Computers in Biology and Medicine 51 (2014) 140–158

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment

Flávio Luiz Seixas [a,*], Bianca Zadrozny [b], Jerson Laks [c], Aura Conci [a], Débora Christina Muchaluat Saade [a]

[a] Institute of Computing, Federal Fluminense University, Rua Passo da Pátria, 156, Niterói, RJ 24210-240, Brazil
[b] IBM Research Brazil, Av. Pasteur, 138, Rio de Janeiro, RJ 22296-903, Brazil
[c] Center for Alzheimer's Disease and Related Disorder, Institute of Psychiatry, Federal University of Rio de Janeiro, Av. Venceslau Brás, 71, Rio de Janeiro, RJ 22290-140, Brazil

ARTICLE INFO

Article history:
Received 29 September 2013
Received in revised form
10 April 2014
Accepted 15 April 2014

Keywords:
Clinical decision support system
Bayesian network
Dementia
Alzheimer's disease
Mild cognitive impairment

ABSTRACT

Population aging has been occurring as a global phenomenon with heterogeneous consequences in both developed and developing countries. Neurodegenerative diseases, such as Alzheimer's Disease (AD), have high prevalence in the elderly population. Early diagnosis of this type of disease allows early treatment and improves patient quality of life. This paper proposes a Bayesian network decision model for supporting diagnosis of dementia, AD and Mild Cognitive Impairment (MCI). Bayesian networks are well-suited for representing uncertainty and causality, which are both present in clinical domains. The proposed Bayesian network was modeled using a combination of expert knowledge and data-oriented modeling. The network structure was built based on current diagnostic criteria and input from physicians who are experts in this domain. The network parameters were estimated using a supervised learning algorithm from a dataset of real clinical cases. The dataset contains data from patients and normal controls from the Duke University Medical Center (Washington, USA) and the Center for Alzheimer's Disease and Related Disorders (at the Institute of Psychiatry of the Federal University of Rio de Janeiro, Brazil). The dataset attributes consist of predisposal factors, neuropsychological test results, patient demographic data, symptoms and signs. The decision model was evaluated using quantitative methods and a sensitivity analysis. In conclusion, the proposed Bayesian network showed better results for diagnosis of dementia, AD and MCI when compared to most of the other well-known classifiers. Moreover, it provides additional useful information to physicians, such as the contribution of certain factors to diagnosis.

1. Introduction

Global population aging has been one of the main concerns of the twentieth century with great economic, political and social consequences. Moreover, dementia is prevalent among the elderly observed both in developed and developing countries [1]. Dementia is a clinical state characterized by loss of function in multiple cognitive domains. The most commonly used criteria for diagnosis of dementia were established by DSM-IV (Diagnostic and Statistical Manual for Mental Disorders) by the American Psychiatric Association [2].

There are various specific types of dementia, often showing slightly different symptoms. The most common is Alzheimer's Disease (AD), accounting for between 60% and 80% of dementia cases [1]. AD is a degenerative disease causing lesions in the brain. Early clinical symptoms of AD are often related to difficulty in remembering new information, and later symptoms include impaired judgment, disorientation, confusion, behavioral changes and difficulty in speaking and walking.

According to the annual report published by the Alzheimer's Association [3], patients with AD impact more than 10 million of health care providers in the United States in 2013. AD was considered as the sixth-leading cause of death across all ages groups in the United States. The Alzheimer's Association estimated in the United States that there will be over 6 million people over 65 years old affected by AD in 2025 [4].

The prevalence of dementia in Latin America is similar to that reported in developed countries [5]. An epidemiological study

* Corresponding author.
E-mail addresses: fseixas@ic.uff.br (F.L. Seixas), biancaz@br.ibm.com (B. Zadrozny), jlaks@centroin.com.br (J. Laks), aconci@ic.uff.br (A. Conci), debora@ic.uff.br (D.C. Muchaluat Saade).

http://dx.doi.org/10.1016/j.compbiomed.2014.04.010
0010-4825/

with the Brazilian population estimated a prevalence of dementia of 8% for the elderly population [6]. The screening tools, assessment and diagnostic criteria applied for such studies are synthesized by Nitrini et al. [5]. Among the diseases causing dementia, AD was the most frequent in all studies, ranging from 50% to 55% of all cases.

The diagnosis of individuals with earliest stage of AD motivates a number of research initiatives [7,8]. Mild Cognitive Impairment (MCI) is usually associated to a preclinical stage of AD and causes cognitive impairments severe enough to be noticed by patient's relatives, or other people, without producing any changes in patient's daily activities [9]. A deficit in episodic memory is the most common symptom in MCI patients [10].

An AD incidence study revealed that, annually, 10–30% of patients who had received an MCI diagnosis converted to AD, while the conversion rate of normal elderly subjects is 1–2% [11].

The first broadly accepted criteria for the clinical diagnosis of probable AD were established by the NINCDS-ADRDA (National Institute of Neurological, Communicative Disorders and Stroke - Alzheimer's Disease and Related Disorders Association) workgroup [12]. These criteria were revised by the National Institute on Aging (NIA) and the Alzheimer's Association recently [13], including advances in medical imaging techniques and biomarkers to support AD diagnosis, and better understanding of the AD process. The criteria were published in three documents: the core clinical criteria of the recommendations regarding all-cause dementia [14], dementia due to AD [13] and MCI due to AD [10].

Regarding healthcare services, there is a growing global concern about patient safety and healthcare system effectiveness [15]. In the same way, diagnostic errors are an important source of preventable harm to health systems [16]. A diagnostic error can be defined as a diagnosis that is missed, wrong, or delayed, as detected by some definitive test or finding. The awareness and understanding of medical errors have promoted safer health care through health information systems solutions [17]. Clinical decision support systems (CDSS) are considered as an important category of health information systems designed to improve clinical decision-making [18]. Characteristics of individual patients are matched to characteristics from a knowledge base, and an algorithm generates patient-specific assessments or recommendations. Studies have indicated that CDSS can reduce diagnostic error rates [19].

This paper proposes a Bayesian Network (BN) decision model for supporting the diagnosis of dementia, AD and MCI. The proposed BN can be used for building clinical decision support systems (CDSS) to help in the diagnosis of such diseases. BNs are well-suited for representing uncertainty and causality, which are both present in the clinical domain. The proposed BN was modeled using a combination of expert knowledge and data-oriented modeling. The network structure was built based on current diagnostic criteria and input from physicians who are experts in this domain. The network parameters were estimated using a supervised learning algorithm from a dataset of real clinical cases. In this text, we define an instance as an ordered list of values representing a set of observations of a patient. Such observations include symptoms, signs and results of neuropsychological exams of patients, and normal controls from the Duke University Medical Center (Washington, USA) as well as the Center for Alzheimer's Disease and related disorders, at the Institute of Psychiatry of the Federal University of Rio de Janeiro (Rio de Janeiro, Brazil). We define an attribute as a labeled element (or observation) of an instance.

The next sections are organized as follows. Section 2 discusses CDSS and related work. Section 3 describes some concepts of Bayesian networks, including the Bayesian learning methods and the inference algorithms. Section 4 shows the process used to build the BN model, the dataset of patients and normal controls used for training and the proposed BN. In Section 5, we show the experimental results. Section 6 discusses the results. Finally, in Section 7, we present conclusions and directions for future work.

2. Clinical decision support systems and related work

CDSSs are computational systems designed to support high-level cognitive functions involving clinical diagnosis, such as reasoning, decision-making and learning [20]. Wright et al. [21] provided a taxonomic description of CDSS, grouping them according to their purpose regarding clinical activities, including prevention and screening, drug dosing, chronic disease management, diagnosis and treatment, considering patients from both outpatient and inpatient settings. CDSS can be designed for a number of user profiles, including physicians, general practitioners, nurses, or even patients seeking prevention or other health-related behaviors.

CDSSs are also grouped by their prevailing inference engine, including rule-based systems, clinical guideline-based systems and semantic network-based systems [22]. Rule-based systems use simple conditional expressions (e.g., if-then-else clauses) for making deductions and aiding clinical decisions. Guideline-based systems indicate the most likely clinical decision or pathway from a set of predetermined options, guided by a workflow that describes, for example, diagnosis rules or a treatment process. Semantic network-based systems use semantic relations between concepts to perform an inference algorithm. BN-based CDSSs fall into the semantic network category, since BN nodes represent clinical concepts, and edges (or arcs) represent causal relations. Fig. 1 depicts some CDSSs found in the literature and their prevailing inference engine. In the next paragraphs, we comment

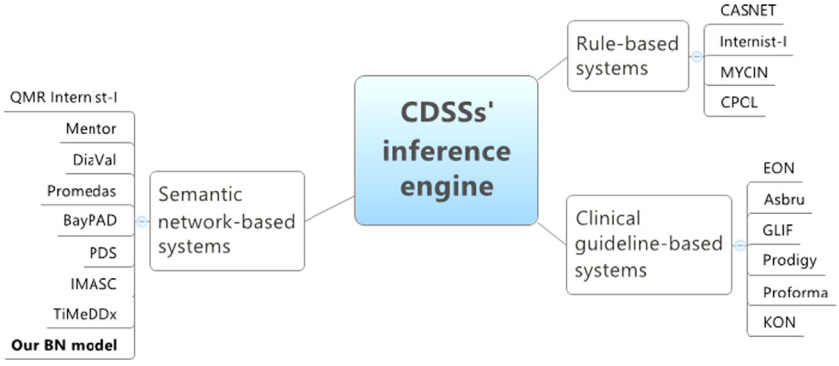


Fig. 1. Clinical Decision Support Systems (CDSSs) grouped by their inference engine.

Table 1
Clinical decision support systems.

Inference engine	Name	Disease domain	Reference
Rule-based systems	CASNET	Glaucoma	Weiss et al. [23]
	Internist-I	572 diseases linked to more than 4000 symptoms	Miller et al. [24]
	MYCIN	Meningitis	Buchanan and Shortliffe [25]
	unnamed	257 mental disorders	Amaral et al. [28]
	CPOL	Diabetes, cardiac diseases and respiratory disorders	Beliakov and Warren [26]
	unnamed	Alzheimer's disease using SPECT (Single-Photon Emission Computed Tomography) images and classifier based on SVM (Support Vector Machine)	Salas-Gonzalez et al. [29]
	unnamed	Alzheimer's disease using fMRI (functional Magnetic Resonance Image) and classifier based on PCA (Principal Component Analysis)	Tripoliti et al. [30]
Clinical guideline-based systems	EON	Hypertension	Musen et al. [31]
	Asbru	Diabetes, jaundice, breast cancer	Shahar et al. [32]
	GLIF	Depression, diabetes and hyperglycemia	Peleg et al. [33]
	Prodigy	Chronic diseases, including asthma and hypertension	Johnson et al. [34]
	Proforma	Not specified.	Fox et al. [35]
	KON[3]	Oncology	Ceccarelli et al. [36]
Semantic network-based systems	QMR Internist-I	572 diseases linked to more than 4000 symptoms	Miller et al. [24]
	Mentor	Mental retardation in newborns	Mani et al. [37]
	DiaVal	Cardiovascular diseases	Diez et al. [38]
	Unnamed	Head-injury	Sakellaropoulos and Nikiforidis [39]
	Unnamed	Mammography	Burnside et al. [40]
	Unnamed	Pneumonia	Aronsky and Haug [41]
	Unnamed	Mild cognitive impairment using MRI (Magnetic Resonance Image) and neuropsychological assessments	Yan et al. [42]
	Promedas	2000 diseases linked to 1000 symptoms	Wemmenhove et al. [43]
	BayPAD	Pulmonary embolisms	Luciani et al. [44]
	PDS	Pediatric diseases described in ICD-10 (International Classification of Diseases – release 10)	Pyper et al. [45]
	IMASC	Cardiac diseases	Czibula et al. [46]
	TiMeDDx	Infectious and noninfectious diarrhea	Denekamp and Peleg [47]

some CDSSs, focusing on BN-based ones, and conceptually compare their decision models to ours.

Table 1 lists some CDSSs grouped also by their prevailing inference engine and their corresponding disease domain. The earliest CDSSs were rule-based, such as CASNET [23], Internist-I [24], and MYCIN [25] and CPOL [26]. CASNET and MYCIN were developed for aiding the diagnostic of Glaucoma and Meningitis, respectively, and Internist-I and CPOL, for aiding the diagnostic of diseases of various domains. Korb and Nicholson [27] provided an exhaustive list of CDSSs applied to different areas. Some of those CDSSs are described in the following paragraphs.

QMR (Quick Medical Reference)/Internist-I is one of the earliest systems using an inference engine based on BN. A BN is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG) [48]. QMR/Internist-I used a BN with two-level structure representing the probabilistic relationships between diseases and symptoms [49]. The conditional probabilities of symptoms and diseases were estimated by combining probabilities from disease profiles and statistical data from hospitals [50], deriving statistical parameters from health information systems and the judgment of human experts from the corresponding disease domain. Furthermore, QMR used a noisy-OR model, described by Pearl [48], to simplify the conditional probability estimation and expert elicitation. Hence, based on given symptoms, the network can be used to compute the probabilities of the presence of various diseases by a suitable inference engine, supporting the physicians in their diagnostic process. One version of QMR network described by Pradhan et al. [51] included background nodes, which they called predisposing factors, needing prior probabilities, while the remaining nodes required probabilities to be assessed for each of their values. Thus, we called such Bayesian structure a three-level structure, where each level

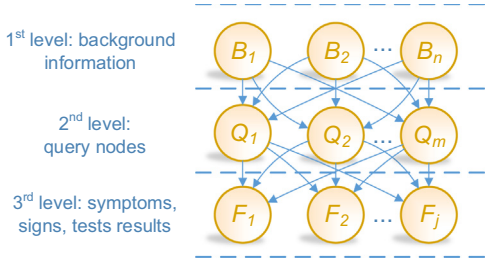


Fig. 2. Three-level Bayesian network structure.

represents (1) background information, (2) diseases and (3) symptoms, signs and neuropsychological test results, respectively, as depicted in Fig. 2. One difficulty of using such modeling structure is that of estimating the conditional probabilities that reflect real clinical cases and, at the same time, meeting the random variables independence assumption from the noisy-OR model [52].

Mentor is a CDSS to predict mental retardation in newborns [37]. That system is based on a BN whose structure was discovered from data using an algorithm proposed by [53] and validated by domain experts.

DiaVal is a CDSS for diagnosis of cardiovascular diseases using a BN [38]. The BN structure was built by domain experts using a causal representation of the cardiac pathophysiology and, afterwards, they incorporated some results, mainly from echocardiography.

