not adequately trained in this field. Therefore efficient decision support systems would be very useful for supporting clinicians to make better heart sound diagnosis. In this study a rule-based method, based on decision trees, has been developed for differential diagnosis between "clear" Aortic Stenosis (AS) and "clear" Mitral Regurgitation (MR) using heart sounds.

Methods: For the purposes of our experiment we used a collection of 84 heart sound signals including 41 heart sound signals with "clear" AS systolic murmur and 43 with "clear" MR systolic murmur. Signals were initially preprocessed to detect 1st and 2nd heart sounds. Next a total of 100 features were determined for every heart sound signal and relevance to the differentiation between AS and MR was estimated. The performance of fully expanded decision tree classifiers and Pruned decision tree classifiers were studied based on various training and test datasets. Similarly, pruned decision tree classifiers were used to examine their differentiation capabilities. In order to build a generalized decision support system for heart sound diagnosis, we have divided the problem into sub problems, dealing with either one morphological characteristic of the heart-sound waveform or with difficult to distinguish cases.

Results: Relevance analysis on the different heart sound features demonstrated that the most relevant features are the frequency features and the morphological features that describe S1, S2 and the systolic murmur. The results are compatible with the physical understanding of the problem since AS and MR systolic murmurs have different frequency contents and different waveform shapes. On the contrary, in the diastolic phase there is no murmur in both diseases which results in the fact that the diastolic phase signals cannot contribute to the differentiation between AS and MR.

We used a fully expanded decision tree classifier with a training set of 34 records and a test set of 50 records which resulted in a classification accuracy (total corrects/total tested) of 90% (45 correct/50 total records). Furthermore, the method proved to correctly classify both AS and MR

cases since the partial AS and MR accuracies were 91.6% and 88.5% respectively. Similar accuracy was achieved using decision trees with a fraction of the 100 features (the most relevant). Pruned Differentiation decision trees did not significantly change the classification accuracy of the decision trees both in terms of partial classification and overall classification as well.

Discussion: Present work has indicated that decision tree algorithms decision tree algorithms can be successfully used as a basis for a decision support system to assist young and inexperienced clinicians to make better heart sound diagnosis. Furthermore, Relevance Analysis can be used to determine a small critical subset, from the initial set of features, which contains most of the information required for the differentiation. Decision tree structures, if properly trained can increase their classification accuracy in new test data sets. The classification accuracy and the generalization capabilities of the Fully Expanded decision tree structures and the Pruned decision tree structures have not significant difference for this examined sub-problem. However, the generalization capabilities of the decision tree based methods were found to be satisfactory. Decision tree structures were tested on various training and test data set and the classification accuracy was found to be consistently high.

Background

New technologies like Echocardiography, Color Doppler, CT, and MRI provide more direct and accurate evidence of heart disease than heart auscultation. However, these modalities are costly, large in size and operationally complex [1]. Therefore these technologies are not suitable for use in rural areas, in homecare and generally in primary healthcare set-ups. Although heart sounds can provide low cost screening for pathologic conditions, internal medicine and cardiology training programs underestimate the value of cardiac auscultation and junior clinicians are not adequately trained in this field. The pool of skilled clinicians trained in the era before echocardiography continues to age, and the skills for cardiac auscultation is in danger to disappear [2]. Therefore efficient decision support systems would be very useful for supporting clinicians to make better heart sound diagnosis, especially in rural areas, in homecare and in primary healthcare. Recent advances in Information Technology systems, in digital electronic stethoscopes, in acoustic signal processing and in pattern recognition methods have inspired the design of systems based on electronic stethoscopes and computers [3,4]. In the last decade, many research activities were conducted concerning automated and semi-automated heart sound diagnosis, regarding it as a challenging and promising subject. Many researchers have conducted research on the segmentation of the heart sound into heart cycles [5-7], the discrimination of the first from the second heart sound [8], the analysis of the first, the second heart sound and the heart murmurs [9-12], and also on features extraction and classification of heart sounds and murmurs [14,15]. These activities mainly focused on the morphological characteristics of the heart sound waveforms. On the contrary very few activities focused on the exploitation of heart sound patterns for the direct diagnosis of cardiac diseases. The research of Dr. Akay and co-workers in Coronary Artery

Disease [16] is regarded as very important in this area. Another important research activity is by Hedben and Torry in the area of identification of Aortic Stenosis and Mitral Regurgitation [17]. The reason for this focus on the morphological characteristics is the following: Cardiac auscultation and diagnosis is quite complicated depending not only on the heart sound, but also on the phase of the respiration cycle, the patient's position, the physical examination variables (such as sex, age, body weight, smoking condition, diastolic and systolic pressure), the patient's history, medication etc. The heart sound information alone is not adequate in most cases for heart disease diagnosis, so researchers generally focused on the identification and extraction of the morphological characteristics of the heart sound.

The algorithms which have been utilized for this purpose were based on: i) Auto Regressive and Auto Regressive Moving Average Spectral Methods [5,11], ii) Power Spectral Density [5], iii) Trimmed Mean Spectrograms [18], iv) Sort Time Fourier Transform [11], v) Wavelet Transform [7,9,11], vi) Wigner-Ville distribution, and generally the ambiguous function [10]. The classification algorithms were mainly based on: i) Discriminant analysis [19], ii) Nearest Neighbour [20], iii) Bayesian networks [20,21], iv) Neural Networks [1,8,18] (backpropagation, radial basis function, multiplayer perceptron, self organizing map, probabilistic neural networks etc) and v) rule-based methods [15].

In this paper a rule-based method, based on decision trees, has been developed for differential diagnosis between the Aortic Stenosis (AS) and the Mitral Regurgitation (MR) using heart sounds. This is a very significant problem in cardiology since there is very often confusion between these two diseases. The correct discrimination between them is of critical importance for the

determination of the appropriate treatment to be recommended. Previous research activities concerning these diseases were mainly focused on their clinical aspects, the assessment of their severity by spectral analysis of cardiac murmurs [13] and the time frequency representation of the systolic murmur they produce [9,22]. The problem of differentiation between AS and MR has been investigated in [3] and [17]. The method proposed in [3] was based on the different statistic values in the spectrogram of the systolic murmurs that these two diseases produce. The method proposed in [17] was based on frequency spectrum analysis and a filter bank envelope analysis of the first and the second heart sound. Our work aims to investigate whether decision tree-based classifier algorithms can be a trustworthy alternative for such heart sound diagnosis problems. For this purpose a number of different decision tree structures were implemented for the classification and differentiation of heart sound patterns produced by patients either with AS or with MR. We chose the decision trees as classification algorithm because the knowledge representation model that they produce is very similar to the differential diagnosis that the clinicians use. In other words this method does not work as a black box for the clinicians (i.e. in medical terms). On the contrary neural networks, genetic algorithms or generally algorithms that need a lot of iteration in order to converge to a solution are working as a black box for the clinicians. Using decision trees clinicians can trace back the model and either accept or reject the proposed suggestion. This capability increases the clinician confidence about the final diagnosis.

In particular, the first goal of this work was to evaluate the suitability of various decision tree structures for this important diagnostic problem. The evaluated structures included both Fully Expanded decision trees Structures and Pruned decision trees Structures. The second goal was to evaluate the diagnostic abilities of the investigated heart sound features for decision tree – based diagnosis. In both the above evaluations the generalization capabilities of the implemented decision tree structures were also examined. Generalization was a very important issue due to the difficulty and the tedious work of having adequate data for all possible data acquisition methods within the training data set. The third goal of this work was to suggest a way of selecting the most appropriate decision tree structures and heart sound features in order to provide the basis of an effective semi-automated diagnostic system.

Problem definition: differentiation between AS and MR murmurs
A typical normal heart sound signal that corresponds to a heart cycle consists of four structural components:

– The first heart sound (S1, corresponding to the closure of the mitral and the tricuspid valve).

– The systolic phase.

– The second heart sound (S2, corresponding to the closure of the aortic and pulmonary valve).

– The diastolic phase.

