# Heart Murmurs Identification Using Random Forests in Assistive Environments


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

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



## ABSTRACT

The aging population in many countries, in combination with high government deficits and financial resources limitations, necessitates new methods for the home care of the elderly at reasonable costs based on the exploitation of modern information and communication technologies (ICT). This requires the installation of assistive environments at the homes of elderly people, which include various types of sensors, generating bio-signals of other types of signals, which are transferred through networks to local health centers or hospitals in order to be monitored. However, scaling up the application of such ICT-based methods of elderly home care is going to increase tremendously the workload of the medical staff of local health centers or hospitals. Therefore it is of critical importance to develop capabilities for an automated first screening of these signals and identification of abnormal elements and diseases. In this direction the present paper proposes a system for the automatic identification of murmurs in heart sound signals, and also for the classification of them as systolic or diastolic, using a new generation of advanced Random Forests classification algorithms, which are aggregating the prediction of multiple classifiers (ensemble classification). The proposed system has been applied and validated in a representative global dataset of 198 heart sound signals, which come both from healthy medical cases and from cases having systolic and diastolic murmurs. Also, some alternative classifiers have been applied to the same data for comparison purposes. It has been concluded that the proposed systems shows a good performance, which is higher than the examined alternative classifiers.


## Categories and Subject Descriptors

J.3.3 [**Life and Medical Sciences**]: Medical Information Systems – *medical diagnosis system, classification, data mining.*



*PETRA'10, June 23 - 25, 2010, Samos, Greece.*



## General Terms

Algorithms, Measurement, Design, Experimentation, Verification


## Keywords

Assistive environments, bio-signals processing, heart sounds diagnosis, random forests.


## 1. INTRODUCTION

The aging population in many countries, in combination with high government deficits and therefore serious financial resources limitations, necessitate new methods for the home care of the elderly at reasonable costs, based on the exploitation of modern information and communication technologies (ICT). This requires the installation of assistive environments at the homes of elderly people, which include various types of sensors, generating bio-signals of other types of signals, which are transferred through networks to local health centers or hospitals in order to be examined and monitored, and based on them appropriate action to be taken if necessary (i.e. send a nurse or doctor to elderly person's home if these signals reveal a problem or disease). However, scaling up the application of such ICT-based methods of in-home elderly care is going to increase tremendously the workload of the medical staff of local health centers or hospitals. Therefore it is of critical importance to develop capabilities for an automated first screening of these signals and identification of abnormal elements and diseases; this is going to be absolutely necessary for the practical large scale application of such ICT-based methods and assistive environments for significant portions of elderly people needing care in a cost-efficient manner.

One of the basic examinations that elderly people should undergo has been for long time the heart sound auscultation, since it has high sensitivity to many important heart diseases that elderly people often suffer from. It is an operationally simple, low cost and non-invasive examination, which can be easily performed in the context of home care. The development of digital electronic stethoscopes allows at regular time intervals (defined by the responsible doctors, according to each elderly person's healthiness), or in case of a health crisis, the easy acquisition of heart sound at home (by just pressing a button by the elderly person), and then its digitization, storage, transmission to remote systems of local health centers or hospitals (e.g. using wireless technologies, the Internet, etc.), where it can be digitally processed, included in medical records and presented on screen. If this heart sound shows that there is probably a serious heart problem, then the appropriate action can be taken, such as a visit of a nurse or doctor at elderly person's home, a more sophisticated examination, such as Echocardiography or Medical Imaging (e.g. Ultrasound imaging US, Computed Tomography CT, Magnetic Resonance Imaging MRI, etc.). These examinations can provide more direct and accurate evidence of heart disease than heart auscultation, however they require highly complex and expensive equipment and specialized personnel, so they are much more costly and operationally complicated, and can be made only in well organized healthcare environments.

However, the large scale application of this 'home digital heart auscultation' approach will create a huge workload to the medical staff of the local health centers or hospitals, who will be responsible to examine these numerous heat sounds, diagnose possible problems and prescribe appropriate actions. For this reason it is of critical importance for the practical applicability of it to develop a system that can perform an automated first screening of the numerous incoming heart sound signals, identify which of them have some abnormal elements (e.g. murmurs or additional heart sounds) and draw medical staff attention on them. In this direction the present paper proposes a system for the automatic identification of murmurs in heart sound signals, and also classification of them as systolic or diastolic, using a new generation of advanced Random Forests classification algorithms, which are aggregating the prediction of multiple classifiers (ensemble classification). The proposed system has been applied and validated in a representative global dataset of 198 heart sound signals, which come both from healthy medical cases and from cases having systolic and diastolic murmurs. Also, some alternative classifiers have been applied to the same data for comparison purposes.

