Artificial Intelligence In Medicine 117 (2021) 102108

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

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

A comprehensive scoping review of Bayesian networks in healthcare: Past, present and future

Evangelia Kyrimi [a,*], Scott McLachlan [a,b], Kudakwashe Dube [b,c], Mariana R. Neves [a], Ali Fahmi [a,1], Norman Fenton [a]

[a] School of Electronic Engineering and Computer Science (EECS), Queen Mary University of London, London, United Kingdom
[b] Health Informatics and Knowledge Engineering Research (HiKER) Group
[c] School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

A R T I C L E I N F O

Keywords:
Bayesian networks
Healthcare
Clinical decision support
Scoping review
Survey

A B S T R A C T

No comprehensive review of Bayesian networks (BNs) in healthcare has been published *in the past*, making it difficult to organize the research contributions *in the present* and identify challenges and neglected areas that need to be addressed *in the future*. This unique and novel scoping review of BNs in healthcare provides an analytical framework for comprehensively characterizing the domain and its current state. A literature search of health and health informatics literature databases using relevant keywords found 3810 articles that were reduced to 123. This was after screening out those presenting Bayesian statistics, meta-analysis or neural networks, as opposed to BNs and those describing the predictive performance of multiple machine learning algorithms, of which BNs were simply one type. Using the novel analytical framework, we show that: (1) BNs in healthcare are not used to their full potential; (2) a generic BN development process is lacking; (3) limitations exist in the way BNs in healthcare are presented in the literature, which impacts understanding, consensus towards systematic methodologies, practice and adoption; and (4) a gap exists between having an accurate BN and a useful BN that impacts clinical practice. This review highlights several neglected issues, such as restricted aims of BNs, ad hoc BN development methods, and the lack of BN adoption in practice and reveals to researchers and clinicians the need to address these problems. To map the way forward, the paper proposes future research directions and makes recommendations regarding BN development methods and adoption in practice.

1. Introduction

Bayesian Networks (BNs), also described as *causal probabilistic models* or *belief networks*, are directed acyclic graphical (DAG) models [1]. Early attempts to use Bayesian analysis for medical problems were generally considered unsuccessful due to the necessary Bayesian inference being computationally intractable [2]. Development of efficient BN inference propagation algorithms [1,3] and advances in computational power since, have made it possible to develop BNs capable of addressing real-world decision support problems. This led to a renewed research interest in BNs which we found has resulted in many thousands of publications describing Bayesian solutions in the context of healthcare.

The immense volume of published works describing Bayesian solutions for application in clinical decision-suport demonstrates an enormous and rapidly increasing appetite for BNs as a reasoning tool to support healthcare delivery. With such significant research interest it is important to properly organize and summarize the range of research contributions which will aid identification of research gaps and development of future research directions. However, apart from a small number of micro-reviews focused on BNs for specific medical conditions [4], one epidemiology-focused review on DAGs that is yet to be peer reviewed [5] and a contemporaneous review published by the authors using the same literature collection but focusing only on the distribution of medical conditions for which BNs have been developed [6], the domain lacks any formal systematic or scoping review. There is a significant gap between developing a claimed accurate model and demonstrating its clinical usefulness and actual impact on healthcare decision-making [7–11]. Yet, evidence for the adoption of BNs in clinical practice remains extremely limited.

This work presents a scoping review that examines and evaluates BNs

* Corresponding author.
E-mail address: e.kyrimi@qmul.ac.uk (E. Kyrimi).
[1] Current affiliation: School of Health Sciences, University of Manchester.

https://doi.org/10.1016/j.artmed.2021.102108
Received 10 September 2020; Received in revised form 7 May 2021; Accepted 10 May 2021
Available online 13 May 2021
0933-3657/© 2021 Elsevier B.V. All rights reserved.

proposed for use in healthcare with specific regard to identifying the: (1) targeted *decision support*, (2) *modelling approach*, and (3) *usefulness* in clinical practice. We believe that combined, these can provide a global assessment of the research domain. To the best of our knowledge, this is the first scoping review of *BNs in healthcare*. Moreover, a taxonomy is used to classify concepts related to each of the above three objectives. This work does not focus merely on presenting the findings of the literature review. Useful recommendations directed to researchers for improving both *reproducibility* of their proposed methods and overall *usefulness* of the resulting BNs in clinical practice are made. Finally, crucial future directions are also highlighted.

The remainder of this paper is organized as follows: Section 2 presents necessary background material related to BNs. Section 3 provides the motivation for undertaking this scoping review. The methodology followed and the obtained results are described in Sections 4 and 5, respectively. A list with all the major findings and a detailed discussion on the findings as well as the strength and limitation of this review are described in Sections 6 and 7, respectively. Section 8 offers a list of recommendations for the authors, while Section 9 shows future research directions. Finally, a summary is provided in Section 10.

2. Bayes theorem and Bayesian networks

As this paper focuses on a review of BNs in healthcare, it is beyond its scope to provide a detailed formal description of the underlying theory of BNs and Bayesian inference. In this section we provide only a brief overview necessary to ground our objectives for the review. We refer readers who need more detail to texts such as [12] and [13], while those familiar with BNs may skip this section.

Bayes' theorem [14] was developed in the 18th century by the English Mathematician and Reverend Thomas Bayes. It is a formula that enables us to calculate how to update our initial belief in the probability of an unknown hypothesis *H* (called the *prior probability* of *H* and written as P(*H*)) when we observe new evidence *E* about *H*. The updated probability (called the *posterior probability* of *H* and written P(*H*|*E*)) represents the conditional probability of *H* given *E*. Specifically the formula is:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Where P(*E*), the probability of the evidence E is calculated as:

$$P(E|H)P(H) + P(E|not\ H)P(not\ H)$$

This initial and updated belief process is common to everyday life. For example, when clinicians perform differential diagnosis in circumstances where multiple diseases present with similar symptoms. As new evidence from diagnostic tests and clinical examination of the patient become available, clinicians update their belief on what they consider to be the illness causing the patient's poor health. There are other methods of updating beliefs, but Bayes theorem provides a normative and universally valid rule for such updating.

