Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 75–79
ISBN 978-83-60810-51-4

The importance of handling multivariate attributes in the identification of heart valve diseases using heart signals

Ahmed Hamdy, Hesham Hefny
Department of Computer Sciences and Information,
Institute of Statistical Studies and Researches,
Cairo University, Egypt
Email: ahmed.h.elsayed@gmail.com, hehefny@hotmail.com

Mostafa A. Salama
Department of Computer Science,
British University in Egypt,
Cairo, Egypt
Email: mostafa.salama@gmail.com

Aboul Ella Hassanien
Faculty of Computers and Information,
Cairo University,
Cairo, Egypt
Email: aboitcairo@gmail.com

Tai-hoon Kim
Hannam University,
Korea
Email: taihoonn@hannam.ac.kr

***Abstract*—Automated detection of heart valve disease through heart sound has a great requirement due to its inexpensive and non-invasive availability. Extensive research has been conducted recently on applying different classification and features selection techniques. Heart sound data sets represent a real life data that contains continuous attributes and a large number of features that could be hardly classified by most of classification techniques. Data mining techniques including the feature evaluation and classification techniques that ignore the important characteristics that may exist in the heart sound data set may not be applicable on this case. In this context, the present paper initially surveys the research that has been conducted concerning the exploitation of heart sound signals for automated detection of heart conditions. Then, A comparative study is applied to determine the most effective data mining techniques that are capable for the detection of heart valve disease with a high accuracy. The results shows that the techniques that are capable of the handling the multivariate data sets that has continuous nature show the highest classification accuracy.**

I. Introduction

THE DIAGNOSIS of diseases like heart valve diseases using data mining tools is an essential requirement in daily life. Most of heart valve diseases have an effect on the heart sound of patients [1]. Classification can be applied to detect whether the patient's heart sound signal is patient or not, and also can detect the type of the heart disease in sick patients [2]. Such an approach could be useful in the diagnosis of heart disease remotely, just by sending a record of the heart signals to a medical back end system that replies automatically by the problem in heart. Also it is considered as a low cost approach rather than the high cost medical examinations. Also a computerized system could provide physicians with suggestions about the diagnostic suggestions about the diseases. Due to the sensitivity of heart diagnosis results, a high classification accuracy and performance are required with the least error percentage. After extracting features from heart sound signals, preprocessing is applied on these features [3].

This paper proves that that all the terminals of the data mining process should put into consideration the characteristics of the internal structure of the input data. The most important preprocessing step in data mining is the feature reduction of the input data set. The data set contain features that are considered as noisy or irrelevant features, these features could have a negative impact on the classification accuracy of the instances, patients. Feature reduction methods are either feature extraction or feature selection method. Feature extraction method applies operation on the original features and extracts a lower number of features that carries the same characteristics. Feature selection methods rank and select the most important features, where if only a subset of features with the highest rank are used in classification, high classification accuracy could be achieved.

The extracted heart sound data are three different data sets, each of 100 features where they are sliced into six different parts. The first data set is required to classify whether the heart of the patients are normal or not. The second and third data set is required for the detection of the heart valve disease. The heart valve diseases under investigation in this paper are the aortic stenosis AS, the aortic regurgitation AR, the mitral stenosis MS and the mitral regurgitation MR. This disease classification is performed in two steps where the first step is applied on the second data set for determining the type of the systolic murmur which means AS or MR, and the second step is applied on the third data set of a diastolic murmur diseases which means AR or MS. The second importance of feature selection method is to determine which stage of the heart sound could have the greatest indication to heart valve disease in the case of each murmur type. The four stages of a heart sound are the first heart signal S1, the systolic period, the second heart signal and the diastolic period [4].

The rest of this paper is organized as follows: Section II gives an overview of the feature reduction and classification techniques. Section III describes the experimental results and analysis, while the conclusion is presented in Section IV.

II. Feature selection and classification techniques

Machine learning approach for the analysis of the heart valve diseases contains three phases, the feature selection and evaluation phase, the classification phase, and finally the analysis phase. These three phases are described in detail in this section along with the steps involved and the characteristics feature for each phase.