Sakellaropoulos and Nikiforidis [39] used a BN with discrete probability distribution for the assessment of prognosis after 24 h for patients who had head injuries. The BN structure and parameters were learned from cases of patients and the prognosis results were compared to those made by an expert. They achieved a BN success rate closer to the success rate of the expert. Burnside

et al. [40] built a BN for aiding radiologists in their decision-making, integrating mammogram findings using BI-RADS (Breast Imaging Reporting and Data System), a standardized lexicon developed for mammography. Their BN model provided probabilities associated to benign, pre-malignant and malignant breast cancer disease. Aronsky and Haug [41] showed the modeling and the evaluation of a BN for the diagnosis of pneumonia.

Although using a BN structure similar to QMR, Promedas [43] used as an inference engine a novel method called Loop Corrected Belief Propagation (LCBP) [54] described in Section 3.2.

BayPAD (Bayes Pulmonary embolism Assisted Diagnosis) used a BN for diagnosis of pulmonary embolisms [44]. The BN structure was constructed from expert elicitation and the conditional probabilities estimated from patient data provided by PISA-PED (Prospective Investigative Study of Acute Pulmonary Embolism Diagnosis). In order to improve the model accuracy, the results were reviewed using the Cox Regression-Calibration Model [55].

PDS (Pediatric Decision Support system) is a BN-based CDSS deployed at the Children's Hospital of Eastern Ontario [45]. The system allows a resident physician to define a patient's case based on symptoms and generates a list of possible diagnoses based on the World Health Organization. The BN structure was manually built based on expert elicitation and BN conditional probabilities were estimated using a set of patient cases.

IMASC (Intelligent MultiAgent System for Clinical Decision Support) is a CDSS for diagnosis of cardiac diseases using the artificial neural network (ANN) [46]. Considering the contributions, the IMASC extends the intelligent agent concept, sharing messages among other agents in a modular approach.

In TiMeDDx [47], the BN was modeled using a main clinical manifestation or MCM-oriented model. MCM-oriented diagnosis is a problem-oriented process that starts with a chief clinical problem, reasons about possible diagnoses that would be manifested as MCM, and suggests the clinical data items that should be collected in order to differentiate among alternative diagnoses. Its BN was built based on expert elicitation.

Yan et al. [42] used a BN for classifying patients with MCI using Magnetic Resonance Images (MRI) and other demographic and health data from patients. The BN structure was discovered using a greedy search algorithm.

Another popular technique for constructing decision models is applying multicriteria decision aiding (MCDA) methods. MCDA involves a set of methods for aggregation of multiple evaluation criteria to one or more potential actions [56]. There are conceptual similarities between MCDA approaches and Bayesian learning methods. Both perspectives consider the problem of learning a decision from data as a maximization of an empirical utility function [57]. In a Bayesian learning perspective, the utility function can be maximizing some Bayesian Estimation score, like Maximum-Likelihood Estimation (MLE) [27]. MLE score will be described in Section 3.1.

There are works that propose integrating approaches based on MCDA and statistical learning methods, i.e., the implementation of MCDA concepts in a statistical learning framework and the development of hybrid methodologies. Castro et al. [58] showed an MCDA approach called MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) integrated with a BN. Their decision model aims at showing what assessment items from neuropsychological batteries are more attractive given the CDR (Clinical Dementia Rating) scale result. They used an extended discrete BN with utility and decision nodes, so-called Influence Diagram [59]. An Influence Diagram is a generalization of a BN, in which not only probabilistic inference problems but also decision-making problems can be modeled and solved. The utility nodes represent a set of objectives or preferences considered by a decision-maker. The decision nodes represent a set of outgoing branches considered in a decision-making process. In an MCDA approach, BN nodes represent the internal or external factors that can affect the decision criteria. Their BN structure allowed calculating the conditional probability tables (CPTs) of BN nodes from dataset without applying a complex statistical learning method. Their dataset is composed by patient cases set, where each attribute is related to an assessment item from neuropsychological batteries. In addition, missing attribute values were treated as a discrete state, being replaced by a hypothetical value. In our BN modeling, we address a different treatment for missing attribute values, as described in Section 3.1.

Menezes et al. [60] proposed a hybrid model combining MCDA and BN for diagnosis of Diabetes type 2. Pinheiro et al. [61] used a similar approach and showed a ranking model based on MCDA and BN for aiding the diagnosis of AD. Their objective indicates what assessment patient items have the highest impact for determining the AD diagnosis. Their BN was manually built, where the Bayesian nodes were semantically related to each assessment item. Assessment items were composed by neuropsychological batteries provided by CERAD [62].

Instead of considering each neuropsychological question separately, our work used final neuropsychological test results. We designed a decision model to predict AD diagnosis, and other related diseases, such as dementia and MCI. Our decision model allows indicating the most relevant items based on a sensitive analysis instead of MCDA.

Furthermore, our CDSS differs from other BN-based CDSSs in terms of BN structure building and its parameter estimation. QMR and Promedas used a manually built structure oriented by the three-level structure proposed by Pradhan et al. [51]. BayPAD, PDS and TiMeDDx also have a BN structure manually built based on domain expert elicitation, but they did not use the three-level structure template described before. In contrast, the BN structure designed for Mentor was automatically discovered from data. Regarding BN parameter estimation, QMR and TiMeDDx have probability distribution estimated by a domain expert. Other BN-based CDSSs use some learning algorithm from data to estimate the BN parameters. In contrast, our CDSS has a BN structure oriented by Pradhan's three-level structure template, which was manually built with the support of domain experts. In addition to that, our BN parameter values are estimated using a Bayesian learning method that is described in Section 3.1.

Fig. 3 depicts some components of the CDSS described in this paper. The knowledge base contains the BN in which each node is associated to a clinical concept from the diagnosis criteria of the disease of interest. The inference engine estimates the posteriori probability distribution of non-evidence random variables, namely variables that are not observed in the patient dataset. Each BN random variable must be associated to a concept from the healthcare domain, including demographic data. The patient dataset is used to estimate the BN conditional probabilities by a supervised learning algorithm. Each patient dataset attribute must also be associated to a healthcare concept during the BN modeling phase. The supervised learning results and BN performance measures are evaluated by a system analyst.

The patient clinical data are commonly stored in a Health Information System (HIS) using an Electronic Health Record (EHR) model. A BN evidence is derived from the matching between the healthcare concept assigned to the patient clinical datum and the concept assigned to the BN random variable. Hence, both concept domains must be the same for system interoperability. The CDSS results can be viewed through HIS's user interface, indicating the most probable diagnosis and the most sensitive non-observed items that should be evaluated by the physician, in order to confirm or refute an initial diagnosis hypothesis.

In the next section, we will describe some BN concepts used in this work.

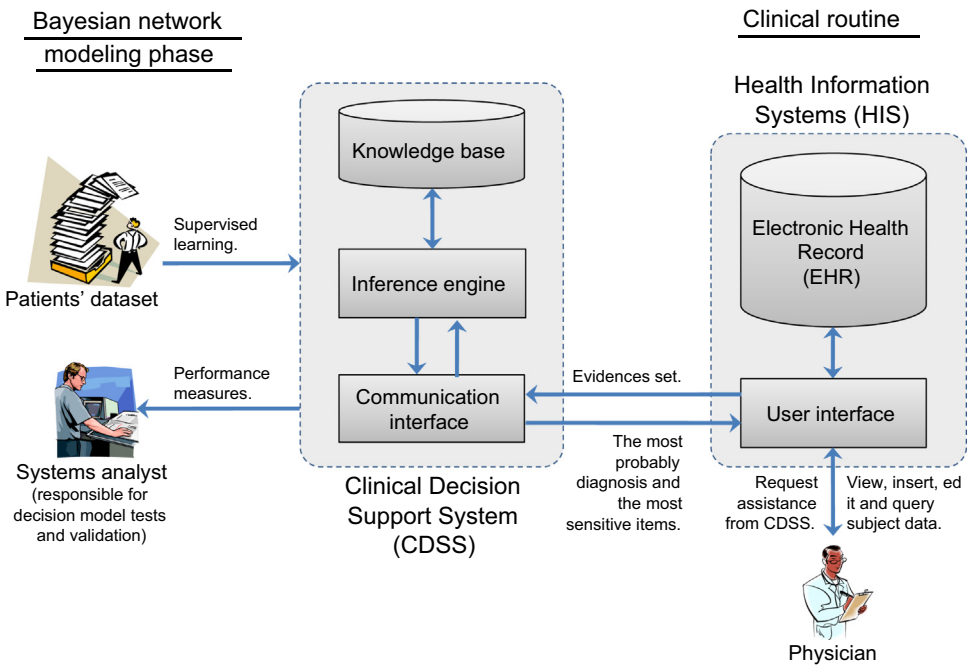


Fig. 3. Clinical decision support system components.

3. Bayesian network concepts

BNs represent a domain in terms of random variables and explicitly model the interdependence between them [63]. BNs are graphically represented by directed acyclic graphs (or DAGs), whose nodes represent random variables and arcs express direct dependencies. In our BN model, we consider only discrete random variables. Suppose X_i is the ith random variable containing r_i discrete states. Also suppose that X_{ik} is the kth state of random variable X_i. Consider a BN containing a set of nodes or random variables $X^G = \langle X_1, X_2, \ldots, X_N \rangle$, a pair $\langle G, \Theta \rangle$, where G represents a DAG and Θ the set of conditional probabilities variables $\Theta = \langle \theta_{ijk} : \forall ijk \rangle$. θ_{ijk} is called the joint probability, i.e., $\theta_{ijk} = \Pr(X_i = x_i^k | pa(X_i) = p_i^j)$ where p_i^j represents the parent (pa) node set of X_i from G and p_{ij} the jth combination of parent nodes of X_i.

A Bayesian network modeling involves a number of computational algorithms. Bayesian learning algorithms are commonly applied to discover the Bayesian structure and/or estimate parameter values from data. The structure and parameters necessary for characterizing a BN can be provided either by experts, or by automatic learning from a dataset. Section 3.1 describes the Bayesian learning methods and compares them to different approaches. The Bayesian inference algorithm is responsible for calculating marginal probabilities, given an evidence set. Section 3.2 lists some Bayesian inference algorithms.

3.1. Bayesian parameter learning

The BN learning task can be divided into two subtasks: structural learning and parametric learning, i.e., the estimation of the numerical parameters (conditional probabilities) for a given BN structure. A number of different methods were proposed for estimating a BN from a dataset. Such methods are usually classified into two approaches: constrained-based methods and scored-based methods. A well-known constrained-based method is the Peter–Clark (PC) algorithm used for discovering BN structure [64]. Scored-based methods have two major components: a scoring metric that measures every candidate BN using a score function with respect to a dataset, and a search procedure to move through a solution space composed by possible BNs. Examples of such learning methods are Expectation-Maximization (EM) algorithm [65], evolutionary algorithms [66], Gibbs sampling based algorithms [67].

We used a BN model whose structure is known and parameters (conditional probabilities) are unknown. Thus, we used the EM algorithm for parametric learning. EM mainly differs from evolutionary algorithms in its searching method, i.e., EM uses an ascendant gradient algorithm, while evolutionary algorithms use genetic algorithms. EM differs from Gibbs sampling in the following aspects: EM is a deterministic algorithm and converges in Maximum-Likelihood, while Gibbs sampling is a non-deterministic method and converges towards a posterior distribution. Nevertheless, all these learning methods are relatively similar.

Given a data set $D = (D_1, \ldots, D_M)$, where $D_i \in R^d$ composed by M independent and identically distributed observations from a distribution $\Pr(D|\theta)$. Let a BN node represents a random variable X with R discrete states or multiple values. So, the Bayesian parameter learning consists of estimating parameter set $\theta = (\theta_1, \ldots, \theta_R)$ that best represents data. The parameters that best represents a dataset is known as θ_{MLE}, where MLE stands for Maximum Likelihood Estimation, whose equation is shown below:

$$\theta_{MLE} = \arg\max_{\theta \in \Theta} \Pr(D|\theta) \tag{1}$$

Let $\theta_i = \Pr(D = x_i)$. Suppose in a dataset D there are m_i instances where D takes value x_i. Then, the multinomial likelihood is

$$L(\theta|D) = \Pr(D|\theta) = \prod_{j=1}^{N} \Pr(D_j|\theta) = \prod_{i=1}^{r} \theta_i^{m_i} \tag{2}$$

The conjugate family for multinomial likelihood is a Dirichlet distribution. A Dirichlet distribution is parametrized by R hyperparameters $\alpha_1, \ldots, \alpha_R$. So, assuming as prior probability a Dirichlet distribution, the posterior probability $\Pr(\theta|D)$ is given by

$$\Pr(\theta|D) \sim \prod_{i=1}^{R} \theta_i^{\alpha_i + m_i - 1} \tag{3}$$

where $\alpha_i + m_i$ are parameters from Dirichlet distribution $Dir(\alpha_1 + m_1, \ldots, \alpha_R + m_R)$.

Let us assume that

$$\theta^{i}_{MLE} = \frac{m_i}{m_1 + m_2 + \cdots + m_R} \tag{4}$$

In the case of a BN with discrete distribution that contains N nodes or random variables $X_1, \ldots, X_N$, number of states of $X_i : 1, 2, \ldots, r_i$, number of configurations of parents (pa) of $X_i : 1, 2, \ldots, q_i$, so the parameters to be estimated are $\theta_{ijk} = \Pr(X_i = j | pa(X_i) = k)$, $i = 1, \ldots, n;\ j = 1, \ldots,\ r_i; k = 1, \ldots, q_i$ in general BNs:

$$\theta^{ijk}_{MLE} = \frac{\alpha_{ijk} + m_{ijk}}{\sum_j \alpha_{ijk} + m_{ijk}} \tag{5}$$

Eq. (5) is used to compute θ parameters for BN. However, if the dataset contains instances with partial observations, i.e., attributes with missing values, it will be necessary to use a more appropriate learning method. Missing attribute values (partial observations) are caused by a number of reasons, either the patient might not have performed such neuropsychological test, or the physician might not have included such results in the dataset. A parametric learning method usually applied in the case of partial observations is the EM algorithm [68]. EM is an iterative algorithm, which tries to maximize θ_{MLE} at each iteration. The algorithm contains two steps: E-Step computes a posterior distribution over D_i using a BN inference engine and the M-step maximizes the log-likelihood from a Dirichlet distribution using a Newton-based (or other ascendant gradient) algorithm detailed by Minka [69].

Some approximation methods for EM, as the MCMC algorithm (Markov Chain Monte Carlo) [70], try to reduce the computational cost for Bayesian learning. We did not use any approximation method because our BNs are fairly small in terms of the amount of nodes and their corresponding parents, which makes the computational cost irrelevant.