Heart sound signals with various additional sounds are observed in patients with heart diseases. The tone of these sounds can be either like murmur or click-like. The murmurs are generated from the turbulent blood flow and are named after the phase of the heart cycle where they are best heard, e.g. systolic murmur (SM), diastolic murmur (DM), pro-systolic murmur (PSM) etc. The heart sound diagnosis problem consists in the diagnosis from heart sound signals of a) whether the heart is healthy, or not and b) if it is not healthy, which is the exact heart disease. In AS the aortic valve is thickened and narrowed. As a result, it does not fully open during cardiac contraction in systolic phase, leading to abnormally high pressure in the left ventricle and producing a systolic murmur that has relatively uniform frequency and rhomboid shape in magnitude (Figure 1a). In MR the mitral valve does not close completely during systole (due to tissue lesion) and there is blood leakage back from the left ventricle to the left atrium. MR is producing a systolic murmur that has relatively uniform frequency and magnitude slope (Figure 1b). The spectral content of the MR systolic murmur has faintly higher frequencies than the spectral content of the AS systolic murmur. The closure of aortic valve affects the second heart sound and the closure of the mitral valve affects the first heart sound.

It can be concluded by the visual insplection of Figure 1 that the AS systolic murmur has very similar characteristics with the MR systolic murmur and therefore the differentiation between these two diseases is a difficult problem in heart sound diagnosis, especially for young inexperienced clinicians. All these facts define the problem of differentiation between the AS and the MR murmurs. In the following sections we propose a method based on time-frequency features and decision tree classifiers for solving this problem.

Methods
Preprocessing of heart sound data
Cardiac auscultation and diagnosis, as already mentioned, are quite complicated, depending not only on the heart sound but also on other factors. There is also a great variability at the quality of the heart sound affected by factors related to the acquisition method. Some important factors are: the type of stethoscope used, the sensor that
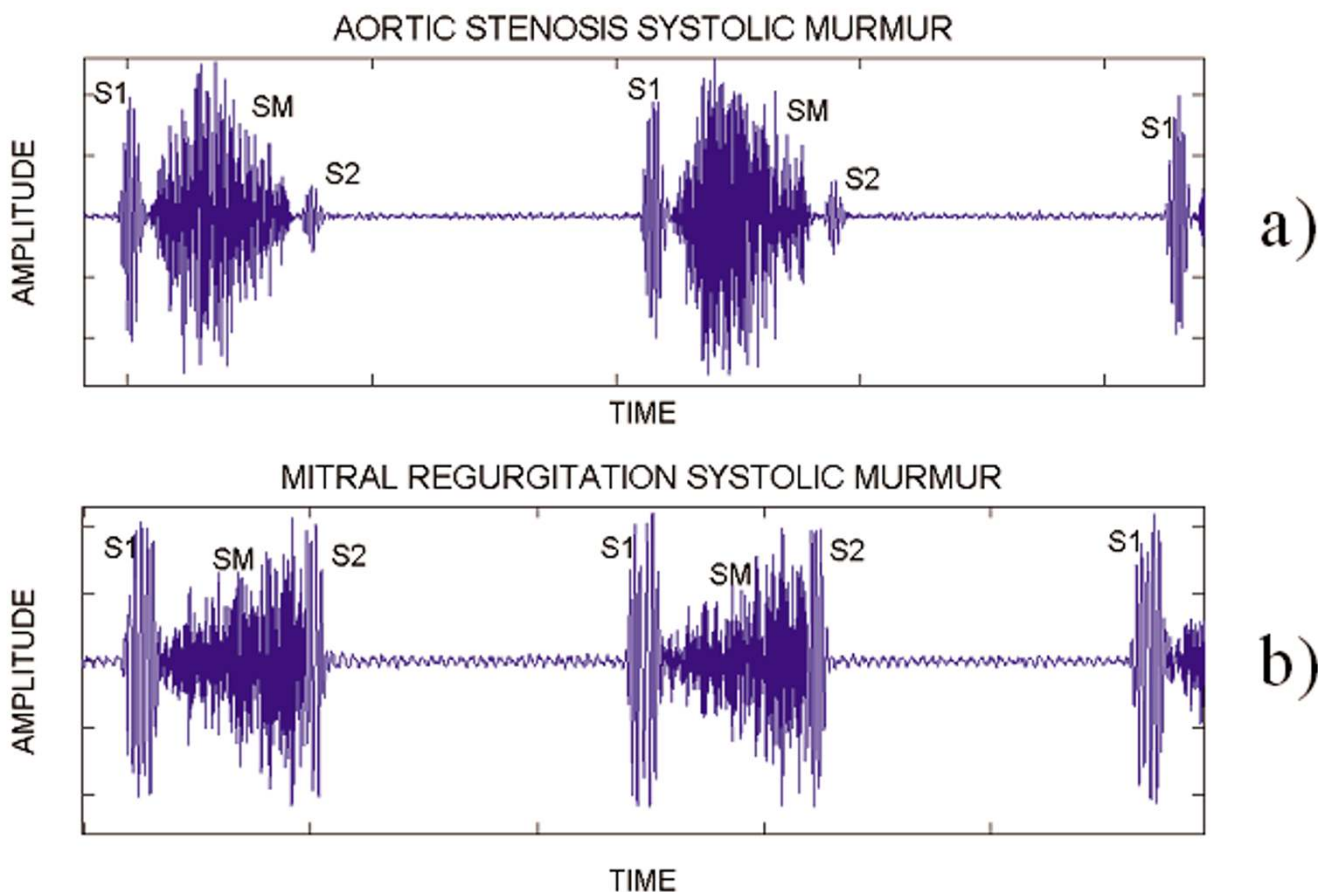
Figure 1
Two heart cycles of AS and MR heart sounds.

the stethoscope has (i.e. microphone, piezoelectric film), the mode that the stethoscope was used (i.e. bell, diaphragm, extended), the way the stethoscope was pressed on the patients skin (firmly or loosely), the medication the patient used during the auscultation (i.e. vasodilators), the patient's position (i.e. supine position, standing, squatting), the auscultation areas (i.e. apex, lower left sternal border, pulmonic area, aortic area), the phase of patients' respiration cycle (inspiration, expiration), the filters used while acquiring the heart sound (i.e. anti-tremor filter, respiratory sound reduction, noise reduction). The variation of all these parameters leads to a large number of different heart sound acquisition methods. A heart sound diagnosis algorithm should take into account the variability of the acquisition method, and also be tested in heart sound signals from different sources and recorded with different acquisition methods. For this purpose we collected heart sound signals from different heart sound sources [31-39] and created a "global" heart sound database. The heart sound signals were collected from educa-

tional audiocassettes, audio CDs and CD ROMs and all cases were already diagnosed and related to a specific heart disease. For the purposes of the present experiment 41 heart sound signals with "clear" AS systolic murmur and 43 with ' "clear" MR systolic murmur were used.

This total set of 41 + 43 = 84 heart sound signals were initially pre-processed in order to detect the cardiac cycles in every signal, i.e. detect S1 and S2, using a method based on the following steps:

a) Wavelet decomposition as described in [7] (with the only difference being that the 4th and 5th level detail was kept, i.e. frequencies from 34 to 138 Hz).

b) Calculation of the normalized average Shannon Energy [6].

c) A morphological transform action that amplifies the sharp peaks and attenuates the broad ones [5].

d) A method, similar to the one described in [6], that selects and recovers the peaks corresponding to S1 and S2 and rejects the others.

e) An algorithm that determines the boundaries of S1 and S2 in each heart cycle [30].

f) A method, similar to the one described in [8], that distinguishes S1 from S2

In a second phase, every transformed (processed in the first phase) heart sound signal was used to calculate the standard deviation of the duration of all the heart cycles it includes, the standard deviation of the S1 peak value of all its heart cycles, the standard deviation of the S2 peak value of all its heart cycles, and the heart rate. These were the first four scalar features (F1-F4) of the feature vector of the heart sound signal.

In a third phase, the rest of the features 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 for the S1, two for the systolic phase, two for the S2 and two for the diastolic phase. The first of these mean signals focused on the frequency characteristics of the heart sound; it 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 mean signal focused on the morphological time characteristics of the heart sound; it 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. Then the second S1 mean signal was divided into 8 equal parts. For each part we calculated the mean square value and this value was used as a feature in the corresponding heart sound vector. In this way we calculated 8 scalar features for S1 (F5-F12), and similarly we calculated 24 scalar features for the systolic period (F13-F36), 8 scalar features for S2 (F37-F44) and 48 scalar features for the diastolic period (F45-F92). The systolic and diastolic phase components of the above first mean signal were also passed from four band-pass filters: a) a 50–250 Hz filter giving its low frequency content, b) a 100–300 Hz filter giving its medium frequency content, c) a 150–350 Hz filter giving its medium-high frequency content and d) a 200–400 Hz filter giving its high frequency content. For each of these 8 outputs, the total energy was calculated and was used as a feature in the heart sound vector (F93-F100). With the above three phases of preprocessing every heart sound signal was transformed in a heart sound vector (pattern) with dimension 1×100.