A. Feature evaluation and selection phase

Data sets may contain irrelevant or redundant features that are considered as misleading to the classification technique applied. These features could lead to what is known as curse of dimensionality problem [5]. The application of feature selection techniques greatly reduces the computational cost and increases the classification accuracy of classifying high dimensional data. The selection of the most important features to the classification problem could be based on two phases, feature evaluation or ranking and feature selection. There are a lot of possible combinations between each feature search and each feature evaluation algorithms [6]. Feature evaluation technique involves the evaluation of each feature according to the target class labels, while feature selection techniques perform the evaluation of subset of feature explicitly via a predictive model, classifier, built from just those features. Feature selection is grouped in two ways according to the attribute evaluation measure: depending on the type (filter or wrapper techniques) or on the way that features are evaluates (individual or subset evaluation) [7], [8]. The used feature selection in this context is the sequential floating selection, filter model, as it leads to the highest classification accuracy. The first requirement in this study to find the feature evaluation techniques that leads to the highest classification accuracy in the current heart sound data sets. The following set of feature evaluation techniques are used here in the experimental work:

- **Consistency subset evaluation** Evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the set.
- **ChiSquaredAttributeEval** Evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class attribute.
- **FilteredSubsetEval** Class for running an arbitrary subset evaluator on data that has been passed through an arbitrary filter
- **InfoGainAttributeEval** Evaluates the worth of an attribute by measuring the information gain with respect to the class
- **GainRatioAttributeEval** Evaluates the worth of an attribute by measuring the gain ratio with respect to the class
- **SVMAttributeEval** Evaluates the worth of an attribute by using an SVM classifier. Attributes are ranked by the square of the weight assigned by the SVM. Attribute selection for multiclass problems is handled by ranking attributes for each class separately using a one-vs-all method and then "dealing" from the top of each pile to give a final ranking [9].
- **FilteredAttributeEval** Class for running an arbitrary attribute evaluator on data that has been passed through an arbitrary filter.

B. Classification phase

This phase includes the training and testing of the classifiers. Every classification technique has its own strong and weak points. Generally, SVMs and neural networks tend to perform much better when dealing with multi dimensions and continuous features [10]. On the other hand, logic-based systems like decision trees tend to perform better when dealing with discrete/categorical features. For neural network models and SVMs, a large sample size is required in order to achieve its maximum prediction accuracy whereas NB may need a relatively small data set. Another problem appears in univariate models like Bayes belief models that assumes features are independent [11] where it will be computationally intractable unless an independence assumption (often not true) among features is imposed [12]. The main judge of which machine learning technique to select depends on the nature of the input data set, as selecting inappropriate algorithm may lead to either high processing cost or low classification accuracy. The classification techniques used in this study is as follows:

- **Voted perceptron** The key point of the voted version is that, while training, it stores information in order to make more robust predictions on test data. While training, it stores information in order to make more robust predictions on test data by set weight for the prediction vector which survives from number of iteration applied on it. All prediction vectors combined by a weighted majority vote [13].
- **IB1** Use a simple distance measure to find the training instance of closest data set to the given test instance, and predicts the same class as this training instance. If multiple instances are the same (smallest) distance to the test instance, the first one found is used [14].
- **KStar** it is instance based classifier, that is the class of a test instance is based upon the class of those training instances similar to it [15].
- **LWL** Locally-weighted learning. Uses an instance-based algorithm to assign instance weights which are then used by a specified WeightedInstancesHandler. A good choice for classification is NaiveBayes. LinearRegression Is suitable for regression problems [16].
- **IBk** is one from Lazy methods as IB1, such that K-nearest neighbours classifier [14].

- **Decision Table** Class for building and using a simple decision table majority classifier. It employs the wrapper method to find a good subset of attributes for inclusion in the table. By eliminating attributes that contribute little or nothing to a model of the dataset, the algorithm reduces the likelihood of over-fitting and creates a smaller and condensed decision table [17].
- **Conjunctive Rule Learner** Induce a set of rules from data that captures all generalizable knowledge within that data, and at the same time being as small as possible [18]
- **LADTree** Class for generating a multi-class alternating decision tree using the LogitBoost strategy [19].
- **ZeroR** Predicts the majority class in the training data. It is for building and using a 0-R classifier, it predicts the mean (for a numeric class) or the mode (for a nominal class) [20].
- **NBTree** A Naive Bayes/Decision-Tree. It is class for generating a decision tree with naive Bayes classifiers at the leaves [21].
- **Multilayer Perceptron (MLP)** is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate output. MLP utilizes a supervised learning technique called backpropagation for training the data set as nodes in a network [22].
- **SMO** Implements sequential minimal optimization algorithm for training a support vector classifier [23].