In formal probabilistic terms Bayes theorem enables us to calculate the joint probability distribution of two dependent variables (H and E in the above example). In general, when there are multiple variables, the chain rule in probability theory enables us to calculate the joint probability. For example, in Fig. 1 we have three variables smoking (A), lung cancer (B) and X-ray result (C). The chain rule tells us that the joint probability distribution of A, B, C is equal to

$$P(C|A,B) \times P(B|A) \times P(A)$$

However, because there is no direct dependency between A and C, we can reduce this full joint probability distribution to the simpler expression:

$$P(C|B) \times P(B|A) \times P(A)$$

In other words we can use information about independence between variables make the joint probability distribution calculation more efficient. As indicated by Fig. 1, BNs formalize this idea by representing the variables as a directed acyclic graph with qualitative and quantitative parts [12]. The qualitative part is the BN structure comprised of nodes representing random variables (discrete or continuous) and directed arcs representing causal or influential relationships. If a directed arc connects variables A and B, such as A→B, then A is called parent node or ancestor of B and B is a child node or a successor of A. The qualitative

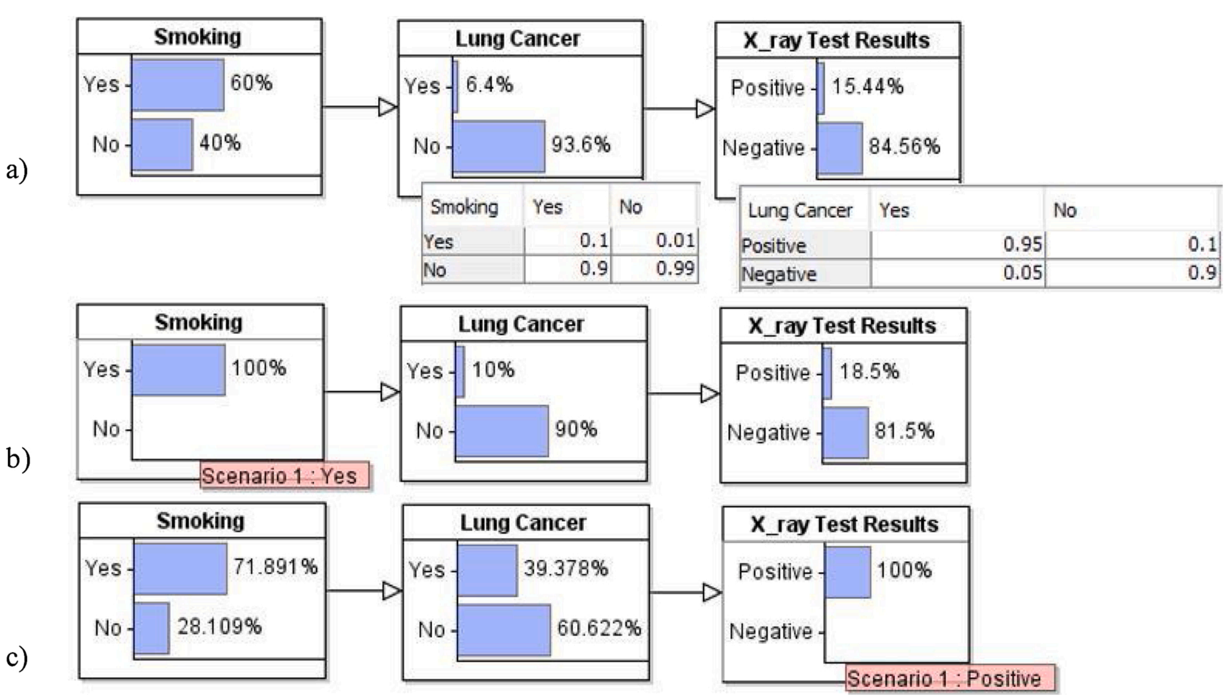

Fig. 1. a) A three-node BN example with Node Probability Tables shown, b) A BN example, where causal reasoning is performed, c) A BN example, where diagnostic reasoning is performed.

part is the BN parameters comprised of a set of conditional probability functions associated with each node – captured by a Node Probability Table (NPT) - to represent the conditional probability distribution of each node in the BN given its parents.

Both BN structure and parameters can be prepared using: (a) automated learning from data if sufficient data are available [15]; (b) a manual "by-hand" approach using knowledge elicitation methods to capture domain expert knowledge and extracting the necessary information from literature [16]; or, (c) through a combination of both [17].

Once the values for all NPTs are provided the BN becomes fully parameterized (Fig. 1a) and it can be used for reasoning from evidence (*observational reasoning*), which describes a joint distribution over possible observed events [18]. Evidence represents the real world, what happened. Once evidence is observed into the BN, the probabilities of remaining unobserved variables can be updated. There are two approaches to reasoning from evidence:

1 *Causal reasoning*: where the reasoning process follows the direction of the arc. For example: knowing that the patient is a smoker increases the probability of lung cancer from 6.4% (Fig. 1a) to 10% (Fig. 1b), which in turn increases the probability of a positive X-ray result from 15.4% to 18.5%.
2 *Diagnostic reasoning*: where the reasoning process is counter to the direction of the arc. For example: knowing that the patient's X-ray is positive increases the probability of lung cancer (Fig. 1c).

What makes BNs especially powerful is that, in addition to observational reasoning, they can also be used to answer hypothetical questions such as *what will happen if an intervention is made. Interventional reasoning* can only be performed when relationships among the variables are causal because an intervention is an exogenous action that fixes the state of the variable we have intervened upon, thus, making it independent of its causes [12,19,20]. Contrary to reasoning about evidence, *interventional reasoning* does not allow diagnostic reasoning from the intervened variable [21]. For instance, when we observe a high body temperature on the thermometer, we can argue that we have a fever. However, if we arbitrarily start rubbing the thermometer to reach a specific temperature, we can no longer argue that we have a fever. To perform interventional reasoning in a BN we simply make the intervened variable independent of its causes by removing all the edges pointing towards that variable. This proces (known as *graph surgery*) is equivalent to the so-called *do operator* as defined by Pearl [19].

Moreover, BNs can also be used to answer *counterfactual questions* such as *what would have happened if events other than those observed had occured*. Considering contrary-to-facts scenarios means imagining alternatives to reality. Although sometimes claimed otherwise [22], counterfactual reasoning is a natural reasoning process and is believed to be an ability unique to the human mind [23]. In BNs, counterfactual reasoning combines both evidence and interventions. Using the process of *twin-networks* proposed by Pearl [19], the actual world is modelled based on evidence while the hypothetical counterfactual world is altered using interventions. In a twin network both networks have identical structures except for the arrows towards the variable that we intervene upon, which are missing in the hypothetical world. The posterior probabilities of the variables, which remain the same in both worlds, are called *background variables,* and are shared between the two networks. The main difference between counterfactuals and interventions is that for the former we know the values that some or all the variables had in the actual world. By contrast, when we intervene, we are unaware of the values of the predecessor variables in the network.

3. Motivation

Clinical decision making is a complex and evolving process where evidence is gathered and diagnostic and treatment decisions are made [24]. Through consideration of the patient's history, as well as any signs, symptoms and test results, clinicians try to resolve two main questions: '*What is the problem*?' and '*How can we best solve it*?'. A plethora of potential symptoms, diagnostic tests, diseases, treatment options, our complex human physiology and increasing uncertainty can make clinical decision-making challenging. While clinicians can be good decision-makers, it can be difficult to combine all available evidence and accurately reason under conditions of such uncertainty [25].

