DIAVAL, a Bayesian expert system for echocardiography*

F. J. Díez† J. Mira† E. Iturralde‡ S. Zubillaga§

†Dept. of Artificial Intelligence. UNED.
Avda. Senda del Rey. 28040 Madrid. Spain
E-mail: <{fjdiez,jmira}@dia.uned.es>
WWW: http://www.dia.uned.es/{~fjdiez,~jmira}

‡Section of Echocardiography. Hospital de la Princesa.
Diego de León, 62. 28006 Madrid. Spain

§Intensive Care Unit. Hospital Universitario.
Ctra. Toledo, Km. 12'5. 28905 Getafe. Madrid. Spain

Keywords: Knowledge-based systems. Uncertainty management. Bayesian networks. AI in Medicine. Echocardiography.

Abstract

DIAVAL is an expert system for the diagnosis of heart diseases, based on several kinds of data, mainly from echocardiography. The first part of this paper is devoted to the causal probabilistic model which constitutes the knowledge base of the expert system in the form of a Bayesian network, emphasizing the importance of the OR gate. The second part deals with the process of diagnosis, which consists of computing the a posteriori probabilities, selecting the most probable and most relevant diagnoses, and generating a written report. It also describes the results of the evaluation of the program.

1 Introduction

1.1 Echocardiography. History of DIAVAL

The **diagnosis** of a patient begins by registering his/her personal data and medical history, and proceeds with the physical examination, which consists of inspection, auscultation, palpation and percussion. In the case that a suspect disease exists, it is necessary to carry out complementary tests, beginning with those of lower risk and lower cost [27]. In cardiology, it is usual to carry out first an electrocardiogram and then, if necessary, an echocardiogram or an angiogram. Other complementary diagnostic techniques include radiography, analytical exams, radioisotopical tests, nuclear magnetic resonance (NMR), etc. [4].

The main advantage of *echocardiography* is that it is a non-invasive technique that offers a lot of valuable information without any risk to the patient. There are two basic types of echocardiography: transthoracic (placing the probe on the patient's chest) and trans-esophageal (introducing the probe through his/her mouth), and for each one, there are several

*The final version of this manuscript appeared in *Artificial Intelligence in Medicine.*

modes: M-mode, bidimensional and Doppler study (with continuous or pulsed wave, color-codified, color map, etc.) [32]. The main difficulty of this technique lies in the blurriness of the images produced by the echocardiograph, especially in transthoracic echocardiography if the patient has a "bad window", due, for example, to obesity. The interpretation of these images and the measurement of relevant parameters[1] is a time-consuming task that requires the expertise of a trained cardiologist.

In view of the advantages that artificial intelligence and computer vision might contribute to this problem, the Section of Echocardiography at the Hospital de la Princesa, in Madrid, and the Department of Computer Science and Automatics at the UNED decided in 1989 to initiate a research project that would be the subject of F. J. Díez's doctoral thesis [9]. The objective was to build a vision system incorporating a rule-based expert system, possibly extended with fuzzy logic.

Nevertheless, since the first knowledge-acquisition session, doctors preferred to use a causal representation of the cardiac pathophysiology. The difficulty of converting a causal model into a set of rules and the infeasibility of implementing all possible inferences by chaining such rules, led to the investigation of a new approach which used a probabilistic causal network as a knowledge base, and Bayes' theorem as the inference method. In the process, Díez rediscovered some of the results already published in the literature on Bayesian networks, such as *d*-separation, the binary OR-gate and the propagation of evidence through message-passing in singly-connected networks [35]. Afterwards, he incorporated into his work some of the published results and made new discoveries: the generalized OR-gate (see sec. 2.2), the local conditioning algorithm (sec. 3.1), a learning model [8], a rudimentary explanation method (unpublished work) and a technique for selecting diagnoses and writing a report (sec. 3).

The result was DIAVAL, a program that assists cardiologists in the interpretation of the data obtained from echocardiography. Its name stems from "DIAgnóstico de VALvulopatías", since its main objective was the diagnosis of valvular heart diseases, although in the end it also addressed other cardiac anomalies, such as acute myocardial infarction and the different forms of pericarditis. Our Department of Artificial Intelligence is now engaged in the project of building a vision system for echocardiography, whose output will be the input of the expert system, hence alleviating the doctor from the burden of introducing the data manually.

1.2 Normative expert systems in medicine

Expert systems arose in the 70's in the field of artificial intelligence as computer programs that, like human experts, possess a deep knowledge about a narrow domain, which allows them to solve problems by reasoning and explaining this reasoning [5, 41]. A great proportion of the most famous expert systems were built as diagnosis assistants and therapy advisors in different medical areas (MYCIN, PIP, CENTAUR, INTERNIST, ONCOCIN, etc.), and almost all of them based their reasoning totally or partially on IF-THEN rules, which, combined with frames or objects, constitute at present the standard method for building expert systems.