In the following paragraphs previous relevant research is briefly reviewed, while in section 3 our basic methodology based on Random Forests is presented. The data we used for the abovementioned first application and validation of the proposed system and the preprocessing of them are described in section 4. In section 5 the results of this application are presented, while in the final section 6 the conclusions are outlined.

## 2. PREVIOUS RESEARCH

There has been considerable previous research concerning the automated detection of various heart pathological conditions and diseases from heart sound signals. It can be broadly divided into two research streams: the first deals with the preprocessing, removal of noise and segmentation of heart sound signals, while the second deals with the detection of heart pathological conditions and diseases. The present paper belongs to the second research stream, so we are going to focus our review on it. Some of the studies of this stream are dealing with the discrimination between normal and abnormal (i.e. from subjects having a disease) heart sound signals [1-5], or with the discrimination between innocent and pathological murmurs in children [6-11]. Other studies are dealing with the detection from heart sound signals of particular heart diseases, such as coronary artery diseases [12-16] and heart valve diseases or murmurs [17-27]. With respect to the type of classification algorithms used it should be emphasized that in most of the studies of this research stream the diagnostic classification of the heart sound signals is based on neural networks of various types (e.g. back-propagation, radial basis function, self organizing map, probabilistic neural networks etc.) [1,2, 4,6,8, 10-15, 19, 22, 23]. However, there are only a few studies using other classification algorithms, such as discriminant functions [7, 21], decision trees [24, 25], Bayesian networks [3], Support Vector Machines [26] and Hidden Markov Models [27]. Therefore the diagnostic potential of other classifiers than the neural networks for the automated detection of heart pathological conditions and diseases from heart sound signals has not been sufficiently explored yet, so further research is required in this direction.

## 3. BASIC METHODOLOGY: RANDOM FORESTS

Recently, there have been significant approaches towards improving the prediction ability of classification algorithms. The idea of aggregating the prediction of multiple classifiers, known as ensemble classification, has created classification methods that have the potential to outperform single classifiers. The following example illustrates how an ensemble method can improve a classifier's performance. Suppose that one has constructed a set of twenty-five binary classifiers, each of which predicts the class with an error rate of 0.35. As mentioned above, an ensemble classifier performs the classification based on the majority vote of each base classifier. In the case that all base classifiers are identical, the error rate of the ensemble will remain 0.35, while, if the base classifiers are independent (their error is not correlated) then the ensemble will make false prediction if more than half of the base classifiers predict wrongly. From a mathematical perspective the error rate of the ensemble is expressed by the following equation:

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

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

The most representative ensemble methods are Adaboost, Bagging and Random Forests. Random Forests (RF) [28] are a conglomeration of tree base classifiers, where each tree depends on the values of a random input vector sampled independently from the total data set (with replacement) and with the same distribution for all trees in the forest. The generalization error of a forest of tree classifiers depends on the strength of the individual trees within the forest and the average correlation among trees. The main characteristics of Random Forests that favor its use as an ensemble method are [29]:

- they are relatively robust to outliers and noise,
- they provide useful internal estimates of error, strength, correlation and variable importance,
- they are simple and easily parallelized.

A Random Forest classifier Θ(x) contains of a set of trees, with each tree grown using some form of randomization, where x is an input instance. The leaf nodes of each tree are labeled by estimates of the posterior distribution over the data class labels.

Each internal node contains a test that best splits the space of data to be classified. A new, unseen instance is classified by traversing itself down in every tree and aggregating the reached leaf class labels [30]. The process is depicted in Figure 1. Each tree is grown as follows:

- If the number of cases in the training set is N, sample N cases at random but with replacement, from the original data. This sample will be the training set for growing the tree.
- If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
- Each tree is grown to the largest extent possible. Therefore, no pruning is applied.