Many *clinical decision support systems* (CDSS) have been developed to help clinicians in their decision-making role [26–28,11,29]. These can range from simple scoring systems to complicated multivariate regression models, neural networks, decision trees and graphical probabilistic models like BNs. BNs are increasingly recognized as powerful tools in risk analysis and decision support for real-world problems, and have been been proposed for healthcare applications [30,31]. The interest in BNs in healthcare is due to their ability to: (i) model complex problems with causal dependencies where a significant degree of uncertainty is present; (ii) combine different sources of information such as data and experts' judgement; (iii) be presented in an interpretable graphical structure; and, (iv) model interventions and reason both diagnostically and prognostically.

A preliminary review of BNs in healthcare [32] based only on a very small number of published papers was initially conducted by the authors to identify potential research gaps and justify the conduct of this broader scoping review. Our conclusions from that preliminary review were:

1 The body of literature on BNs in healthcare is large and rapidly increasing, indicating a significant research interest.
2 It is rare for the process for developing a BN to be clearly and comprehensively described in the literature, and certainly not to a degree that would support repeatability.
3 Despite the fact that many authors identified obvious and widely accepted benefits that could arise from the use of BNs to support decision-making in healthcare, and the enormous number of publications, there has been negligible clinical adoption of the BNs described in the reviewed academic articles.
4 Despite the intense interest as reflected in the academic literature, a systematic review of the literature is missing and this is likely to have led to duplication of effort and repetition of methodological mistakes, while at the same time allowed important research gaps to remain hidden.

These conclusions provided a strong case for conducting a larger and far more comprehensive scoping review to summarize the literature, highlight research gaps and future directions, and propose a way towards their resolution.

4. Methodology

4.1. Literature search

A comprehensive search of health and health informatics literature databases including PubMed, MedLine, ScienceDirect, Scopus, DOAJ and Elsevier was performed using a selection of keywords arranged in the following general search query:

> "(((Bayes OR Bayesian) AND network) OR (probabilistic AND graphical AND model)) AND (medical OR clinical)"

Terms such as *Bayesian networks* or *probabilistic graphical models* were used here because they are widely observed in the targeted literature. Different ways for explaining the medical condition do occur, in that in some papers the exact condition is mentioned while in others broader terms such as *medical* or *clinical application*, *medical* or *clinical condition*, or *medical* or *clinical setting* are used. Our scoping review settled on the broader terms *medical* or *clinical* as they were found in a wider collection of papers. Searching for specific medical conditions would have been impractical as there are many thousands of distinct known conditions.

Due to the high number of articles returned, further scrutiny was applied to narrow the collection to the most relevant articles for this study. This was achieved by selecting only papers where the described keywords were present in their abstract. Additional screening was conducted to exclude papers published outside the period 2012–2018, those that were not published in English, or those whose primary content was not healthcare related. The remaining papers were those possessing all the following identified characteristics. They:

1 Describe a genuine BN model or BN modeling process or BN adoption process.
2 Are targeted clearly at a medical condition or application.
3 Are intended to support clinicians or patients in decision making or prediction

Papers whose focus was one of the following were classified as ineligible and they were excluded from the literature collection.

1 Bayesian statistics, as opposed to Bayesian networks
2 Simple naive Bayesian or neural networks
3 Network meta-analysis
4 The predictive performance of multiple machine learning algorithms (of which BNs were simply one of many)

4.2. Analytical framework for characterizing BNs in healthcare

Fig. 2 presents the analytical framework that has been developed in this work for use in characterizing the domain of BNs in healthcare. This framework also guides the review and analysis of the body of literature in this study. In alignment with this framework, our review plan identified six primary objectives shown in Fig. 2 that cover important aspects critical to harnessing BNs in healthcare. Addressing all six objectives in one paper would result in multiple areas of focus such that none could be described with sufficient depth. This could also have been confusing to the reader. Objective 5, which focuses merely on gaining an insight in the distribution of medical conditions for which BNs have been developed, has already been addressed in detail in our recent publication [6]. Objectives 4 and 6 are also separately addressed in [33,34]. Thus, each zone of our framework (Fig. 2) results in a paper focusing on specific research objectives. While each paper draws data from the same literature collection, different research objectives are investigated and therefore different findings are reported, meaning each paper stands alone. This paper provides in-depth investigation of the first three objectives:

1 *Objective 1–Decision support*: This objective involves investigating published BN models focusing on the identification of: (1) the *reasoning processes* which have been defined in Section 2; and (2) the *type of decision support* associated with the reasoning process, which influences and is influenced by the *temporal aspect* of the medical problem.
2 *Objective 2–Modelling approach*: BN development was identified and broken into two parts, each determined by: (1) the BN structure (variables and arcs), and (2) the conditional probabilities for each variable (parameters). The source for variables, arcs and parameters was identified and recorded. Additionally, we explored whether the paper provided a repeatable overall BN development process along with a clear description of the structure and parameter elicitation or learning process. Finally, the distribution of BN software tools used by authors, where identified, was captured.
3 *Objective 3–Model usefulness*: The potential usefulness of published BNs to clinical practice was examined. Many researchers have investigated the gap between developing an accurate model and having a useful model that has or can have an actual impact on the clinical decision-making process [9,35,36,11]. Based on existing literature [7,8,37] the properties presented in Table 1 are considered necessary and should be present to ensure a useful model.

Each objective identified in Fig. 2 captures a unique attribute from the literature on BNs in healthcare. The first three objectives when combined provide a global assessment of the BNs in the healthcare domain. As shown in Fig. 2, each objective has been broken into a set of attributes, which represents data recorded from the studied literature. These attributes were developed with the input of two decision scientists and refined inductively during a preliminary review conducted by the first two authors (EK and SM) and reported in [32]. In addition, two decision scientists experienced in developing BNs and who were not involved in the preliminary study were asked to further refine the selected attributes.

The proposed framework was also used to ensure a consistent review of the literature. To further empower a consistent review process, a set of tools that included a systematically designed online literature review questionnaire and a manual for researchers on how to interpret and respond to the review questions was distributed.[2] The manual provided a detailed definition for each attribute and a clear explanation of how to identify that attribute in the literature being reviewed. Additional training sessions were conducted during which three papers were reviewed by the six reviewers involved in this study, followed by a thorough discussion on the given answers and how they were derived. Each paper in this review was examined by two reviewers out of a team of six, who entered their responses into the secure online survey tool. In cases where responses differed, two senior reviewers (EK and SM) reviewed the paper collaboratively and resolved the conflict by consensus.

4.3. Development of taxonomies

During the review a taxonomy was deductively resolved that classifies concepts related to each of the three objectives. Various approaches exist for resolving knowledge and defining the taxonomy, but at their core taxonomies are systematic classifications within a domain of interest that typically present as hierarchical, multidimensional arrangements of linked categories of concepts [38–40]. Taxonomies are represented in almost every conceivable shape, including: linear [41], circular [42], flow diagrams [43] and so on [44]. While most taxonomies are observed to be unidirectional and therefore arranged similar to a concept hierarchy, from most general to most specific, our review resolved a bi-directional taxonomy for each of the three objectives presented in this work. Each taxonomy also provides secondary information in that the weight of the connecting line, as described in Fig. 3, represents the frequency of papers (as an overall percentage of papers in the review) whose work is represented by the relationship between the two identified nodes.