However, the use of rules raises serious problems with regard to knowledge representation, and furthermore, with regard to uncertainty management, which may lead to erroneous conclusions [1, 14, 17]. The origin of the conflict consists of applying a formalism developed for classical logic, in which every proposition was either true or false, to the management of uncertainty through numerical factors [18],[35, sec. 1.2].

Bayesian networks appeared in this scenario in the 80's as a normative method for uncertain reasoning. According to Heckerman, Horvitz and Nathwani [19], "the word 'normative'

[1]In this context, it is usual to call "parameters" some magnitudes measured on the echocardiogram; this differs from the mathematical meaning of the term.

comes from decision analysts and cognitive psychologists who emphasize the importance of distinguishing between *normative behavior*, which is what we do when we follow the desiderata of decision theory, and *descriptive behavior*, which is what we do when unaided by these desiderata". Therefore, the main advantages of Bayesian networks is that they are grounded on a solid mathematical theory, in which all the assumptions of conditional independence are explicit, and they constitute a causal model from which it is possible to obtain all sound inferences, performing abductive, deductive and intercausal reasoning at the same time [13].

The assertion that many of the most famous expert systems have been developed in the field of medicine is also true in the subdomain of Bayesian expert systems. Pathfinder, for example, is a normative expert system that assists surgical pathologists with the diagnosis of lymph-node diseases [19, 20]. Other research groups at Aalborg University [2], in Denmark, and at the University of Pavia[2], in Italy, have developed several Bayesian networks for medicine, and the number of this kind of systems is growing exponentially in America and Europe. Even some programs that were initially implemented in other formalisms have been converted into Bayesian networks, such as Internist-1/QMR [39, 38] and Iliad [28]; in particular, the latter study suggests that "the Bayesian network model [of Iliad] is more reliable and discriminative than the [original] model" and "had more positive influence on physician's diagnosis". These results agree with empirical studies showing that Bayesian networks perform better than other approaches to expert systems [15, 22, 24, 42].

Besides the emphasis on echocardiography, this normative character is the main difference between DIAVAL and the Heart Failure Program [29, 30], which also uses a probabilistic causal network —it is not an authentic Bayesian network because it contains cycles— but applies a heuristic method to find the hypothesis (diagnosis) that better explains the available findings.

1.3 Overview

The rest of this papers is organized as follows. The second section is devoted to the knowledge base of the expert system, i.e., to the cardiological causal model; after reviewing the general properties of Bayesian networks (sec. 2.1) and the multivalued OR-gate (sec. 2.2), it discusses the process of knowledge acquisition and representation in DIAVAL (sec. 2.3). The third section deals with the process of diagnosis, which consists of propagating evidence (sec. 3.1) and then selecting the diagnoses (sec. 3.2), and with the generation of a written report (sec. 3.3). Finally, section 4 discusses the evaluation of the program and section 5 summarizes the conclusions. The graphical interface and the explanation of reasoning will be described in a future paper because of the limit of space.

The different aspects of Bayesian inference were also presented with a different perspective in [12, 11], by considering for each one of them the three levels proposed by David Marr: theory, algorithm and implementation [31], with special interest on distributed implementations.

2 Cardiological model

2.1 Bayesian networks as a knowledge-representation scheme

Bayesian networks are probabilistic models based on graphical representations [35, 33]. More precisely, in a Bayesian network every node represents a variable, such as sex, age, body-surface area, dyspnea, hemoptysis, atrial diameter, mitral annulus dilatation, pulmonary hypertension, subaortic stenosis, etc. For example, the node "mitral regurgitation" (X) in figure 1, can take on four values: "absent" (x_0), "mild" (x_1), "moderate" (x_2) and "severe" (x_3).[3] Nodes

[2]See `http://ipvaimed9.unipv.it/lab/publications.html` on Internet.

[3]We follow the convention of representing nodes or variables in uppercase and their values in lowercase.

are connected by directed arcs, such that a link from A to B indicates that the value taken on by A *influences* the value that B takes (later we will define in more detail the meaning of "influences"). Then, A is said to be a *parent* of B, and B to be a *child* of A. A limitation of Bayesian networks is that, by definition, their graphs can not contain cycles,[4] and this prevents them from representing feedback effects.