Finally these preprocessed data feature vectors were stored in a database table. This table had 84 records, i.e. as many records as the available heart sound signals; each record describes the feature vector of a heart sound signal and has 102 fields. Each field corresponds to one feature of the feature vector or in other words to one attribute of the heart sound. One attribute for the pattern identification code named ID (used as the primary key of this database table), one attribute named *hdisease* for the characterization of the specific heart sound signal as MR or AS and 100 attributes for the above 100 heart sound features (F1-F100).

The decision tree-based method
Relevance analysis
Before constructing and using the decision tree Classifiers a Relevance Analysis [24] of the features was performed. Relevance Analysis aims to improve the classification efficiency by eliminating the less useful (for the classification) features and reducing the amount of input data to the classification stage. For example from the previous description of the database table attributes it is obvious that the pattern identification code (primary key) is irrelevant to the classification, therefore it should not be used by the classification algorithm.

In this work we used the value of the Uncertainty Coefficient [23-25] of each of the above 100 features to rank these features according to their relevance to the classifying (dependent variable) attribute which in our case is the hdisease attribute. In order to compute the Uncertainty Coefficients, the 100 numeric attributes were transformed into corresponding categorical ones. The algorithm that has been used for optimizing this transformation for the specific classification decision is described in [23]. Then for each of these 100 categorical attributes, its Uncertainty Coefficient was calculated, as described in the following paragraphs.

Initially we have a set P of p data records (84 data records in our case). The classifying attribute (dependent variable) for this differential diagnosis problem has 2 possible discrete values: AS or MR, which define 2 corresponding subsets discrete classes Pi (i = 1, 2) of the above set P. If P contains pi records for each Pi (i = 1, 2), (p = p1 + p2) then the Information Entropy (IE) of the initial set P, which is a measure of its homogeneity concerning the classifying attribute-dependent variable (higher homogeneity corresponding to lower values of IE) is given by:

$$IE(p_1, p_2) = -\sum_{i=1}^{2} \frac{p_i}{p} \log_2 \frac{p_i}{p} \qquad (1)$$

Any categorical attribute CA (from the 100 ones created via the above transformation) with possible values ca_1,

$ca_2, \dots ca_k$ can be used to partition P into subsets Ca_1, Ca_2, $\dots Ca_k$, where subset Ca_j contains all those records from P that have CA = ca_j. Let Ca_j contain $p1_j$ records of class P1 and $p2_j$ from class P2. The Information Entropy of the set P, after this partition, is equal to the weighted average of the information Entropies of the subsets Ca_1, Ca_2, $\dots Ca_k$ and is given by:

$$IE(CA) = \sum_{j=1}^{k} \frac{p_{1j} + p_{2j}}{p} I(p_{1j}, p_{2j}) \qquad (2)$$

Therefore, the Information Entropy gained by partitioning according to attribute CA, namely the improvement in homogeneity concerning the classifying attribute-dependent variable is given by:

$$Gain(CA) = IE(p_1, p_2) - IE(CA) \qquad (3)$$

The Uncertainty Coefficient U (CA) for a categorical attribute CA is obtained by normalizing this Information Gain of CA so that U (CA) ranges from 0% to 100 %.

$$U(CA)\% = \frac{IE(p_1, p_2) - IE(CA)}{IE(p_1, p_2)} * 100\% \qquad (4)$$

A low value of U(CA) near 0% means that there is no increase in homogeneity (and therefore in classification accuracy) if we partition the initial set according to CA therefore there is low dependence between the categorical attribute CA and the classifying attribute-dependent variable, while a high value near 100% means that there is strong relevance between the two attributes.

Construction of decision tree classifier structures

A decision tree is a class discrimination tree structure consisting of non-leaf nodes (internal nodes) and leaf nodes (final-without child nodes), [23,25,28]. For constructing a decision tree we use a training data set, which is a set of records, for which we know all feature-attributes (independent variables) and the classifying attribute (dependent variable). Starting from the root node, we determine the best test (= attribute + condition) for splitting the training data set, which created the most homogeneous subsets concerning the classifying attribute and therefore gives the highest classification accuracy. Each of these subsets can be further split in the same way etc. Each non-leaf node of the tree constitutes a split point, based on a test on one of the attributes, which determines how the data is partitioned (split). Such a test at a decision tree non-leaf node usually has exactly two possible outcomes (binary decision tree), which lead to two corresponding child nodes. The left child node inherits from the parent node the data that satisfies the splitting test of the parent node, while the right child node inherits the data that does not satisfy the splitting test. As the depth of the tree increases,

the size of the data in the nodes is decreasing and the probability that this data belongs to only one class increases (leading to nodes of lower impurity concerning the classifying attribute-dependent variable). If the initial data that is inherited to the root node are governed by strong classification rules, then the decision tree is going to be built in a few steps, and the depth of the tree will be small. On the other hand, if the classification rules in the initial data of the root node are weak, then the number of steps to build the classifier will be significant and the depth of the tree will be higher.

During the construction of the tree, the goal at each node is to determine the best splitting test (= attribute + condition), which best divides the training records belonging to that leaf into most homogeneous subsets concerning the classifying attribute. The value of a splitting test depends upon how well it separates the classes. For the evaluation of alternative splitting tests in a node various splitting indexes can be used. In this work the splitting index that has been used is the Entropy Index [28]. Assuming that we have a data set P which contains p records from the 2 classes and where p_j/p is the relative frequency of class j in data set P then the Entropy Index Ent(P) is given by:

$$Ent(P) = -\sum_{j=1}^{2} \frac{p_j}{p} \log_2(\frac{p_j}{p}) \qquad (5)$$

If a split divides P into two subsets P_1 and P_2 where P_1 contains p_1 examples, P_2 contains p_2 examples and $p_1 + p_2 = p$, then the values of Entropy Index of the divided data is given by:

$$Ent_{split}(P) = \frac{p_1}{p} Ent(P_1) + \frac{p_2}{p} Ent(P_2) \qquad (6)$$

In order to find the best splitting test (= attribute + condition) for a node we examine for each attribute data set all possible splitting tests based on these attributes. We finally select the splitting test with the lowest value for the above Ent_{split} to split the node.

The expansion of a decision tree can continue, dividing the training set in to subsets (corresponding to new child nodes), until we have subsets-nodes with "homogeneous" records having all the same value of the classification attribute. Although theoretically this is a possible scenario in practical situations usually leads to a decision tree structure that is over-fitted to the training data set and to the noise this data set contains. Consequently such a decision tree structure has not good generalization capabilities (= high classification accuracy for other test data sets). Stopping rules have been therefore adopted in order to prevent such an over-fitting; this approach is usually referred to as decision tree Pruning [27]. A simple stopping rule that has

been examined in this work was a restriction concerning the minimum node size. The minimum node size sets a minimum number of records required per node; in our work for the pruned decision tree structure described in the Results section a node was not split if it had fewer records than a percentage of the initial training data set records (percentages 5%, 10%, 15%, and 20% were tried).

Finally the constructed Decision tree structure is used as a classifier in order to classify new data sets. In our work we used this classifier to classify both the training data set and the test data set.

Selection of training and test pattern sets

In practical situations, where the acquisition of heart sound signals for all probable cases is time consuming and almost impossible, we are very much interested in the capability of the constructed-trained decision tree structure to generalize successfully. In order to examine the generalization capabilities of the constructed decision tree structures, the complete pattern set was divided in two subsets. The first subset included 34 patterns, 17 with AS systolic murmur and 17 with MR systolic murmur (40% of each category of the heart sound patterns set) randomly selected out of each pattern category (AS and MR) of the patterns set. This subset was used as the training set. The other subset, that included the remaining 50 patterns (24 belonging to the AS class and 26 belonging to the MR class), was used as the test set. In this way the first training-test sets scheme was developed. The division of the pattern set was repeated, keeping the same proportions (40% training set – 60% test set), but using different patterns giving a second scheme. In the same way we created some more schemes with different proportions in order to examine the impact of the training and test data set size on the performance of the decision tree classifier. These schemes are presented in Table 1.