5. Results

5.1. Literature search and collection results

The results of the literature selection process are presented in diagrammatic form in Fig. 4. Our initial search, looking at keywords in the abstract, identified a collection of 3810 papers for screening. After applying the screening process described in Section 4.1, 462 papers remained. From these, 123 papers met the eligibility criteria described in Section 4.1 and remained for inclusion in this review.[3]

[2] The questionnaire and the manual can be found in https://pambayesian.org/wp-content/uploads/2020/09/BNs-In-Healthcare-FINAL.docx.

[3] The full reporting of the literature reviewed in this work is available in the supplementary material at https://pambayesian.org/wp-content/uploads/2020/09/Results_Data_Outputs.xls

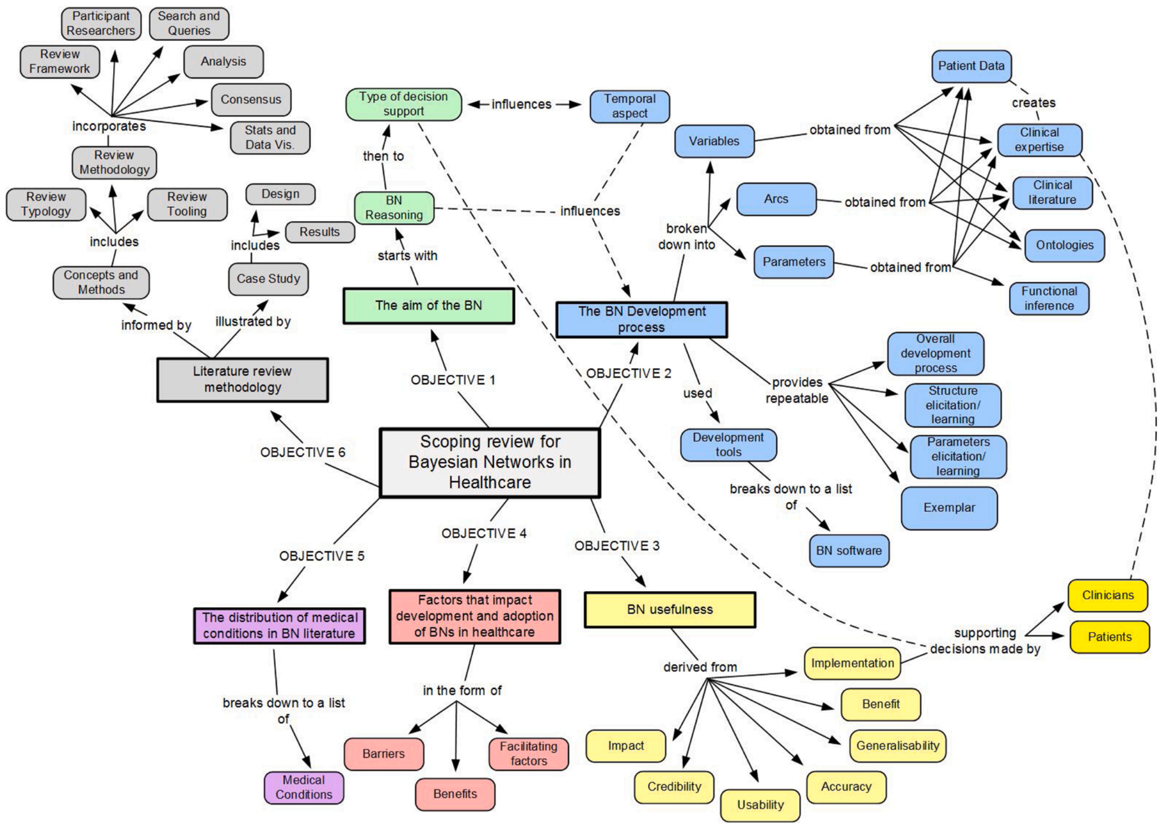

Fig. 2. Framework for characterizing the domain of Bayesian networks in healthcare.

Table 1
Review framework for model usefulness.

Properties	Description
Benefit	The BN addresses a clinical question of enough importance that has a margin of improvement to justify time by practitioners
Credibility	The BN has a clear logic and appears to describe what it intends to describe
Accuracy	The BN has an acceptable predictive performance (internal validation)
Generalizability	The BN has an acceptable accuracy in datasets other than the original training dataset (temporal and/or external validation)
Usability	There is an interface adjusted to the user's needs
Impact	The model must have potential to change patient care

5.2. Results of survey on Bayesian networks in healthcare

This section presents results from analysis of the included literature with respect to the attributes identified for each of the first three objectives of the review as illustrated in Fig. 2.

5.2.1. Decision support

As described in Section 4.2 and illustrated in Fig. 2, the first objective of this review was to identify the aim of medical BN models described in the literature. Data were collected regarding (1) the *reasoning processes* and (2) the *type of decision support* associated with the reasoning process, which influences and is influenced by the *temporal aspect* of the medical problem.

In 90 of the papers (73%) *observational reasoning* was performed, for example: [45–49,15]. Fig. 5 shows that there is a range of decisions for which observational reasoning may be applied. Diagnosis of a medical condition given a set of known risk factors and observed signs and symptoms is the most frequent, as may be seen in examples such as the works of [50,46,51–54]. For a review and detailed description of the scope of targeted *medical conditions* identified from our literature collection, we refer the reader to [6]. *Interventional reasoning* was performed in only 10 of the studied papers (8%) and was mainly used for treatment selection, such as seen in [55–59]. In three papers, namely, the two included works of Constantinou et al. [58,60] and single work from Xu et al. [61], interventional reasoning was performed for managing acceptable risk through consideration of a number of relevant interventions. In one paper by Neapolitan et al. [62], interventional reasoning was used to predict prognosis: that is, the likelihood of treatment success. In the papers [63,64,51,65–67], interventional reasoning was mentioned as a potential future use for their developed

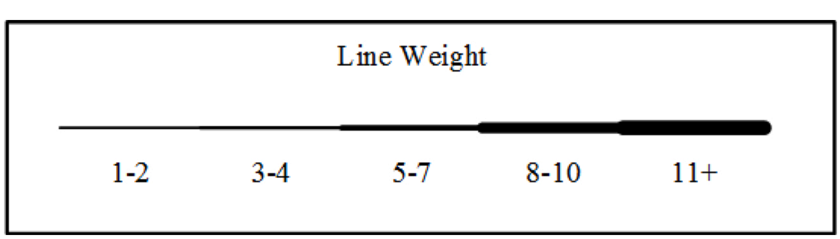

Fig. 3. Line weight and percentage of papers.

