Research on medical decision analysis at the CISIAD, UNED

Francisco Javier Díez Manuel Luque Jorge Pérez-Martín Manuel Arias
Dept. Artificial Intelligence
Universidad Nacional de Educación a Distancia (UNED)
Madrid, Spain
{fjdiez, mluque, jperezmartin, marias}@dia.uned.es

***Abstract*—The Research Centre for Intelligent Decision-Support Systems (CISIAD) has been doing research on probabilistic graphical models applied to medicine for almost three decades. In this paper we summarise the contributions we have made, analyse the main difficulties we have found, and present the main failures and successes we have had in those years.**

***Index Terms*—artificial intelligence, probabilistic graphical model, Bayesian network, influence diagram, Markov model, medical decision making, cost-effectiveness analysis**

I. Introduction

One of the features of artificial intelligent systems is the ability to draw conclusions in uncertain domains. In medicine uncertainty is ubiquitous, mainly due to limited knowledge about the causal mechanisms and to the non-determinism of the real world. For this reason medicine has been one of the main fields of application since the beginning of artificial intelligence. The first methods for reasoning with uncertainty were based on the theory of probability, more specifically on what we now call *the naïve Bayes method*, which relies on two assumptions: diseases are mutually exclusive and findings are conditionally independent given the diagnosis. With these simplifying hypotheses is was possible to build several models that succeeded in solving several diagnostic problems in the 1960's and 1970's [1]–[6]. However, the assumptions required by this method are usually unrealistic in practice, which led many researchers to assert that probabilistic methods could not be used to solve large AI problems—see [7], [8] for a discussion.

The situation changed significantly in the next decade with the advent of probabilistic graphical models (PGMs). Howard and Matheson, two economists of the Stanford Research Institute (SRI) developed influence diagrams as a compact representation of decision problems, alternative to decision trees [9], and Judea Pearl, an artificial intelligence researcher at UCLA, developed Bayesian networks as an extension of the naïve Bayes [10], [11]. Very soon other authors proposed efficient algorithms for the evaluation of influence diagrams [12], [13] and Bayesian networks [14]. The first PGMs for real-world medical problems were developed in the next years [15]–[18] and the number of applications has grown so fast afterwards that now it is impossible to have a registry of all the medical applications that use PGMs.

This work has been supported by grant TIN2016-77206-R of the Spanish Government, co-financed by the European Regional Development Fund. J.P. received a predoctoral grant from the Spanish Ministry of Education.

In this paper we review some of the applications developed at the Centre for Intelligent Decision-Support Systems (CISIAD) of the National University for Distance Education (UNED), in Madrid, Spain, summarise the contributions we have made, analyse the main difficulties we have found, and present the main failures and successes we have had in almost three decades of research.

II. Probabilistic models for medical problems

A. PGMs for medical diagnosis

In 1989 Javier Díez began a doctoral thesis in artificial intelligence for medicine under the supervision of Prof. José Mira at UNED. The topic was the construction of an expert system for echocardiography, in collaboration with some doctors of the Hospital de la Princesa, in Madrid. In those years most expert systems were built using rules, and fuzzy logic was more and more popular. Prof. Mira had supervised several PhD theses that had applied these techniques to different medical problems [19]–[22], so they seemed to be the obvious choice for Díez's thesis. However, in the first knowledge elicitation sessions one of the doctors proposed building a causal network: mitral stenosis causes left atrium hypertension, which back-propagates to the lungs, and so on. However, when Díez tried to encode this causal model into a set of rules, it was impossible, because a piece of knowledge such as "A causes B" can be used either to infer B from A, or vice versa; but rule-based reasoning is unidirectional. Additionally, when A and B are causes of C and C is observed, the presence of A rules out B (this phenomenon is called *explaining away* [10]) and, conversely, discarding A increases the suspicion that B has caused C. Due to these limitations of rule-based reasoning and to his training as a physicist, Díez began exploring probabilistic reasoning for causal models, without being aware of the landmark contributions made by Pearl a few years earlier [10], [11], [23], [24]. He then rediscovered Bayesian networks, the noisy OR gate and its generalization to multivalued variables [25], which he called the noisy MAX [26], [27], and developed a new algorithm for evidence propagation [28]. In 1992 he spent three months at UCLA invited by Judea Pearl, and was able to catch up with the avant-garde of the research in this field, which was led by Pearl's group.