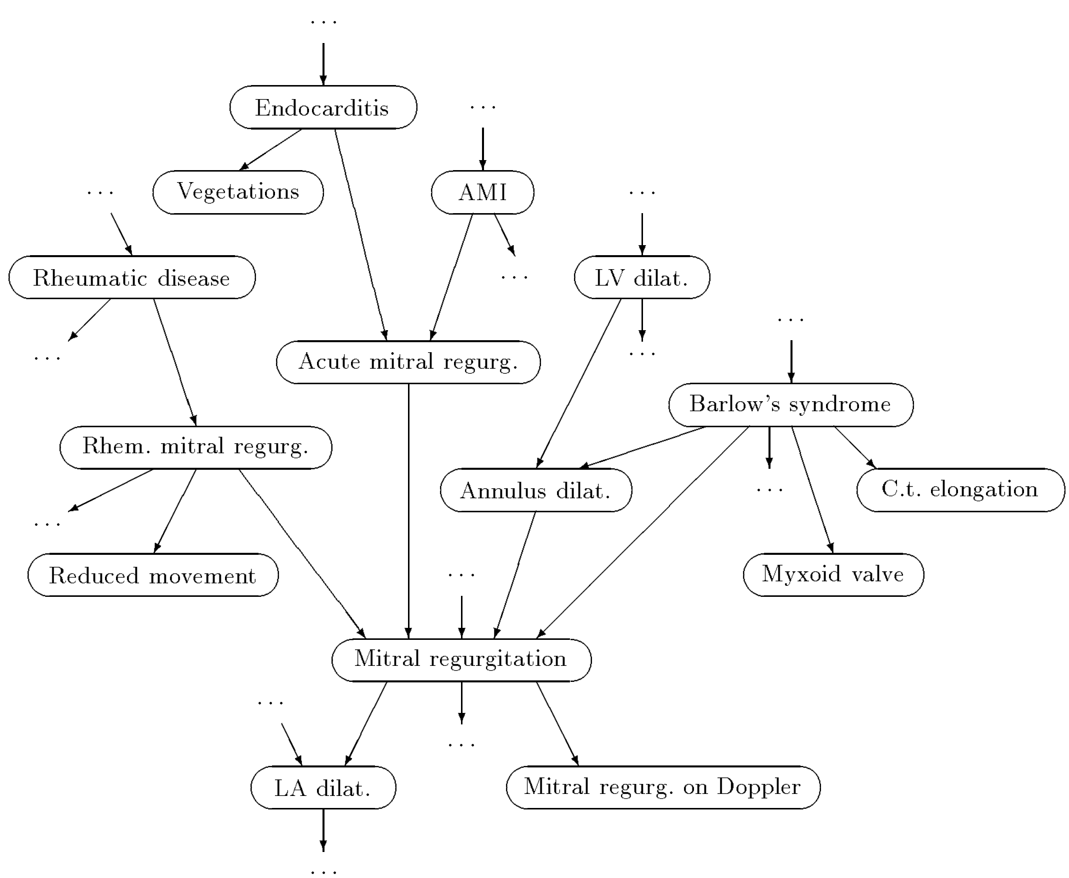

Figure 1: A portion of the network for mitral regurgitation.

Besides this structural information, a Bayesian network also contains numerical data, usually expressed in the form of conditional probabilities: if the parents of X are $pa(X) = \{U, V, W\}$, the conditional probability is given by a table $P(x|pa(x)) = P(x|u, v, w)$; for a node without parents, its conditional probability is simply the a priori probability: $P(x|\emptyset) = P(x)$. These conditional probabilities induce a joint probability given by

$$P(x_1, \ldots, x_n) = \prod_i P(x_i|pa(x_i)) \tag{1}$$

In summary, a Bayesian network is defined as a connected[5] acyclic directed graph in which every node has an associated probability table $P(x|pa(x))$, such that x represents each of the values taken on by the variable corresponding to the node, and $pa(x)$ represents a configuration

It is also possible to have continuous variables in a Bayesian network, but in DIAVAL all of them are discrete.

[4]In directed graphs, closed paths can be either loops or cycles. The difference is that in a cycle one always arrives at the starting point by following the direction of its links, while in a loop it is impossible. Bayesian networks may contain loops, but no cycles.

[5]In our opinion, the connectedness of the graph must be included in the definition of a Bayesian network, because an unconnected graph does not represent one model but several of them.

of its parent nodes in the graph. The joint probability given by eq. (1) holds some probabilistic independence properties relative to the structure of the graph, called d-separation, which can be used as an alternative definition of Bayesian networks [35].

