Machine learning helps physicians in diagnosing of mitral valve prolapse

P. Povalej[1], M. Lenič[1], M. Zorman[1], P. Kokol[1], L. Lhotska[2], R. Pišot[3]
[1] Laboratory for system design, Faculty of Electrical Engineering and Computer Science,
University of Maribor, Slovenia
{Petra.Povalej, Mitja.Lenic, Milan.Zorman, Kokol}@uni-mb.si

[2] Gerstner Lab, Czech Technical University in Prague, Faculty of Electrical Engineering, Czech Republic
Lhotska@feld.cvut.cz

[3] Science and Research Centre of Koper, University of Primorska, Slovenia
Rado.Pisot@zrs-kp.si

Abstract

In this paper we present a multimethod approach for induction of a specific class of classifiers, which can assist physicians in medical diagnosing in the case of mitral valve prolapse. Mitral valve prolapse is one of the most controversial prevalent cardiac condition and may affect up to ten percent of the population and in the worst case results in sudden death. MultiVeDec is a general framework enabling researchers to generate various intelligent tools based on machine learning. In this paper we focused on various decision tree methods, which are capable of extracting knowledge in a form closer to human perception, a feature that is very important in medical field. The experiment included classifiers with various classical single method approaches, evolutionary approaches, hybrid approaches and also our newest multimethod approach. The main concern of the latest approach is to find a way to enable dynamic combination of methodologies to the somehow quasi unified knowledge representation. The proposed multimethod approach was capable to outperform all other tested approaches by producing classifier for diagnosing mitral valve prolapse with the highest overall and average class accuracy. More importantly, it was also capable to find some new knowledge important in diagnosing of mitral valve prolapse.

1 INTRODUCTION

Many real-world medical problems are nowadays being treated with tools for automatic intelligent data analysis. Various methods have been developed to improve the quality of analysis for specific domains. Application of any method in a specific domain requires specials characteristics, like for example instance methods based on artificial neural networks are capable of generalisation of nonlinearly separable problems, but have poor explanatory power not suitable for human understanding. While medical experts are not "very good acustomised with numbers" we as informaticians focused on methods, which are capable of extracting knowledge in a form closer to human perception, e.g. methods that induce decision trees, classification rules, etc.

In the project presented in this paper we focus on inducing a user-friendly classifier, which can assist physicians in diagnosing mitral valve prolapse (MVP) [1]. That is very important while MVP is one of the most prevalent cardiac conditions, which may affect from five up to ten percent of adult population, and is also one of the most controversial ones, and thereafter hard to diagnose without the use of expensive technology normally not available to general practitioners.

With a selection of knowledge representation and concentrating on decision trees we have only narrowed down the potential set of methods. Thereafter in order to find most appropriate classifier from above we tried out different methods for decision tree induction. First we made analysis with widely known tools for decision tree induction C4.5 and C5/See5 [2]. Despite acceptable results we wanted to find alternative solutions i.e. evolutionary and multimethod approach.

Main contributions of our paper are:

- the throughout analysis and comparison of different decision tree induction methods on a medical case of diagnosing mitral valve prolapse
- the introduction of the multimethod decision tree induction method
- the introduction of the general multimethod approach
- conformation of some well known medical pathways for diagnosing mitral valve prolapse
- new knowledge discovered which helps physicians in diagnosing mitral valve prolapse.

2 MITRAL VALVE PROLAPSE

Prolapse is defined as the displacement of a bodily part from its normal position. The term mitral valve prolapse (MVP)[1][3][4], therefore, implies that the mitral leaflets are displaced relative to some structure, generally taken to be the mitral annulus. The silent prolapse is the prolapse which can not be heard with the auscultation diagnosis and is especially hard to diagnose. The implications of the MVP are the following: disturbed normal laminar blood flow, turbulence of the blood flow, injury of the chordae tendinae, the possibility of thrombus' composition, bacterial endocarditis and finally hemodynamic changes defined as mitral insufficiency and mitral regurgitation.

Mitral valve prolapse is one of the most prevalent cardiac conditions, which may affect from five up to ten percent of population and is one of the most controversial one. The most common cause is probably myxomatous change in the connective tissue of the valvular leaflets that makes them excessively pliable and allows them to prolapse into the left atrium during ventricular systole. The clinical manifestations of the syndrome are multiple. The great majority of patients are asymptomatic. Other patients, however may present atypical chest-pain or supraventricular tachyarrhythmyas. Rarely, patients develop significant mitral regurgitation and, as with any valvular lesions, bacterial andocarditis is a risk.