After finishing his doctorate in 1994 [29], Díez supervised several PhD theses that built Bayesian networks for several problems: Carmen Lacave [30], [31] built Prostanet for urology, Severino Fernández Galán [32] built Nasonet for nasopharyngeal cancer and Nuria Alonso Santander [33], an ophthalmologist of Hospital de la Princesa (Madrid), built Catarnet for cataract surgery.

B. PGMs for decision analysis

Because of our contact with health professionals, we realised that in medicine the final goal is not to issue a diagnosis, but to make decisions. In many cases finding out the disease with the higher probability or obtaining a list of variables whose posterior probability exceeds a certain threshold is not enough, because very often a low probability of a serious disease is more relevant that a high probability for a relatively-benign disease. In fact, newspapers from time to time tell the story of a patient who died after presenting at the urgency room of a hospital and being discharged because the doctors just gave the most likely diagnosis, without taking into account that the symptoms were compatible with an infrequent but mortal disease. Clearly, in medicine a false positive and a false negative have very different costs: the former usually leads to performing additional tests and sometimes starting an unnecessary treatment, which has an economic cost and may cause anxiety and discomfort to the patient, but a false negative may lead to delaying a treatment necessary to save the patient's life.

For this reason we were interested in building models that explicitly took into account the decisions and the cost of tests and treatments, including the cost of giving the wrong treatment or no treatment. This is how we started investigating influence diagrams (IDs), which differ from Bayesian networks in that they do not only have chance nodes, but also decision and utility nodes [9]. During his doctoral work, Manuel Luque built Mediastinet, an influence diagram for the mediastinal staging of non-small cell lung cancer, in collaboration with Dr. Carlos Disdier, of Hospital San Pedro de Alcántara (Cáceres) [34], [35], and Diego León built Arthronet, an influence diagram for total knee arthroplasty, during his master thesis, in collaboration with a doctor of Valladolid [36]. Every influence diagram is equivalent to a decision tree, but IDs have the advantage of being much more compact and, consequently, much easier to build and modify. In particular, both Mediastinet and Arthronet are equivalent to decision trees containing more than 10,000 leaves, and there is an algorithm that can transform each of them into the corresponding tree—provided that the computer has enough working memory.

In these models the main criterion was clinical effectiveness, measured in quality adjusted life years (QALYs), but they also represented the economic cost. In health economics the standard way of combining cost and effectiveness into a single criterion is to compute the net monetary benefit,

$$NMB = \lambda \cdot e - c \,,$$

or, alternatively, the net health benefit,

$$NHB = \lambda^{-1} \cdot NMB = e - \lambda^{-1} \cdot c \,,$$

where e is the effectiveness (measured in QALYs or in another clinical unit), c is the cost (in monetary units, such as dollars or euros) and λ, known as *willingness to pay* or *cost effectiveness threshold*, is a parameter that converts effectiveness into monetary units [37], [38]. This way it was possible to evaluate Mediastinet taking into account just the effectiveness (i.e., disregarding the cost, by making $\lambda^{-1} = 0$) or for specific values of λ. The problem is that λ is different for each decision maker. In Spain most health economists accept $\lambda = 30,000$ €/QALY as a value of consensus for our public health system, but there are also some experts claiming that this value is too high. Therefore, it would be desirable to perform a true cost-effectiveness analysis (CEA) in order to find out the values of λ (the thresholds) that determine the most beneficial intervention for each decision maker. Unfortunately, the algorithms available ten years ago were only able to perform CEA for decision trees containing just one decision node, at the root, and both Mediastinet and Arthronet had several decisions.

For this reason our group first developed a CEA algorithm for trees with several decisions [39] and then for IDs [40]. After implemented them in OpenMarkov, an open-source software tool that we describe below, it was possible to evaluate these IDs in a few seconds.