Every *finding* determines the value of one variable; for example, the fact of verifying that the patient presents with severe dyspnea can be translated into the assignment of value d_3 to variable D. The set of findings is termed *evidence* and is usually represented by $\mathbf{e}$. In the probabilistic paradigm, the problem of *diagnosis* consists of computing the a posteriori value of one or several variables given the available evidence: $P(x|\mathbf{e})$.

Therefore, a fundamental research issue in the field of Bayesian networks consists of designing efficient algorithms for finding the a posteriori probabilities. For singly-connected networks, there is an elegant algorithm whose complexity grows linearly with the number of nodes in the network (assuming a limit for the number of parents in every family) [34]. For general networks, the problem is NP-hard [6], but some algorithms have been developed which can deal with real-world models in reasonable time [26, 23]; in particular, DIAVAL employs a new algorithm [10], called local conditioning, which will be described in section 3.1.

2.2 The OR-gate

The noisy OR-gate was introduced by Pearl [34] and by Peng and Reggia [36] as a simplified model of interaction among the parents of node, which considers each one of them as one of the possible *causes* of X and assigns to each link $U_i \rightarrow X$ the probability that U_i produces X when the other causes of X are absent, unlike the general model, which consists of a table of conditional probabilities $P(x|u_1, \ldots, u_n)$. The main advantage of the OR-gate is that it simplifies the process of **knowledge acquisition**, not only because it reduces the number of parameters for each family from exponential to linear in the number of parents, but also because it is easier to answer a few questions like "What is the probability that U_i produces X" instead of a great number of questions involving a complex casuistry, such as "What is the probability of X when U_1 is present, U_2 is present, U_3 is absent and U_4 is present?", etc. This is true when probabilities are elicited from human experts as well as when they are extracted from a database, because in general only one of the causes of X is present in each patient.

A second advantage of the OR-gate, also related to the reduction in the number of parameters, is that the **computational complexity** of evidence propagation is proportional to the number of parents, while the general model requires an exponential amount of time.

Finally the fact of considering the parents of a node as the possible *causes* that may *produce* a certain anomaly and not only as *factors* that *influence* the node, gives rise to **explanation patterns** that are not applicable in the general case, such as trying to determine which is the cause that has produced a certain anomaly. This is a consequence of the different *semantics* of the general model with respect to the OR-gate, which is deeper than the computational differences.

So far we have discussed the binary OR-gate. Nevertheless, in real applications, especially in medicine, most of the variables are not binary but multivalued; for many of them, the typical values are "present, mild, moderate, severe". For this reason, Henrion [21] proposed a generalized gate that simplified the process of knowledge acquisition, but the propagation of evidence required the explicit construction of the probability table $P(x|u_1, \ldots, u_n)$, which led again to exponential time complexity. In the construction of DIAVAL we arrived independently at the same model as Henrion,[6] proposed the name MAX-gate and formalized the model by

[6]The same model is also discussed in [38]. A different generalization of the OR gate for multivalued variables was developed by Srinivas [40] as a modeling tool for digital circuits diagnosis, network reliability analysis and similar problems.

establishing two axioms and introducing the c_x^u parameters,

$$c_x^u = P(X = x|U = u, \text{the other causes of } X \text{ are absent}), \tag{2}$$

We also developed an algorithm for computing the probability in linear time [8] even in the case of multiply-connected networks, provided that the global propagation algorithm is local conditioning (see sec. 3.1).[7]

Because of the advantages of the OR-gate, in the construction of DIAVAL we apply this simplified model to each family that satisfies the following conditions:

1. both the child node and its parents must be variables indicating the degree of presence of an anomaly, i.e., the range of values must be "absent/present" or "absent/mild/moderate/severe" or a similar set [8]; this prevents the application of the OR-gate when the parents represent other kind of variables, such as age, sex or race;

2. each of the parent nodes represents a cause that can produce the effect (the child variable) in the absence of the other causes;

3. there is no significant synergy among the causes, i.e., the mechanism leading U to produce X is independent of the mechanisms of the other causes of X.

   Therefore, conditions 2 and 3 prevent the application of the OR-gate when the parents represent risk factors, such as smoking, obesity, hypercholesterolemia... none of which is able to produce the effect (myocardial infarction, for example) independently of the others.

As an exception, we apply the general model instead of the OR-gate when there is only one explicit cause of a node, because in this case the probability table explicitly contains the sensitivity and specificity of the link, which, for a doctor, are more meaningful than a leaky probability (i.e. the probability that causes not explicit in the model produce the effect) [21, 8].

2.3 Bayesian network of DIAVAL

The construction of a Bayesian network is usually divided into two phases. The *structural phase* consists of selecting the variables and establishing causal links among them. The *numerical phase*, even more difficult than the first one, consists of assigning the corresponding numerical probabilities. In the development of DIAVAL, we have employed two sources of information: interviews with experienced cardiologists, especially echocardiographists, and medical literature, both books and journals.