The parameters for the EM algorithm are the training set D and the Dirichlet initial hyperparameters set α. We considered $\alpha = 1$ for all attributes, assuming a prior uniform distribution, or non-informative priors [71]. The routine goes on until it reaches the convergence or stopping criterion. As a result of the iterative process, we reach θ_{MLE}. We used the EM implementation available in the BN toolbox developed in Mathworks Matlab© [72].

3.2. Bayesian inference engine

A Bayesian inference engine computes the posterior probability distribution, as shown in Eq. (6) in terms of $\theta_{i'j'}$, admitting a statistically independent distribution θ_{ij} with ith random variable and jth combination of parent node of $\theta_{i'j'}$, and G representing a DAG.

$$\Pr(\theta_{i'j'}|G) = \prod_{i=0}^{N} \prod_{j=1}^{q_i} \Pr(\theta_{ij}|G), \tag{6}$$

where N is the number of BN nodes, q_i is the number of combination of parent nodes from $\theta_{i'j'}$.

Our interest is computing the marginal probabilities of the query node. The query node is a BN node that represents the diagnosis for a disease, as shown in Fig. 2 (nodes Q_i). Such probabilities can be either positive or negative for diagnosis of the disease of interest. There are inference engines that use exact methods and approximate methods. Furthermore, approximate methods can be deterministic or stochastic. The most suitable algorithm will depend on the computational cost and accuracy level required.

Junction Tree (JT) [63] is a message passing algorithm for performing inference on graphical models, such as BNs. In contrast to other inference algorithms, JT transforms the BN to a junction tree representation to perform an exact inference.

Belief Propagation (BP) [73] is also a message passing based algorithm that operates on a factor graph. A factor graph is a bipartite graph with nodes, factor nodes, and an edge between variables and factors. BP produces exact results on tree-like factor graphs. However, if the factor graph contains one or more loops, results are approximate. A novel algorithm proposed a method that corrects BP for the presence of loops in the factor graph and obtains improvements in accuracy using a loop expansion scheme called Loop Correct Belief Propagation (LCBP). LCBP is used in cases which the error becomes unacceptable [54]. In short, LCBP removes a node from the graph and factorizes the probability distribution on its neighbors, so-called cavity distribution. For more details see [43].

Since our discrete BNs are simple as previously mentioned, we used JT as an inference engine.

4. Bayesian network modeling

As mentioned in Section 1, the training dataset used in our work for Bayesian learning was composed by normal controls and patient cases set provided by two Institutions: CERAD (Consortium to Establish a Registry for Alzheimers Disease) of Duke University Medical Center (Washington, USA) [62], and CAD (Center for Alzheimer's Disease and related disorders) of Federal University of Rio de Janeiro (Rio de Janeiro, Brazil) with the consent of the ethics committee for medical research, project identification number 284/2010.

Table 2 lists the selected attributes of the CERAD dataset. The cited neuropsychological tests are detailed at http://cerad.mc.duke.edu/ (visited at February 22th, 2014). Table 3 lists the selected attributes of the CAD dataset.

Table 4 shows the number of cases from both training datasets. CERAD's dataset contains patients diagnosed with dementia and AD composing two subsets. The dementia subset is composed of normal control subjects and patients with dementia. The AD subset is composed of patients previously classified as positive for dementia. The patients from the AD subset can be positive for AD or negative. Thirty cases from the AD subset (CERAD) had the diagnosis not classified. The MCI subset is composed of patients previously classified as negative for dementia. The patients from the MCI subset can be either positive or negative for MCI.

As shown in Table 4, because some subsets were unbalanced (e.g. the number of negative cases was much smaller than the number of positive cases), we applied a balancing technique called SMOTE (Synthetic Minority Oversampling Technique) for over-sampling the negative cases without modifying the training dataset characteristics [74].

As mentioned in Section 3.1, we have some attributes in both datasets with missing values. The attributes with completely missing values were removed from both datasets. We also used a filter based on information gain score to remove attributes of datasets that have score values below a threshold [75]. We used information gain implementation available on WEKA[1] (Waikato Environment for Knowledge Analysis), a popular suite of machine learning software. The information gain score is between 0 and 1. The attributes with information gain score below 0.0001 were rounded to zero by WEKA and thus removed from both datasets. The attributes of both datasets showed before (Tables 2 and 3)

[1] http://www.cs.waikato.ac.nz/ml/weka/.

Table 2
List of CERAD dataset selected attributes.

Level[a]	Attribute[b]	States[c]	Diagnosis[d]	
			D	AD
D	Dementia	1:negative; 2:positive	✔	
D	Alzheimer's Disease (AD)	1:negative; 2:positive		✔
B	Education	1:[0–9]; 2:[10–16]; 3:[17–19]; 4:[20–inf]	✔	✔
F	Clinical Dementia Rating (CDR) scale	1:0; 2:0.5; 3:1; 4:2; 5:3; 6:4; 7:5	✔	✔
B	Ethnicity	1:caucassian; 2:afrodescendent	✔	✔
F	Mini Mental State Exam (MMSE) score	1:[0–27]; 2:[28–30]	✔	✔
F	Boston naming	1:[0–14]; 2:[15–inf]	✔	✔
F	Short Blessed score	1:[0–5]; 2:[6–inf]	✔	✔
F	Cognitive impairment reported by patient or informant	1:negative; 2:positive	✔	✔
F	Activities for Daily Living (ADL) score	1:0; 2:0.5; 3:1; 4:2	✔	✔
F	Recall Word List score	1:0; 2: > 0	✔	✔
B	Cerebrovascular disease	1:negative; 2:positive		✔
F	Difficulties in the use of language elements	1:without difficulty; 2:mild difficulty; 3:moderate difficulty	✔	✔
B	Gender	1:male; 2:female	✔	✔
F	Gradual decline of cognitive functions	1:negative; 2:positive	✔	✔
F	Verbal fluency	1:negative; 2:positive		✔
F	Functional abilities	1:negative; 2:positive	✔	
F	Changes in personality and behavior	1:negative; 2:positive		✔
F	Word List Recognition score	1:0; 2: > 0	✔	✔
F	Memory Word List score	1:0; 2: > 0	✔	✔
F	Primary progressive aphasia	1:negative; 2:positive	✔	

[a] B=Background information; D=Disease of interest; F=Findings. Attributes classified as Background information are located at the 1st level of BN. Attributes classified as Findings are located at the 3rd level of BN.
[b] Each selected attribute was assigned to a BN random variable. Attributes can represent symptoms, signs, neuropsychological tests, demographic data and predisposal factors.
[c] Represent the possible discrete states that random variable can assume. The syntax for some multinomial attributes is N:[J–K], where N=multinomial value, J=lower limit of interval; K=upper limit of interval. Inf=infinite. The limit values were estimated by the discretization algorithm used.
[d] If the attribute was selected for Dementia (D) subset or Alzheimer's disease (AD) subset.

Table 3
List of CAD dataset selected attributes.

Level[a]	Attribute[b]	States[c]
D	Dementia	1:negative; 2:positive
D	Alzheimer's Disease (AD)	1:negative; 2:positive
D	Mild Cognitive Impairment (MCI)	1:negative; 2:positive
F	Mini Mental State Exam (MMSE) score	D and AD: 1:[0-17]; 2:[18-26]; [27-30] MCI: 1:[0-28]; 2:[29-30]
F	Clinical Dementia Rating (CDR) scale	1:0; 2:0.5; 3:1; 4:2; 5:3
F	Pfeffer questionnaire score	D and AD: 1:0; 2:[1-2]; 3:[3-inf] MCI: 1:[0-1]; 2:[2-3]; 3:[3-inf]
F	Verbal Fluency Test (VFT) score	D and AD: 1:[0-4]; 2:[5-11]; 3:[12-inf] MCI: 1:[0-15]; 2:[16-inf]
F	Clock Drawing Test (CDT) scale	1:0; 2:1; 3:2; 4:3; 5:4; 6:5
F	Trial Making Test (TMT) score	D and AD: 1:[0-16]; 2:[17-59]; 3:[60-inf] MCI: 1:[0-36]; 2:[37-inf]
B	Age	D and AD: 1:[0–72]; 2:[73–inf] MCI: 1:[0–69]; 2:[70–inf]
F	Lawton scale	D and AD: 1:[0–9]; 2:[10–inf] MCI: 1:[0–14]; 2:[15–inf]
F	IQCode (Informant Questionnaire on Cognitive Decline in the Elderly)	D and AD: 1:[0–3.55]; 2:[3.56–inf] MCI: 1:[0–3.32]; 2:[3.33–inf]
F	Stroop color word test score	D and AD: 1:[0–15]; 2:[16–inf] MCI: 1:[0–17]; 2:[18–inf]
F	Berg balance scale	D and AD: 1:[0–51]; 2:[52–inf] MCI: 1:[0–54]; 2:55; 3:[56–inf]
B	Gender	1:male; 2:female
F	Depression	1:negative; 2:positive
B	Education	D and AD: 1:[0–13]; 2:[14–inf] MCI: 1:[0–15]; 2:[16–inf]

[a] B=Background information; D=Disease of interest; F=Findings. Attributes classified as Background information are located in the 1st level of BN. Attributes classified as Findings are located in the 3rd level of BN.
[b] Each selected attribute was assigned to a BN random variable. Attributes can represent symptoms, signs, neuropsychological tests, demographic data and predisposal factors.
[c] Represent the possible discrete states that random variable can assume. The syntax for some multinomial attributes is N:[J–K], where N=multinomial value, J=lower limit of interval; K=upper limit of interval. inf=infinite. The limits values were estimated by the discretization algorithm used.

Table 4
Patients' cases datasets.

Institution	Diseases	Dementia		AD			MCI	
	Diagnosis	N[a]	P	N	P	U	N	P
CERAD	Number of cases	463	1094	502	562	30	–	–
	Number of cases after applying the oversampling technique for balancing the base	926	1094					
	Missing values ratio[b]	48%	44%	47%	42%	–	–	–
CAD	Number of cases	67	180	45	135	–	35	32
	Number of cases after applying the oversampling technique for balancing the base	134	180	90	135	–		
	Missing values ratio	29%	24%	29%	22%	–	37%	21%

[a] N=negative; P=positive; U=unclassified.
[b] The missing values ratio was calculated dividing the total of missing attribute values by the number of instances multiplied by the number of attributes of the corresponding subset. The number of attributes was obtained by counting the remaining attributes after applying the attributes filter.

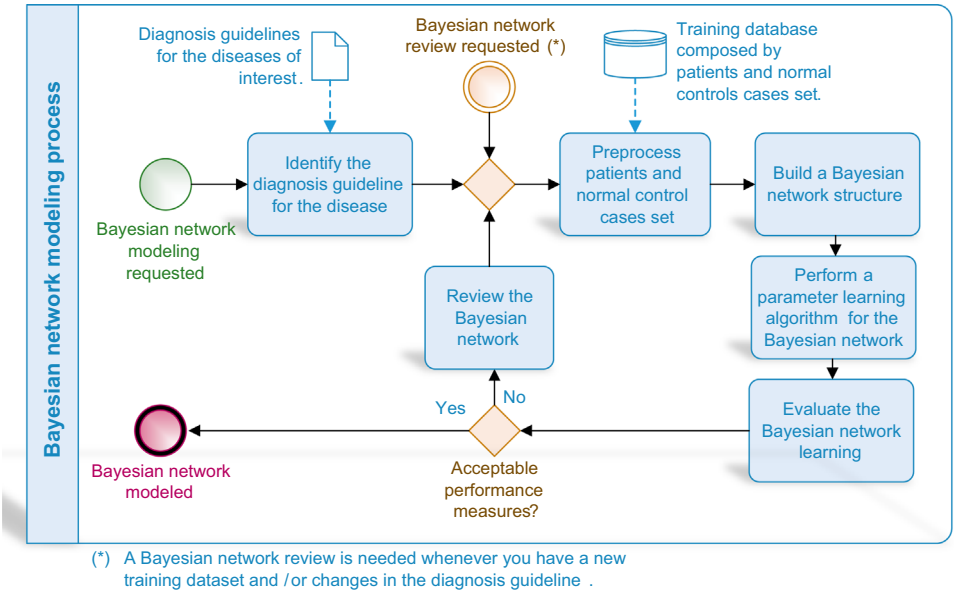


Fig. 4. Bayesian network modeling process.

were ordered by information gain score. We also removed attributes with missing values ratios greater than 70%. Values below such limit have measure of confidence within expected [76]. We carried out all those attribute filters aiming at improving the Bayesian learning algorithm performance.

The dataset attributes have data with numerical (or continuous) and multinomial (or binomial) types. Since the Bayesian learning algorithm and the inference engine used for modeling our BN are suitable for discrete random variables, those attributes values are converted to multinomial or binomial data types. Furthermore, some Bayesian classifiers have difficulty in dealing with continuous variables [77]. Using an objective function and a search algorithm, the discretization algorithm estimates the cutting points for numerical attributes, splitting them into well-defined numerical ranges covering the whole numerical domain. We used the discretization algorithm based on MDL (Minimum Description Length) described by Kononenko [78], using its implementation in WEKA.

We used a data-oriented approach as BN modeling. Fig. 4 shows the BN modeling process using BPMN (Business Process Modeling Notation)[2] workflow notation. The process starts with mapping the diagnosis criteria guideline for the disease of interest. Due to the data-oriented approach, the BN depends on the attributes from patients' case set. Furthermore, the healthcare concepts from diagnosis criteria are compared to concepts from BN random variables. The objective is building a BN decision model as close as possible to the diagnosis criteria. Then, the training dataset is preprocessed, in order to filter the attributes with high missing values ratio, or those that are not relevant for the diagnosis. After that, the BN is manually built with parameters estimated by a supervised learning algorithm from data. Then, the Bayesian learning is evaluated based on well-established performance metrics. If the performance measures are acceptable, the BN is modeled and ready to use. Otherwise, a System Analyst should review the BN. A decision model review is triggered whenever a domain expert or physician needs to add new findings to the decision model, or repeats the BN learning algorithm using a more complete patients' cases set.

Fig. 5 shows the diagnosis guideline for dementia, AD and MCI, represented also using BPMN notation. In a first moment, the general practitioner asks the patient about his/her clinical history and carries out clinical tests or exams for dementia screening. If the patient has possible dementia, the physician carries out a number of neuropsychological tests, to confirm the initial hypothesis of dementia. If the diagnosis of Dementia is confirmed, the

[2] http://www.bpmn.org.