Uncertainty persists about how it should be diagnosed and about its clinical importance. Historically, MVP was first recognized by auscultation of a mid systolic "click" and late systolic murmur, and its presence is still usually suggested by auscultatory findings. However, the recognition of the variability of the auscultatory findings and of the high level of skill needed to perform such an examination has prompted a search for reliable laboratory methods of diagnosis. M-mode echocardiography and 2D echocardiography have played an important part in the diagnosis of mitral valve prolapse because of the comprehensive information they provide about the structure and function of the mitral valve.

Medical experts propose [1][3][4] that echocardiography enables properly trained experts armed with proper criteria to evaluate mitral valve prolapse (MVP) almost 100%. Unfortunately however, there are some problems concerned with the use of echocardiography. The first problem is that current MVP evaluation criteria are not strict enough. The second problem is the incidence of the MVP in the general population and the unavailability of the expensive ECHO - machines to general practitioners. According to above problems we have decided to develop a decision support system enabling the general practitioner to evaluate the MVP using conventional methods and to identify potential patients from the general population.

3 METHODS

Machine learning community has a long tradition in knowledge extraction that can be traced at least as far as the mid-1960. Through the time different approaches evolved [2], such as symbolic approaches, computational learning theory, neural networks, etc. Most of the strength has been concentrating in finding a way to extract generalized knowledge from the examples.

The selection of appropriate method for analysis of data can be crucial for success. Therefore, for a given problem, different methods should be applied to increase quality of extracted knowledge. A brief overview of selected methods is presented in the following subsections.

3.1 Classical approaches

Decision trees [2] are easy understandable to the human and can be used even without a computer, but they have difficulties expressing complex non-linear problem. On the other hand, connectivistic approaches (such as neural nets), that simulate cognitive abilities of the brain, can extract complex relation, but are not understandable to humans, and therefore in such way not directly usable for data mining.

There are many other approaches, like representation of the knowledge with rules, rough-sets, case based reasoning, support vector machines, different fuzzy methodologies, ensemble methods [6] and they all try to answer the basic question: How to find an optimal solution, i.e. learn how to learn.

3.2 Evolutionary approach

Evolutionary approaches (EA) to knowledge extraction are also a good alternative, because they are not inherently limited to local solution. They are based on the evolutionary ideas of natural selection and genetic processes of biological organisms. As the natural populations evolve according to the principles of natural selection and "survival of the fittest", first laid down by Charles Darwin, so by simulating this process, genetic algorithms are able to evolve solutions to real-world problems, if they have been suitably encoded [7]. They are often capable of finding optimal solutions even in the most complex of search spaces or at least they offer significant benefits over other search and optimization techniques.

3.3 Hybrid approach

The hybrid approaches rest on the assumption that only in the synergetic combination of single models can unleash their full power [8]. Each of the single method has its advantages, but also inherent limitations and disadvantages, which must be taken into account when using the particular method. Therefore the logical step is to combine different methods (classic approaches) to overcome the disadvantages and limitations of a single method. For example in 1999 Boz [15] presented a method for converting a trained backpropagation neural network into a decision tree.

3.4 Multimethod approach

While studying presented approaches we were inspired by the idea of hybrid approaches and evolutionary algorithms. Both approaches are very promising in achieving the goal to improve the quality of knowledge extraction and are not inherently limited to sub-optimal solutions. We also noticed that almost all attempts to combine different methods use loose coupling approach. The methods work almost independent of each other and therefore a lot of luck is needed to make them work as a team.

Each of those methods uses its own internal knowledge representation (symbolic, connectivistic) that other methods cannot reuse, because of the incompatibility of knowledge representations. That incompatibility presents a major obstacle when trying to combine different methods using conventional hybrids. Opposed to the conventional hybrids described in the previous section, our idea is to dynamically combine and apply different methods in not predefined order to the same problem or the decomposition of the problem.

The main concern of the mutlimethod approach [9] is to find a way to enable dynamic combination of methodologies to the somehow quasi-unified knowledge representation. Multiple equally qualitative solutions like in EA approach, where each solution is obtained using application of different methodologies with different parameters are used. Therefore we introduce a population composed out of individuals / solutions that have the common goal to improve their classification abilities on a given environment/problem. We also enable coexistence of symbolic and cognitive representation in the same population. The most common knowledge representation models have to be standardized to support the use different methods on individuals. In that manner the transformation support between each individual method does not need to be provided. The action is based on the assumption that it is highly improbable to find unified representation for all knowledge representations, therefore we decided to standardize the most popular representations like neural nets, decision trees, rules, etc. Standardization brings in general greater modularity and interchangeability, but it has the following disadvantages - already existing methods cannot be directly integrated and have to be adjusted to the standardized representation.

