Diagnostic Test Suggestion for Emergency Room via Bayesian Network of Non Expert-assisted Knowledge Base

Jorge Quinteros*, Nelson Baloian*, Jose A. Pino*, Álvaro Riquelme*,
Sergio Peñafiel*, Horacio Sanson**, Douglas Teoh**
*Department of Computer Science, Universidad de Chile, Santiago, Chile
**Allm Inc., Tokyo, Japan
jorge.quinteros@ug.uchile.cl, **nbaloian@dcc.uchile.cl**, **jpino@dcc.uchile.cl**, **alvaroriquelme@uchile.cl**,
spenafie@dcc.uchile.cl, **horacio@allm.net**, **d.teoh@allm.net**

***Abstract*—The Japanese public health system relies upon a mandatory insurance scheme that subsidizes every medical procedure. This causes some practitioners in doubt to order unnecessary exams, especially in departments like the emergency room (ER) (where time and personnel constraints apply), generating additional costs for the public health system. In this context arises the need and challenge of developing a computer application based on Artificial Intelligence that, given a patient's symptoms upon entering the ER, recommends the most appropriate exams to increase the accuracy of the diagnosis. This paper presents the preliminary results on the development of such a tool using a Bayesian Network (BN). Although there is a lot of literature on BN for medical diagnosis, this work is innovative as it is focused on suggesting useful exams based on pre-test probabilities, and that it was built using only medical data and other freely available information sources. A fundamental disease list was established using a Human Symptom-Disease Network (HSDN) containing symptom-disease relationships. The co-occurrence between disease and symptom terms on the HSDN was translated into rough sensitivity and specificity estimates and used to set the conditional probabilities of the BN. Prior probabilities of diseases were estimated using hospital data of regular and emergency visits. Information about findings (exams) and their sensitivity-specificity data was scraped from web databases and mapped into the network. Preliminary tests for inspecting the accuracy of the developed tool were made with the help of a medical expert, based on relevant literature. Obtained results show that the tool is able to find differential diagnoses for most cases. This work opens the door for future improvements of the system.**

***Keywords*—Exam Recommendation, Medical cause, Bayesian Networks, CDSS, e-Health**

I. INTRODUCTION

The Japanese public health system requires every resident in the country to enroll in one of the two main types of insurance programs: the National Health Insurance and the Employees' Health Insurance [1]. Both cover a great percentage of almost every medical procedure including consultations, examinations and medicines. In particular, these insurances cover great part of the costs of the often very expensive diagnostic tests.

The Emergency Room (ER) of a hospital is often an overcrowded place, where medical procedures are tied to time and personnel constraints [2], [3]. Practitioners have to make decisions towards diagnosing as fast as possible based on patients' *findings*: acute symptoms, medical history, physical examinations and laboratory/radiology tests [4]. Due to lack of specific knowledge and/or fear of making medical mistakes, practitioners sometimes order exams that may not be necessary to reach a precise diagnostic, which translates into heavy costs for the public health system. This is a recurrent problem in many countries that decreases the quality of health care, increase its cost and even exposes patients to secondary effects of unnecessary exams [5].

To reach a diagnosis practitioners need to explore symptoms and signs of a patient, on what is called a *clinical examination*. In most cases this procedure is enough to reach a right diagnosis but sometimes it only brings a spectrum of possible diagnoses, which is known as *differential diagnosis*. Diagnostic tests are needed then to achieve a more specific diagnosis within all the possibilities given by the differential diagnosis.

It is in this scenario where a software tool based on Artificial Intelligence could be of great help: it may be able to suggest practitioners to perform certain diagnostic tests based on symptoms and vital signs obtained from a clinical examination, and whose results could help to clarify the diagnosis, reducing the number of possibilities in a differential diagnosis. The combination of a patient's information taken on arrival to the ER along with pre-test probabilities could enable a computer-based system to suggest the most useful diagnostic tests for that specific scenario.

The approach taken in this work is common in the literature of Clinical Decision Support Systems (CDSS): *Bayesian networks*. The baseline design is typical for diagnosis-support models: a set of findings which can be present or not due to the manifestation of different diseases is established and used to calculate the most probable ones via multiple instances of the Bayes' theorem.