The medical knowledge elicited was represented in a network which, in its current version, consists of 324 nodes and 335 links; it means that the network contains a certain number of loops, which complicate the propagation of evidence. This set of nodes is composed of 3 personal data (age, sex and country of origin [8]), 86 anomalies and 235 data. The basic difference between anomalies and data is that the former are not directly observable, while the latter correspond to medical findings; as a consequence, in the network every anomaly

[7] Clustering algorithms can not take advantage of this property directly, not even in the case of singly-connected networks; they can propagate evidence in linear time only by introducing dummy nodes [16] which complicate the explanation of reasoning because they do not represent any real-world entity.

[8] The country of origin is relevant for the diagnosis of rheumatic diseases: while they are already erradicated in some zones, in others they affect most of the population and constitute the main cause of cardiopathies. Consequently, we assigned three values to this node, corresponding to low, medium or high risk of suffering from rheumatic diseases.

node has children (its effects) while data are leaf nodes. Among them, there are 52 numerical parameters, corresponding to measures performed in the different echocardiographic modes; the rest of the data are qualitative observations.

One of the unique features of knowledge representation in DIAVAL, which differentiates it from other Bayesian networks, is the extensive use of the OR-gate whenever the conditions stated above make it possible. Figure 1 shows a small portion of DIAVAL's model.

Besides this causal network, there exists another organization of nodes, also hierarchical but orthogonal to the first one, in the form of a classification tree, which clusters the nodes in 21 *chapters*: 8 for the patient's antecedents (previous diseases, symptoms, risk factors, ECG, etc.) and 13 for the echocardiographic findings: mitral valve, aortic valve, segmentary contractility, pericardium, etc. The objective of classifying the nodes into chapters is to generate a written report in which entered data and resulting diagnoses are presented in an ordered fashion (see sec. 3.3).

In our opinion, the weakest side of DIAVAL's knowledge base is the inaccuracy of conditional probabilities, since the large size of the network made impossible a thorough study of each parameter by analyzing its relative weight with respect to its neighbor parameters and its influence on the diagnosis (see sec. 4).

3 Diagnosis

3.1 Probabilistic inference

After introducing all the available findings, the expert systems propagates evidence on the Bayesian network, in order to compute the probability of each node. Several algorithms exist for this process, some of them exact, some approximate (based on stochastic simulation). Nowadays, the standard method consists of a clustering algorithm known as "click-tree propagation" [26, 23]; nevertheless, during the construction of DIAVAL we developed a new algorithm, called *local conditioning* [10]. Conditioning methods break the loops in the network by instantiating a set of variables (called "cutset"), propagate evidence in the resulting tree and combine the a posteriori probabilities by taking into account the a priori probability of each cutset instantiation. Local conditioning, in particular, is very similar to Kim and Pearl's [25, 34] algorithm for the polytree, because it exchanges two messages for every link, with the only difference that if link $X \rightarrow Y$ belongs to loops broken by variables $\{V_1, \ldots, V_n\}$, then messages $\pi_Y(X)$ and $\lambda_Y(X)$ are conditioned on those variables. As a consequence, local conditioning is more efficient than other methods [35, 37, 7] which apply more conditioning than necessary [10]. More specifically, DIAVAL uses the integrated version of local conditioning, which propagates evidence directly on the original network, unlike clustering algorithms, which must compile the network by triangulating it and forming clusters.

Obviously, it would be possible to replace local conditioning with any other algorithm for computing the probability without affecting the appearance of the expert system. The user would at most notice a change in the time spent in finding the diagnosis.

3.2 Selection of diagnoses

In some Bayesian expert systems, the process of diagnosis is limited to computing and showing the a posteriori probability of every node. This poses the problem that the user must search for the relevant information in the network. As a solution, we have endowed DIAVAL with a mechanism for selecting diagnoses and showing them in an ordered fashion, which consists of assigning a *relevance factor* to every node and establishing two thresholds:

- *Certainty threshold*: it pays for selecting out only the most certain diagnoses; its default value is 0.65, which means that a necessary condition for a node to be included in the list of diagnoses is that its most probable value (a posteriori) is over 65%.
- *Relevance threshold*: its objective is to only select those diagnoses that are significant for a cardiologist; its default value is 7, in a subjective scale ranging from 0 to 10.

After propagating evidence, DIAVAL examines every node in the succession of chapters (sec. 2.3) and selects the nodes which overcome both thresholds, thus generating an ordered list in which diagnoses relative to the same part or function of the heart appear together.