Initial population of intelligent systems is generated using different methods. In each generation different operations appropriate for individual knowledge are applied to improve existing and create new intelligent systems. That enables incremental refinement of extracted knowledge, with different views on a given problem. For example, using different induction methods such as different purity measures can be simply combined in decision trees. That is also true for combining different learning techniques for neural networks. As long as knowledge representation is the same a combination of different methods is not a big obstacle.

The main problem is how to combine methods that use different knowledge representation (for example neural networks and decision trees). In such cases we have two alternatives: (1) to convert one knowledge representation into another using different already known methods or (2) to combine both knowledge representations in a single intelligent system.

The first approach requires implementation of knowledge conversion (for example there are numerous methods that can convert neural networks into decision trees and vice versa). Such conversions are not perfect and some of the knowledge can be lost. But on the other hand it can give a different view on presented problem that can lead to better results.

The second approach, which is based on combining knowledge, requires some cut-points where knowledge representations can be merged. For example in decision tree such cut-points are internal nodes, where condition in an internal node can be replaced by another intelligent system (for example support vector machine - SVM). The same idea can be applied also in decision leafs (Figure 1).

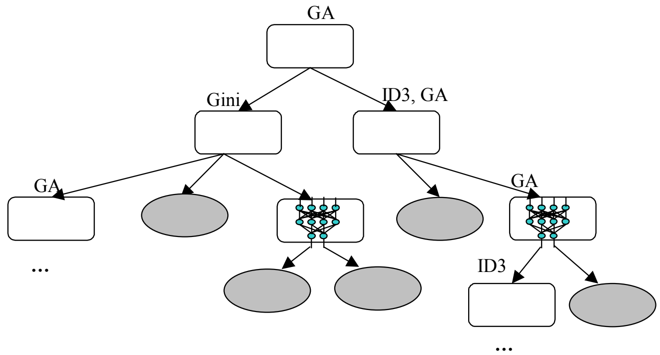

Figure 1: An example of a decision tree induced using multimethod approach. Each node is induced with appropriate method (GA – genetic algorithm, ID3, Gini, Chi-square, J-measure, SVM, neural network, etc.)

Using the idea of multimethod approach we designed a framework that operates on a population of intelligent systems - individuals. Since methods are usually composed out of operations that can be reused in other methods we introduced methods on the basis of operators. Therefore we introduced the operation on an individual as a function that transforms one or more individuals to a single individual. Operation can be a part of one or more methods, like pruning operator, boosting operator, etc. Operator based view provides the ability to simply add new operations to the framework (Figure 2).

The representation with individual operations facilitates an effective and modular way to represent the result as a single individual, but in general the result of operation can be also a population of individuals (for example mutation operation in EA is defined on individual level and on the population level). The single method itself is composed out of population operations that use individual operations and is introduced as a strategy in the framework that improves individuals in a population (Figure 2).

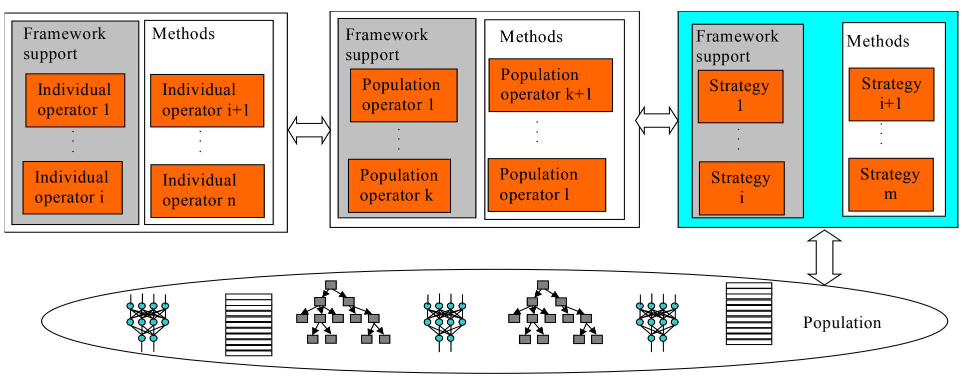

Figure 2: Multimethod framework

4 RESULTS AND DISCUSSION

Using the Monte Carlo sampling method, 900 children and adolescents were selected representing the whole population under eighteen years of life. Routinely they were called for an echocardiography no matter of prior findings. 631 of them

passed an examination of their health state in a form of a carefully prepared protocol specially made for the syndrome of MVP. The protocol consisted of 103 parameters that can possibly indicate the presence of MVP. Distribution of the three decision classes were: 5% "prolapse", 6% "silent prolapse", and 89% "no prolapse".