However, many medical problems involve asymmetries of several types. There is *order asymmetry* when the decisions are not totally ordered; for example, when it is not clear which test to do first, if any, and what tests to do afterwards depending on the result of previous tests. There may be *information asymmetry* due conditional observability; for example, the result of a test is know only when the doctor decides to perform it. And there is *domain asymmetry* when the value of one variable restricts the values of others; for example, when the decision is not to do a test, its result is neither positive nor negative. In IDs the second and third types of asymmetry can be represented—clumsily—by adding dummy states to some variables, but order asymmetry cannot be represented because IDs require a total ordering of the decisions. With the purpose of overcoming these limitations we proposed decision analysis networks (DANs) [41] and developed a CEA algorithm for them [42].

C. PGMs for temporal reasoning

In the same way as our collaboration with medical doctors led us from diagnosis (with Bayesian networks) to unicriterion decision analysis and then to cost-effectiveness analysis, it also showed us the importance of temporal reasoning. Our group had proposed two new types of temporal PGMs, namely networks of probabilistic events in discrete time (NPEDTs) [32] and dynamic limited memory influence diagrams (DLIM-IDs) [43]. The former were developed originally to model the spread of nasopharyngeal cancer [44] and the latter to predict the progression of carcinoid tumours [45]. For different reasons none of these types of networks could solve the typical problems in which health economists are currently interested. Therefore we extended our work on IDs to develop Markov IDs [46]. With them we have been able to conduct CEAs for several medical problems, such as pleural effusion [47], colorectal cancer [47], cochlear implantation [48], etc.

III. Development of open-source tools

These medical applications have been built with two open-source tools: Elvira and OpenMarkov. Elvira was the result of a collaboration of several Spanish universities, mainly Granada, Almería, Castilla-La Mancha, País Vasco and UNED [49]. The high number of developers, with experienced researchers, was the main reason for its rapid development, with the implementation of many algorithms for inference and learning, but, on the other hand, the physical distance between the teams involved in the project and the lack of adherence to the principles of software engineering made the tool difficult to maintain.

For this reason the UNED started the development of a new tool, called OpenMarkov [50]. We departed from the experience gained with Elvira, but the all the code was written from scratch. It is now a large tool, with around 115,000 lines of Java code, excluding blanks, and more than 200,000 lines in total, organised in 44 maven projects. The fact that it is managed by a single team and the use of several software engineering tools (git, maven, nexus, jenkins, etc.) has allowed us to make several redesign decisions and to maintain the code, which is still growing actively.

Both Elvira and OpenMarkov have advanced graphical user interfaces for editing and evaluating PGMs. Elvira has many learning algorithms, while OpenMarkov only implements the two basic algorithms, namely search-and-score and PC, but in general it is much more robust and more efficient in inference, and offers more types of networks (Markov IDs, DANs, etc.), more options for sensitivity analysis and temporal models, CEA, and the possibility of learning Bayesian networks interactively [51].

To our knowledge, Elvira was used in 10 countries, almost exclusively at universities, while OpenMarkov has been used for teaching, research and developing applications at universities, research centres and companies of more than 30 countries in Europe, America, Asia and Africa.

IV. Difficulties, failures and successes

In this section we analyse some of the difficulties we have found when building artificial intelligence applications for medicine, which range from technical challenges to human factors, and describe the main failures and successes we have faced.

A. Building PGMs with expert knowledge

Our group differs from most others in the field of PGMs in that, instead of investigating new learning algorithms, we have specialised in building PGMs with expert knowledge. This process is time consuming and, what makes it much more challenging, requires in general the collaboration of medical doctors. None of the health professionals who have collaborated with us has ever received any economic compensation for their work. Some of them have collaborated actively, but others had a low degree of commitment, to the point that it was difficult for us to arrange the meetings with them. For this reason, some of our attempts to build models for medical problems have failed after having investing a significant amount of time and effort.

B. Use of PGMs for clinical decisions

Clinical decision-support systems can be used in at least two ways. One of them is to guide the diagnosis and the treatment of individual patients at the clinical consultation or at the bedside. Many expert systems have been designed for this purpose, including our first PGMs. However, we do not know of any AI system routinely used this way. We were close to succeed with Catarnet, the above-mentioned Bayesian network for cataract surgery. The Health Department of the regional government of Madrid, who had financed the project, was interested in implanting this system into the new big hospitals it manages. We collaborated with the technicians of one of them to design a protocol for integrating Catarnet into their information system. When we were just about to start the tests, there was a change in the leadership of the hospital and the new person responsible refused to implant the decision-support system unless he could obtain from it some benefit for his professional/political career.