III. Experimental works and discussions

A. the heart sound signals data set: characteristics and declaration

A lot of researches have been applied on heart sound for the detection of heart valve disease. Features are extracted from the heart sound signal into a data set that is composed of 100 features. Then, a classification algorithm is applied on such data set for detection of heart valve disease. Features are extracted in three phases, segmentation [24] [25], transformation and extraction. These extracted feature represent the four stages of a heart signal which are $S1$ signal, systolic period, $S2$ signal and diastolic period as shown in figure 1. These features are divided into six groups as follows:

- F1:F4 are the standard deviation of all heart sounds, $S1$, $S2$ and average heart rate.
- F5:F12 represents signal $S1$.
- F13:F36 represents the systolic period.
- F37:F44 represents signal $S2$.
- F45:F92 represents the diastolic period.
- F93:F100 the four stage of a heart signal are passed from four band-pass frequency filters. The energy of each output is calculated to form these last 8 features.

The identification of heart valve diseases proposed system were applied on three different data sets of heart sound signals with the same number of instances in every class. The first data set "HS_AS_MR" is about systolic diseases where it contains 37 instances of aortic stenosis AS cases and 37 instances of mitral regurgitation MR cases. The second data set "HS_AR_MS" is about diastolic diseases where it contains 37 instances of aortic regurgitation AR cases and 37 instances of mitral stenosis MS cases. The third data set "HS_N_S" contains 64 instance, where 32 instances represent healthy patients and the other 32 represents unhealthy, murmur diseased patients.

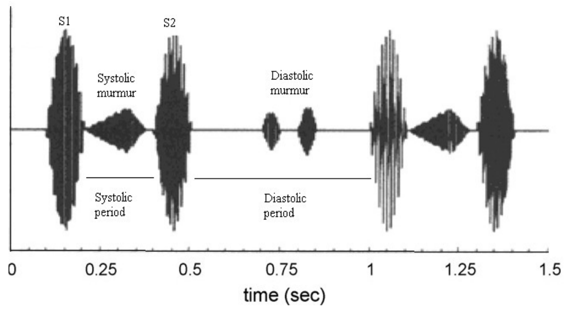

Fig. 1. Heart signal: systolic period and diastolic period [4]

B. Analysis, results and discussion

The comparative analysis will be applied on the three data sets of heart sound. It will include a all the combinations of the seven feature selection techniques and the twelve classification techniques. Figures 2, 3 and 4 shows the feature selection and classification results of each the tree data sets. In the healthy sick data set "HS_N_S", the best classification results appears when the Consistency Subset Evaluation feature selection technique is applied followed by the NBTree classification technique. In the AR MS data set "HS_AR_MS", the best classification results appears when the Consistency Subset Evaluation feature selection technique is applied followed by the KStar classification technique. In the healthy sick data set "HS_N_S", the best classification results appears when the ChiSquare, Gain Ration and Information Gain Evaluation feature selection technique is applied followed by the KStar classification technique.

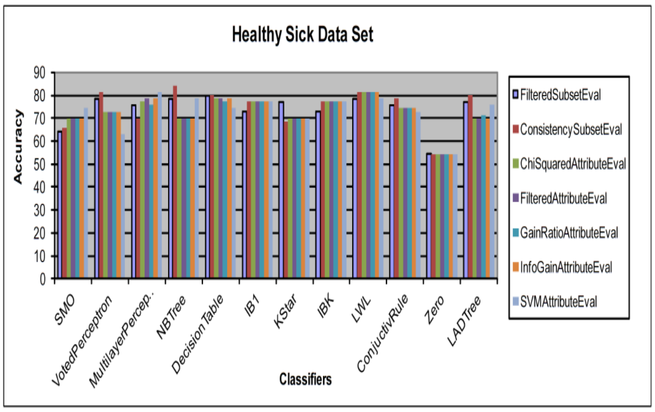

Fig. 2. Feature selection and classification results of the Healthy Sick Data Set