The following basic purity metrics for greedy induction of decision trees where used: Information gain ratio (ID3), χ^2 (Chi square), Gini, J Measure [11,12,13], IBR (iBARET) [14] and newly developed Mgain metric. iBARET represents a learning tool based on principles of instance-based learning and is derivative of the well-known kNN method. iBARET implements a batch feature weighting method with performance bias, i.e., it uses feedback from the performance function during training. This function is calculated for each setting of feature weights and a genetic algorithm is used to suggest new sets of feature weights. Receiver operating curve methodology is used to calculate this fitness function, the methodology is able to optimize classifier performance in domains with non-uniform class distribution.

Different linear combinations of classic purity metrics were introduced and evaluated. In addition AdaBoost method [13] was used in order to improve the quality of classically induced classifiers. In order to make objective assessment of classifier quality we used average class accuracy, which on unbalanced data sets is more informative measure than overall accuracy. The classification results on testing set are presented in table 1.

Table 1: MVP classification accuracy (□), average class accuracy (Δ) on test set. The methods are ranked on the basis of average class accuracy

Method	□	Δ
Multimethod	93.84	84.31
Greedy linear	91.54	83.47
Greedy Chi square + ID3	91.54	83.47
Greedy Gini + Chi square	91.54	83.47
IBR - reduced (10) attributes	91.54	83.47
Greedy MGain + Chi square	90.77	83.19
Greedy Chi square	90.00	82.91
Greedy Gini + ID3	92.31	79.27
Greedy MGain * ID3	90.77	78.71
Greedy Voting	90.00	78.43
Greedy Gini * ID3	90.00	78.43
Greedy MGain * Chi square	89.23	78.15
Greedy MGain * Gini	89.23	78.15
Greedy Gini	88.46	77.87
Greedy MGain + ID3	88.46	77.87
Greedy Boost linear	91.54	74.51
Greedy J measure + Chi square	88.46	74.30
IBR – all attributes	77.69	73.95
Greedy ID3	88.46	73.39
Boost Greedy ID3	91.54	70.94
Greedy Chi square * ID3	90.77	70.66
Boost Greedy Chi square * ID3	93.08	70.59
Boost Greedy MGain * ID3	92.31	70.31
Boost C5	92.30	70.00
Boost Greedy Gini + ID3	90.77	69.75
Boost Greedy MGain + ID3	90.77	69.75
Greedy Gini * Chi square	87.69	69.54
Greedy Mgain	87.69	68.63
Greedy MGain + Gini	87.69	68.63
Boost Greedy J measure + Chi square	91.54	66.46
Boost Greedy MGain + J measure	90.77	66.18
Genetic	92.31	65.83
C4.5	91.50	65.50
C5	91.50	65.50
Boost Greedy Gini	87.69	64.15
Nested Boost	93.08	62.54
Boost Greedy Gini * Chi square	90.00	61.42
Boost Greedy MGain * Gini	90.00	61.42
Boost Greedy MGain + Chi square	85.39	59.73
Boost Greedy Gini * ID3	92.31	57.77
Boost Greedy Chi square	90.77	57.21
Boost Greedy Chi square + ID3	89.23	56.65
Boost Greedy Gini + Chi square	89.23	56.65
Boost Greedy MGain * Chi square	89.23	56.65
Greedy J measure * Chi square	83.85	55.60
Boost Greedy MGain	85.39	55.25
Boost Greedy MGain + Gini	85.39	55.25
Boost Greedy J measure * ID3	90.77	53.64
Boost Greedy J measure	89.23	53.08
Greedy J measure + ID3	86.15	51.96
Greedy J measure + Gini	85.39	51.68
Greedy MGain * J measure	73.08	51.68
Greedy J measure	83.08	50.84
Greedy J measure * ID3	81.54	50.28
Boost Greedy J measure * Chi square	87.69	48.95
Boost Greedy J measure + ID3	89.23	48.60
Greedy J measure * Gini	80.77	41.95
Greedy MGain + J measure	80.00	41.67
Boost Greedy J measure + Gini	89.23	40.55
Boost Greedy J measure * Gini	89.23	40.55
Boost Greedy MGain * J measure	87.69	31.93

The methods in table 1 were ranked according to the average class accuracy of the induced classifier. It can be seen, that the average class accuracy varied from 31.93% (Boost Greedy MGain * J measure) in the worse case to 84.31% in the best case (multimethod). On the other hand the overall accuracy varied from 73.08% (Greedy MGain * J measure) to 93.84% (multimethod). Considering both evaluation criteria we established that the average class accuracy is far more distinctive compared to overall accuracy. Nevertheless, the multimethod approach outperformed other methods in both categories.