AI might also be applied to developing public health policies. However, these policies are based, in the best scenarios, on epidemiological studies, economic evaluations of health technologies and the consensus of experts; there seems to be no room for expert systems. However, in our group we have combined PGMs, an AI technique, with cost-effectiveness analysis, as mentioned above. One of the models we have built is a Markov ID for analysing the cost-effectiveness of paediatric bilateral cochlear implantation (BCI), i.e., for determining whether it is worth putting two implants instead of one to babies who are born with severe to profound deafness. The preliminary study we conducted, which included a thorough review of the literature, contributed to convincing the Ministry of Health that it is cost-effective, and Spain became the first country in the world—to the best of our knowledge—to include BCI for both children and adults in the portfolio of health services (cf. Orden SSI/1356/2015, de 2 de julio). In spite of this law, several regional governments still refuse to cover it in practice, even for newborns. We wrote a detailed report, based on our cost-effectiveness analysis [48], which proved beyond any reasonable doubt that this intervention is clearly cost-effective for children, and submitted it to the Ministry of Health and to 11 regional health departments. In May 2018 the Ministry of Health sent a letter to F. J. Díez in which it explicitly rectified its previous stance and confirmed that BCI must be covered by all health providers in Spain. More recently, Catalonia and Andalusia, two of the regions that had steadfastly refused covering it have announced that they will start putting two implants to the children that need them. This is the first time that our research on medical AI has had an impact on the life of patients.

We are working on two models for finding the optimal screening patterns for breast cancer and colorectal cancer. Even though there are several studies about these topics, we intend to develop new models and new algorithms for finding the optimal screening pattern for each patient based on his/her personal features. We will soon begin a CEA of screening for cytomegalovirus in newborns; if our study concludes that it is cost-effective, as some experts have recently claimed, health authorities should include it in the battery of tests for neonatal screening, which would have an impact on the life of many children and families.

C. A probabilistic expert system for programming cochlear implants

Our interest for cochlear implants led us to contact Dr. Paul Govaerts, who had been investigating the application of AI to programming them. This is a difficult task, because an implant has more than 100 electronic parameters that can be fitted. Improving the quality of hearing in one setting (for example, in a quiet room) may deteriorate the hearing in others (for example, in a noisy street). He had built a rule-based system that improved the performance of human audiologists, but the results were far from impressive. For this reason he started a new project, financed by a EU grant, aimed at building a new version of the expert system. A prestigious research group specialised in machine learning joined the project as a partner. However, some technical problems made it impossible to obtain the data they counted on, and even with them it would have been virtually impossible to build a model using learning algorithms, due to the complexity of the task. Seeing that the project had run aground, Dr. Govaerts contacted our group. The combination of his knowledge of audiology and cochlear implant technology with our expertise in building PGMs from human knowledge made it possible to create a probabilistic model based on a causal graph and subjective estimates of the probabilities [47], [52], [53]. In a few months it gave the first useful results and two years later impressed some experts in Europe and the USA for its performance. The main manufacturer of cochlear implants, who has a market share of more than 50%, has bought the rights to exploit it exclusively.

V. Conclusions and future work

Our group has been doing research on artificial intelligence applied to health decision making for almost three decades. We have contributed new algorithms [28], [39], [40], [42], [50], [54]–[61], new types of probabilistic graphical models (NPEDTs [32], DLIMIDs [43], [45], tuning networks [52], Markov IDs [46], DANs [41]), new canonical models [26], [27], [52] and several methods for the explanation of reasoning [62]–[66]. Each of them was motivated by a specific medical problem for which we were building a probabilistic network, but all of them can be applied to other domains. Similarly, the software tools we have developed [49], [50] are designed mainly for medicine, but other groups have used them to build applications in very different fields. These software tools have also been very useful for teaching PGMs to our students [67], [68].