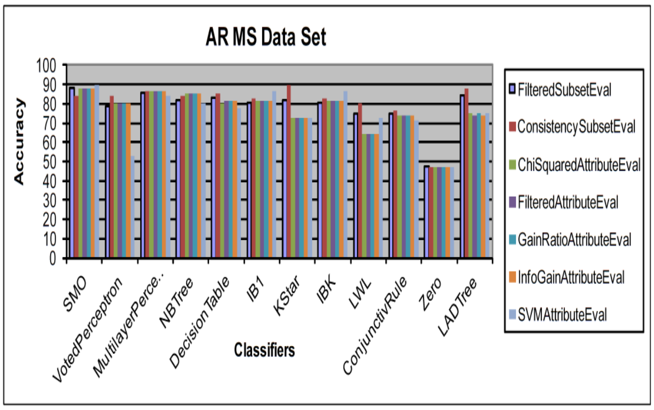

Fig. 3. Feature selection and classification results of the AR MS Data Set

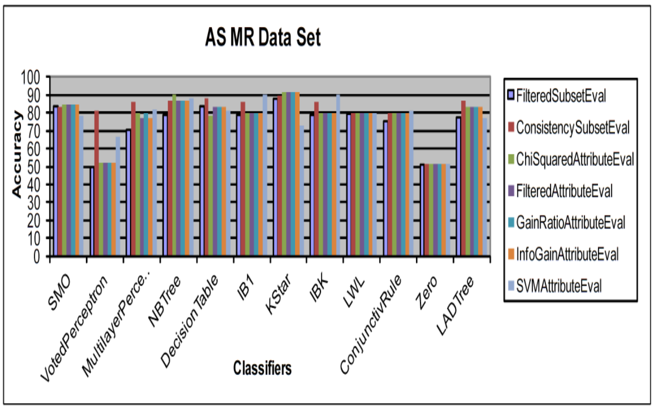

Fig. 4. Feature selection and classification results of the AS MR Data Set

The results shows that for Healthy-sick data set, the NBTree classifier shows the highest classification accuracy after applying Consistency subset evaluation technique (84%). while in the AR-MS and AS-MR data set, the KStar classifier shows the highest classification accuracy after applying Consistency subset evaluation (90%) and the entropy based feature evaluation techniques (91 %) respectively.

The hybrid combination between consistency subset evaluation technique and the classification techniques, NBTree and KStar shows the highest classification accuracy. Unlike the commonly used univariate measures (e.g., distance, information, and dependence measures), consistency subset evaluation is a multivariate measure which checks a subset of features at a time [26]. Where it evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes [28]. Consistency of any subset can never be lower than that of the full set of attributes; hence the usual practice is to use this subset evaluator in conjunction with a Random or Exhaustive search. As it looks for the smallest subset with consistency equal to that of the full set of attributes, as shown in the following equation 1:

$$Consistency_s = 1 - \frac{\sum_{i=0}^{j} \mid D_i \mid - \mid M_i \mid}{N} \qquad (1)$$

Where:

- s is an attribute subset,
- J is the number of distinct combinations of attribute values for s,
- $\mid D_i \mid$ is the number of occurrences of the ith attribute value combination, i
- $\mid M_i \mid$ is the cardinality of the majority class for the ith attribute value combination,
- N is the total number of instances in the data set.

On the other hand, NBTree classifier is hybrid technique that eliminate the assumption of independence among features [27]. On the other hand, One of the advantages of the KStar approach is that both real attributes and symbolic attributes can be dealt with together within the same framework [15]. Also, the independence assumption among attributes does not take place in the KStar learning technique. These two combinations, between the consistency-subset feature evaluation and either of the two classifiers NBTree and KStar, proves that the relativity among features is an important characteristic. And any feature evaluation technique and learner used in the data mining process should be handled this important characteristic. Accordingly, the average of the covariance between attributes has been calculated for the three data sets, the results shows that covariance average of the first data set is 0.11, while the next two data sets are 0.08 and 0.07. This explains the reason of NBTree to show the classification accuracy higher than KStart in the first data set, while the inverse appears in the next two data sets.

IV. CONCLUSIONS AND FUTURE WORKS

The selection of the most applicable feature selection technique and classification has a great effect in the enhancement of the classification accuracy results. Every technique has its own methodology in dealing with the input data set, and according to the data, this technique could be effective or not. The results in this experiment show that the effectiveness of the feature selection and the classification techniques is also dependent on the problem under consideration. Three data sets from the same source, which is the heart sound for heart valve disease detection, is tested by different feature selection and classification technique. The extensive experimental results on the heat sound signals data set demonstrate that each data set shows the highest classification accuracy by feature selection and classification technique different from the other data sets.

REFERENCES

[1] T. Chen, K. Kuan, L. Celi and G. Clifford, “Intelligent heart sound diagnostics on a cellphone using a hands-free kit", Proceedings of AAAI Artificial Intelligence for Development (AI-D’10). Stanford University, California, (2010).

[2] J.E. Hebden and J.N. Torry, “Neural network and conventional classifiers to distinguish between first and second heart sounds", Artificial Intelligence Methods for Biomedical Data Processing IEE Colloquium (Digest), vol 3, pp. 1–6, (1996).