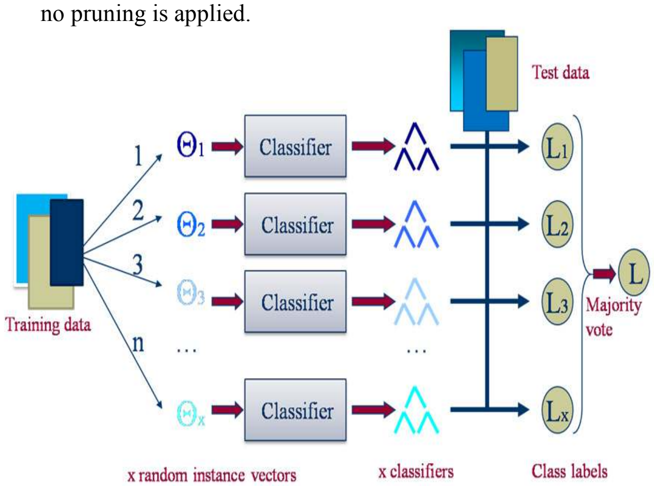


**Figure 1. Random Forests training and testing process**.

In order to formally illustrate the Random Forests classification process, suppose that the joint classifier $\Theta(x)$ contains x individual classifiers $\Theta_1(x)$, $\Theta_2(x)$,..., $\Theta_x(x)$. Let us also assume that each data instance is a pair (x, y), where x denotes the input attributes, taken from a set $_{Ai}$, i=1,...,M, and y symbolizes the set of class labels $L_j$, j=1,...,c (c is the number of class values). For reasons of simplicity, the correct class will be denoted as y, without any indices. Each discrete attribute $A_i$ takes values from a set $V_i$, i=1 to mi (mi is the number of values attribute $A_i$ has). Finally, the probability that an attribute $A_i$ has value $v_k$ is denoted by $p(v_i, k)$, the probability of a class value $y_j$ is denoted by $p(y_j)$ and the probability of an instance with attribute Ai having value $v_k$ and class label $y_j$ is symbolized by $p(y_j|v_i,k)$. We used unpruned decision trees as base classifiers and introduced two different additional methods for randomness to the trees.

Following Breiman's approach [28], we utilized the Random Input Forests and the Random Combination Forests algorithms. Nevertheless, for both methods, the evaluation metric on which tree nodes are chosen to split is the Gini index, taken from the CART algorithm. The formula of the Gini index is as follows:

$$Gini(A_i) = -\sum_{i=1}^{c} p(y_i)^2 - \sum_{j=1}^{m_i} p(v_{i,j}) - \sum_{j=1}^{c} p(y_i / v_{i,j})^2$$

In order to address the significant issue of uncorrelated features, at each step of the feature subset selection, a Bayesian network structure learning algorithm (K2 [31]) is performed to verify the correlation between features. In the case where features are not connected via a link in the Bayesian network, the resulting tree learned from such features in not fully grown, in order to reduce its probability of affecting the classification procedure in the same degree as a tree which has been learned from a correlated feature set.

## 4. DATA & PREPROCESSING

In order to investigate the usefulness and performance of Random Forests for the automatic identification of murmurs in heart sound signals, a global and representative heart sounds dataset has been created with heart sound signals from various different heart sound sources (educational audiocassettes, audio CDs, CD ROMs, files of existing heart sound databases, etc.), which are described in [24-26]. It includes heart sound signals acquired with various types of stethoscopes, sensors and filters, in various modes, subjects' positions and auscultation areas, from subjects of various ages, heart conditions and medical treatments. This dataset is more 'noisy' and therefore more 'difficult' for the classifiers, than the more 'homogeneous' ones used by most similar studies, but it enables a more realistic investigation of classifiers' performance in conditions better approximating the 'real-life' medical practice. For the purposes of the present study from this dataset we used 198 heart sound signals: 38 normal heart sounds, 41 heart sound signals with AS systolic murmur, 43 ones with MR systolic murmur, 38 ones with a AR diastolic murmur and 38 signals with a MS diastolic murmur. Each of these heart sound signals had been diagnosed by a specialized cardiologist and classified to one of the above four basic heart valve diseases.