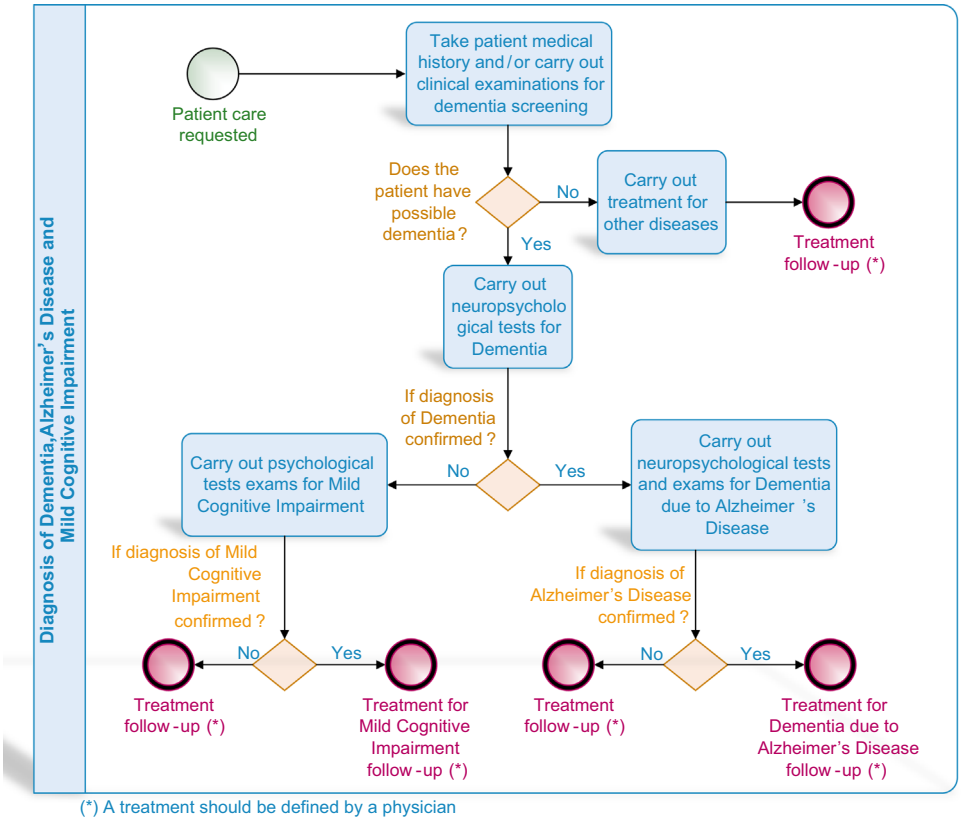


Fig. 5. Diagnosis process of dementia, Alzheimer's disease and mild cognitive impairment.

physician will carry out neuropsychological test batteries for AD, to investigate if the dementia is due to AD. Otherwise, the physician will carry out tests for MCI.

As shown in Fig. 5, since the decisions related to diagnosis are made in a non-concurrent way and at different time, we designed one BN for supporting each decision-making procedure. So, in contrast to Fig. 2, each BN has one target node. An interpretation of marginal probability distribution of the target node can indicate the uncertainty about a diagnosis. For example, if probability values are equilibrated (e.g. 50% for negative and 50% for positive diagnosis, respectively), one can consider such distribution as a maximum uncertainty. The uncertainty can be reduced by gathering more data from the patient, making them BN evidences.

In our BN model, we also included a utility node linked to the BN target node, and a decision box linked to the utility node, transforming such model into an Influence Diagram. As a decision model, the utility function purpose is modeling the relative cost to both healthcare system and patient's health perspectives, of a false-positive and false-negative diagnosis. We used symmetric values as parameters for the utility function. This point can be reviewed in a future work.

The Bayesian inference is based on statistical independence of X_i, given its parent node set, where X_i is the ith node from G. Hence, we evaluated the independence between two sets of attribute data associated with random variables that have the same parent node. A quantitative measure for evaluating the statistical independence is the correlation coefficient. As it approaches zero, they are closer to uncorrelated, meaning data independence. The closer the coefficient is to either -1 or 1, the stronger the correlation between the attribute data, meaning data dependence.

In the case of apparent dependence between a pair of nodes with a common parent node was identified, there might be a causal connection between them. As a result, the BN structure should be reviewed and an edge should be inserted between them. Fig. 6 shows the box plot of correlation coefficient distribution evaluated in pairs. The box plot is a graphical representation of data that shows a dataset's lowest value, highest value, average value, and the values of first and third quartile. In CERAD, most values were focused on zero, meaning that the attribute data are fairly independent. In CAD, values varied widely around zero, meaning that the attribute data are moderately independent. Despite both datasets having almost the same amount of attributes, CERAD has a number of instances significantly higher than CAD. Furthermore, we noticed some outliers in the box plot of both datasets, which lead us to the conclusion that attribute data under analysis are partially correlated. In order to carry out a more consistent statistical independence test, it would be better to consider a dataset with a higher number of instances. Moreover, although some pairs of names for neuropsychological tests seem to be similar (e.g. “Recall word list score” and “Word list recognition score”), all neuropsychological tests from both datasets (CERAD and CAD) assess different dimensions of cognition, quality of life and health of the patient, according to domain experts. Hence, such unexpected correlation between tests could be explained by the few number of instances available on the training dataset.

Figs. 7 and 8 show the BN for diagnosis of dementia and AD respectively, using patients' cases set from CERAD. Figs. 9–11 show the BN for diagnosis of dementia, AD and MCI respectively, using patients' cases set from CAD. The BN was drawn and evaluated using the GeNIe/SMILE[3] (Graphical Network Interface/Structural Modeling Inference and Learning Engine) authoring and inference

[3] http://genie.sis.pitt.edu/.

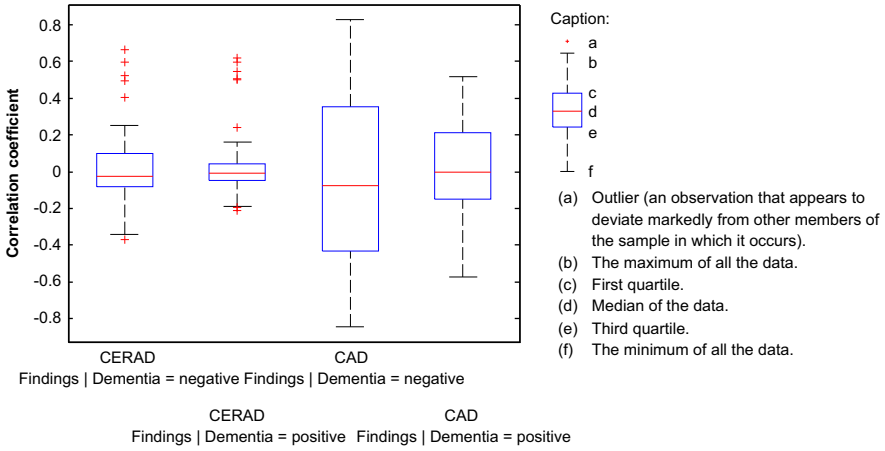


Fig. 6. Box plot for representing the correlation coefficient values.

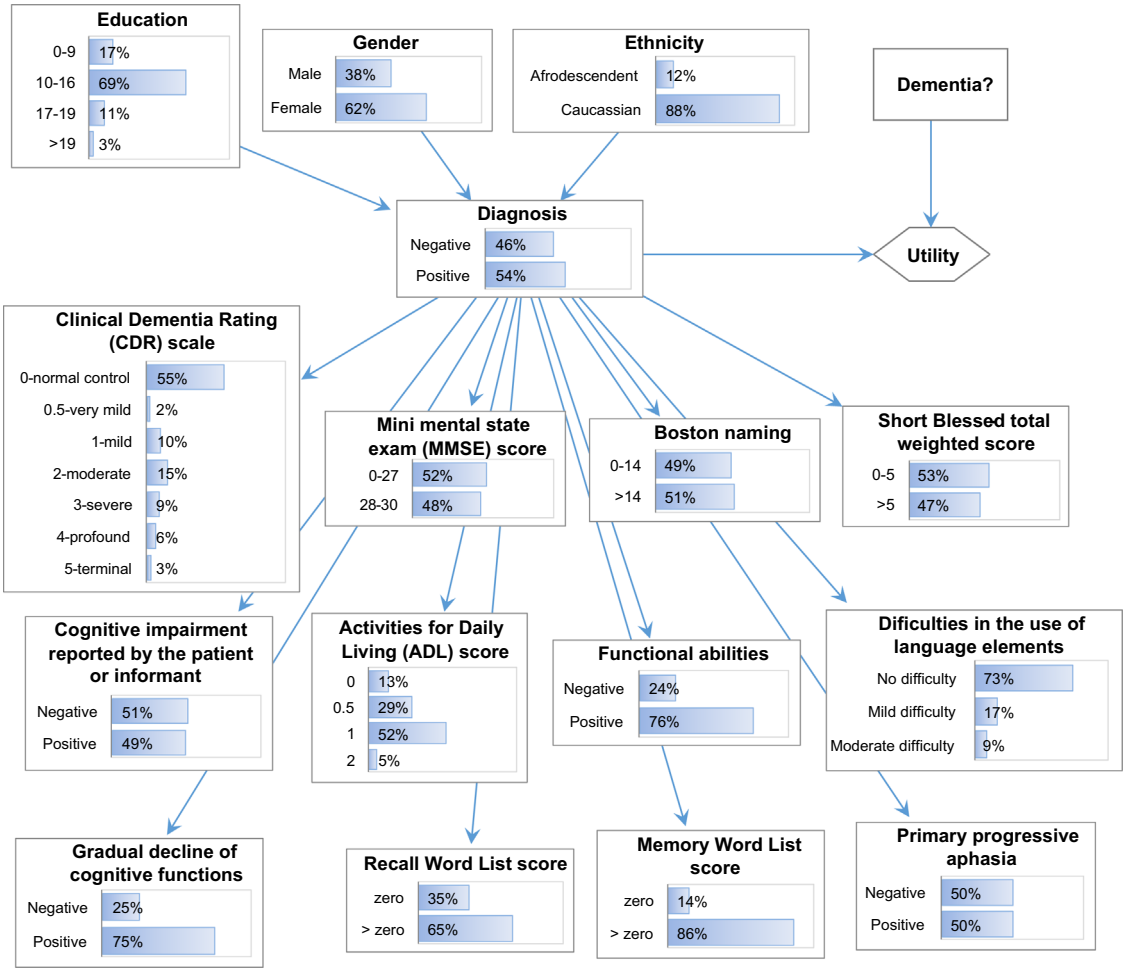


Fig. 7. Bayesian network for diagnosis of dementia based on CERAD patients' cases set.

toolbox developed by Pittsburg University. As mentioned before, the BN nodes stand for random variables that are associated to dataset attributes, as shown in Tables 2 and 3 for CERAD and CAD patients' dataset, respectively. In total, we propose five data-oriented BNs: two BNs for diagnosis of dementia and AD using CERAD patients' dataset, and three BNs for diagnosis of dementia, AD and MCI using CAD patients' dataset (Figs. 7–11).

The following section evaluates the Bayesian learning through performance measures.

5. Evaluation of results

Our BN was evaluated using performance measures for BN classification and measures for evaluating the BN robustness. The classification performance was evaluated using measures based on discrimination and based on probability, as summarized in Table 5. The discrimination measures evaluate how well the algorithm differentiates between two classes. We used the area under the ROC (Receiver Operating Characteristic) curve, known by the

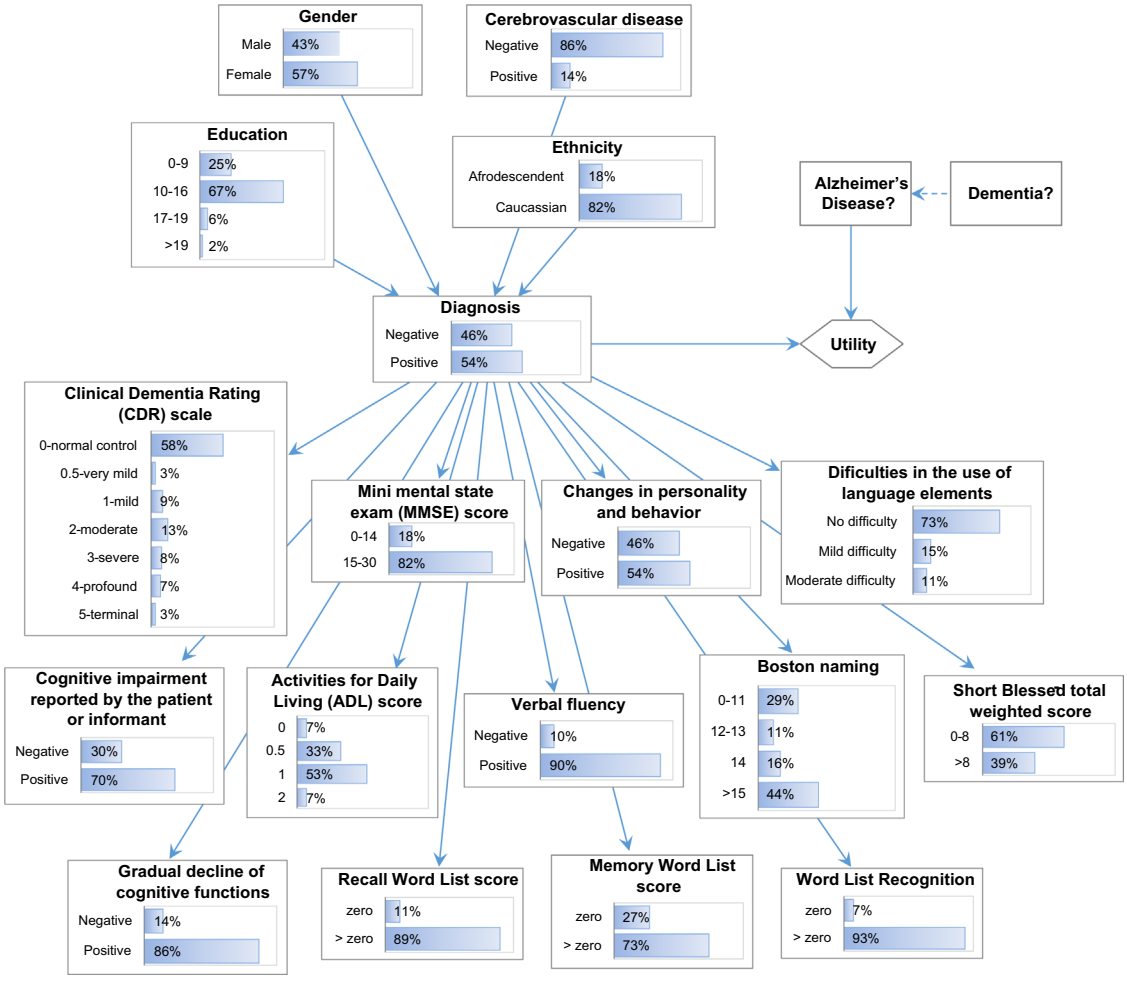


Fig. 8. Bayesian network for diagnosis of AD based on CERAD patients' cases set.

acronym AUC [79], and the F_1-score obtained by the harmonic mean of precision and recall [80]. The Bayesian classification rule is shown in equation below:

$$Z = \text{positive if } \Pr(z_{pos}|E) = \max_{i\,=\,neg,pos} \Pr(z_i|E), \tag{7}$$

where Z is the target node, which has two states for diagnosis, positive or negative, as mentioned before.