Looking retrospectively, we can see that our efforts to build decision-support systems for clinical consultation have failed far from obtaining the benefits we expected. Building a probabilistic model manually takes a lot of time and requires the commitment of medical doctors, who in some cases collaborate enthusiastically but in others are poorly motivated and abandon the project far before arriving at the goal. Similarly, we have invested lots of time in developing software tools with advanced graphical user interfaces, in spite of our scarcity of funding and human resources. These tools have been very useful for our research and teaching, and also for many other universities in four continents. Several institutes and large companies of different countries have used OpenMarkov to build real-world applications. This has brought us the personal satisfaction of having offered the AI community a useful tool, but so far we have not obtained any economic return from it, and in the academic world, governed by the "publish or perish" principle, it is a risk to devote much time to tasks that yield poor results in terms of journal papers. Sometimes we ask ourselves if we made a mistake by following these lines of research instead of working on other areas, such as machine learning, in which the productivity is much higher.

Nonetheless, our research has also brought us other rewards. We have been pioneers in the application of AI to cost-effectiveness analysis, which is more and more relevant for medical decision making. Our economic study of cochlear implantation has contributed to convincing the Spanish health authorities that profoundly deaf people should receive two cochlear implants instead of one, especially in the case of children. Our experience in building probabilistic models from human knowledge and our software tool, OpenMarkov, had been essential in the construction of an expert system that is routinely used for programming cochlear implants; given that there are hundreds of thousands of cochlear implant users in the world, we are happy to know that our work will contribute to improving the quality of life of so many people. This tool is superior in several aspects to the commercial products developed for this task—and also inferior in others, clearly—and even though it is open-source, there are several possibilities of obtaining monetary returns from it: distributing it under dual-licensing, offering consultancy (mainly to pharmaceutical companies and manufacturers of medical devices), doing under-contract developments, etc. We are currently exploring these possibilities in order to obtain financial resources for our research activity.

Acknowledgements

We thank our colleagues and friends who collaborated in the Elvira project, from whom we learnt so much, and all the students and former members of our group who have collaborated in the development of OpenMarkov. We also thank the reviewers of the First Workshop of Spanish AI Research Groups in Biomedicine, at the CAEPIA-2018 conference, for their corrections and suggestions.

References

[1] F. T. de Dombal, J. R. Leaper, J. R. Staniland, A. McCann, and J. Horrocks, "Computer-aided diagnosis of acute abdominal pain," *British Medical Journal*, vol. 2, pp. 9–13, 1972.

[2] G. A. Gorry, "Computer-assisted clinical decision making," *Methods of Information in Medicine*, vol. 12, pp. 45–51, 1973.

[3] G. A. Gorry and G. O. Barnett, "Experience with a model of sequential diagnosis," *Computers and Biomedical Research*, vol. 1, pp. 490–507, 1968.

[4] J. Habbema, "Models for diagnosis and detection of combinations of diseases," in *Decision Making and Medical Care*, F. de Dombal and F. Gremy, Eds. New York: North-Holland, 1976, pp. 399–411.

[5] H. R. Warner, A. F. Toronto, and L. G. Veasy, "Experience with Bayes' theorem for computer diagnosis of congenital heart disease," *Annals of the New York Academy of Sciences*, vol. 115, pp. 558–567, 1964.

[6] H. R. Warner, A. F. Toronto, L. G. Veasy, and R. Stephenson, "A mathematical approach to medical diagnosis: Application to congenital heart disease," *Journal of the American Medical Association*, vol. 177, pp. 177–183, 1961.

[7] P. Szolovits, Ed., *Artificial Intelligence in Medicine*. Boulder, CO: Westview Press, 1982.

[8] P. Szolovits and S. G. Pauker, "Categorical and probabilistic reasoning in medicine," *Artificial Intelligence*, vol. 11, pp. 115–144, 1978.

[9] R. A. Howard and J. E. Matheson, "Influence diagrams," in *Readings on the Principles and Applications of Decision Analysis*, R. A. Howard and J. E. Matheson, Eds. Menlo Park, CA: Strategic Decisions Group, 1984, pp. 719–762.

[10] J. Pearl, "Fusion, propagation and structuring in belief networks," *Artificial Intelligence*, vol. 29, pp. 241–288, 1986.

[11] ——, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[12] S. M. Olmsted, "On representing and solving decision problems," Ph.D. dissertation, Dept. Engineering-Economic Systems, Stanford University, CA, 1983.

[13] R. D. Shachter, "Evaluating influence diagrams," *Operations Research*, vol. 34, pp. 871–882, 1986.

[14] 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*, vol. 50, pp. 157–224, 1988.