Actually, every node has not only one relevance factor but two: one for positive and one for negative diagnosis (PDR/NDR). The former is always higher than the latter, since in general it is more important to indicate the diseases suffered by the patient than those discarded from the diagnosis.[9]

The diagnoses selected are displayed on a screen. By clicking the mouse on one of them, the user can access a menu which offers several possibilities such as:

1. consulting the a priori and a posteriori probability of each value of the variable;
2. overriding a diagnosis by removing it or by selecting another value;
3. finding the most probable cause (only for OR-gates);
4. navigating across the network by visiting the parents and children of this node.

The second option responds to the principle that the doctor's opinion must always prevail over the computer's advice, owing to legal and ethical reasons, even if the expert system were able to diagnose better than a human being (which is not our case). Options 3 and 4 are part of a rudimentary explanation capability described in the next section.

The menu bar of the graphical interface gives access to the possibility of modifying the thresholds: by reducing the certainty threshold, more uncertain results may appear; by lowering the relevance threshold, the system will show some diagnoses it had deemed less relevant. The same menu also allows the doctor to "manually" add new diagnoses to those selected by the system.

3.3 Written report

Finally, DIAVAL asks for the name of the doctor who did the echocardiogram (it maintains a small database of names) and generates a text file made up of different sections which summarize the findings and the conclusions. The diagnoses are the same as shown on the corresponding window of the interface, with the only difference that doctors collaborating in our project did not want the probabilities to appear in the report. Then, the system opens this file in a text editor, so that the user can add the final changes before sending it to the printer; again, this is an attempt to comply with the principle that the doctor must be completely free in the composition of a report for which he/she will take on the responsibility.

[9]Nevertheless, in medicine this is not always the case: sometimes a negative diagnosis is as important as a positive one; therefore, it would be desirable in future versions of DIAVAL that the relevance factor of each node was adjusted automatically depending on the observed evidence and the reason for which the echocardiogram was requested. Anyway, this is a complex issue which requires further research.

4 Evaluation

Due to external constraints, the development of DIAVAL was halted in 1994, while we were in the process of debugging the knowledge base by refining the conditional probabilities of the Bayesian network. Only an informal evaluation of the system was performed, with the collaboration of six cardiologists: five of them were experts in adult echocardiography and one in pediatric echocardiography. We applied three of the seven evaluation methods proposed by Berry and Hart [3]: walkthroughs, questionnaires and interviews.

Every doctor was asked to introduce the findings of one or two echocardiograms —he/she could decide either to take the data from a real patient or to imagine a patient in a certain pathological state— and afterwards he/she filled in a questionnaire addressing six aspects of the expert system: the cardiological model, the interface, the diagnosis, the explanation capability, the written report and his/her overall judgment of the program. It contained seventeen closed questions, which consisted of assessing some features of the program from 1 (positive) to 5 (negative), and seven open questions. We also interviewed every doctor to further inquire his/her opinion about the expert system. The results are as follows:

Interface: All the doctors rated very positively the ease to use the program, the clear presentation of results and the context-sensitive help. The agreement about the comfortability and flexibility of the program was slightly lower. The interviews carried on suggested that —as expected— doctors who had previously used windows environments deemed the program more user-friendly than those who had not.

Cardiological model: The doctors appreciated the existence of an interval table for each parameter; in contrast, they paid little attention to the underlying Bayesian network. Therefore, in the future we will make an effort to explain the users the importance of the probabilistic causal model and of Bayesian reasoning in the process of diagnosis.

Diagnosis: In general, the doctors were favorably impressed by the diagnosis offered by the expert system, in spite of the errors, which were never foolish or absurd. The evaluation showed that DIAVAL tends to *overdiagnose* in three ways: by overestimating the severity of an anomaly, such as diagnosing moderate or severe mitral stenosis instead of mild stenosis; by offering too specific results, such as diagnosing acute regurgitation when there was not enough evidence to determine whether it was acute or chronic; and, in general, by overestimating the probability of anomalies, which occasionally lead to including some diagnoses that should have remained under the certainty threshold (sec. 3.2). This means that the conditional probabilities estimated by the experts who collaborated in the construction of the Bayesian network were too high and must be assessed more accurately.

When asked about the utility of the diagnosis offered by the expert system, there was a significant disagreement among doctors. In our opinion, these divergences stem from different conceptions of echocardiography (and of medicine, in general): some of them, who try to make echocardiography a technique as objective as possible, rate highly a program that helps them in the mathematical evaluation of the echocardiographic findings; on the contrary, other experts, who regard medicine as an art, rely on their subjective judgment above all and mistrust the advice of a computer. Naturally, residents appreciate this advice more than expert echocardiographists.