Let a target variable Y takes values in [0, 1] and let y_i be classification for a given case i. Let y_i=0 for negative state and y_i=1 for positive state, and $\Pr(y_i = 1)$ the probability of y_i being 1. So, the mean square error (MSE) for a dataset of n cases is defined as shown in equation below [81]:

$$MSE = \frac{1}{n} \cdot \sum_{i\,=\,1}^{n} [y_i - \Pr(y_i = 1)]^2 \tag{8}$$

The mean cross-entropy for n cases is calculated by Eq. (9) [82].

$$MXE = \frac{1}{n} \cdot \sum_{i\,=\,1}^{n} -y_i \cdot \log[\Pr(y_i = 1)] - (1 - y_i) \cdot \log[1 - \Pr(y_i = 1)] \tag{9}$$

The performance measures were estimated based on the k-folds cross-validation method [83]. Due to the number of cases of the CAD dataset (314 cases after carrying out the oversampling method), we used fourfolds for cross-validation, i.e., each fold contained around 78 cases. Fig. 12 shows a comparison of performance measures' results for BN and other well-known classifiers found in the literature: (1) Näive Bayes [84], Logistic regression model [85], Multilayer Perceptron ANN [82], Decision table [83], Decision stump optimized by AdaboostM1 algorithm [86] and J48 Decision tree [87]. Those classifier implementations were available in the WEKA data mining tool. Table 6 synthesizes the parameters used by classifiers.

The purpose of the comparison of performance results shown in Fig. 12 is to show that BN's performance is very close to other well known classifiers, not to prove that they give best results. The comments about the results will be addressed in Section 6.

We also compared the performance results for our BNs, shown in the previous section, to corresponding BNs with structure automatically discovered from the CAD dataset. Such comparison results are shown in Fig. 13. The discovery method used for estimating the BN structure and parameters was described by Dash and Druzdzel [88]. Such method is based on the independence test between dataset attributes, and applies the Greedy Thick Thinning algorithm as a search method. Its implementation is available on the GeNIe/SMILE authoring and inference toolbox [89]. Such discovered topologies of BNs are shown in Figs. 14–16 for diagnosis of dementia, AD and MDI, respectively. We will discuss their results in Section 6.

We carried out a sensitivity analysis for testing the robustness of the results of our proposed BNs. Its purpose is to test how the uncertainty in the output can be impacted by the evidence (partial observations) in the input of BN [27]. Entropy is the most common measure used in sensitivity analysis to evaluate uncertainty. We measured the entropy $H(X)$ in the target node X, where X represents the probability of the patient to have the disease of interest (positive for diagnosis of disease). The target node is

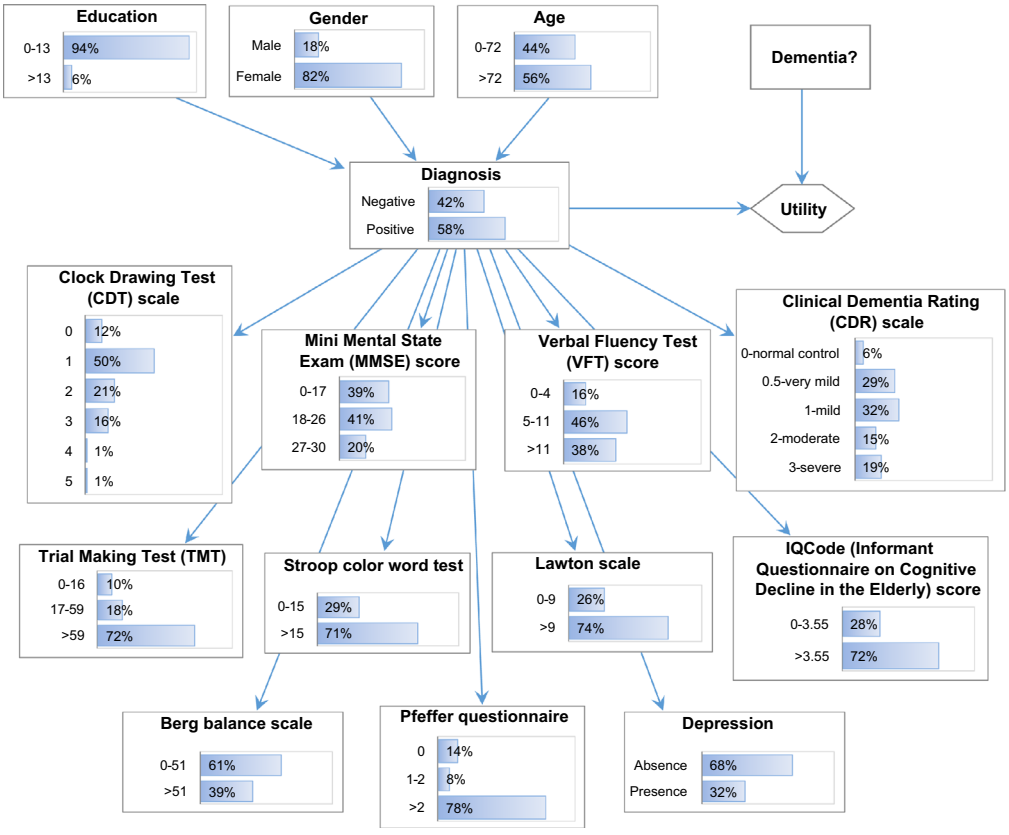


Fig. 9. Bayesian network for diagnosis of dementia based on CAD patients' cases set.

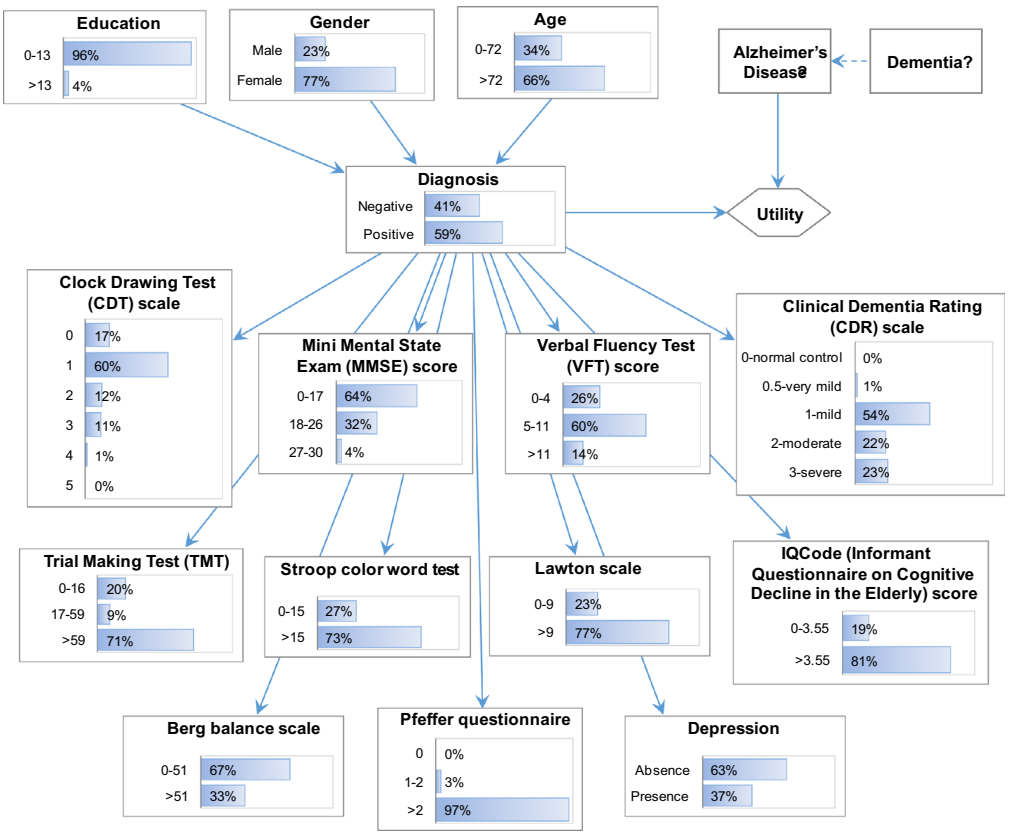


Fig. 10. Bayesian network for diagnosis of AD based on CAD patients' cases set.

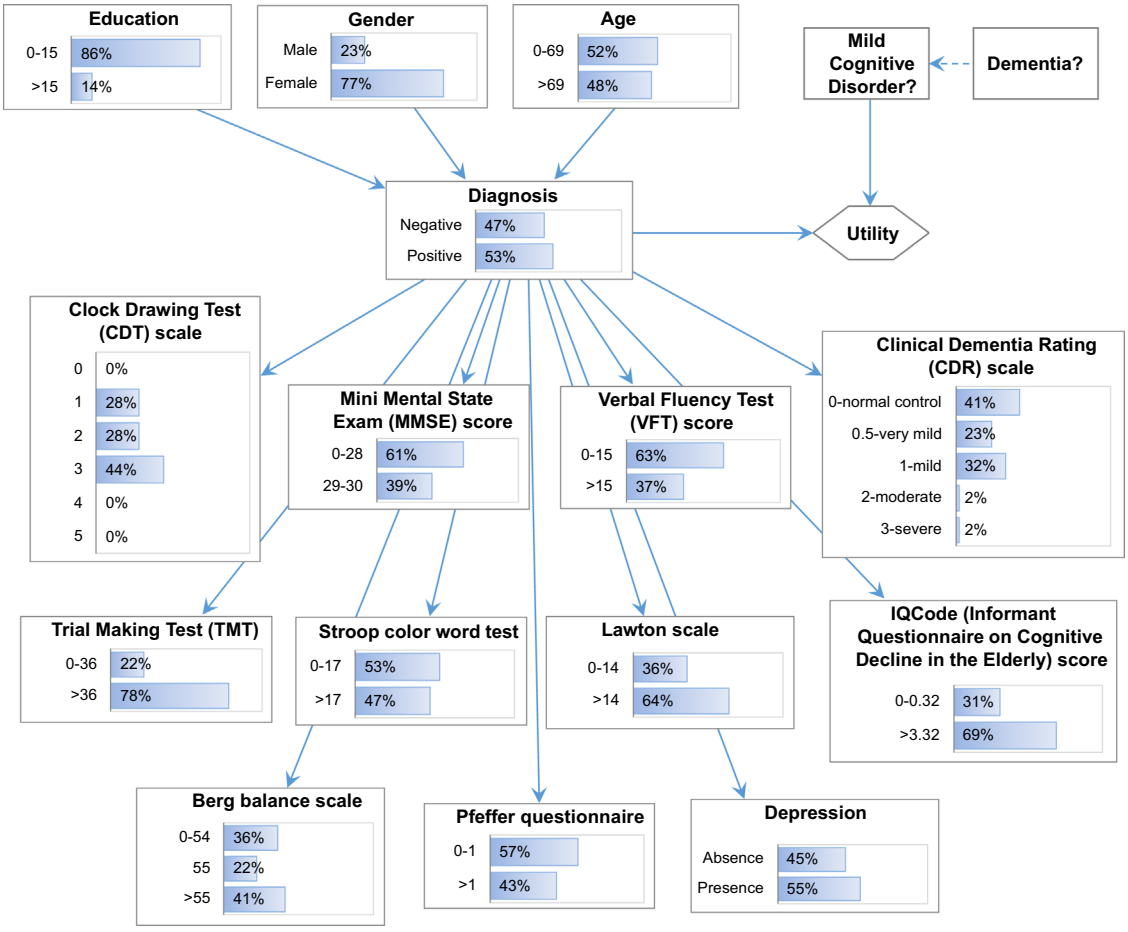


Fig. 11. Bayesian network for diagnosis of MCI based on CAD patients' cases set.

Table 5
Performance measures used to evaluate the proposed BNs.

Performance measure	Acronym	Domain	Best score
Area under ROC curve	AUC	[0, 1]	1
Harmonic mean of precision and recall	F_1	[0, 1]	1
Mean square error	MSE	[0, 1]	0
Mean cross-entropy	MXE	$[0, \infty)$	0

labeled as "Diagnosis" in the BN, as shown in Figs. 9–11, related to diagnosis of dementia, AD and MCI, respectively

Table 7 shows the sensitivity analysis results using BNs modeled from the CAD dataset. The entropy values were obtained in the target node, calculated after performing the inference algorithm using each piece of evidence alone. Then, the evidence was grouped and sorted in decreasing order by entropy value, to understand which evidence was more relevant to corresponding positive or negative diagnosis for the disease. An evidence was considered in a positive diagnosis group if $\Pr(X = Positive|E) > 0.5$, where E stands for corresponding evidence, otherwise it was considered in a negative diagnosis group.

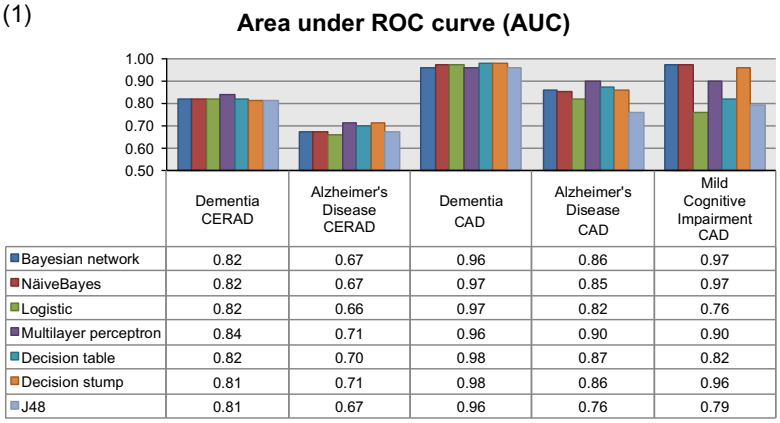
6. Discussion

Regarding the comparison of our BN performance results and other well-known classifiers for the CERAD dataset (Fig. 12), the best results for AUC were obtained by the Multilayer Perceptron ANN classifier (0.84 and 0.71 for dementia and AD, respectively). Although ANNs have the ability to learn complex patterns directly from observations, their reasoning process is inaccessible to human understanding, which is hard for physicians to accept the advice from CDSS without knowing the basis for the system decision [90]. The ANNs are commonly treated as a black-box, not offering a knowledge representation that allows domain experts to criticize the diagnosis criteria. On the other hand, BN allows the diagnostic criteria to be represented graphically using a human-oriented causal diagram that facilitates the communication between domain experts and knowledge engineers. The Decision Stump algorithm optimized by AdaboostM1 showed the best results for F_1-score, MSE and MXE.