Although there is a lot of work done using Bayesian networks for medical diagnosis, not much has been done with a focus on exam recommendation. In addition, most of the research on Bayesian networks as CDSS has been built along medical experts, thus incorporating their knowledge directly. The work presented here is a contribution in the sense that it focuses on suggesting useful examinations based on a pre-test probability and that it was built only using medical data and other sources of freely available information, mostly taken from the Web.

II. BACKGROUND

Some fundamental concepts used in medicine research are closely related-or can be directly applied—to statistical inference. It is partly this "algorithmic" view of medical processes along with the idea that it is possible to automate them that has motivated so much research in this area, especially within Clinical Decision Support Systems [6].

Some basic medical and expert system terminology is introduced here to help clarify these links and then used to contextualize both, old and new works in this area.

A. *Important Concepts*

According to Wyatt et al. (1991) Clinical Decision Support Systems are "active knowledge systems which use two or more items of patient data to generate case-specific advice" [7]. The idea behind these kind of systems aimed at clinical support is old but this definition is still relevant nowadays.

As stated earlier, a Bayesian network was the chosen approach for this problem. A Bayesian network [8] is an extensively applied probabilistic modeling tool, which is especially used in domains involving uncertainty. BNs have been used in a wide range of areas like information retrieval [9], risk analysis [10], sports betting [11], hardware and system diagnostics [12] and many more [13].

According to the definition presented by Zagorecki et al. (2013) [14] a Bayesian network is a probabilistic graphical model that represents a set of random variables (nodes) and their conditional dependencies (arcs) via a directed acyclic graph. Nodes with no incoming arcs are associated with simple probability distributions called prior probabilities, while nodes with incoming arcs or parents, have multiple conditional probabilities associated and stored in tables (CPTs). Figure 1 shows the values required to specify a simple network. The key feature of BNs is they represent a compact version of the joint probability distributions (JPD) of the random variables in the model. Following the independence of the variables implied from the model, one can reduce the size of CPTs by trimming the distributions required to specify each JPD. Thus, BNs allows for efficient answering of queries related to arbitrary conditional probabilities involving the variables, such as the probability of a disease given a set of findings, or the findings that maximize the probability of some disease.

For this work, a *finding* is defined as a symptom, the outcome of a laboratory examination, analysis of an image or any other kind of diagnostic test that can move probabilities closer or away from a certain diagnosis. Findings usually have two associated numerical values: *sensitivity*, defined as the likelihood of a person who is ill to have a positive test result, and *specificity*, defined as likelihood of a healthy person receiving a negative test result [15], [16]. These sensitivity and specificity concepts are widely studied in the field of evidence based medicine.

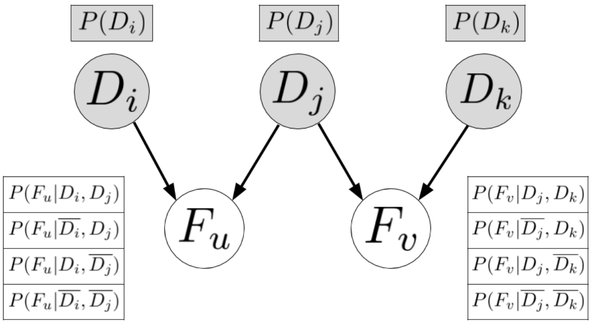


Fig. 1: A simple model of a BN with 3 diseases leading to 2 different findings. Values on top represent prior probabilities for each disease and at the side of each finding node the corresponding CPT which, in first instance, grows exponentially with the number of parent nodes.

Let D_i represent disease i and F_u represents finding u. Sensitivity is the probability that finding F_u appears when disease D_i is present, i.e.:

$$S_e := P(F_u|D_i).$$

Equivalently, specificity is the probability that finding u does not appear when disease i is not present, i.e.:

$$S_p := P(\overline{F_u}|\overline{D_i}).$$

Based on the definition of BN stated earlier, we need to establish these CPTs as prior probabilities of diseases (parent nodes) and the conditional probabilities of findings (child nodes) being manifested given that one or more diseases are present. We will see that these last elements can be simplified greatly if one assumes independence of diseases (the usual assumption in this kind of models), thus sensitivity and specificity information being the only required values.

B. *Related Work*

There is an extensive history of collaboration between Artificial Intelligence and Medical experts, dating as far as 1954 [17]. Among the most notorious projects in this collaborative area are IBM's Watson [18] as well as Quick Medical Reference (QMR) [19] and DxPlain [20] decision-support systems.