Table 1: The training-test schemes used in this study

| Schemes	Number of records of Training dataset	Number of records of Test dataset
40% a	34	50
40% b	34	50
50% a	42	42
50% b	42	42
60% a	51	33
60% b	50	34
70% a	59	25
70% b	59	25
80% a	67	17
80% b	67	17

Results
Relevance analysis

In order to investigate the relevance and the contribution to the differentiation between AS, and MR for each of the above mentioned 100 heart sound features, the Uncertainty Coefficients were calculated for each one of them considering the hdisease field as the classifying attribute. The calculation was made separately for the training data set of each of the 10 schemes outlined in Table 1 and for each heart sound feature. Then the average value and the standard deviation of the Uncertainty Coefficient were calculated taking into account the 10 values that were calculated from these 10 schemes. The average values and the standard deviations of the Uncertainty Coefficient for all the 100 features are presented in Figure 2. Note that the most relevant features are the frequency features (i.e. E_dias_hf = High Frequency Energy in diastolic phase, E_dias_mf = Medium Frequency Energy in diastolic phase, E_sys_hf = High Frequency Energy in systolic phase, E_dias_mh = Medium High Frequency Energy in diastolic phase, etc) and the morphological features that describe the S1 (i.e. s1_1...s1_8), the S2 (i.e. s2_1...s2_8) and the systolic murmur (sys1, ... sys24). These results are compatible with our physical understanding of the problem; the AS and MR systolic murmurs have different frequency content and different envelope shape. On the contrary, in the diastolic phase there is no murmur in both diseases, therefore the diastolic phase of heart sound signals cannot contribute to the differentiation between AS and MR. Additionally the closure of Mitral valve affects the S1 and the closure of Aortic valve affects the S2.

The standard deviation values are generally smaller that 10%, showing that the Uncertainty Coefficients calculated from each scheme separately, especially the ones of the most relevant features, are similar and consistent.

Fully expanded decision tree

According to the methodology described in the Methods section, we initially construct the decision tree structure with no restriction to the nodes (without pruning). The training data-set used was from the scheme 40%a (Table 1) and had 34 records (heart sound patterns). The rest 50 patterns were used as a test set. The Decision tree was constructed based on the training data set and afterwards using this decision tree the patterns of the test data set were classified in order to investigate the generalization capabilities of the decision tree. Then the percentage of the correctly classified patterns of the test data set (according to the hdisease attribute) was calculated. The classification performance for the 40%a scheme was: 45 patterns were classified correctly (22 AS and 23 MR) while for 5 patterns the classification was wrong (2 AS and 3 MR); therefore the percentage of classification accuracy, is defined as
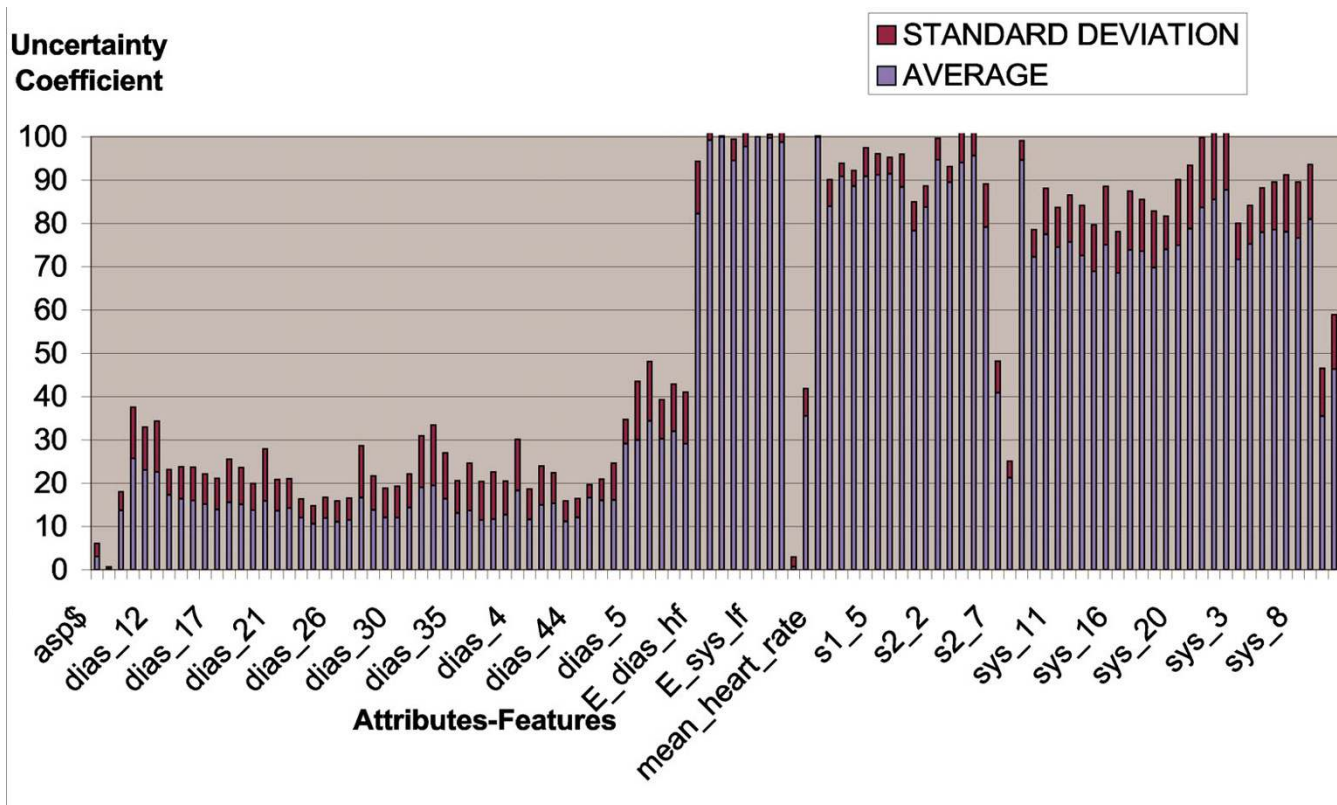
Figure 2
Average values and standard deviations of the uncertainty coefficient for the 100 features regarding the disease attribute.

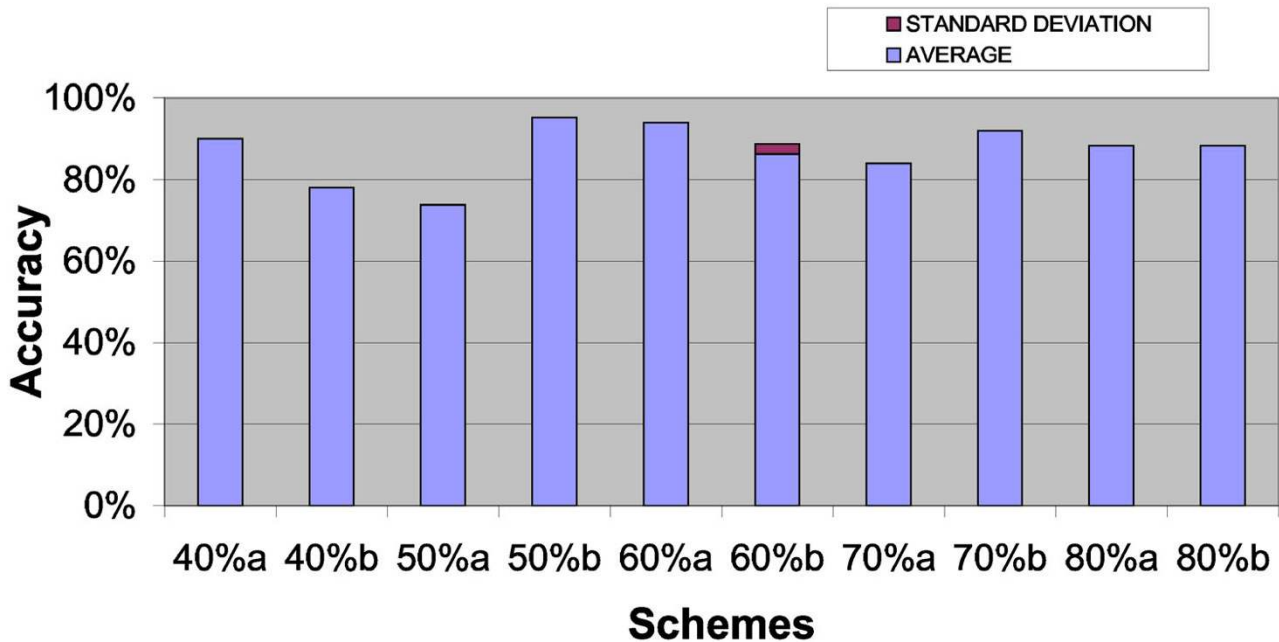$$Accuracy = \frac{correctly\ classified\ (AS\ sounds + MR\ sounds)}{tested\ (AS\ sounds + MR\ sounds)} \quad (7),$$

is 45/50*100 = 90%.