Explanation: As expected, the explanations offered by the expert system were unsufficient for the doctors, who were not very interested in navigating across the network to find out the origin of inaccurate diagnoses. This explanation capability —like that of MYCIN

[5], which consisted of a trace of the rules chained— is very useful for the knowledge engineer but offers little help to the end user.

Written report: The feature of DIAVAL that produced the most favorable impression in the physicians who evaluated it is the elaboration of a report which gathers the findings and diagnoses in a ordered concise way. This capability is a consequence of the classification of the nodes of the network into chapters, as mentioned in section 2.3.

5 Conclusions

This paper has discussed some of the aspects of the DIAVAL expert system for the diagnosis of heart diseases through echocardiography. In particular, it has tackled two issues: the **probabilistic model** (the Bayesian network), and the **operation** of the program.

With regard to the first point, the feature that differentiates this program from most expert systems is the use of a probabilistic causal model instead of using rules; as a consequence, inference consists on applying Bayes theorem in accordance with the conditional separation axioms of the model, instead of chaining rules. In comparison with other Bayesian networks, DIAVAL is distinguished by the use of the multivalued OR-gate as the standard interaction model, and by the application of the local conditioning algorithm for propagating evidence. On the other hand, one of the weaknesses of the current version of the program is the inaccuracy of conditional probabilities, whose assessment requires a more detailed study.

However, DIAVAL is not only a Bayesian network, but a true Bayesian expert system, endowed with a flexible graphical *interface* which allows the user to introduce information orderly, according to a mixed-initiative approach. The *diagnosis* offered by the system is not limited to showing the probability of every variable, but it selects the diagnoses in accordance with two user-adjustable thresholds, standing for certainty and relevance. It also incorporates a rudimentary *explanation* capability that distinguishes six different kinds of links as an attempt to present information in the most natural way for a physician. Since the explanation of reasoning is one of the essential factors for the acceptance of the expert system by doctors and for the debugging of its knowledge base, we are currently investigating this topic. Finally, the composition of a written report that contains an ordered account of the echocardiogram is another attempt to facilitate the work of medical professionals.

In summary, although DIAVAL is still a prototype, it already includes novel features as an expert system and as a Bayesian network.

6 Acknowledgments

We thank Drs. Río Aguilar, Antonio Baño, Enrique González, Javier Jiménez, Mónzer Khanji, Íñigo Lozano, Javier Lozano, Gregorio Rodríguez and Sonia Rodríguez for their assistance in the design and evaluation of DIAVAL. We also thank Marisa Alonso, a nurse at the Section of Echocardiography of the Hospital de la Princesa, in Madrid, for her invaluable support.

References

[1] J. B. Adams. Probabilistic reasoning and certainty factors. In B. G. Buchanan and E. H. Shortliffe, editors, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, chapter 12, pages 263–271. Addison-Wesley, Reading, MA, 1984.

[2] S. Andreasen, F. V. Jensen, and K. G. Olesen. Medical expert systems based on causal probabilistic networks. *International Journal of Biomedical Computing*, 28:1–30, 1991.

[3] D. C. Berry and A. E. Hart. Evaluating expert systems. *Expert Systems*, 7:199–208, 1990.

[4] E. Braunwald, editor. *Heart Disease*. Saunders & Co., Philadelphia, fifth edition, 1996.

[5] B. G. Buchanan and E. H. Shortliffe, editors. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA, 1984.

[6] G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[7] A. Darwiche. Conditioning algorithms for exact and approximate inference in causal networks. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 99–107, Montreal, 1995. Morgan Kaufmann, San Francisco, CA.

[8] F. J. Díez. Parameter adjustement in Bayes networks. The generalized noisy OR–gate. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 99–105, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[9] F. J. Díez. *Sistema Experto Bayesiano para Ecocardiografía*. PhD thesis, Dpto. Informática y Automática, UNED, Madrid, 1994. In Spanish.

[10] F. J. Díez. Local conditioning in Bayesian networks. *Artificial Intelligence*, 87:1–20, 1996.

[11] F. J. Díez and J. Mira. Distributed reasoning and learning in Bayesian expert systems. In C. A. Ntuen, editor, *Advances in Fault-Diagnosis Problem Solving*. CRC Press, Boca Raton, FL. To appear.

[12] F. J. Díez and J. Mira. Distributed inference in Bayesian networks. *Cybernetics and Systems*, 25:39–61, 1994.