Knowledge-based expert systems, like the work of Sánchez et al. (1979) on fuzzy set theory [21] or others based on sets of rules [22] have historically dominated over systems that work only with data, like those based on Artificial Neural Networks [23] or Decision Trees [24]. This is mainly because the latter methods are more recent and also because medical data has always been more difficult to acquire, due to legal and ethical reasons, [25] and to manipulate [26].

With the recent boom of the so called *Big Data* field some new powerful machine learning algorithms have been developed. *Deep Learning* networks or *Random Forests* can attain good results in classification and regression problems, among others.

Although these algorithms can solve problems in a variety of domains, getting access to enough medical data of reasonable quality is still a difficult task. Thus, not many revolutionary solutions using these tools have been developed yet, especially in the area of general medical diagnosis, where the set of possible outcomes is not limited to the classification of 1 or 2 diseases, but hundreds. For this reason, the subject of diagnosis-support systems is still very linked to knowledge-based methods.

In particular, Bayesian networks appear as one of the most recurrent approach to these kind of problems: Quick Medical Reference's significant adaptation to Bayesian Networks (QMR-DT) [27] appeared only a few years after BNs networks were first introduced by Pearl in 1985 [28]. Also other works applying very similar foundation were developed in this decade [14]. The big challenge with this approach is setting up the knowledge base. Most of the systems based on BNs build their knowledge base with strong help of medical experts or are adaptations of other system's knowledge.

We focus on building a system for test recommendation whose knowledge base is gathered from medical data and freely available information, i.e., without input from medical experts.

III. METHODOLOGY

In order to build the recommendation system a *2-level* Bayesian network was designed, consisting of a bipartite graph where finding nodes have incoming arcs from one or more disease nodes. This kind of network is called B2NO [29] and is similar to the one shown in figure 1 but on a much larger scale.

Normally, one of the problems with BNs is that the size of CPTs (number of probabilities) grows exponentially with the number of parent nodes, which makes them impractical for domains requiring a large number of nodes. Additionally, let's say there are two diseases D_1 and D_2 related to finding F. Even though joint conditional probabilities (like $P(F|D_1, D_2)$ are required to specify its local table, the information that can be normally obtained from medical literature are, at most, sensitivity $P(F|D)$ and specificity $P(\overline{F}|\overline{D})$ values [30]. For this reason, actually specifying the 2^n probability values required for each node with n parents becomes too complex, so additional independence constraints and other techniques are applied.

A common approach to this problem in the case of BNs for diagnosis is the noisy-OR model [31]. An underlying assumption of this model is that different diseases act independently to produce certain findings. Under this model, it is the added effect of all diseases that combined can produce the cumulative probability of the finding. Nikovski (2000) explains this thoroughly and shows how this can be applied to prior, sensitivity and specificity values [30].

To establish the diseases the system is going to handle a Human Symptoms-Disease Network (HSDN) [32] was used as baseline knowledge source. This network contains thousands of symptoms-disease relationships obtained from lexical co-occurrence of MeSH terms [33] in medical publications. It contains around 4000 thousand disease terms linked to more than 300 hundred symptom terms, considering the MeSH vocabulary and including the number of co-occurrences along with their corresponding TF-IDF score.

Diseases with too few studies were filtered out, considering them to be either synonyms of others already in the list or simply extremely rare, thus unlikely to occur. Some symptom-disease links with too low a score in relation to others for the same disease were considered not representative and therefore ignored. These links were translated into arcs for the BN, being the symptoms the first set of findings.

As usual, diseases were modeled as binary variables drawn from a Bernoulli distribution over the prior probabilities. To obtain the parameters for these priors a large database of approximately 1 million medical records (of which about 190000 correspond to ER) from a hospital in Okayama, Japan was used. These records included patients' information, visit date, outcome disease name in Japanese and—most of the time—department and ICD-10 [34] codes. In order to map these diseases, the UMLS [35] service, which is a large database of medical terminology, along with Wikipedia's WikiData information and other mappings [36] were used.

This mapping was made by reading disease names (MeSH terms with associated MeSH codes) and selecting most of the various ICD-10 codes that could be found in any of the databases for each particular disease. Thus, a list of possible ICD-10 codes was attached to every disease in the BN. Diseases with no mapping in any of the databases were filtered out.