[15] S. Andreassen, M. Woldby, B. Falck, and S. K. Andersen, "MUNIN — a causal probabilistic network for interpretation of electromyographic findings," in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI'87)*, J. P. McDermott, Ed. Milan, Italy: Morgan Kaufmann, 1987, pp. 366–372.

[16] R. F. Nease and K. D. K. Owens, "Use of influence diagrams to structure medical decisions," *Medical Decision Making*, vol. 17, pp. 263–275, 1997.

[17] K. G. Olesen, U. Kjærulff, F. Jensen, F. V. Jensen, B. Falck, S. Andreassen, and S. K. Andersen, "A MUNIN network for the median nerve. a case study on loops," *Applied Artificial Intelligence*, vol. 3, pp. 385–403, 1989.

[18] D. K. Owens, R. D. Shachter, and R. F. Nease, "Representation and analysis of medical decision problems with influence diagrams," *Medical Decision Making*, vol. 17, pp. 241–262, 1997.

[19] R. Marín, "Un sistema experto para el diagnóstico y tratamiento anteparto del estado materno-fetal," Ph.D. dissertation, Universidad de Santiago, Santiago de Compostela, 1987.

[20] R. Marín, M. Taboada, J. Mira, A. Barreiro, and R. P. Otero, "Design and integration of a graphic interface for an expert system in oncology," *International Journal of Biomedical Computing*, vol. 33, pp. 25–43, 1993.

[21] R. P. Otero, "Medtool: Una herramienta para el desarrollo de sistemas expertos," Ph.D. dissertation, Universidad de Santiago, Santiago de Compostela, 1991.

[22] R. P. Otero and J. Mira, "Medtool: A teachable medical expert system development tool," in *Proceedings of the Third International Symposium on Knowledge Engineering*, Madrid, Spain, 1988, pp. 191–200.

[23] J. Pearl, "Reverend Bayes on inference engines: A distributed hierarchical approach," in *Proceedings of the Second National Conference on Artificial Intelligence (AAAI'82)*, D. L. Waltz, Ed. Pittsburgh, PA: AAAI Press, 1982, pp. 133–136.

[24] J. H. Kim and J. Pearl, "A computational model for combined causal and diagnostic reasoning in inference systems," in *Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI'83)*, A. Bundy, Ed. Karlsruhe, Germany: William Kauffmann, 1983, pp. 190–193.

[25] M. Henrion, "Some practical issues in constructing belief networks," in *Uncertainty in Artificial Intelligence 5 (UAI'89)*, M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, Eds. Amsterdam, The Netherlands: Elsevier Science Publishers, 1989, pp. 161–173.

[26] F. J. Díez, "Parameter adjustment in Bayes networks. The generalized noisy OR–gate," in *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence (UAI'93)*, D. Heckermann and E. Mamdani, Eds. San Francisco, CA: Morgan Kaufmann, 1993, pp. 99–105.

[27] F. J. Díez and M. J. Druzdzel, "Canonical probabilistic models for knowledge engineering," UNED, Madrid, Spain, Technical Report CISIAD-06-01, 2006.

[28] F. J. Díez, "Local conditioning in Bayesian networks," *Artificial Intelligence*, vol. 87, pp. 1–20, 1996.

[29] ——, "Sistema experto bayesiano para ecocardiografía," Ph.D. dissertation, Dpto. Informática y Automática, UNED, Madrid, 1994, in Spanish.

[30] C. Lacave and F. J. Díez, "Knowledge acquisition in Prostanet, a Bayesian network for diagnosing prostate cancer," *Lecture Notes in Computer Science*, vol. 2774, pp. 1345–1350, 2003.

[31] C. Lacave, "Explicación en redes bayesianas causales. aplicaciones médicas," Ph.D. dissertation, Dept. Inteligencia Artificial. UNED, Madrid, 2003.

[32] S. F. Galán and F. J. Díez, "Networks of probabilistic events in discrete time," *International Journal of Approximate Reasoning*, vol. 30, pp. 181–202, 2002.

[33] N. Alonso-Santander, F. J. Díez, N. F. de Larrea, C. Margalef, and M. Arias, "A probabilistic decision-support system for cataract surgery," in *XXVIII Congress of the ESCRS (European Society of Cataract and Refractive Surgery)*, Paris, France, 2010.