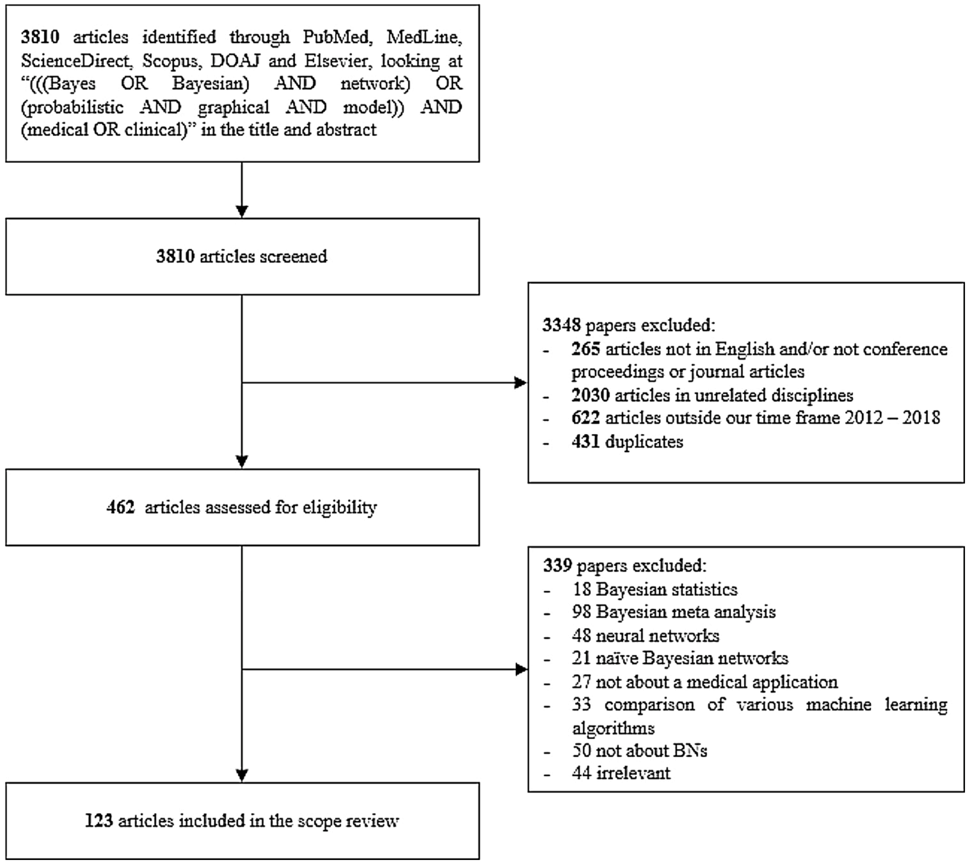

Fig. 4. PRISMA diagram for literature selection.

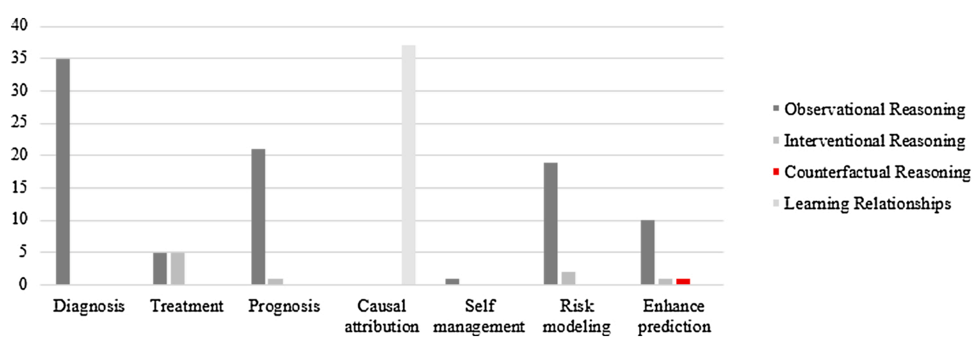

Fig. 5. Distribution of decision types and reasoning processes identified in BN literature.

BNs. While the use of the published BNs to answer *counterfactual questions* was discussed by Solomon et al. [66] as a potential future direction, it is only in the work of Constantinuou et al. [68] that the proposed BN was actually applied in a counterfactual setting. Finally, in 37 papers (30%) of the literature, particularly in the works of [63,69–74], no true reasoning process was performed. Rather, the BN was used to learn associations or causal relationships from data in order to understand and provide causal attribution to a disease, which relates more to BN model development rather than BN model use or its potential for application in clinical practice.

The results on reviewing the *time dynamics* of proposed medical BNs are presented in Fig. 6. In 70 of the papers (57%) static BNs without a time element were developed. The type of decision support possible influences, and is influenced, by the temporal aspect of the medical problem. Static BNs, such as those presented by [75,76,51,48,53,49,77] were mainly proposed for diagnosing a medical condition. In 22 of the 70 static BNs (31%) the process of identifying causal attributions was applied, for example [57,78–81]. In other words, static BNs were used in situations where time was not a significant factor. In 25 of the papers (20%) we identified BNs that implicitly captured time either through specific temporal nodes, like in the works of [82,64,83,84,61], or by modeling prior and/or post treatment variables such as [85,60,86,55,56]. Temporal BNs were identified in only 16 of the papers (13%) such as [87,88,16,89]. While temporal BNs were mainly used in the performance of outcome prediction, a process described medically as prognosis [90–93,88,89], we were also able to observe that prognosis models were quite uniformly distributed across the entire temporal scope. Finally, in 12 of the studied papers (10%), including the works of [94–96] the presence of a time dynamic in the problem or proposed model was not clearly described and therefore classification of this element was not possible.

5.2.2. Modelling approach

As explained in Section 4.2 and shown in Fig. 2 the second objective

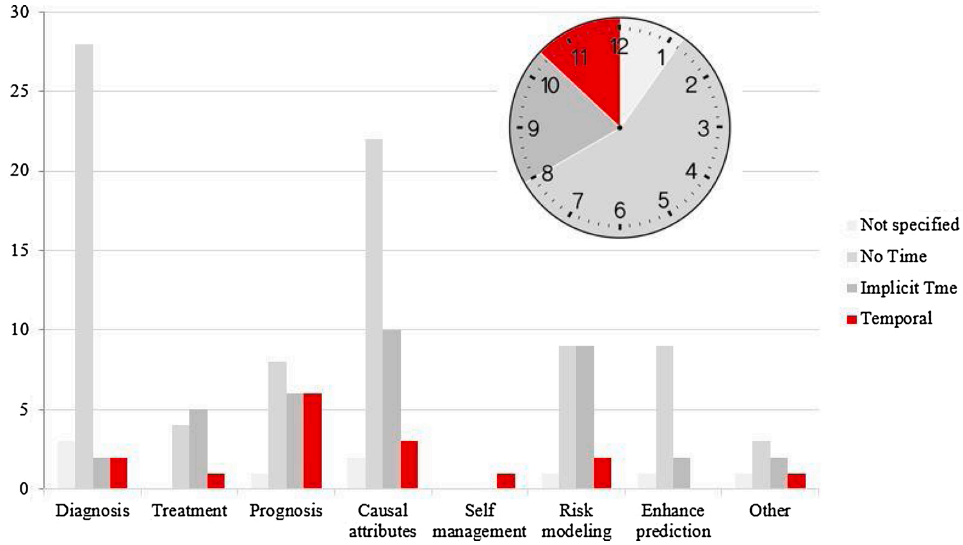

Fig. 6. Distribution of decision types and time dynamics identified in BN literature.

of the review was to investigate the BN development process with the goal of establishing current state-of-the-art from the literature. Data were collected regarding (1) the source for variables, arcs and parameters, (2) the assessment of the overall BN development process along with a description of the structure and parameter elicitation or learning process, and (3) the BN software tools used by authors.