Using this, a frequency measure of the diseases contained in the medical records was obtained, first for all patients making medical visits to the hospital and then for those arriving at the ER only. Table I shows the top 15 most frequent diseases for both all medical visits and only the ones for the ER. As there are ICD-10 codes in the system matching more than one disease, it is possible to see that, for instance, all gastrointestinal diseases have the same frequency as a consequence. Most of these related diseases share the same diagnostic tests so the recommendation system is not greatly affected by this.

These values were directly translated into estimates of prior probabilities for each disease (under the assumption that this probability represents that of a patient having the given disease when entering the hospital and having no additional information) using the following hyperbolic tangent as smoothing function in order to reduce the gap between common and uncommon diseases:

$$P_s(x_i) = A \cdot tanh(Bx_i - C) + D \tag{1}$$

where x_i is the frequency value of disease i, and A, B, C and D are empirically adjusted constants.

As was mentioned, the HSDN is based on co-occurrence of disease and symptom terminology found in academic papers, i.e., a measure of how often the terms appear together in literature. This was used as a general measure of the symptom-disease relevance links and therefore translated into empirically estimated sensitivity values using the simple assignment:

N^o	Disease Name	Frequency
1	Diabetes Mellitus	0.03106
2	Gastroesophageal Reflux	0.01937
3	Esophagitis, Peptic	0.01932
4	Pneumonia	0.01865
5	Hypertension	0.01804
6	Rhinitis, Allergic, Seasonal	0.01560
7	Dehydration	0.01549
8	Spondylarthropathies	0.01525
9	Heart Failure	0.01511
10	Bronchitis, Chronic	0.01471
11	Back Pain	0.01379
12	Stomach Ulcer	0.01375
13	Bronchitis	0.01372
14	Respiratory Tract Infections	0.01264
15	Liver Diseases	0.01256
16	Intestinal Polyposis	0.01201
17	Low Back Pain	0.01142
18	Influenza, Human	0.01114
19	Cerebral Infarction	0.01111
20	Angina Pectoris	0.01081

(a) Top diseases counting all medical visits.

N^o	Disease Name	Frequency
1	Gastrointestinal Diseases	0.04168
2	Ileitis	0.04167
3	Colitis	0.04167
4	Inflammatory Bowel Diseases	0.04167
5	Enteritis	0.04167
6	Enterocolitis	0.04167
7	Gastroenteritis	0.04167
8	Influenza, Human	0.04131
9	Respiratory Tract Infections	0.03354
10	Pharyngitis	0.03127
11	Pharyngeal Diseases	0.03125
12	Craniocerebral Trauma	0.03055
13	Cerebral Hemorrhage	0.02707
14	Dehydration	0.02555
15	Brochitis, Chronic	0.02524
16	Brochitis	0.02511
17	Pneumonia	0.02326
18	Nasopharyngitis	0.01777
19	Common Cold	0.01777
20	Status Asthmaticus	0.01629

(b) Top diseases counting emergency room visits only

TABLE I: Tables showing the most frequent diagnoses in the data consisting of approximately 10^6 medical records. (a) considering all kinds of medical visit and (b) considering only emergency room visits.

$$S_e(\bar{S}_j) = \begin{cases} 0.9 & \text{if } \bar{S}_j \geq 0.5 \\ 0.8 & \text{if } 0.12 \leq \bar{S}_j < 0.5 \\ 0.65 & \text{else} \end{cases} \qquad (2)$$

where $\bar{S}_j$ is the TF-IDF score of symptom j normalized against the highest TF-IDF score for that disease. As for the specificity S_p, a default value representing the average was also empirically adjusted and used for all symptoms. This difference comes from the idea that, given the nature of this value, directly estimating them using only co-occurrence scores is not that simple. Both of these estimates then were used as conditional probabilities for the Bayesian network.

Now, for the other part of the findings (consisting mainly of diagnostic test results) sensitivity and specificity information for all diagnoses available from two different *web databases* were collected. These can be found in the www.sensitivityspecificity.com [37] and www.getthediagnosis.org [38] websites. Both websites are constructed with the cooperation of many medical experts who contribute to organize information and numbers that could be of use to the community. They get their specific information from medical publications and contain hundreds of findings with their respective diagnosis, including both sensitivity and specificity values. This information was matched against the disease list again using the UMLS service, manually checked and cleaned up, and finally integrated with symptoms to obtain a big network of diseases, symptoms and examination results.