In order to exploit the results of the Relevance Analysis described earlier, we constructed five more decision trees using only the heart sound features that have Uncertainty Coefficient above a) 50%, b) 60%, c) 70%, d) 80% and e) 90%. For example in the case e) of 90%, the features that used were the following sixteen: 1) E_dias_lf, 2) E_dias_mf, 3) E_dias_mhf, 4) E_sys_hf, 5) E_sys_lf, 6) E_sys_mf, 7) E_sys_mhf, 8) s1_2, 9) s1_4, 10) s1_5, 11) s1_6, 12) s2_2, 13) s2_4, 14) s2_5, 15) sys_23, 16) sys_24. These five decision trees were used for classifying the patterns of the test data set, with results identical to the ones described above. Therefore using only 16 out of the 100 features we can get identical levels of classification accuracy with much less computational effort. For each of the data schemes of Table 1, we repeated the above calculation and finally we calculated the average classification accuracy and its standard deviation for the above five cases a) to e) (= using different number of features, according to their Uncertainty Coefficients). The results are shown in Figure 3. We notice from Figure 3 that the

classification accuracy results for cases a) to e) are identical; therefore the standard deviation is very low (zero in most of the cases) confirming the above-mentioned relevant conclusions. Also we can see that for the schemes with the largest training data sets (= 80%a and 80%b schemes, with the training data set of 67 record-patterns) we have a consistent (= the same for both schemes) level of classification accuracy about 88.3%. For other schemes of the classification accuracy e.g. for the 50%a scheme the classification accuracy is 75% while for the 50%b scheme it is 95%. This is probably due to the small size of the corresponding training sets. Therefore it is concluded that decision trees if properly trained can give a high and consistent level of classification accuracy concerning the differentiation between AS and MR, using auscultation findings. Also it is worth mentioning that the Classification Accuracy for the training data set was 100% for all the examined cases.

In order to investigate the performance of the decision tree classifier in more detail we calculated, in addition to the above Total Accuracy, also the partial accuracy for AS heart sounds, referred to as AS_Accuracy:

Figure 3
Average classification accuracy for all data schemes and cases for the Fully Expanded decision tree.

$$AS\_Accuracy = \frac{\text{correctly classified (AS sounds)}}{\text{tested (AS sounds)}} \qquad (8)$$

and the partial accuracy for MR heart sounds, referred to as MR_Accuracy:

$$MR\_Accuracy = \frac{\text{correctly classified (MR sounds)}}{\text{tested (MR sounds)}} \qquad (9).$$

The same calculations described above were repeated six times for every data scheme, using each time the features that had an Uncertainty Coefficient value above the predefined percentages we mentioned before: a) 0%, b) 50%, c) 60%, d) 70%, e) 80%, f) 90%. The final average classification accuracy (total and partial) achieved for each data scheme is shown in Figure 4.

The standard deviation of all these average values was very low, proving once more that the 16 most relevant features (out of the initial 100 features) contain all the information we need for the differentiation between AS and MR. The conclusions derived from Figure 4 concerning the partial Classification Accuracies are similar to the ones derived above concerning the total Classification Accuracy; with the exception of schemes 80%a and 80%b, in which the test data set for each disease is very small (8–9 patterns), therefore the corresponding partial Classifica-

tion Accuracies are less reliable. Also we can see that for 6 out of 10 schemes (40%a, 50%a, 50%b, 70%a, 70%b, 80%b) there are no significant difference between AS Classification Accuracy and MR Classification Accuracy, while for the remaining 4 schemes (40%b, 60%a, 60%b, 80%a) the MR Classification Accuracy is significantly higher than the AS Classification Accuracy; therefore in general the classification performance of the decision trees for the MR heart sounds is higher than it is for AS heart sounds. In order to confirm statistically this first conclusion, a t-test was performed with the null hypothesis statement H0:

Mean   (Scheme_Average_AS_Accuracy)   =   Mean (Scheme_Average_MR_Accuracy),

and the alternative hypothesis statement H1:

Mean   (Scheme_Average_AS_Accuracy)   <   Mean (Scheme_Average_MR_Accuracy).

Assuming that the error probability of rejecting H0 while H0 is truth is a = 1%, we found that the null hypothesis H0 should be rejected (with confidence interval 99%). This statistical result is consistent with the above first conclusion; therefore it confirms statistically the above first conclusion. Again the classification total Accuracy,
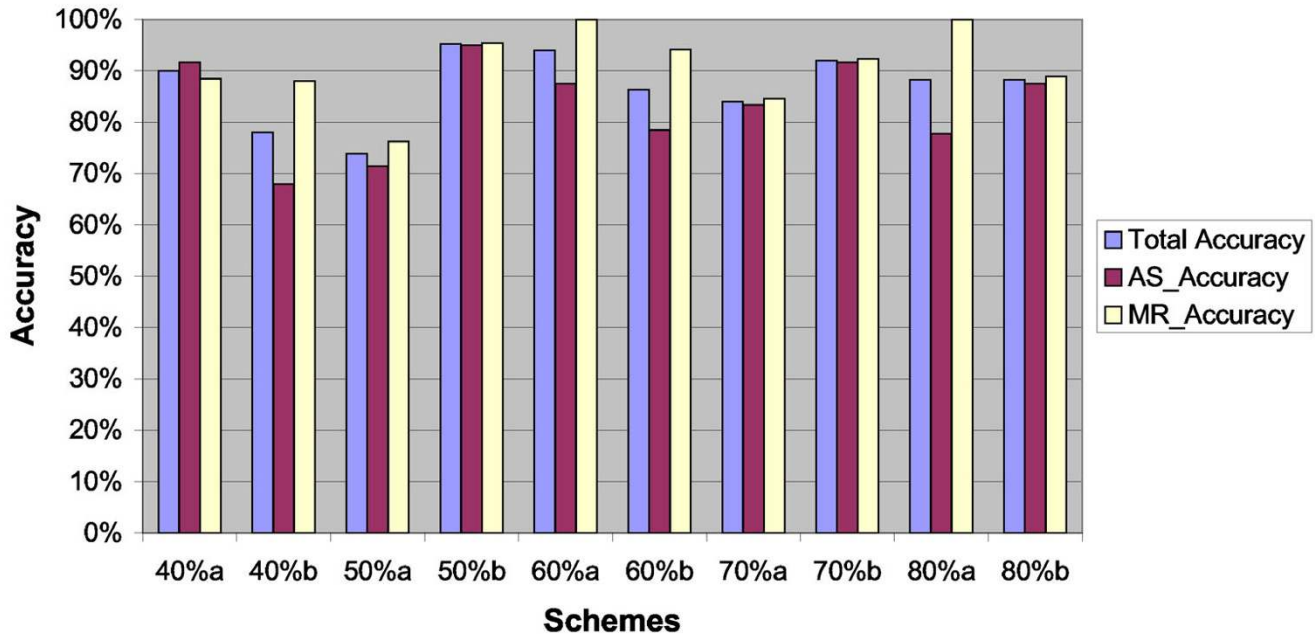
Figure 4
Average classification total accuracy, AS_accuracy, and MR_accuracy for the fully expanded decision tree classifier.

AS_Accuracy and MR_Accuracy for the training data set was 100% for all the examined cases.

Pruned decision tree
In this part of our work we tried to investigate if we could improve the generalization capabilities of the decision trees by placing a restriction during the training phase concerning the creation of new nodes of the decision tree and not allowing it to be fully expanded as in 5.2. The restriction we placed was that the number of records-patterns in a node as a percentage of the total number of records of the initial training data set should be at least a predefined percentage, usually referred to as the support of the node. The predefined percentages (minimum support levels) we examined were 5%, 10%, 15% and 20%. In particular, in the construction of a decision tree structure during the training phase for each parent node, the size (number of records-patterns) of its child nodes was examined. If the size of one of these child nodes was less than, for example, 5% of the total number of records-patterns of the initial training set i.e. the node had support less than 5%, then the parent node was not further split and was converted to a leaf node and the decision tree was not further expanded from that path. With this exception the same calculations described in paragraph 5.2 were repeated (namely with the only difference that the leaf

nodes of the decision tree had at least the predefined support). Figure 5 shows an example of a resulting decision tree following the procedure described.