As shown in Table 2, in 77 of the papers (63%), including those of [97,98,51,99–101], the *BN variables* originated purely from *data*. However, in 24 of the papers, the BN variables were obtained from *knowledge* elicited from medical experts or extracted from published scientific evidence. Further, in one-third of the knowledge-driven BNs such as [102,87,103,60,16], the BN variables were elicited from experts through a series of interviews. In two papers [104,84] the variable selection was based on a previous study, while in another published by Chang et al. [95] an ontology was used to resolve the BN structure. In 13 papers, the variables were elicited either from both experts and literature [105–107,76,92,108–111] or from a combination of clinical expertise and templates, such as ontologies and fragments [58,52,112]. Finally, in 15 of the papers the BN variables originated from both data and knowledge, such as [55,56,70,62,113,114]. Referring now to *the arcs* that connect the variables in the BN, the majority were learned only from data using different constraint or score based learning algorithms (Table 2), such as those found in [63,50,57,96,115,116]. In 30 of the published BNs, the arcs were elicited from knowledge, where clinical expertise was the main source, as in the work of [87,59,60,16]. In a paper published by Zarringhalam et al. [117] the arcs between variables were elicited from literature, while in 5 papers *rules*, described as arising from predefined *templates* or *ontologies*, were used [75,82,95,48,118]. In addition, a combination of *knowledge and rules*, specified as templates and *idioms* [119] was also used to elicit the BN arcs [105,102,92,88,111]. Furthermore, *mixed approaches*, where structure learning algorithms alongside causal constraints provided by experts, literature or causal rules, were used in 14 of the papers, auch as [65,67,56,120,61]. Regarding the BN parameters, in 96 of the papers, including those [45,82,51,83,121,116], the *parameters* were learned purely from *data*. In only 6 papers the parameters were learned from *knowledge* with clinical expertise being the main source of it [48,122,16]. In two papers the BN parameters were elicited only from literature [123,110], while in another one the parameters were elicited from a combination of experts, literature and the use of mathematical functions [92]. In 13 of the published BNs their parameters were based on both data and knowledge, such as [102,58,124,100,111].

Each paper was evaluated regarding how detailed the overall *BN development process* was. In 59 of the studied papers (48%), for example [125,109,72,97] and [61], the steps followed for developing and validating the BN were either completely omitted or could only be implied. In 26 of the papers (21%), the overall development process was either described in detail in the text, for instance in the works of [103] and [111], or visually illustrated as seen in the works of [102,82,58,67,126].

Apart from analyzing the overall development process, we further investigated whether the *BN structure and parameter learning, or elicitation process* was detailed enough to be considered as repeatable. The degree of repeatability was clustered as *negligible* detail (no information available), *insufficient* detail (some information provided but not enough for dublicating the method), *sufficient* detail (enough information to repeat the described process). As shown in the inner pie of the doughnut

Table 2
Origin of BN inputs; variables, arcs, parameters.

Origin of BN inputs	BN inputs Variables (%)	Arcs (%)	Parameters (%)
Knowledge	19%	24%	5%
Experts	6%	12%	2%
Literature	2%	1%	2%
Templates/ Rules	1%	4%	0%
Combinations	10%	7%	1%
Data	63%	54%	78%
Data and Knowledge	12%	11%	11%
Not specificied	6%	11%	6%

graph in Fig. 7a, from the papers where the origin of the BN structure was specified, 80 BNs were learned from data (either completely or partially) and 44 from knowledge (either completely or partially). The outer layer of Fig. 7a identifies that in situations where the BN structure was learned from data, papers were almost uniformly distributed between those with negligible, insufficient, and sufficient detail regarding the structured learning process. In 30 of the papers, such as [127,125,109,128–130], the authors gave only simple statements such as *the BN structure was learned from data or the structure was learned using supervised (or unsupervised) learning algorithms*, which are incomplete descriptions that do not allow the reader to fully understand the process used. In 28 of the data-driven BNs greater detail, such as the name of the learning algorithm used, was provided. Examples include: [90,64,131,115] and [54]. Finally, in 22 of the data-driven BNs a detailed description of the structured learning approach was provided, such as those found in [132,65,133] and [134]. Conversely, in 29 out of the 44 of papers where the BN structure was elicited from knowledge no description of the elicitation process was given, as in the works of [107,47,88,113,61]. Statements such as *the model was constructed by two experts or structure was based on expert domain knowledge or the structure was determined by combining known relationships from the medical literature with input from our clinical collaborations* provided the totality of methodological description. In situations where the BN structure was derived from ontologies, the process was usually described with sufficient detail [82,95,48]. In two papers published by Kalet et al. [103] and Seixas et al. [135] a detailed description of the elicitation process was provided. Regardless of the origin of the BN structure, the complete BN structure was provided in 65 of the papers (53%), while a detailed description of the variables included in the BN and their states were provided in only 34 of the studied papers (28%), such as [55,46,58,88,121,77].

As shown in the inner pie of the doughnut graph in Fig. 7b, from the papers where the origin of the BN parameters was specified, 109 were learned from data (either completely or partially) and 19 from knowledge (either completely or partially). In 74 of the papers, where BN parameters were learned from data, no description of the learning process was provided, as in [63,45,57,70,52,49,126]. In such cases, the authors simply mention that the BN parameters were learned from data without any further information. Only three papers provided a detailed description of the parameter learning process ([136,137,58]). Regarding the knowledge driven BN parameters, in 12 of the papers, the parameter elicitation process was not described, for example Tylman et al. [110] and Lou et al. [122]. Finally, in only three papers the parameter elicitation process was described in depth [105,102,85]. Regardless of the origin of the BN parameters, marginal probabilities or complete NPTs were rarely provided. More specifically, in 27% of the papers some parameters were presented, while a full list of parameters was available in only 7 papers [50,46,104,83,62,15,138].

The last modelling approach element studied in this part of the review was to identify the *BN software* used by authors. We see from Fig. 8 that in 47 of the papers either the software used was not specified or the authors implemented their own ad-hoc software [63,90,94,109,123,49,120]. The two software tools most often used were GeNIe™ such as in [104,113,111] and R™ as in [131,139,74].