With all this, the intended work-flow of the network is as follows:

First, a short list of symptoms both present and confirmed and not present is used as input. Next, the system generates a list of pre-test diagnoses along with a score indicating their probability. Thus, based on this list and for each disease, a list of findings that could increase the probability of a diagnosis is generated as the output.

In the next section results of an expert validation are presented.

IV. PRELIMINARY EXPERT VALIDATION

In order to test if the tool serves as a guide onto what examinations would better help reach a more accurate diagnosis, the input of a medical expert was used. The medical expert assembled a list of 14 different diagnoses, each one consisting of a set of up to six general symptoms a patient could present. This was considered fairly simple data that could be acquired during a preliminary medical examination and was taken from medical literature [39]. The 14 diagnoses chosen, which represent a good percentage of ER entries are the following:

1) Allergic Rhinitis
2) Laryngitis
3) Pneumonia
4) Pulmonary Embolism
5) Migraines
6) Cerebrovascular Accident (Stroke)
7) Myocardial Infarction
8) Gastroesophageal Reflux
9) Gastroenteritis
10) Urinary Tract Infections
11) Cholelithiasis
12) Appendicitis
13) Atrial Fibrillation
14) Meningitis

As was mentioned earlier, the system first builds a pre-test diagnoses list based on the estimated prevalence and the symptoms on the input. Of the formerly over 4000 diagnoses, a list of 20 possible diseases comes out as initial output, along with a score indicating the probability of each disease. Figure 2 shows, for each one of the diseases, if the correct diagnosis was present on the list generated by the system along with how many of the rest correspond to valid differential diagnoses

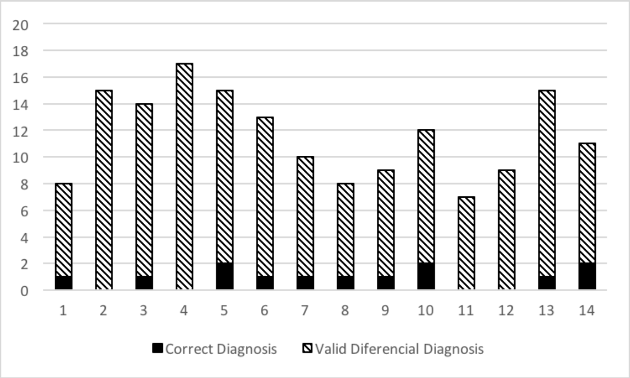


Fig. 2: For each one of the chosen diseases, the bottom pattern shows if it is found in the list of 20 possible diagnoses and the top pattern shows how many of the rest correspond to valid differential diagnoses.

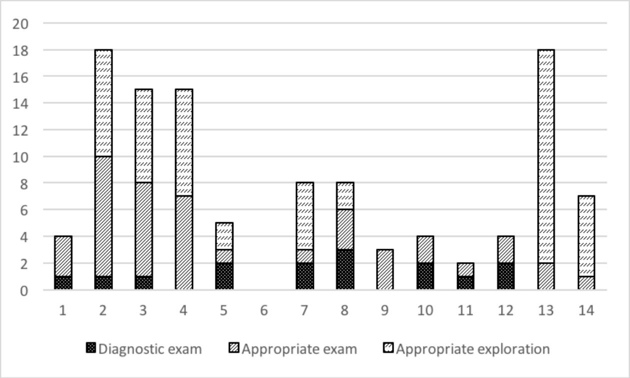


Fig. 3: For each one of the chosen diseases, the bottom pattern shows the gold standard or best exam test to diagnose the described disease, the middle pattern shows other useful tests and the top pattern represents physical examination findings that help to reach the diagnosis.

which should not be discarded until further examination is made.

Based on this pre-test possible diagnoses, a similar list of findings that could increase the probability of having a certain disease is elaborated. Given that there are additional findings for just a fraction of the total diseases, not every disease on the list outputs directly the examinations that could increase its probability. However, since there are many closely related diseases in the database that usually appear simultaneously on the list, some of the findings are useful for the diagnosis of more than one disease.

For this final output a similar evaluation was made. Figure 3 shows, for each one of the chosen diseases, if the *gold standard* (the best available diagnostic test [40]) is present on the finding list along with how many of the rest of the findings correspond to appropriate tests to reach the desired diagnoses and how many correspond to other types of findings (like results from a physical examination) that could also be of use.

These results are briefly discussed in the next section.