Figure 6 shows the Total Classification Accuracies achieved with Pruned decision trees having minimum leaf node support at least 5%, 10%, 15% and 20%, for all the data schemes described in Table 1, compared with the previous results from the Fully expanded decision tree (0% minimum support at leaf nodes). From Figure 6 we can see that in 5 out of the 10 schemes (40%a, 40%b, 50%a, 70%a, 80%a) Pruned decision trees give exactly the same Total Classification Accuracy results with the corresponding Fully-Expanded decision trees, in 2 schemes (60%b, 80%b) pruning slightly increases the Total Classification Accuracy, while in the remaining 3 schemes (50%b, 60%a, 70%b) pruning slightly decreases the Total Classification Accuracy. Therefore it is concluded that for our specific differentiation problem pruning does not significantly change the Classification Accuracy of the decision trees. We performed four individual t-tests, to confirm statistically this conclusion. The null hypothesis statement H0X% was defined as:

Mean (Scheme_Average_Accuracy_support_0%) = Mean (Scheme_Average_Accuracy_support_X%),
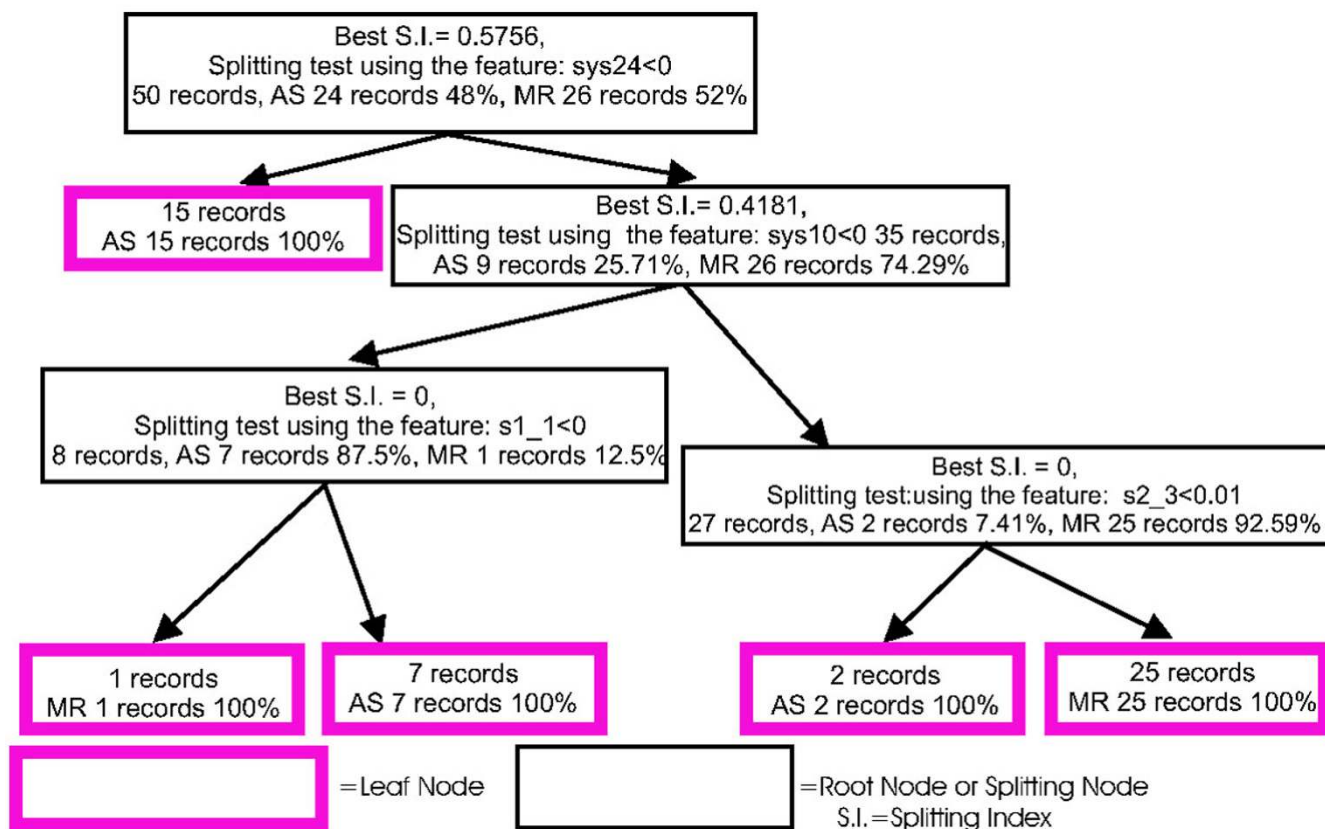
Figure 5
Example of a resulting pruned decision tree.

and the alternative hypothesis statement H1X% was defined as:

Mean (Scheme_Average_Accuracy_support_0%) > Mean (Scheme_Average_Accuracy_support_X%).

The X% stands for the minimum support levels (i.e. 5% or 10% or 15% or 20%) used individually in each one of the four t-tests. Assuming that the error probability of rejecting H0X% while H0X% is truth is a = 1%, we found that there was a 99% confidence that the null hypothesis should be accepted. This statistical result confirms once more the above conclusion for the decision tree pruning.

This conclusion is also confirmed by the corresponding partial classification Accuracies for AS and MR shown in Figure 7 and Figure 8 respectively. Finally it is worth mentioning that using Pruned decision trees the Classification

Accuracy for the training data sets was not 100% for all schemes / cases (as it was when we used Fully Expanded decision trees); for many schemes / cases it was much lower reaching the level of 88% for some of them.

Conclusions
We have investigated the applicability and the suitability of a number of decision tree structures for the AS MR differentiation problem. Criteria for this evaluation ware considered the classification Accuracy (both the total one and the partial ones) for the training set and the test set. The main conclusions from this work are the following:

– The decision trees can be used with high levels of success for the differentiation between AS and MR. The decision tree model is simple and clinicians are familiar with it, because they use a similar way when they make differential diagnosis. Therefore the decision trees can be used as
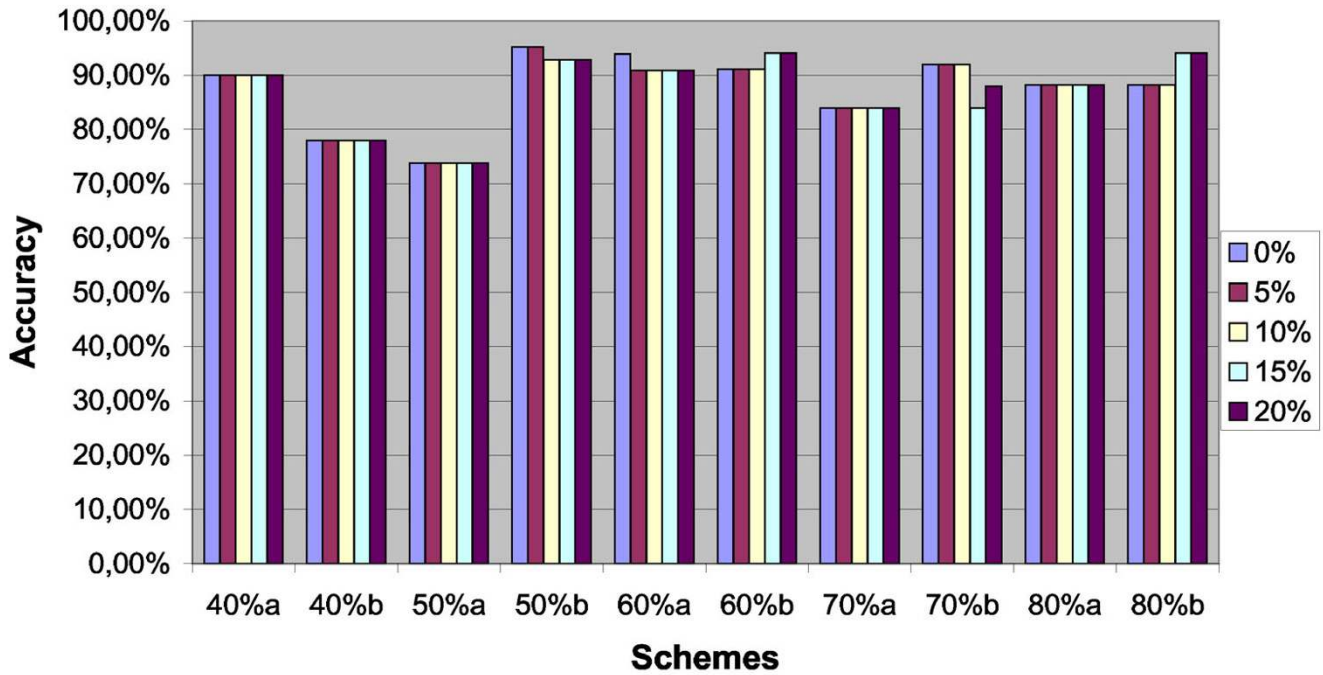
Figure 6
Total classification accuracy results for all data schemes for pruned decision trees.
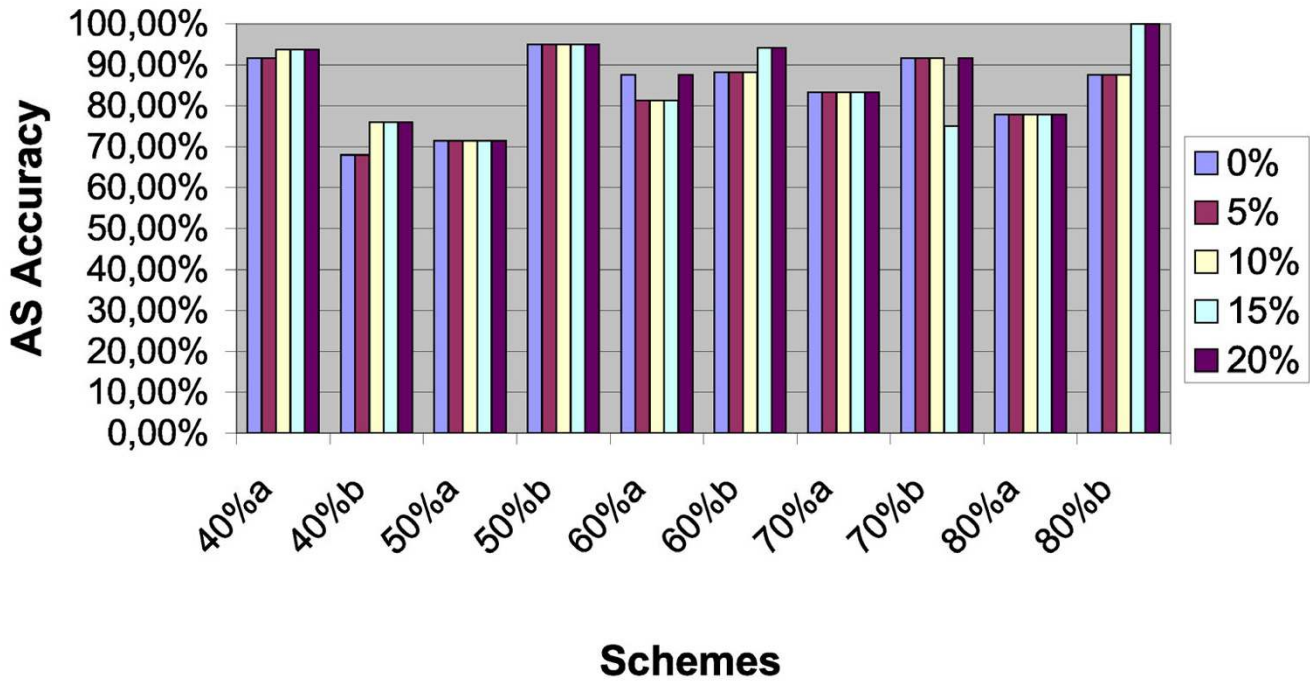
Figure 7
AS classification accuracy for all data schemes for the pruned decision trees.
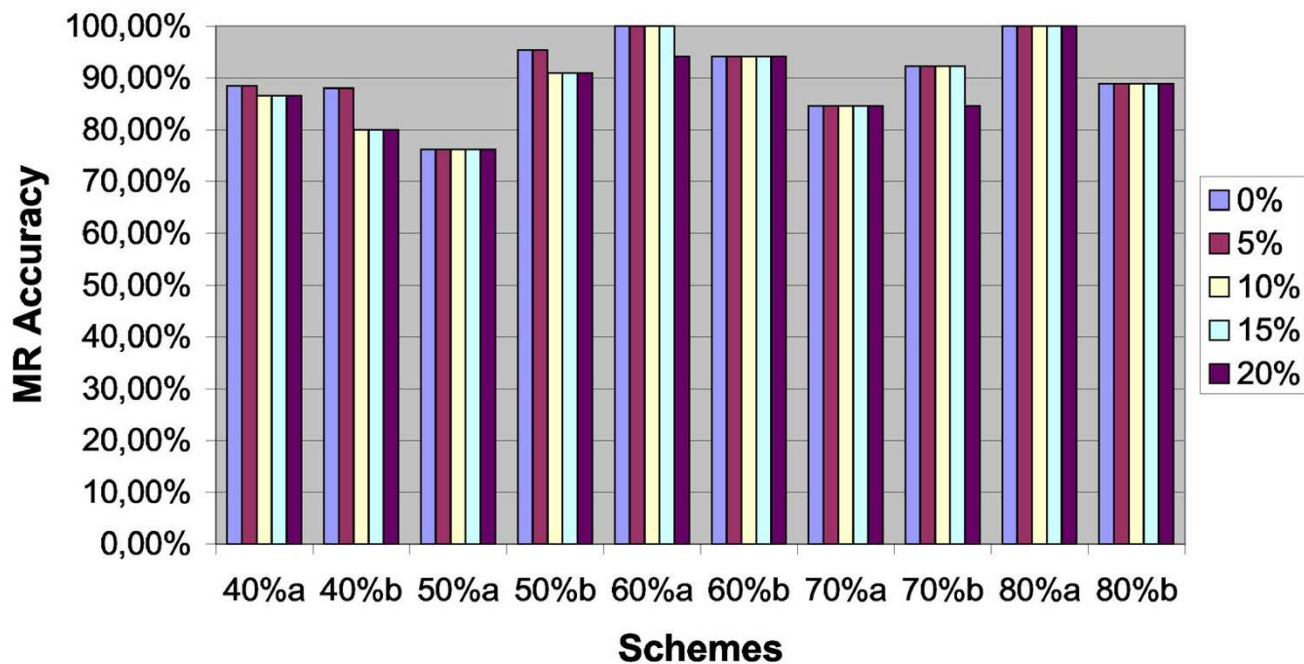
Figure 8
MR Classification for all data schemes for pruned decision trees

a basis for decision support systems, giving to the clinicians dealing with the heart sound diagnosis, especially to the ones in rural areas, in homecare, the primary healthcare etc, almost immediately an advice that helps them to make better heart sound diagnosis. This capability can reduce the costs and improve the quality of the healthcare for the cardiological problems.

– There is a small subset of the initial features that contain most of the information required for the differentiation. Using Relevance Analysis we can determine this critical subset of features, which then can be used for constructing the decision tree. Also the diagnosis for any new data set will be based only on this subset of features. In this way the required computational effort for training and using these decision trees significantly decrease.

– In the specific discrimination problem the Fully Expanded decision tree structures have similar levels of Generalization and Classification Accuracy for new data in comparison with the Pruned decision tree structures.

– Increasing the size of the training data sets (more patterns) improves the Classification Accuracy and the general reliability of the system. This is reasonable

because the decision tree is trained with more samples that cover more cases.

– The generalization capabilities of the decision tree based methods for problems similar to the one examined in this paper were found to be satisfactory. This is very important, due to the difficulty and the high cost of having enough training data for every possible case. The decision tree structures were tested on various training and test data set and the Classification Accuracy was found to be consistently high. However having more training data the decision trees can produce more trustworthy predictions.