[34] M. Luque, "Probabilistic graphical models for decision making in medicine," Ph.D. dissertation, UNED, Madrid, 2009.

[35] M. Luque, F. J. Díez, and C. Disdier, "Optimal sequence of tests for the mediastinal staging of non-small cell lung cancer," *BMC Medical Informatics and Decision Making*, vol. 16, pp. 1–14, 2016.

[36] D. León, "A probabilistic graphical model for total knee arthroplasty," Master's thesis, Dept. Artificial Intelligence, UNED, Madrid, Spain, 2011.

[37] M. F. Drummond, M. J. Sculpher, K. Claxton, G. L. Stoddart, and G. Torrance, *Methods for the Economic Evaluation of Health Care Programmes*, cuarta ed. Oxford, UK: Oxford University Press, 2015.

[38] A. A. Stinnett and J. Mullahy, "Net health benefit: A new framework for the analysis of uncertainty in cost-effectiveness analysis," *Medical Decision Making*, vol. 18, pp. S68–S80, 1998.

[39] M. Arias and F. J. Díez, "Cost-effectiveness analysis with sequential decisions," UNED, Madrid, Spain, Technical Report CISIAD-11-01, 2011.

[40] ——, "Cost-effectiveness analysis with influence diagrams," *Methods of Information in Medicine*, vol. 54, pp. 353–358, 2015.

[41] F. J. Díez, M. Luque, and I. Bermejo, "Decision analysis networks," *International Journal of Approximate Reasoning*, vol. 96, pp. 1–17, 2018.

[42] M. Arias, M. Luque, J. Pérez-Martín, and F. J. Díez, "Cost-effectiveness analysis with decision analysis networks," in *Annual Meeting of the Society for Medical Decision Making*, Pittsburgh, PA, 2017.

[43] F. J. Díez and M. A. J. van Gerven, "Dynamic LIMIDs," in *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions*, L. E. Sucar, J. Hoey, and E. Morales, Eds. Hershey, PA: IGI Global, 2011, pp. 164–189.

[44] S. F. Galán, F. Aguado, F. J. Díez, and J. Mira, "NasoNet. Modelling the spread of nasopharyngeal cancer with temporal Bayesian networks," *Artificial Intelligence in Medicine*, vol. 25, pp. 247–254, 2002.

[45] M. A. J. van Gerven, F. J. Díez, B. G. Taal, and P. J. F. Lucas, "Selecting treatment strategies with dynamic limited-memory influence diagrams," *Artificial Intelligence in Medicine*, vol. 40, pp. 171–186, 2007.

[46] F. J. Díez, M. Yebra, I. Bermejo, M. A. Palacios-Alonso, M. Arias, M. Luque, and J. Pérez-Martín, "Markov influence diagrams: A graph-

ical tool for cost-effectiveness analysis," *Medical Decision Making*, vol. 37, pp. 183–195, 2017.

[47] I. Bermejo, "New types of probabilistic graphical models: Applications to medicine," Ph.D. dissertation, Dept. Artificial Intelligence, UNED, Madrid, Spain, 2015.

[48] J. Pérez-Martín, M. A. Artaso, and F. J. Díez, "Cost-effectiveness of pediatric bilateral cochlear implantation in Spain," *The Laryngoscope*, 2017. [Online]. Available: http://dx.doi.org/10.1002/lary.26765

[49] Elvira Consortium, "Elvira: An environment for creating and using probabilistic graphical models," in *Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM'02)*, J. A. Gámez and A. Salmerón, Eds., Cuenca, Spain, 2002, pp. 1–11.

[50] M. Arias, M. A. Artaso, I. Bermejo, F. J. Díez, M. Luque, and J. Pérez-Martín, "Advanced algorithms for medical decision analysis. Implementation in OpenMarkov," in *Proceedings of the 16th Conference on Artificial Intelligence in Medicine (AIME 2017)*, Vienna, Austria, 2017.

[51] I. Bermejo, J. Oliva, F. J. Díez, and M. Arias, "Interactive learning of Bayesian networks with OpenMarkov," pp. 27–34.

[52] I. Bermejo, "Probabilistic graphical models for the tuning of systems," Master's thesis, Dept. Artificial Intelligence, UNED, Madrid, Spain, 2012.