V. DISCUSSION AND CONCLUSIONS

Even though these are preliminary results, if we consider that the original list consists of about 4000 diseases it is possible to observe that in most cases specifying the presence of only a few symptoms leads to finding the specific diagnosis in the list of candidates. The fact that there are also many other diseases that should not be discarded on a first inspection (because they belong to the differential diagnosis) evidences that the symptom-disease links in the HSDN, after filtering and mapping their TF-IDF score to conditional probability values, are significant.

As for the actual exam recommendation process, given that it is based on the 20 pre-test most probable diseases, results show again that in most cases the diagnostic exams can be found in the list generated by the system. As for the rest of the findings, some were considered to be appropriate exams to reach the final diagnosis and some to be of another type of finding that could be of use (if *found*). The results here are much more disperse and are good for some diseases, like for Laryngitis, or very poor, like for Stroke. This is most probably because the findings obtained include only a fraction of the diseases in the total list, and some of the ones that are present still lack many useful associated findings.

These results indicate that it is possible to build an exam recommendation tool which could potentially improve greatly the efficiency and work-flow of the ER even without the input of medical experts. That being said, we believe that these results open the door to new ideas on how it can be improved:

First, a good filtering of diseases and symptoms could be made with the help of one or more medical experts to reduce the number of diseases that are the same in an emergency context and to be more specific with some symptoms that are too vague or simply missing (e.g. *Trauma* is missing and there are not many body part-specific symptoms). Second, the functions that map prevalence and conditional probability values from frequency and co-occurrence data, respectively, could also be somehow optimized to include other factors such as severity of a disease or uniqueness of a symptom related to some disease. Third, additional information sources of findings are known to be available and would be useful to integrate it to the knowledge source, which after cleaning and restructuring could become more complete and robust. Fourth, some other widely used mathematical assumptions, like *leak probabilities* [30] or the *soft approach evidence* model for disease restrictions and risk factors [29], have not been considered yet but are believed to be able to improve considerably the performance of BNs for decision support.

From a medical expert point of view, the results on the pre-test diagnoses show that the tool works quite well proposing differential diagnoses. Results on the proposed exams not only show the exam to be made but also the expected result on the exam. At this moment the system only pretends to suggest what exam should be made but in the future it could guide the practitioner to the most appropriate diagnosis using this results information.

Emergency Room services require decisions to be made in a fast manner, so a diagnostic test suggestion system, previously validated by an expert team, could help practitioners get orientation in real time about the most adequate diagnostic exams to conduct to reach the best diagnosis.

REFERENCES

[1] J. H. Info. (2011) Japanese health insurance. [Online]. Available: http://japanhealthinfo.com/japanese-healthcare-services/japanese-health-insurance/

[2] M. J. Ward, H. Farley, R. K. Khare, E. Kulstad, R. L. Mutter, R. Shesser, and S. Stone-Griffith, "Achieving efficiency in crowded emergency departments: a research agenda," *Academic Emergency Medicine*, vol. 18, no. 12, pp. 1303–1312, 2011.

[3] O. Ben-Assuli, M. Leshno, and I. Shabtai, "Using electronic medical record systems for admission decisions in emergency departments: examining the crowdedness effect," *Journal of medical systems*, vol. 36, no. 6, pp. 3795–3803, 2012.

[4] I. Lapić and D. Rogić, "Laboratory utilization in the emergency department–are the requested tests patient-oriented?" *Signa Vitae*, vol. 10, no. Suppl. 1, pp. 81–83, 2015.

[5] R. W. Derlet and J. R. Richards, "Ten solutions for emergency department crowding," *Western Journal of Emergency Medicine*, vol. 9, no. 1, p. 24, 2008.

[6] R. S. Ledley and L. B. Lusted, "Reasoning foundations of medical diagnosis," *Science*, vol. 130, no. 3366, pp. 9–21, 1959.

[7] J. Wyatt and D. Spiegelhalter, "Field trials of medical decision-aids: potential problems and solutions." in *Proceedings of the annual symposium on computer application in medical care*. American Medical Informatics Association, 1991, p. 3.

[8] J. Pearl, "Probabilistic reasoning in intelligent systems: Networks of plausible inference," 1988.

[9] L. M. de Campos, J. M. Fernández-Luna, and J. F. Huete, "Bayesian networks and information retrieval: an introduction to the special issue," *Information processing & management*, vol. 40, no. 5, pp. 727–733, 2004.