– The general heart sound diagnosis problem can be divided into a number of simpler problems [29] such as: detection of systolic murmur, detection of diastolic murmur, determination of the type of the murmurs (crescendo, decrescendo), determination of the frequency content (low, high, medium), detection of Mid-systolic click, arrhythmia, of premature ventricular contraction, differentiation between heart diseases with similar heart sound signals, differentiation between Opening Snap, 2nd heart sound split, and 3rd heart sound [30], differentiation between the 4th heart sound, ejection clicks and the split of S1etc. All these simpler problems can be solved

by separate specialized decision support systems, which can be based on different methods, algorithms and features. The partial diagnosis given by these decision support systems can then be combined to give a total diagnosis. The combination of all these decision support systems can lead to an integrated decision support system architecture for Heart Sound Diagnosis.

Further research is required for the development of a methodology for the selection of the most appropriate decision tree structure and for improving the Classification Accuracy. Also further research is required in order to investigate the applicability and suitability of decision tree-based methods for other significant problems in the area of heart sound diagnosis.

Acknowledgements

References

1. DeGroff CG, Bhatikar S, Hertzberg J, Shandas R, Valdes-Cruz L, Mahajan RL: **Artificial neural network-based method of screening heart murmurs in children.** *Circulation* 2001, **103:**2711-2716.
2. Criley SR, Criley DG, Criley JM, Beyond Heart Sound: **An Interactive Teaching and Skills Testing Program for Cardiac Examination.** *Blaufuss Medical Multimedia, San Francisco, CA, USA, Computers in Cardiology* 2000, **27:**591-594.
3. Myint WW, Dillard B: **An Electronic Stethoscope with Diagnosis Capability.** In *Proceedings of the 33rd IEEE Southeastern Symposium on System Theory* 2001. 18–20 March 2001
4. Lukkarinen S, Noponen A-L, Sikio K, Angerla A: **A New Phonocardiographic Recording System.** *Computers in Cardiology* 1997, **24:**117-120.
5. Haghighi-Mood A, Torry JN: **A Sub-Band Energy Tracking Algorithm for Heart Sound Segmentation.** *Computers in Cardiology* 1995:501-504. 10–13 September 1995
6. Liang H, Lukkarinen S, Hartimo I: **Heart Sound Segmentation Algorithm Based on Heart Sound Envelogram.** *Computers in Cardiology* 1997:105-108. 7–10 Sept 1997
7. Liang H, Lukkarinen S, Hartimo I: **A heart sound segmentation algorithm using wavelet decomposition and reconstruction.** *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 1997, **1:**1630-1633.
8. Hebden JE, Torry JN: **Neural network and conventional classifiers to distinguish between first and second heart sounds.** *IEE Colloquium (Digest)* 1996:3/1-3/6.
9. Tovar-Corona B, Torry JN: **Time-frequency representation of systolic murmurs using wavelets.** *Computers in Cardiology* 1998:601-604. 13–16 Sep 1998
10. White PR, Collis WB, Salmon AP: **Analysing heart murmurs using time-frequency methods, Time-Frequency and Time-Scale Analysis.** *Proceedings of the IEEE-SP International Symposium* 1996:385-388. 18–21 June 1996
11. Wu Yanjun, Xu Jingping, Zhao Yan, Wang Jing, Wang Bo, Cheng Jingzhi: **Time-frequency analysis of the second heart sound signals.** In *Proceedings of IEEE 17th Annual Conference Engineering in Medicine and Biology Society* 1995, **1:**131-132. 20–23 Sept. 1995
12. Wang W, Guo Z, Yang J, Zhang Y, Durand LG, Loew M: **Analysis of the first heart sound using the matching pursuit method.** *Med Biol Eng Comput* 2001, **39(6):**644-648.
13. Nygaard H, Thuesen L, Terp K, Hasenkam JM, Paulsens PK: **Assessing the severity of aortic valve stenosis by spectral analysis of cardiac murmurs (spectral vibrocardiography). Part II: Clinical aspects.** *The Journal of Heart Valve Disease* 1993, **2(4):**468-475.
14. Liang H, Hartimo I: **A heart sound feature extraction algorithm based on wavelet decomposition and reconstruction.** In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 1998, **3:**1539-1542. 29 Oct. – 1 Nov. 1998
15. Sharif Z, Zainal MS, Sha'ameri AZ, Salleh SHS: **Analysis and classification of heart sounds and murmurs based on the instantaneous energy and frequency estimations.** *TENCON 2000. Proceedings IEEE* **2:**130-134.
16. Akay Y, Akay M, Welkowitz W, Kostis J: **Noninvasive detection of coronary artery disease.** *IEEE Engineering in Medicine and Biology* 1994, **13(5):**761-764.
17. Hebden JE, Torry JN: **Identification of Aortic Stenosis and Mitral Regurgitation by Heart Sound Analysis.** *Computers in Cardiology* 1997, **24:**109-112.
18. Leung T, White P, Collis W, Brown E, Salmon A: **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. 23–28 July 2000
19. Leung T, White P, Collis W, Brown E, Salmon A: **Analysing paediatric heart murmurs with discriminant analysis.** In *Proceedings of the 19th Annual conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong* 1998:1628-1631.
20. Durand L, Guo Z, Sabbah H, Stein P: **Comparison of spectral techniques for computer-assisted classification of spectra of heart sounds in patients with porcine bioprosthetic valves.** *Med Biol Eng Comput* 1993, **31(3):**229-36.
21. Wu CH: **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* 1997, **4(4):**343-350.
22. Haghighi-Mood A, Torry JN: **Time-frequency analysis of systolic murmurs.** *IEE Colloquium (Digest)* 1997:7/1-7/2.
23. Koundourakis G, En Visioner: **A Data Mining Framework Based On Decision Trees.** *Doctoral Thesis* University of Manchester Institute of Science and Technology; 2001.
24. Kamber M, Winstone L, Gong W, Cheng S, Han J: **Generalisation and Decision Tree Induction: Efficient Classification in Data Mining.** In *Proc. of 1997 Int'l Workshop on Research Issues on Data Engineering (RIDE'97)* Birmingham, England; 1997:11-120.
25. Stasis A: **Decision Support System for Heart Sound Diagnosis, using digital signal processing algorithms and data mining techniques.** *PhD Thesis* Athens, National Technical University of Athens; 2003.
26. Liu B, Xia Y, Yu P: **Clustering through decision tree construction.** In *Proceedings of the ACM International Conference on Information and Knowledge Management* Washington, DC; 2000.
27. Weiss SM, Indurkhya N: **Small sample decision tree pruning.** *Proceedings of the 11th International Conference on Machine Learning, Morgan Kaufmann* 1994:335-342.
28. Han J, Kamber M: *Data Mining: Concepts and Techniques* Morgan Kaufman Publisher; 2001.
29. Stasis A, Loukis E, Pavlopoulos S, Koutsouris D: **Using decision tree algorithms as a basis for a heart sound diagnosis decision support system.** *Information Technology Application in Biomedicine Conference Birmingham UK, IEEE-EMBS* 2003.
30. Stasis A, Skarpalezos D, Pavlopoulos S, Koutsouris D: **Differentiation of opening snap, second heart sound split and third heart sound, using a multiple Decision Tree Architecture.** *Computational Management Science Conference Crete, Journal of the CMS* 2003.
31. Karatzas N, Papadogianni D, Spanidou E, Klouva F, Sirigou A, Papavasiliou S: *Twelve Recorded Lessons with heart sound simulator, Merk Sharp & Dohme Hellas* Medical Publications Litsas, Athens; 1974.
32. Criley JM, Criley DG, Zalace C: *The physiological Origins of Heart Sounds and Murmurs* Harbor UCLA Medical Center, Blaufuss Medical Multimedia; 1995.
33. *Littman, 20 Examples of Cardiac & Pulmonary Auscultation* .
34. Cable C: *The Auscultation Assistant* 1997 [http://www.wilkes.med.ucla.edu/intro.html].
35. Glass L, Pennycook B: *Virtual Stethoscope* 1997 [http://sprojects.mmip.mcgill.ca/mvs/mvsteth.htm]. McGill University, Molson Medical Informatics Project
36. **VMRCVM Class of 2002, Virginia Maryland regional College of veterinary medicine.** .

37. Kocabasoglu YE, Henning RH: *Human Heart Sounds* [http://www.lf2.cuni.cz/Projekty/interna/heart_sounds/h12/index.html].
38. **Frontiers in Bioscience, Normal and Abnormal EKGs and Heart Sounds** [http://www.bioscience.org/atlases/heart/sound/sound.htm]
39. **Student Internal Medicine Society, Class 2000 Cardiology Heart Sounds Wave Files, American College of Physicians, Baylor College of Medicine.** .