Initially a pre-processing of these heart sounds was performed, in order to remove noise and extract features from them, which consisted of three phases. In the first phase the segmentation of the heart sound signal is performed, i.e. the cardiac cycles in every signal are detected by locating the S1 and S2 peaks. In the second phase, for each of the heart sounds produced in the first phase were calculated the standard deviation of the duration of all the heart cycles it includes, the standard deviation of the S1 peak values of all heart cycles, the standard deviation of the S2 peak values of all heart cycles and the average heart rate; these values are the first four scalar features (F1-F4) of the feature vector of each heart sound signal. In the third phase, the rest of the features used for classification are extracted. For this purpose we calculated for each transformed heart sound signal two mean signals for each of the four structural components of the heart cycle, namely two signals for the S1, two for the systolic phase, two for the S2 and two for the diastolic phase. The first signal focused on the frequency characteristics and was calculated as the mean value of each component, after segmenting and extracting the heart cycle components, time warping them and aligning them. The second signal focused on the morphological time characteristics and was calculated as the mean value of the normalized average Shannon Energy Envelope of each component, after segmenting and extracting the heart cycles components, time warping them and aligning them. The second S1 mean signal is then divided into 8 equal parts, for each part the mean square value is calculated and the resulting 8 values are used as features (F5-F12). Similarly 24 scalar features for the systolic period (F13-F36), 8 scalar features for S2 (F37-F44) and 48 scalar features for the diastolic period (F45-F92) were calculated. Finally the systolic and diastolic phase components of the first mean signal were passed from four band-pass filters: a) a 50–250 Hz filter giving its low frequency content, b) a 100–300 Hz filter giving its medium frequency content, c) a 150–350 Hz filter giving its medium-high frequency content and d) a 200–400 Hz filter giving its high frequency content. For each of these 8 outputs, the total energy was calculated and was used as a feature in the heart sound vector (F93-F100). The above pre-processing produced for each heart sound signal a feature vector consisting of 100 components. These feature vectors of our heart sound signals were used for the identification of murmurs using Random Forests described in the next section.

## 5. RESULTS

Two different stages of classifications were carried out. In the former stage, we classified the heart sound signals into either healthy (without murmurs) or unhealthy (with murmurs) class labels, while, in the latter case, we have separated the unhealthy dataset into systolic and diastolic murmurs and performed classification upon each. As for evaluation metrics, the standard parameters of precision and recall, borrowed from the domain of Information Retrieval were chosen. As regards to the proposed methodology, three Random Forest variations, as explained in Section 3, have been utilized, namely Random Input Forests, Random Combination Forests and Random Forests using Bayesian Networks for feature correlation evaluations. Amongst the three variations, the first two were proposed by Breiman [28], while the third approach is novel. In the Random Input Forests variation, a maximum number of 10 features was taken into consideration, while in the Random Combination Forests approach 7 features were considered. Finally, using the K2 Bayesian network learning algorithm, for feature correlations, 14 features were found to be the most effective choice. Several Machine Learning algorithms that have been reported to behave robustly for the task at hand have been compared to the proposed implementations. More specifically, a Naïve Bayesian classifier, which incorporated a Gaussian distribution for tackling with numerical features, along with Radial Basis Functions Neural Networks (using a Gaussian kernel), together with Support Vector Machines (with polynomial kernel), K-Nearest Neighbor and the C4.5 decision trees algorithm formed our comparison framework. As Figure 2 illustrates, the classification performance regarding the prediction of a healthy or unhealthy heart signal reaches a peak precision and recall of 84.4% and 84.3% when utilizing the proposed Bayesian Network variation for dealing with uncorrelated features. Furthermore, all Random Forests implementation surpass the performance of all other implementations by a varying factor of 1.7% to 28.7% for precision and 2.3% to 28.6% for recall. The above precision and recall levels are satisfactory, taking into account that our dataset is highly 'noisy' and therefore more 'difficult' for classifiers, since as described in section 4 it includes heart sound signals acquired with various types of stethoscopes, sensors and filters, in various modes, subjects' positions and auscultation areas, from subjects of various ages, heart conditions and medical treatments. Therefore, if some of these factors can be controlled in the context of a particular elderly home care project (e.g. using a standard stethoscope, sensor and filters, in a predefined common mode, subjects' position and auscultation area) then precision and recall levels will be higher. Besides, as mentioned in Section 3, another advantage of Random Forests is that they could be straightforwardly parallelized and achieve computational costs comparable to that of the Naïve Bayesian classifier.