[10] I. C. Cárdenas, S. S. Al-jibouri, J. I. Halman, and F. A. van Tol, "Capturing and integrating knowledge for managing risks in tunnel works," *Risk Analysis*, vol. 33, no. 1, pp. 92–108, 2013.

[11] A. C. Constantinou, N. E. Fenton, and M. Neil, "pi-football: A bayesian network model for forecasting association football match outcomes," *Knowledge-Based Systems*, vol. 36, pp. 322–339, 2012.

[12] K. W. Przytula and D. Thompson, "Construction of bayesian networks for diagnostics," in *Aerospace Conference Proceedings, 2000 IEEE*, vol. 5. IEEE, 2000, pp. 193–200.

[13] O. Pourret, P. Naïm, and B. Marcot, *Bayesian networks: a practical guide to applications*. John Wiley & Sons, 2008, vol. 73.

[14] A. Zagorecki, P. Orzechowski, and K. Holownia, "A system for automated general medical diagnosis using bayesian networks." in *MedInfo*, 2013, pp. 461–465.

[15] A. K. Akobeng, "Understanding diagnostic tests 1: sensitivity, specificity and predictive values," *Acta paediatrica*, vol. 96, no. 3, pp. 338–341, 2007.

[16] R. Parikh, A. Mathai, S. Parikh, G. C. Sekhar, and R. Thomas, "Understanding and using sensitivity, specificity and predictive values," *Indian journal of ophthalmology*, vol. 56, no. 1, p. 45, 2008.

[17] R. A. Miller, "Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary," *Journal of the American Medical Informatics Association*, vol. 1, no. 1, pp. 8–27, 1994.

[18] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager *et al.*, "Building watson: An overview of the deepqa project," *AI magazine*, vol. 31, no. 3, pp. 59–79, 2010.

[19] R. Miller, F. E. Masarie, and J. D. Myers, "Quick medical reference (qmr) for diagnostic assistance." *MD computing: computers in medical practice*, vol. 3, no. 5, pp. 34–48, 1986.

[20] G. O. Barnett, J. J. Cimino, J. A. Hupp, and E. P. Hoffer, "Dxplain: an evolving diagnostic decision-support system," *Jama*, vol. 258, no. 1, pp. 67–74, 1987.

[21] E. Sanchez, *Solutions in composite fuzzy relation equations: application to medical diagnosis in Brouwerian logic*. Faculté de Médecine de Marseille, 1977.

[22] B. G. Buchanan, E. H. Shortliffe *et al.*, *Rule-based expert systems*. Addison-Wesley Reading, MA, 1984, vol. 3.

[23] M.-C. Su, "Use of neural networks as medical diagnosis expert systems," *Computers in Biology and Medicine*, vol. 24, no. 6, pp. 419–429, 1994.

[24] I. Kononenko, "Inductive and bayesian learning in medical diagnosis," *Applied Artificial Intelligence an International Journal*, vol. 7, no. 4, pp. 317–337, 1993.

[25] Y. Coppieters and A. Levêque, "Ethics, privacy and the legal framework governing medical data: opportunities or threats for biomedical and public health research?" *Archives of public health*, vol. 71, no. 1, p. 15, 2013.

[26] B. Reiner, "Strategies for medical data extraction and presentation part 1: Current limitations and deficiencies," *Journal of digital imaging*, vol. 28, no. 2, pp. 123–126, 2015.

[27] M. A. 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*, vol. 30, no. 4, pp. 241–255, 1991.

[28] J. Pearl, "Bayesian networks: A model of self-activated memory for evidential reasoning," in *Proceedings of the 7th Conference of the Cognitive Science Society, 1985*, 1985, pp. 329–334.

[29] B. D'Ambrosio, "Symbolic probabilistic inference in large bn20 networks," in *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1994, pp. 128–135.

[30] D. Nikovski, "Constructing bayesian networks for medical diagnosis from incomplete and partially correct statistics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 4, pp. 509–516, 2000.

[31] F. J. Diez, "Parameter adjustment in bayes networks. the generalized noisy or-gate," in *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1993, pp. 99–105.

[32] X. Zhou, J. Menche, A.-L. Barabási, and A. Sharma, "Human symptoms–disease network," *Nature communications*, vol. 5, p. 4212, 2014.

[33] C. E. Lipscomb, "Medical subject headings (mesh)," *Bulletin of the Medical Library Association*, vol. 88, no. 3, p. 265, 2000.