The most reliable decision tree classified MVP with very high total accuracy of 93.84% and average class accuracy 84.31% was induced with multimethod approach. As expected some well-known medical diagnostic pathways were shown in the decision tree but surprisingly (from the clinical experts' point of view) the most informative attribute for diagnosing silent prolapse in the decision tree was SI (systolic index) value. The medical explanation is that we could speculate about the SI values, which are directly correlated with systolic and diastolic diameters. Systolic and diastolic parameters represent indirect functional values of the heart muscle. In a heart with prolapsed mitral valve the end systolic pressure could be a little bit higher than in normal hearts since the mitral valve descend in the left atrium. We can confidently speculate that because of this unusual movement of the leaflets of the prolapsed mitral valve the SI is actually higher in the normal hearts.

The explanation for better performance of multimethod approach can be found in the concept of multimethod approach. Classical approaches always produce one deterministic result that is not necessarily an optimal solution. Evolutionary approach overcomes limitation of classical approaches and can find global optima, but solution search space is huge and the optima is not necessarily found. Multimethod approach tries to combine benefits of classical and evolutionary approaches and reduces search space with help of conventional methods.

5 CONCLUSION

Since mitral valve prolapse is one of the most controversial prevalent cardiac condition which can affect up to ten percent of population, the aim of presented research was to identify important factors for diagnosis. Many well known methods for machine learning for induction of decision trees were used. Also a new approach – multimethod approach for induction of classifiers was presented. The results showed that our multimethod approach outperformed other methods in classifying unseen mitral valve prolapse cases in the terms of overall and average class accuracy. The multimethod decision tree showed many well known decision pathways, but more importantly, some new knowledge was extracted which can be helpful in better diagnosing of mitral valve prolapse.

REFERENCES

[1] R. Devereoux, "Diagnosis and Prognosis of Mitral Valve Prolaps", *The New England Journal of Medicine*, vol. 320, no. 16 1989, pp. 1077-1079.

[2] J.R. Quinlan. *C4.5: Programs for Machine Learning*, Morgan Kaufmann publishers, San Mateo, 1993.

[3] H.R. Anderson et al, "Clinicians Illustrated Dictionary of Cardiology", *Science Press*, London, 1991.

[4] J.B. Barlow et al, "The Significance of Late Systolic Murmurs", *American Health Journal*, vol. 66, pp. 443-452, 1963.

[5] S. Thrun and L. Pratt, *Learning to Learn*. Kluwer Academic Publishers, 1998

[6] T.G. Dietterich, "Ensemble Methods in Machine Learning". in *First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science* New York: Springer Verlag, 2000, pp. 1-15.

[7] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison Wesley, Reading MA, 1989.

[8] C.J. Iglesias, "The Role of Hybrid Systems in Intelligent Data Management: The Case of Fuzzy/neural Hybrids", *Control Engineering Practice*, vol. 4, no. 6, 1996, pp. 839-845.

[9] M. Lenič, P. Kokol, "Combining Classifiers with Multimethod Approach", in *Hybrid intelligent systems*, in press.
[10] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, Wadsworth International Group, Belmond, CA, 1984.
[11] G.W. Snedecor and W.G. Cochran, *Statistical Methods*, Eighth Edition, Iowa State University Press, 1989.
[12] M.A. Bramer, "Using J-pruning to reduce overfitting in classification", in *Knowledge Based Systems*, vol. 15, issues 5-6, pp. 301-308, July 2002.
[13] Y. Freund, R.E. Schapire, "Experiments with a new boosting algorithm", in *Proceedings Thirteenth International Conference on Machine Learning*, Morgan Kaufman, San Francisco, pp. 148-156, 1996.
[14] J. Klema, J. Palouš, "iBARET - Instance-Based Reasoning Tool", Eunite Proceedings, Verlag Mainz, pp. 213-221,2001
[15] O. Boz, "Converting a Trained Neural Network to a Decision Tree", [http://www.leigh.edu/~ob00/integrated/report-0.0.html], Januar 1999