Regarding the comparison of our BN performance results for the CAD dataset, evaluating for AUC, Decision table and Decision stump classifiers showed the best results for dementia (0.98). The Multilayer Perceptron classifier showed the best results for AD (0.92). BN and Näive Bayes showed the best results for MCI (0.97), which may indicate very promising results by the usage of probabilistic causal models. As mentioned before, MCI is a clinical condition often related to a preclinical stage of AD. Scientific community efforts are driven by characterizing earliest stages of AD, when its treatment is more efficient. For F_1-score, the best results were obtained by Decision Table (0.95), Multilayer Perceptron (0.86) and BN (0.90) for dementia, AD and MCI, respectively. For MSE and MXE, occurrences of best result were verified among Decision stump, Multilayer Perceptron, Logistic regression and Näive Bayes.

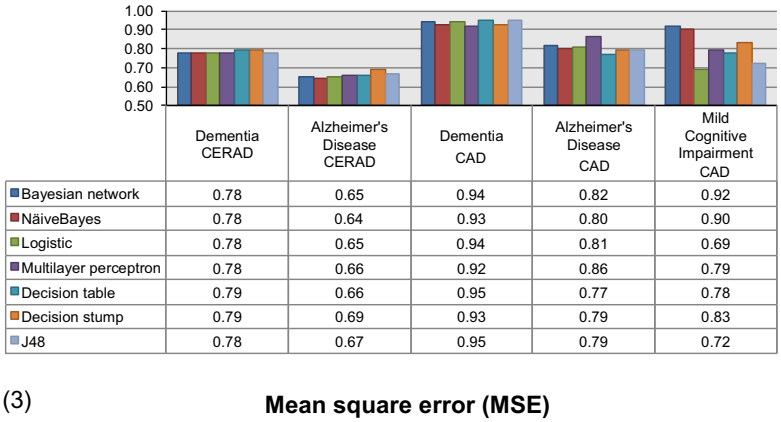
Comparing our proposed BN performance results to the automatically discovered structure BN results, our BN showed results very close to those obtained by BN with structure automatically

(1)



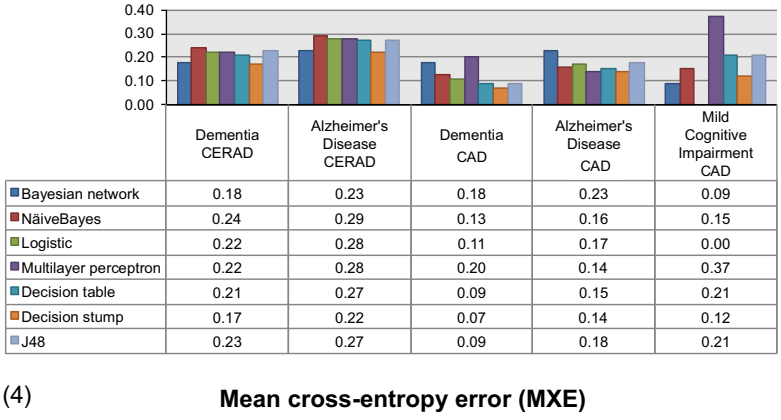
	Dementia CERAD	Alzheimer's Disease CERAD	Dementia CAD	Alzheimer's Disease CAD	Mild Cognitive Impairment CAD
Bayesian network	0.82	0.67	0.96	0.86	0.97
NäiveBayes	0.82	0.67	0.97	0.85	0.97
Logistic	0.82	0.66	0.97	0.82	0.76
Multilayer perceptron	0.84	0.71	0.96	0.90	0.90
Decision table	0.82	0.70	0.98	0.87	0.82
Decision stump	0.81	0.71	0.98	0.86	0.96
J48	0.81	0.67	0.96	0.76	0.79

(2)



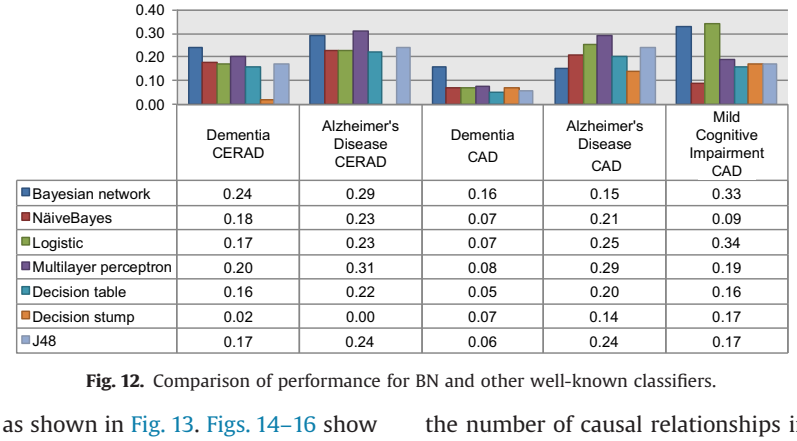
	Dementia CERAD	Alzheimer's Disease CERAD	Dementia CAD	Alzheimer's Disease CAD	Mild Cognitive Impairment CAD
Bayesian network	0.78	0.65	0.94	0.82	0.92
NäiveBayes	0.78	0.64	0.93	0.80	0.90
Logistic	0.78	0.65	0.94	0.81	0.69
Multilayer perceptron	0.78	0.66	0.92	0.86	0.79
Decision table	0.79	0.66	0.95	0.77	0.78
Decision stump	0.79	0.69	0.93	0.79	0.83
J48	0.78	0.67	0.95	0.79	0.72

(3)



	Dementia CERAD	Alzheimer's Disease CERAD	Dementia CAD	Alzheimer's Disease CAD	Mild Cognitive Impairment CAD
Bayesian network	0.18	0.23	0.18	0.23	0.09
NäiveBayes	0.24	0.29	0.13	0.16	0.15
Logistic	0.22	0.28	0.11	0.17	0.00
Multilayer perceptron	0.22	0.28	0.20	0.14	0.37
Decision table	0.21	0.27	0.09	0.15	0.21
Decision stump	0.17	0.22	0.07	0.14	0.12
J48	0.23	0.27	0.09	0.18	0.21

(4)



	Dementia CERAD	Alzheimer's Disease CERAD	Dementia CAD	Alzheimer's Disease CAD	Mild Cognitive Impairment CAD
Bayesian network	0.24	0.29	0.16	0.15	0.33
NäiveBayes	0.18	0.23	0.07	0.21	0.09
Logistic	0.17	0.23	0.07	0.25	0.34
Multilayer perceptron	0.20	0.31	0.08	0.29	0.19
Decision table	0.16	0.22	0.05	0.20	0.16
Decision stump	0.02	0.00	0.07	0.14	0.17
J48	0.17	0.24	0.06	0.24	0.17

Fig. 12. Comparison of performance for BN and other well-known classifiers.

discovered from the dataset, as shown in Fig. 13. Figs. 14–16 show the BN automatically discovered from the CAD dataset for diagnosis of dementia, AD and MCI, respectively. We can notice that the number of causal relationships in BN discovered structures is higher than BNs with structure built manually (Figs. 9–11). Furthermore, when evaluating the Bayesian structure discovered,

Table 6
List of parameters used by classifiers.

Classifier	Parameters description for performance evaluation
Näive Bayes	Do not use a Kernel estimator for numerical attributes Do not use a supervised discretization to convert numeric attributes to nominal ones
Logistic regression	Logistic regression model with a ridge estimator. The missing values are replaced by nominal values. Maximum number of iterations equal to 1 Ridge value set in the log-likelihood equal to 10-8
Multilayer perceptron	Do not add and connect up hidden layers in the network Do not decrease the learning rate by the epoch number Set the number of hidden layers=(number of attributes + number of classes)/2 Initial learning rate set to 0.3 Apply 0.2 as momentum to weights during the network updating Normalize the nominal attributes to values between −1 and 1 Maximum number of epoch equal to 500 Allow the network to reset with a lower learning rate The percentage size of validation set equal to 20
Decision table	Use as validation the leave-one-out method User accuracy as evaluation measure Use Best-First algorithm as search method. The Best-First uses as heuristics method the greedy hill-climbing with backtracking The maximum amount of backtracking is 5
Decision stump	No parameters to configure
J48 Decision tree	Do not use binary splits on nominal attributes when building the trees The confidence factor used for pruning (smaller values incur more pruning) set to 0.25 The minimum number of instances per leaf set to 2 Amount of data used for reduced-error pruning set to 3 Consider the sub-tree raising operation when pruning

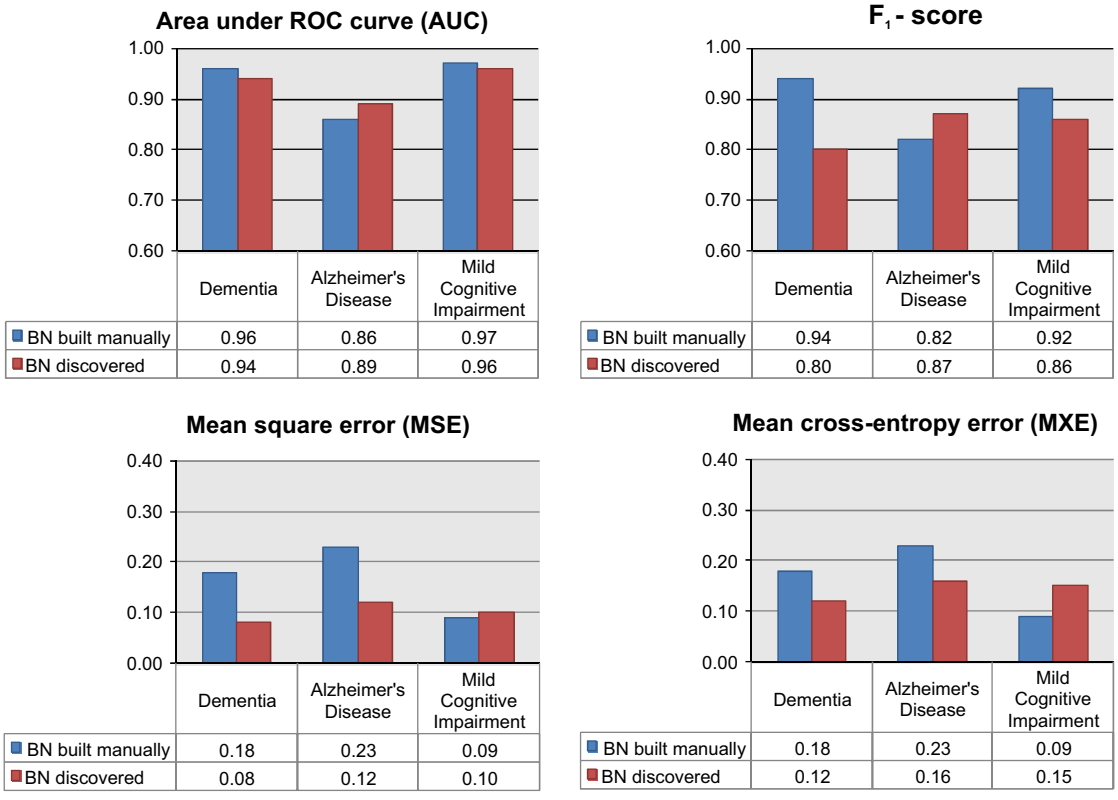


Fig. 13. Comparison of performance measures results for BN with structure built manually and BN discovered from data.

it is possible to conceive some Bayesian nodes of symptoms and neuropsychological tests being addressed as causes of the disease, instead of consequences. For example, Fig. 14 reveals that dementia is caused by the results of CDR (Clinical Dementia Rating), because of the directed arc linking both BN nodes. However, this is incorrect from the semantic point of view. There are many similar examples in Figs. 15 and 16. Hence, BNs with structure built manually are simpler and more readable for physicians and domain experts than BNs discovered from dataset. Therefore, despite showing a slight improvement of some performance

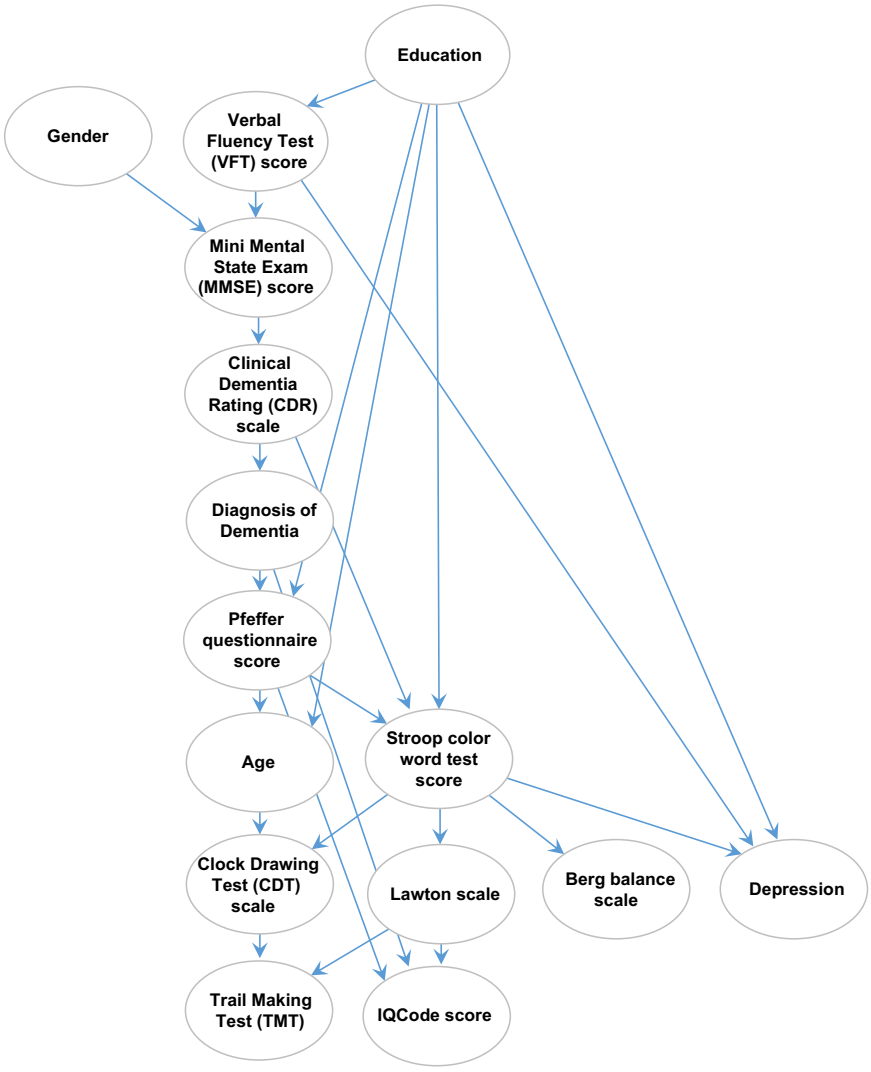


Fig. 14. Bayesian network for diagnosis of dementia discovered from data.

results, BN manually built structure might be more interesting for maintaining the decision model readability for human experts.