[34] W. H. Organization, *International Classification of Functioning, Disability and Health: ICF*. World Health Organization, 2001.

[35] D. A. Lindberg, B. L. Humphreys, A. T. McCray *et al.*, "The unified medical language system," *IMIA Yearbook*, pp. 41–51, 1993.

[36] "Snomed ct to icd-10-cm map." [Online]. Available: https://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html

[37] T. W. Jolis, "Sensitivity and specificity." [Online]. Available: http://www.sensitivityspecificity.com/

[38] "A database of sensitivity and specificity." [Online]. Available: http://getthediagnosis.org/

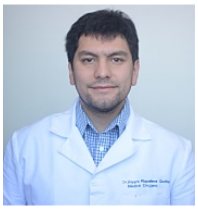
[39] L. Guzmán, F. J. Sánchez, M. F. Martín, A. P. Encinas, I. J. P. Arellano, J. L. Ruiz, and C. M. J. L. Guzmán, *Diagnóstico diferencial en medicina interna*. Elsevier,, 2005.

[40] E. Versi, "" gold standard" is an appropriate term." *BMJ: British Medical Journal*, vol. 305, no. 6846, p. 187, 1992.

Jorge Quinteros was born in Santiago, Chile in 1992. He joined University of Chile's Faculty of Physical and Mathematical Sciences in 2011. He obtained a Bachelor's degree in electrical engineering and a Bachelor's degree in computer science engineering in 2016, both from University of Chile, and is now on his way to obtaining a Master's degree in computer science from the same institution. He has experience as Research Assistant in the fields of football robotics, data mining for astronomy and recently did an internship on early 2017 as a Research Assistant in the field of health informatics in Allm Inc. in Tokyo, Japan. Research interests include robotics, machine learning and more recently blockchain technology and e-Health.

Nelson Baloian is an Associate Professor of Computer Science at the University of Chile. He obtained his Engineering title from the University of Chile and his Ph.D. from the University of Duisburg, Germany. He has been Chairman of the Computer Science Dept. of the University of Chile. His research interests are collaborative computer supported learning, mobile computing and geo-collaboration. He is author of about 80 conference articles and 20 journal papers in these areas. He has been Visiting Professor at the University of Duisburg-Essen (Germany), University of Waseda (Japan) and American University of Armenia (Armenia).

José A. Pino is a full professor of Computer Science at the University of Chile. His research interests include ubiquitous computing, human–computer interaction and socio-technical systems. He has served as President of the Chilean Computer Science Society (SCCC) and President of CLEI (the Latin American Association of Universities Concerning Information Technology). He has co-authored six books and published research papers in international conferences and journals, including Journal of the ACM, Communications of the ACM, ACM Computing Surveys, Decision Support Systems, Group Decision and Negotiation, and Future Generation Computer Systems. He is currently a member of the Editorial Boards of the Journal of Network and Computer Applications and the Journal of Educational Technology and Society.



Alvaro Riquelme is a Physician from University of Concepción, Chile. He is currently coursing a Medical Informatics Master at University of Chile. Until now has worked as a physician on various Clinics and Emergency Departments in Santiago, Chile. His master thesis is related to usability of health information systems on clinical environments.

Sergio Peñafiel was born in Coquimbo, Chile in 1994. He started studying at the Faculty of Physical and Mathematical Sciences of Universidad de Chile in 2013. He obtained the Bachelor's degree in Computer Science in early 2017. He is now a Master student in computer science from Universidad de Chile. He worked as Research Assistant in CEAMOS Santiago Chile in urban crime prediction to optimize police patrolling. He did an internship to Allm Inc. Tokyo Japan to collaborate in disease prediction using machine learning. He is currently working in his thesis about interpretable methods for classification and regression and its applications in real scenarios.

Horacio Sanson obtained his Master's Degree at the Global Information and Telecommunication Institute of Waseda University in Tokyo, Japan, in 2005, where he researched distributed localization on emergency wireless networks. From there, he worked on several engineering positions until he became lead research engineer at Allm Inc. on 2014, where he spends his time applying engineering and machine learning to solve practical medical and health-care issues.

Douglas Teoh graduated from The Univ. of Queensland (Australia) in 2010, receiving a B.Eng (First Class Honours) in software engineering. He joined Allm Inc. as an engineer in 2015, and is currently involved with applying machine learning research in the medical domain.