5.2.3. Model usefulness

As described in Section 4.2 and shown in Fig. 2 the third objective of the scoping review was to identify the BN usefulness in clinical practice. Data were collected regarding BN's (2) benefit, (2) credibility, (3)

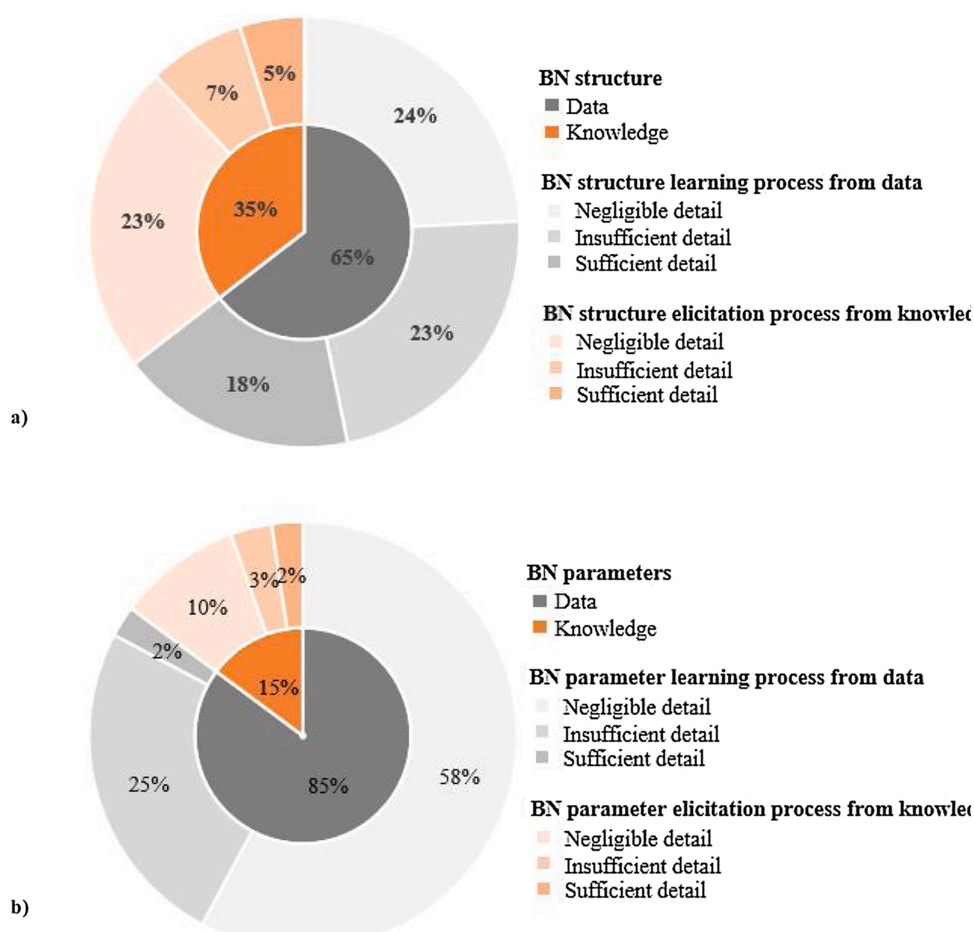

Fig. 7. Detailed description of the BN development process; a) BN structure learning and elicitation process, b) BN parameter learning and elicitation process.

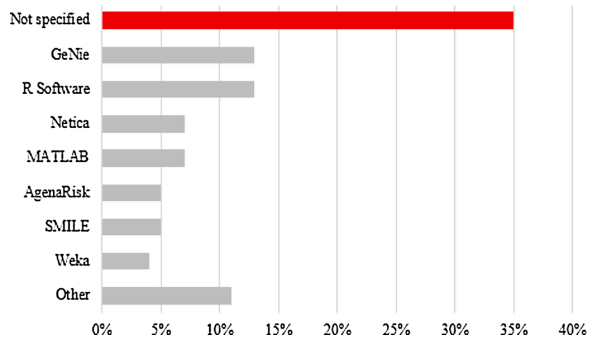

Fig. 8. Software used in BN Development.

accuracy, (4) generalizability, (5) usability, and (5) impact.

Fig. 9 presents the frequency of elements related to BN usefulness in healthcare. In 101 of the papers (82%) clearly described some *clinical benefit* for their medical BN, as in [125,105,92,140,48,121,89]. In 94 of the papers (76%) an accepted approach to evaluate and report the *accuracy* of models and their predictions was provided, like [141,87,142,131,53,120]. However, *generalizability* could be identified in only 6 papers (5%), such as [140,88,80,143]. From the published BNs, only 42 appeared to be *credible* per the definition in Table 1. Two thirds of these were knowledge-driven, a strong indicator that the BN structure was sensible, for example in [102,59,60,118,111]. From the credible BNs only 8 were learned completely from data, including [144,137,71,133,49]. In the majority of data-driven BN structures, the model's credibility was difficult to be identified. Two of the rarest characteristics to identify were the model's *usability* and *impact*. To find usability we sought whether the authors proposed or demonstrated an interface, as well as whether they had also validated that interface using experts. In 12 papers (10%) an interface was proposed. Examples can be found in [75,57,95,48,133,126]. From the papers, where an interface was presented, only one described evaluation of the proposed interface [95]. Two types of *impact* were identified: (1) *potential impact*, indicating that a small study on whether the developed BN can impact clinical decision making was performed; and (2) *actual impact*, indicating that a trial had been conducted to evaluate whether the BN could be applied in a longer-term implementation. In only three papers the *potential impact* of the developed BN was investigated [85,67,133]. The *actual impact* of the BN was not identified in any paper in this survey. *Finally, there was nothing to suggest that any of the BN models in this review were in current clinical use.*

6. Major findings

In this section we present our major findings resulting from investigating Objectives 1–3.

6.1. Decision support

The first taxonomy presented in Fig. 10 contextualizes the relationship between the *problem to be solved* and the type of *decision support* being employed. Three *reasoning types* were identified from the literature. As can be seen by the indicated line weight, the vast majority of developed BNs engaged in a process of *observational* reasoning, leaving *interventional* and *counterfactual* reasoning significantly underexplored. Further, four *clinical decision types* and four *decision contexts* were identified. *Diagnosis* was the most frequent clinical decision, for which only reasoning from evidence was required. The most frequent decisions related to *interventional reasoning* were *treatment recommendations*. From the decision contexts, learning *causal attributes* was more frequently observed. A summary of the major findings related to the first objective of this scoping review is presented in Box 1. A further detailed discussion on these findings can be found in Section 7.