[3] D. Kumar, P. Carvalho, M. Antunes, R.P. Paiva and J. Henriques, “Heart murmur classification with feature selection", In proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, vol. 1, pp. 4566–4569, (2010).

[4] Maglogiannis I, Loukis E, Zafiropoulos E and Stasis A., “Support vectors machine-based identification of heart valve diseases using heart sounds", Journal of Computer Methods and Programs in Biomedicine, Elsevier, North-Holland, vol. 95, pp. 47–61, (2009).

[5] C. Shang and Q. Shen, “Aiding classification of gene expression data with feature selection: a comparative study", Computational Intelligence Research, vol. 1, pp. 68–76, (2006).

[6] Liu, H., Yu, L., “Toward integrating feature selection algorithms for classification and clustering", IEEE Trans. on Knowledge and Data Engineering, vol. 17, pp. 1–12, (2005).

[7] J. Doak, “An Evaluation of Feature Selection Methods and Their Application to Computer Security", University of California at Davis, Tech. Rep. CSE-92-18, (1992).

[8] J. Cabestany, A. Prieto, and D.F. Sandoval, “Heuristic Search over a Ranking for Feature Selection", LNCS, vol. 3512, pp. 742–749, (2005).

[9] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, “Gene selection for cancer classification using support vector machines", Machine Learning, vol. 46, pp. 389–422, (2002).

[10] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques", Informatica, vol. 31, pp. 249-268, (2007).

[11] Carmen Lai, Marcel J. T. Reinders and Lodewyk F. A. Wessels, “Random subspace method for multivariate feature selection", Informatica, Pattern Recognition Letters, vol. 10, pp.1067–1076, (2006).

[12] Pierre Geurts, “Pattern extraction for time series classification", Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, pp. 115-127, (2001).

[13] Xavier Carrearas, et al., “Filtering-Ranking Perceptron Learning for Partiala Parsing", Machine Learning, pp. 1–3, (2005).

[14] D. Aha, D. Kibler, and M. Albert, “Instance-based learning algorithms", Machine learning, vol.6, pp. 37–66(1991).

[15] John, G. Cleary and Leonard, E. Trigg, “K $*$: An Instance- based Learner Using an Entropic Distance Measure", Proceedings of the 12th International Conference on Machine learning, pp. 108–114, (1995).

[16] Eibe Frank, Mark Hall, and Bernhard Pfahringer, 2003, “Locally Weighted Naive Bayes", Department of Computer Science, University of Waikato. Atkeson, C., A. Moore, and S. Schaal, Working Paper 04/03, (1996).

[17] Kohavi R., “The Power of Decision Tables", In Proceeding European Conference on Machine Learning, pp. 174–189, (1995).

[18] Cohen, W., “Fast effective rule induction", In proceedings 12th International Conference on Machine Learning, Morgan Kaufmann. pp. 115–123, (1995).

[19] Geoffrey Holmes, Bernhard Pfahringer, Richard Kirkby, Eibe Frank, Mark Hall: “Multiclass alternating decision trees", In: ECML, pp. 161–172, (2001).

[20] Plamena Andreeva, et al., “Data Mining Learning Models and Algorithms for Medical Applications", 18 Conference Systems for Automation of Engineering and Research (SEAR 2004), Varna, BG, pp. 148–152, (2004).

[21] Ron Kohavi: Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid. Proceeding In: “Second International Conference on Knowledge Discovery and Data Mining", pp. 202–207, (1996).

[22] K. Hornik, M. Stinchcombe and H. White, “Multilayer feedforward networks are universal approximators, Neural Networks", 2, pp. 359–366, (1989).

[23] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, K.R.K. Murthy, “Improvements to Platt’s SMO Algorithm for SVM Classifier Design. Neural Computation", vol. 13, no. 3, pp 637–649, (2001).

[24] H. Liang, S. Lukkarinen and I. Hartimo, “Heart sound segmentation algorithm based on heart sound envelogram", Computers in Cardiology, pp. 105–108. (1997).

[25] H. Liang, S. Lukkarinen and I. Hartimo, “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, Chicago, USA, vol. 4, pp. 1630–1633, (1997).

[26] Manoranjan Dash a,. and Huan Liu, “Consistency-based search in feature selection", Artificial Intelligence, vol. 151, pp. 155–176, (2003).

[27] R. Kohavi, “Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 202–207, (1996).

[28] I. Kononenko. “Estimating attributes: Analysis and extensions of relief". Proceedings of the Seventh European Conference on Machine Learning, pp. 171–182, (1994).