Aiming at evaluating sensitivity analysis results (Table 7), we asked a physician expert to evaluate each BN random variable in terms of its corresponding relevance to the diagnosis of each disease, dementia, AD and MCI, classifying each evidence into three levels of relevance: (1) not relevant, (2) moderately relevant and (3) very relevant. Then, we compared the neuropsychological tests relevance classification done by the physician expert to the sensitivity analysis results. Neuropsychological tests CDR and VFT were classified in both CDSS and by domain expert as highly relevant for diagnosis of dementia. However, CDT and TMT were classified as moderately relevant by the domain expert. Regarding the diagnosis of AD, Pfeffer appeared in both classifications, CDSS and by domain expert, as highly relevant. However, TMT and IQCode were classified as moderately relevant by the domain expert. All neuropsychological tests shown in the sensitivity analysis for MCI diagnosis (TMT, CDT, Berg and Stroop) were classified as moderately relevant by the domain expert. There were no cases of total disagreement, such as a neuropsychological test shown in sensitivity analysis as highly relevant and classified as not relevant by the domain expert.

Furthermore, there were neuropsychological tests classified as highly relevant and moderately relevant by domain expert, which, however, were removed from BN due to its missing data ratio exceeding an acceptable threshold. For example, the CamCog (Cambridge Cognition) exam was classified as highly relevant by domain expert and was excluded in a preprocessing data phase due to its high missing data ratio (77%) in the dataset used. The processing data stage of BN modeling was described in Section 4. NPI (Neuropsychiatric Inventory), Digit Symbol, Digit Spam and Complex Figure are other examples of neuropsychiatric exams classified as moderately relevant for the diagnosis of diseases of interest that were excluded due to their high missing data ratio. Although these results are promising, we expect to obtain even better results with a more complete patient dataset.

Regarding the statistical independence test, despite the neuropsychological tests being different, we noticed some outliers in the box plot shown in Fig. 6. The cause might be use of a training database with too few instances. As future work, we will address a statistical independence test considering more instances and a more complete dataset.

7. Conclusion

This paper proposed a Bayesian Network (BN) decision model for supporting diagnosis of dementia, AD and MCI. Such diseases

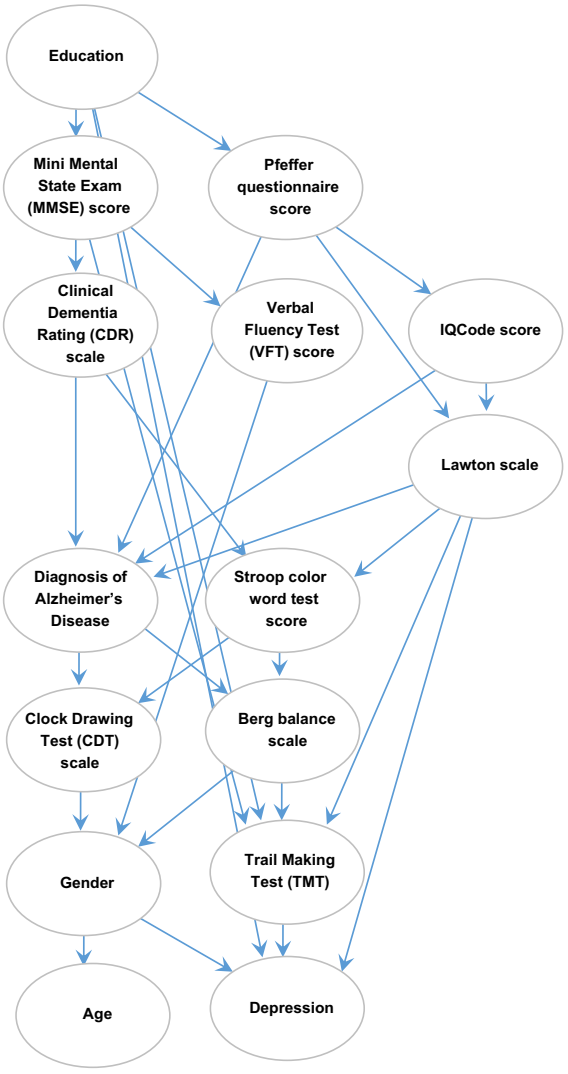


Fig. 15. Bayesian network for diagnosis of AD discovered from data.

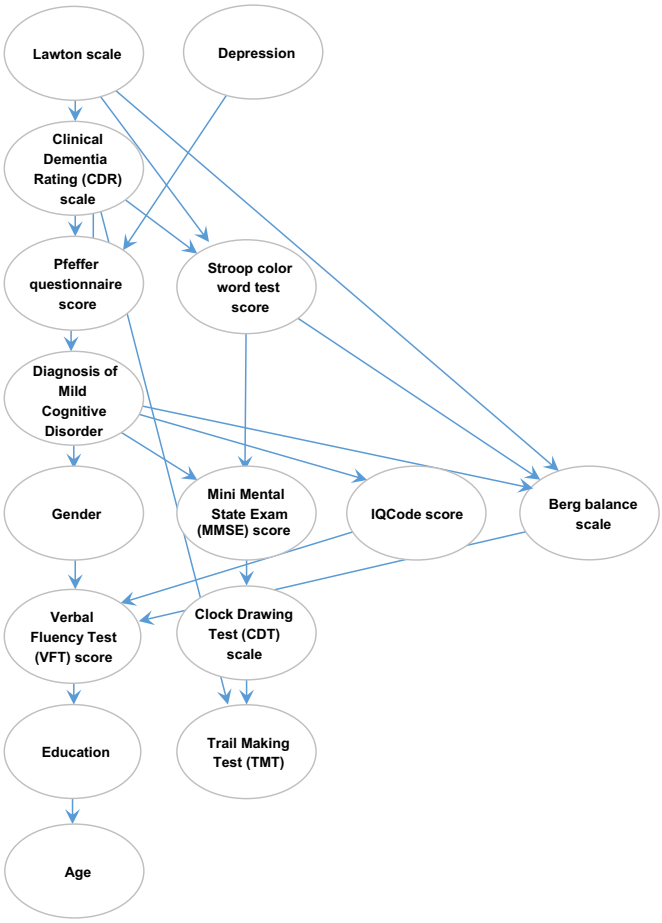


Fig. 16. Bayesian network for diagnosis of MCI discovered from data.

are relevant due to global population the aging phenomenon, their high prevalence among elderly population and their high impact on the family, community and health care system. The proposed decision model can be used to build clinical decision support systems to diagnose such diseases.

The proposed decision model uses a probabilistic approach, where symptoms, signs, test results and background information are associated to random variables linked to each other, giving rise to a causal diagram. BNs can be represented graphically, which facilitates their readability by a domain expert. The proposed BN modeling involved a domain expert elicitation and a learning phase using an available dataset, which makes the decision model more robust and reliable. BN has the ability to deal with partial observations and uncertainty, makes the model suitable for clinical context.

Our BN modeling was data-oriented, i.e., random variables were derived from data attributes of patients' dataset. The BN structure was built manually based on a three-level generic BN structure. The probabilities distribution was estimated from the dataset using the EM supervised learning algorithm. The patients' datasets were provided by two different institutions: CERAD and CAD. In total, five BNs were designed, one for each disease and institution responsible for the patients' dataset.

The main contribution of this paper was the development of a decision model for clinical diagnosis using BN. When comparing the performance results obtained by BN to other well-known classifiers, BNs have shown better results for diagnosing MCI and competitive results for dementia and AD. Another contribution was a description of a BN modeling process. Such a process can be extended to other disease domains.

As future work, we intend to revise the BN using a more complete patients' dataset, including neuropsychological tests that are relevant for diagnosis of the diseases of interest and that were not available in the datasets used. Although it was not described in this paper, the proposed decision model was used for developing a web-based CDSS. We shall also extend the CDSS including other diseases related to aging, using BN modeling.

Another goal is to deploy the CDSS in a clinical environment and evaluate acceptance and feedback reported by physicians. The adoption and more extensive use of CDSS depend on a number of technological developments, including more widespread use of EMRs (Electronic Medical Records) capabilities, development of technologies for healthcare providers to share information across entities, and cheaper, faster or more flexible technology. We are currently working in clinical data modeling using an open-source and multi-level approach.

Conflict of interest statement

None declared.

Table 7
Sensitivity analysis considering the BN for Dementia, Alzheimer's Disease and Mild Cognitive Impairment using the CAD dataset.

Disease[a]	Diagnosis (X)	Evidence		Entropy $H(X)$[d]
		Random variable[b]	State: value[c]	
Dementia	Negative	CDR	1: 0	0.00
		Pfeffer	1: 0	0.00
		MMSE	3: [27–30]	0.27
		Lawton	3: [0–9]	0.55
	Positive	CDT	1: 0	0.00
		VFT	1: [0–4]	0.00
		TMT	1: [0–16]	0.00
		CDR	5: 3	0.18
Alzheimer's Disease (AD)	Negative	CDR	1: 0	0.41
		Lawton	1: [0–9]	0.43
		Stroop	1: [0–15]	0.80
	Positive	Pfeffer	1: 0	0.23
		TMT	1: [0–16]	0.53
		IQCode	1: [0–3.55]	0.65
Mild Cognitive Impairment (MCI)	Negative	Berg	1: [0–54]	0.04
		Lawton	1: [15–inf]	0.06
		CDR	4: 2	0.16
		IQCode	2: [3.33–inf]	0.18
	Positive	TMT	1: [0–36]	0.00
		CDT	1: 0	0.00
		Berg	1: [56–inf]	0.02
		Stroop	1: [0–17]	0.09

[a] Each disease was associated to a Bayesian network.
[b] Each random variable was associated to a Bayesian network node. Table 3 shows their descriptions.
[c] The random variable states are also described in Table 3.
[d] The entropy for a binary random variable ranges from zero to one. Zero stands for minimum uncertainty. One stands for maximum uncertainty.

Acknowledgments

We thank the Center of the Consortium to Establish a Registry for Alzheimer's Disease for kindly providing the patients' cases set used in this study. This work has been partially supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Pesquisa) for supporting J.L., A.C. and D.C.M.S., and Brazilian agencies to promote scientific research (FAPERJ). This work is part of INCT-MACC (National Institute of Science and Technology for Medicine Assisted by Scientific Computing) and SIADE project at CEPE (Centro de Estudo e Pesquisa do Envelhecimento).