As regards to the latter stage, in which the type of heart murmur is detected as either systolic or diastolic, the same Random Forest implementations were confronted against the aforementioned alternative classifiers, in an attempt to retain homogeneity to our evaluation approach. Similarly, depicted by Figure 3, the performance of all Random Forest implementations shows a very good performance and outperforms all other approaches, with the Bayesian Network variation producing the best results for the domain of heart murmur identification. Regarding the other alternative approaches examined, C4.5 behaves more robustly than the other classifiers, while Radial Basis Function Neural Networks are the weakest of them.

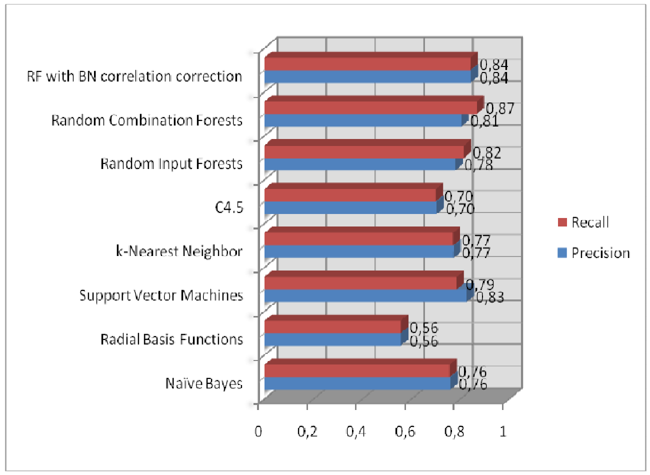


**Figure 2. Classification recall and precision for discriminating between healthy and unhealthy heart signals using three Random Forest implementations against alternative classifiers.**

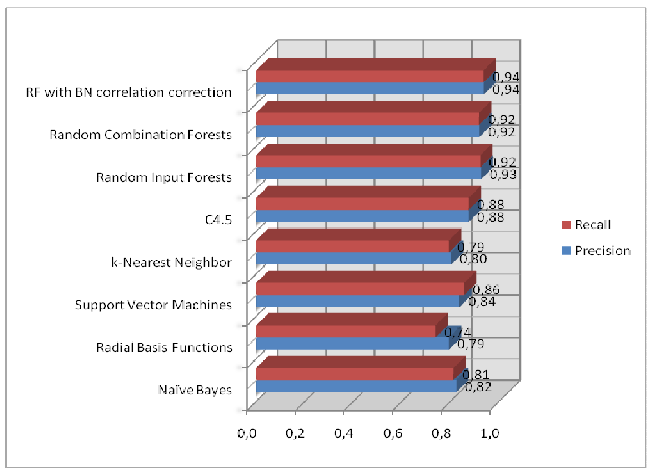


**Figure 3. Classification recall and precision for discriminating between heart sound signals with systolic and diastolic murmurs using three Random Forest implementations against alternative classifiers.**

## 6. CONCLUSIONS

For the practical large scale application of ICT-based method and assistive environments for elderly people home care in a cost-efficient manner it is of critical importance to develop capabilities for an automated first screening of signals generated in subjects' home and then transmitted to the responsible local health centers or hospitals, and identification of abnormal elements and diseases. The present paper contributes to addressing this need. It proposes a system for the automatic identification of murmurs in heart sound signals digitally acquired in home care context, and also for the classification of them as systolic or diastolic, using a new generation of advanced Random Forests classification algorithms, which are aggregating the prediction of multiple classifiers (ensemble classification). Three Random Forest variations have been investigated, namely Random Input Forests, Random Combination Forests and Random Forests using Bayesian Networks for feature correlation evaluations. They have been tested using a highly 'noisy' dataset - and therefore quite 'difficult' for classifiers – including 198 heart sound signals acquired with various types of stethoscopes, sensors and filters, in various modes, subjects' positions and auscultation areas, from subjects of various ages, heart conditions and medical treatments. It has been concluded that the proposed systems shows a good performance in discriminating between healthy and unhealthy heart signals, and a very good performance in discriminating between unhealthy heart sound signals with systolic and diastolic murmurs, which are both higher than the examined alternative classifiers.

## 7. REFERENCES


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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[19] Brusco M., Nazeran H.: Development of an Intelligent PDA-based Wearable Digital Phonocardiograph. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference 2005, 4, pp. 3506-9.

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

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

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

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

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

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

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

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

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

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

[30] Amit, Y., Geman D.: Shape quantization and recognition with randomized trees. Neural Computation, (9), 1997, pp. 1545-1588.

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