[13] M. J. Druzdzel and M. Henrion. Using scenarios to explain probabilistic inference. In *Proceedings of the AAAI-90 Workshop on Explanation*, pages 133–141, Boston, MA, 1990.

[14] D. Heckerman. Probabilistic interpretations for MYCIN's certainty factors. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 167–196. Elsevier Science Publishers, Amsterdam, 1986.

[15] D. Heckerman. An empirical comparison of three inference methods. In R. D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 283–302. Elsevier Science Publishers, Amsterdam, 1990.

[16] D. Heckerman. Causal independence for knowledge acquisition and inference. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 122–127, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[17] D. E. Heckerman and E. J. Horvitz. On the expresiveness of rule-based systems for reasoning with uncertainty. In *Proceedings of the 6th National Conference on AI (AAAI-87)*, pages 121–126, Seattle, WA, 1987.

[18] D. E. Heckerman and E. J. Horvitz. The myth of modularity in rule-based systems for reasoning with uncertainty. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 23–34. Elsevier Science Publishers, Amsterdam, 1988.

[19] D. E. Heckerman, E. J. Horvitz, and B. N. Nathwani. Toward normative expert systems: Part I — The Pathfinder Project. *Methods of Information in Medicine*, 31:90–105, 1992.

[20] D. E. Heckerman and B. N. Nathwani. Toward normative expert systems: Part II — Probability-based representations for efficient knowledge acquisition and inference. *Methods of Information in Medicine*, 31:106–116, 1992.

[21] M. Henrion. Some practical issues in constructing belief networks. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 161–173. Elsevier Science Publishers, Amsterdam, 1989.

[22] M. Henrion and D. R. Cooley. An experimental comparison of knowledge engineering for expert systems and for decision analysis. In *Proceedings of the 6th National Conference on AI (AAAI-87)*, pages 471–476, Seattle, WA, 1987.

[23] F. V. Jensen, K. G. Olesen, and S. K. Andersen. An algebra of Bayesian belief universes for knowledge-based systems. *Networks*, 20:637–660, 1990.

[24] J. Kalagnanam and M. Henrion. A comparison of decision analysis and expert rules for sequential diagnosis. In R. D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 271–281. Elsevier Science Publishers, Amsterdam, 1990.

[25] J. H. Kim and J. Pearl. A computational model for combined causal and diagnostic reasoning in inference systems. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI-83)*, pages 190–193, Karlsruhe, Germany, 1983.

[26] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50:157–224, 1988.

[27] R. S. Ledley and L. B. Lusted. Reasoning foundations of medical diagnosis. *Science*, 130:9–21, 1959.

[28] Y. C. Li. *Automated Probabilistic Transformation of a Large Medical Diagnostic Support System*. PhD thesis, Dept. of Medical Informatics, School of Medicine, University of Utah, 1995.

[29] W. J. Long, S. Naimi, and M. G. Criscitiello. Development of a knowledge base for diagnostic reasoning in cardiology. *Computers in Biomedical Research*, 25:292–311, 1992.

[30] W. J. Long, S. Naimi, and M. G. Criscitiello. Evaluation of a new method for cardiovascular reasoning. *Journal of the American Medical Informatics Association*, 1:127–141, 1994.

[31] D. Marr. *Vision*. Freeman, New York, 1982.

[32] M. J. Monaghan. *Practical Echocardiography and Doppler*. John Wiley & Sons, Chichester, U.K., 1990.

[33] R. E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. Wiley-Interscience, New York, 1990.

[34] J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

[35] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, CA, 1988. Revised second printing, 1991.

[36] Y. Peng and J. A. Reggia. Plausibility of diagnostic hypotheses. In *Proceedings of the 5th National Conference on AI (AAAI-86)*, pages 140–145, Philadelphia, 1986.

[37] M. A. Peot and R. D. Shachter. Fusion and propagation with multiple observations in belief networks. *Artificial Intelligence*, 48:299–318, 1991.

[38] M. Pradham, G. Provan, B. Middleton, and M. Henrion. Knowledge engineering for large belief networks. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, pages 484–490, Seattle, WA, 1994. Morgan Kaufmann, San Francisco, CA.

[39] M. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. *Methods of Information in Medicine*, 30:241–255, 1991.

[40] S. Srinivas. A generalization of the noisy-OR model. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 208–215, Washington D.C., 1993. Morgan Kaufmann, San Mateo, CA.

[41] D. A. Waterman. *A Guide to Expert Systems.* Addison-Wesley, Reading, MA, 1986.

[42] B. P. Wise and M. Henrion. A framework for comparing uncertain inference systems to probability. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 69–83. Elsevier Science Publishers, Amsterdam, 1986.