References

[1] A.L. Sosa-Ortiz, I. Acosta-Castillo, M.J. Prince, Epidemiology of dementias and Alzheimer's disease, Arch. Med. Res. 43 (8) (2012) 600–608.
[2] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (dsm-iv-tr), 2000.
[3] Alzheimer's Association, Alzheimer's Disease Facts and Figures, Technical Report 2, Alzheimer's & Dementia, 2013.
[4] L. Hebert, P. Scherr, J. Bienias, D. Bennett, D. Evans, Alzheimer disease in the us population: prevalence estimates using the 2000 census, Arch. Neurol. 60 (8) (2003) 1119.
[5] R. Nitrini, C.M. Bottino, C. Albala, N.S. Custodio Capunay, C. Ketzoian, J.J. Llibre Rodriguez, G.E. Maestre, A.T. Ramos-Cerqueira, P. Caramelli, Prevalence of dementia in Latin America: a collaborative study of population-based cohorts, Int. Psychogeriatr. 21 (4) (2009) 622–630, using Smart Source Parsing August Epub 2009, June 9.
[6] C.M.C. Bottino, D. Azevedo, M. Tatsch, S.R. Hototian, M.A. Moscoso, J. Folquitto, A.Z. Scalco, M.C. Bazzarella, M.A. Lopes, J. Litvoc, Estimate of dementia prevalence in a community sample from sao paulo, Dement. Geriatr. Cogn. Disord. 26 (4) (2008) 291–299.
[7] J. Morris, M. Storandt, Detecting early-stage Alzheimer's disease in MCI and PREMCI: the value of informants, in: Research and Perspectives in Alzheimer's Disease, Springer, Berlin, Heidelberg, 2006, pp. 393–397.
[8] B. Dubois, H.H. Feldman, C. Jacova, S.T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, K. Meguro, J. O'Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P.J. Visser, P. Scheltens, Research criteria for the diagnosis of alzheimer's disease: revising the NINCDS-ADRDA criteria, Lancet Neurol. 6 (2007) 734–746.
[9] R. Petersen, Mild Cognitive Impairment: Aging to Alzheimer's Disease, Oxford University Press, USA, 2003.
[10] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, A. Gamst, D.M. Holtzman, W.J. Jagust, R.C. Petersen, P.J. Snyder, M.C. Carrillo, B. Thies, C.H. Phelps, The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's Dement.: J. Alzheimer's Assoc. 7 (3) (2011) 270–279.
[11] P. Celsis, Age-related cognitive decline, mild cognitive impairment or pre-clinical Alzheimer's disease? Ann. Med. 32 (1) (2000) 6–14.
[12] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan, Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Workgroup under the auspices of Department of Health and Human Services task force on Alzheimer's disease, Neurology 34 (7) (1984) 939–944.
[13] C.R. Jack, M.S. Albert, D.S. Knopman, G.M. McKhann, R.A. Sperling, M.C. Carrillo, B. Thies, C.H. Phelps, Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's Dement.: J. Alzheimer's Assoc. 7 (3) (2011) 257–262.
[14] G.M. McKhann, D.S. Knopman, H. Chertkow, B.T. Hyman, J. Jack, C. R, C.H. Kawas, W.E. Klunk, W.J. Koroshetz, J.J. Manly, R. Mayeux, R.C. Mohs, J.C. Morris, M.N. Rossor, P. Scheltens, M.C. Carrillo, B. Thies, S. Weintraub, C.H. Phelps, The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's Dement. 7 (3) (2011) 263–269.
[15] D. Newman-Toker, P. Pronovost, Diagnostic errors: the next frontier for patient safety, JAMA: J. Am. Med. Assoc. 301 (10) (2009) 1060–1062.
[16] M. Graber, N. Franklin, R. Gordon, Diagnostic error in internal medicine, Arch. Intern. Med. 165 (13) (2005) 1493.
[17] R. Miller, Computer-assisted diagnostic decision support: history, challenges, and possible paths forward, Adv. Health Sci. Educ. 14 (2009) 89–106.
[18] E.S. Berner, Clinical Decision Support Systems: Theory and Practice, Springer, Birminghain, England, 2007.
[19] R.B. Haynes, N.L. Wilczynski, Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: methods of a decision-maker-researcher partnership systematic review, Implement. Sci. 5 (12) (2010) 12, computerized Clinical Decision Support System (CCDSS) Systematic Review Team Using Smart Source Parsing, February 5.
[20] H. Lindgren, Decision support in dementia care: developing systems for interactive reasoning (Ph.D. thesis), 2007.
[21] A. Wright, D.F. Sittig, J.S. Ash, J. Feblowitz, S. Meltzer, C. McMullen, K. Guappone, J. Carpenter, J. Richardson, L. Simonaitis, R.S. Evans, W.P. Nichol, B. Middleton, Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems, J. Am. Med. Inf. Assoc. 18 (3) (2011) 232–242, using Smart Source Parsing May 1, Epub 2011 Mar 17..
[22] G. Kong, D. Xu, J. Yang, Clinical decision support systems: a review on knowledge representation and inference under uncertainties, Int. J. Comput. Intell. Syst. 1 (2) (2008) 159–167.
[23] S.M. Weiss, C.A. Kulikowski, A. Safir, A model-based consultation system for the long-term management of glaucoma, Spec. Syst. 3 (1976) 826–832.
[24] R.A. Miller, M.A. McNeil, S.M. Challinor, J. Masarie, F. E, J.D. Myers, The internist-1/quick medical reference project: status report, West J. Med. 145 (6) (1986) 816–822, using Smart Source Parsing December.
[25] B. Buchanan, E. Shortliffe, Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley Reading, MA, 1984.
[26] G. Beliakov, J. Warren, Fuzzy logic for decision support in chronic care, Artif. Intell. Med. 21 (1) (2001) 209–213.
[27] K. Korb, A. Nicholson, Bayesian Artificial Intelligence, Chapman & Hall/CRC, Clayton, Victoria, Australia, 2004.
[28] M.B. do Amaral, Y. Satomura, M. Honda, T. Sato, A psychiatric diagnostic system integrating probabilistic and categorical reasoning, Methods Inf. Med. 34 (3) (1995) 232–243.
[29] D. Salas-Gonzalez, J.M. Garriz, J. Ramírez, M. Lopez, I. Alvarez, F. Segovia, R. Chaves, C.G. Puntonet, Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees, Phys. Med. Biol. 55 (2010) 2807–2817.
[30] E.E. Tripoliti, D.I. Fotiadis, M. Argyropoulou, A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment, Artif. Intell. Med. 53 (1) (2011) 35–45, using Smart Source Parsing September; http://dx.doi.org/10.1016/j.artmed.2011.05.005 Epub 2011 June 23.
[31] M. Musen, S.W. Tu, A.K. Das, Y. Shahar, Eon: a component-based approach to automation of protocol directed therapy, J. Am. Med. Inf. Assoc. 3 (6) (1996) 367–388.

[32] Y. Shahar, S. Miksch, P. Johnson, The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines, Artif. Intell. Med. 14 (1) (1998) 29–51.
[33] M. Peleg, A.A. Boxwala, O. Ogunyemi, Q. Zeng, S. Tu, R. Lacson, E. Bernstam, N. Ash, P. Mork, L. Ohno-Machado, E.H. Shortliffe, R.A. Greenes, Glif3: the evolution of a guideline representation format, in: Proceedings of AMIA Symposium, 2000, p. 645–649 (Using Smart Source Parsing 645–649).
[34] P.D. Johnson, S. Tu, N. Booth, B. Sugden, I. N. Purves, Using scenarios in chronic disease management guidelines for primary care, in: Proceedings of the AMIA Symposium, 2000, pp. 389–393.
[35] J. Fox, V. Patkar, R. Thomson, Decision support for health care: the proforma evidence base, Informatics in Primary Care 14.
[36] M. Ceccarelli, A. Donatiello, D. Vitale, Kon: A clinical decision support system, in oncology environment, based on knowledge management, in: 20th IEEE International Conference on Artificial Intelligence Tools with Artificial Intelligence, 2008, ICTAI '08, vol. 2, 2008, pp. 206–210.
[37] S. Mani, S. McDermott, M. Valtorta, Mentor: a Bayesian model for prediction of mental retardation in newborns, Res. Dev. Disabil. 18 (5) (1997) 303–318.
[38] F.J. Diez, J. Mira, E. Iturralde, S. Zubillaga, Diaval: a Bayesian expert system for echocardiography, Artif. Intell. Med. 10 (1) (1997) 59–73.
[39] G. Sakellaropoulos, G. Nikiforidis, Development of a Bayesian network for the prognosis of head injuries using graphical model selection techniques, Methods Inf. Med. 38 (1999) 37–42.
[40] E. Burnside, D. Rubin, R. Shachter, A Bayesian network for mammography, in: Proceedings of the AMIA Symposium, American Medical Informatics Association, 2000, pp. 106–110.
[41] D. Aronsky, P.J. Haug, Automatic identification of patients eligible for a pneumonia guideline, in: Proceedings of the AMIA Symposium, vol. 1067, American Medical Informatics Association, pp. 12–16.
[42] S. Yan, L. Shipin, T. Yiyuan, Construction and application of Bayesian network in early diagnosis of Alzheimer disease's system, in: 2007 IEEE/ICME International Conference on Complex Medical Engineering, CME 2007, 2007, pp. 924–929.
[43] B. Wemmenhove, J. Mooji, W. Wiegerinck, M. Leisink, H. Happen, J. Neijt, Inference in the promedas medical expert system, Artif. Intell. Med. (2007) 456–460.
[44] D. Luciani, S. Cavuto, L. Antiga, M. Miniati, S. Monti, M. Pistolesi, G. Bertolini, Bayes pulmonary embolism assisted diagnosis: a new expert system for clinical use, Emerg. Med. J. 24 (3) (2007) 157–164.
[45] C. Pyper, M. Frize, G. Lindgaard, Bayesian-based diagnostic decision-support for pediatrics, in: 30th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, 2008, EMBS 2008, 2008, pp. 4318–4321.
[46] G. Czibula, I.G. Czibula, G.S. Cojocar, A.M. Guran, Imasc—an intelligent multiagent system for clinical decision support, in: First International Conference on Complexity and Intelligence of the Artificial and Natural Complex Systems, Medical Applications of the Complex Systems, Biomedical Computing, CANS '08, 2008, pp. 185–190.
[47] Y. Denekamp, M. Peleg, TiMeDDX—a multi-phase anchor-based diagnostic decision-support model, J. Biomed. Inf. 43 (2010) 111–124.
[48] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Francisco, the United States of America, 1988.
[49] B. Middleton, M. Shwe, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, G. Cooper, Probabilistic diagnosis using a reformulation of the internist-1/qmr knowledge base, Medicine 30 (1991) 241–255.
[50] A. Onisko, M.J. Druzdzel, H. Wasyluk, Learning Bayesian network parameters from small data sets: application of noisy-or gates, Int. J. Approx. Reason. 27 (2) (2001) 165–182.
[51] M. Pradhan, G. Provan, B. Middleton, M. Henrion, Knowledge engineering for large belief networks, in: Proceedings of the Seventh Conference of Uncertainty in Artificial Intelligence, San Francisco, 1994, pp. 484–490.
[52] A. Zagorecki, M. Druzdzel, An empirical study of probability elicitation under noisy-or assumption, in: Proceedings of the Seventeenth International on Artificial Intelligence Research Society Conference, Florida, FLAIRS 2004, 2004, pp. 880–885.
[53] M. Singh, M. Valtorta, Construction of Bayesian network structures from data: a brief survey and an efficient algorithm, Int. J. Approx. Reason. 12 (2) (1995) 111–131.
[54] A. Montanari, T. Rizzo, How to compute loop corrections to the Bethe approximation, J. Stat. Mech.: Theory Exp. 2005 (10) (2005) 11.
[55] D. Spiegelman, A. McDermott, B. Rosner, Regression calibration method for correcting measurement-error bias in nutritional epidemiology, Am. J. Clin. Nutr. 65 (4) (1997) 1179S–1186S.
[56] E. Jacquet-Lagrèze, Y. Siskos, Preference disaggregation: 20 years of MCDA experience, Eur. J. Oper. Res. 130 (2) (2001) 233–245.
[57] C. Zopounidis, M. Doumpos, Multicriteria classification and sorting methods: a literature review, Eur. J. Oper. Res. 138 (2) (2002) 229–246.
[58] A. Araújo de Castro, P. Pinheiro, M. Pinheiro, A hybrid model for aiding in decision making for the neuropsychological diagnosis of Alzheimer's disease, Rough Sets Curr. Trends Comput. (2008) 495–504.
[59] R. Howard, J. Matheson, Influence diagrams, Decis. Anal. 2 (3) (2005) 127–143.
[60] A.C. Menezes, P.R. Pinheiro, M.C.D. Pinheiro, T.P. Cavalcante, Towards the Applied Hybrid Model in Decision Making: Support the Early Diagnosis of Type 2 Diabetes, Springer Berlin Heidelberg, New York, USA, 2012, pp. 648–655.
[61] P.R. Pinheiro, A. Castro, M. Pinheiro, A multicriteria model applied in the diagnosis of alzheimer's disease: a Bayesian network, in: 11th IEEE International Conference on Computational Science and Engineering, CSE '08, 2008, pp. 15–22.
[62] G. Fillenbaum, G. van Belle, J. Morris, R. Mohs, S. Mirra, P. Davis, P. Tariot, J. Silverman, C. Clark, K. Welsh-Bohmer, Consortium to establish a registry for Alzheimer's disease (cerad): the first twenty years, Alzheimer's Dement.: J. Alzheimer's Assoc. 4 (2) (2008) 96–109.
[63] F.V. Jensen, T.D. Nielsen, Bayesian Networks and Decision Graphs, Springer-Verlag, New York, USA, 2007.
[64] P. Spirtes, C.N. Glymour, R. Scheines, Causation, Prediction, and Search, vol. 81, MIT Press, Cambridge, USA, 2000.
[65] S.L. Lauritzen, The EM algorithm for graphical association models with missing data, Comput. Stat. Data Anal. 19 (2) (1995) 191–201.
[66] P. Larrañaga, H. Karshenas, C. Bielza, R. Santana, A review on evolutionary algorithms in Bayesian network learning and inference tasks, Inf. Sci. 233 (2013) 109–125.
[67] C. Riggelsen, Learning parameters of Bayesian networks from incomplete data via importance sampling, Int. J. Approx. Reason. 42 (1) (2006) 69–83.
[68] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc.: Ser. B Methodol. 39 (1) (1977) 1–38.
[69] T. Minka, Estimating a Dirichlet Distribution, 2000.
[70] R. Neapolitan, Learning Bayesian Networks, Pearson Prentice Hall, Upper Saddle River, NJ, 2004.
[71] G.F. Cooper, E. Herskovits, A Bayesian method for the induction of probabilistic networks from data, Mach. Learn. 9 (4) (1992) 309–347.
[72] K. Murphy, The Bayes net toolbox for Matlab, Comput. Sci. Stat. 33 (2) (2001) 1024–1034.
[73] K.P. Murphy, Y. Weiss, M.I. Jordan, Loopy belief propagation for approximate inference: an empirical study, in: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc., San Francisco, USA, pp. 467–475.
[74] N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, Smote: synthetic minority oversampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[75] D. MacKay, Information-based objective functions for active data selection, Neural Comput. 4 (4) (1992) 590–604.
[76] M. Ramoni, P. Sebastiani, Robust learning with missing data, Mach. Learn. 45 (2) (2001) 147–170.
[77] T. Elomaa, J. Rousu, Efficient multisplitting revisited: optima-preserving elimination of partition candidates, Data Min. Knowl. Discov. 8 (2) (2004) 97–126.
[78] I. Kononenko, On biases in estimating multi-valued attributes, in: International Joint Conference on Artificial Intelligence, vol. 14, Lawrence Erlbaum Associates, Montreal, Quebec, Canada, 1995, pp. 1034–1040.
[79] J. Hanley, Characteristic (ROC) curve, Radiology 743 (1982) 29–36.
[80] D. Olson, D. Delen, Advanced Data Mining Techniques, Springer Verlag, New York, USA, 2008.
[81] G. Brier, Verification of forecasts expressed in terms of probability, Mon. Weather Rev. 78 (1) (1950) 1–3.
[82] I. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers Inc., San Mateo, CA, USA, 2005.
[83] R. Kohavi, The power of decision tables, Mach. Learn.: ECML-95 (1995) 174–189.
[84] G. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, in: Eleventh Conference on Uncertainty in Artificial Intelligence, vol. 1, Citeseer, San Mateo, 1995, pp. 338–345.
[85] S. Le Cessie, J. Van Houwelingen, Ridge estimators in logistic regression, Appl. Stat. (1992) 191–201.
[86] Y. Freund, R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, in: Computational Learning Theory, Springer Berlin Heidelberg, New York, USA, 1995, pp. 23–37.
[87] J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Mateo, CA, USA, 1993.
[88] D. Dash, M.J. Druzdzel, Robust independence testing for constraint-based learning of causal structure, in: Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002, pp. 167–174.
[89] N. Jongsawat, W. Premchaiswadi, A smile web-based interface for learning the causal structure and performing a diagnosis of a Bayesian network, in: IEEE International Conference on Systems, Man and Cybernetics, SMC 2009, IEEE, 2009, pp. 376–382.
[90] Z. Bin, C. Yuan-Hsiang, W. Xiao-Hui, W.F. Good, Comparison of artificial neural network and Bayesian belief network in a computer-assisted diagnosis scheme for mammography, in: International Joint Conference on Neural Networks, IJCNN '99, vol. 6, 1999, pp. 4181–4185.