[53] I. Bermejo, F. J. Díez, P. Govaerts, and B. Vaerenberg, "A probabilistic graphical model for tuning cochlear implants," in *Artificial Intelligence in Medicine*, N. Peek, R. Marín, and M. Peleg, Eds. Springer Berlin Heidelberg, 2013, vol. 7885, pp. 150–155.

[54] F. J. Díez and S. F. Galán, "Efficient computation for the noisy MAX," *International Journal of Intelligent Systems*, vol. 18, pp. 165–177, 2003.

[55] M. Luque and F. J. Díez, "Variable elimination for influence diagrams with super-value nodes," in *Proceedings of the Second European Workshop on Probabilistic Graphical Models (PGM'04)*, P. Lucas, Ed., Leiden, The Netherlands, 2004, pp. 145–152.

[56] M. Arias and F. J. Díez, "Operating with potentials of discrete variables," *International Journal of Approximate Reasoning*, vol. 46, pp. 166 – 187, 2007.

[57] ——, "Computation of cost and effectiveness in decision trees with embedded decision nodes," in *Annual Meeting of the Society for Medical Decision Making*, Pittsburgh, 2007.

[58] M. Luque, T. D. Nielsen, and F. V. Jensen, "An anytime algorithm for evaluating unconstrained influence diagrams," in *Proceedings of the Fourth European Workshop on Probabilistic Graphical Models (PGM'08)*, M. Jaeger and T. D. Nielsen, Eds., Hirtshals, Denmark, 2008, pp. 177–184.

[59] M. Luque and F. J. Díez, "Variable elimination for influence diagrams with super-value nodes," *International Journal of Approximate Reasoning*, vol. 51, pp. 615–631, 2010.

[60] M. Luque, T. D. Nielsen, and F. V. Jensen, "Anytime decision-making based on unconstrained influence diagrams," *International Journal of Intelligent Systems*, vol. 31, pp. 379–398, 2016.

[61] M. Luque, M. Arias, and F. J. Díez, "Synthesis of strategies in influence diagrams," in *Proceedings of the Thirty-third Conference on Uncertainty in Artificial Intelligence (UAI'17)*. Corvallis, OR: AUAI Press, 2017, pp. 1–9.

[62] C. Lacave and F. J. Díez, "A review of explanation methods for Bayesian networks," *Knowledge Engineering Review*, vol. 17, pp. 107–127, 2002.

[63] C. Lacave, A. Oniśko, and F. J. Díez, "Use of Elvira's explanation facilities for debugging probabilistic expert systems," *Knowledge-Based Systems*, vol. 19, pp. 730–738, 2006.

[64] M. Luque and F. J. Díez, "Decision analysis with influence diagrams using Elvira's explanation capabilities," in *Proceedings of the Third European Workshop on Probabilistic Graphical Models (PGM'06)*, M. Studený and J. Vomlel, Eds., Prague, Czech Republic, 2006, pp. 179–186.

[65] C. Lacave, M. Luque, and F. J. Díez, "Explanation of Bayesian networks and influence diagrams in Elvira," *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics*, vol. 37, pp. 952–965, 2007.

[66] F. Elizalde, L. E. Sucar, M. Luque, F. J. Díez, and A. Reyes, "Policy explanation in factored Markov decision processes," in *Proceedings of the Fourth European Workshop on Probabilistic Graphical Models (PGM'08)*, M. Jaeger and T. D. Nielsen, Eds., Hirtshals, Denmark, 2008, pp. 97–104.

[67] F. J. Díez, "Teaching probabilistic medical reasoning with the Elvira software," in *IMIA Yearbook of Medical Informatics*, R. Haux and C. Kulikowski, Eds. Sttutgart, Germany: Schattauer, 2004, pp. 175–180.

[68] F. J. Díez, I. París, J. Pérez-Martín, and M. Arias, "Teaching Bayesian networks with OpenMarkov," in *9th International Conference on Probabilistic Graphical Models*, Prague, Czech Republic, 2018.

[69] M. Jaeger and T. D. Nielsen, Eds., *Proceedings of the Fourth European Workshop on Probabilistic Graphical Models (PGM'08)*, Hirtshals, Denmark, 2008.