6.2. Modeling approach

The second taxonomy shown in Fig. 11 focuses on the interplay between *model development* and the *developers* engaged to the task. Three types of *time dynamics* were identified with more than half of the published BNs identified as *static*: used in circumstances where no time element was essential to the model. Three elements, *variables*, *arcs* and *parameters*, are necessary for representing the developed BN. As can be seen in the taxonomy, *data-driven* approaches were more common than the *knowledge-driven* methods that rely on *experts, literature or ontologies*. The approach of learning, or eliciting, each element was characterized as being *bespoke* to the published case, or *generic* to the degree of being potentially repeatable. On the *developer* side our review identified two primary *developer* actors: *decision scientists*, such as computer scientists and statisticians; and *clinicians*. We were able to classify three potential development approaches: *method-, problem-* and *hybrid-* driven. It was apparent from a thorough reading of the literature that *hybrid-driven* papers provided a more generic development approach. A summary of the major findings related to the second objective of this scoping review, is presented in Box 2. A further detailed discussion on these findings can be found in Section 7.

6.3. Model usefulness

Fig. 12 presents the third and final taxonomy, which classifies the BN's potential or claimed usefulness and connects this with elements that can measure potential for successful engagement of BNs in clinical practice. Six main elements were identified for the potentially useful BN that are grouped into three main categories. These evaluate the BN's: *applicability, validity* and *adoptability*. As can be observed from the thickness of the connecting lines, *benefit, accuracy*, and *credibility* were

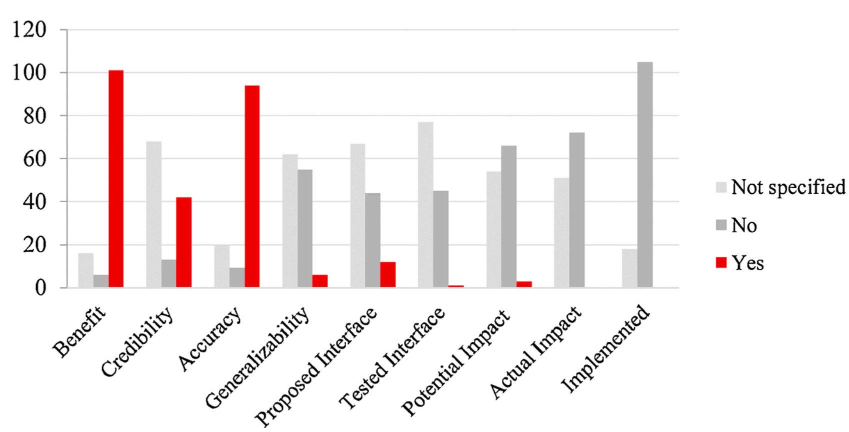

Fig. 9. Frequency of the properties related to the BN usefulness in healthcare.

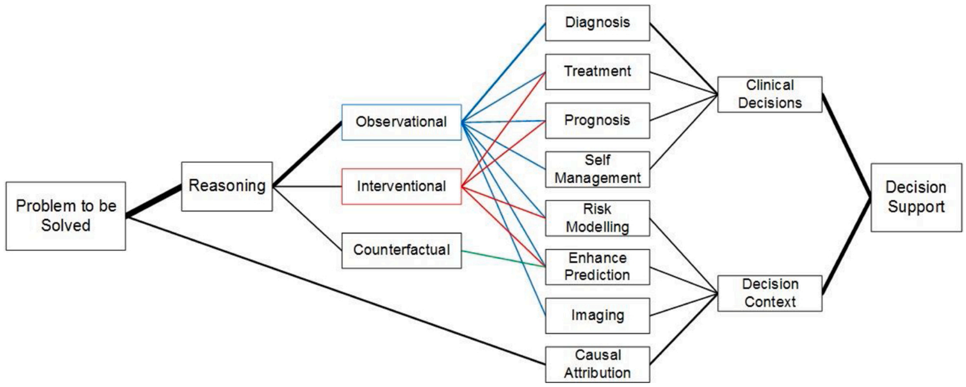

Fig. 10. Taxonomy of the first objective of the scoping review.

Box 1
Major findings related to the first objective - decision support.

1 In 90 of developed BNs reasoning from evidence was performed
2 In 10 papers interventional reasoning was performed
3 Only one BN on counterfactual reasoning
4 The two most common aims of the published BNs were to help diagnosis and learn causal attributes from data

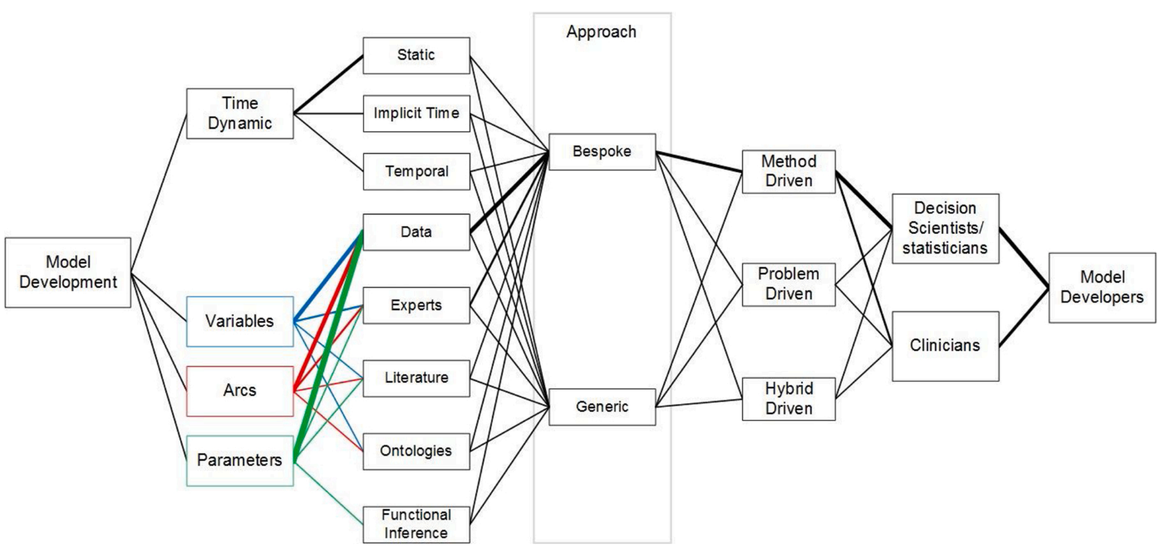

Fig. 11. Taxonomy of the second objective of the scoping review.

the most observed elements in this review. It was interesting to note that the BN model's *adoptability* was identified as the most neglected category, as this was also reflective of the lack of works claiming to present a model that had been adopted in some way into clinical practice. A summary of the major findings related to the third objective of this scoping review, is presented in Box 3. A further detailed discussion on these findings can be found in Section 7.

7. Discussion

This scoping review intended to examine and evaluate the: (1) targeted *decision support*; (2) *modelling approach*; and (3) *usefulness* in practice for BNs in healthcare. We found that the literature on BNs in healthcare could be broadly grouped into: (1) *method-driven*, (2) *problem-driven*, and (3) *hybrid-driven* papers. On one hand, *method driven papers* were seen to be primarily written by decision scientists and their focus was on a specific aspect of the modelling methodology. Hence, the BN *development* and *validation* are usually described in greater detail than the *clinical benefit* and *potential impact* in practice, which appear to have been neglected by most researchers. On the other hand, *problem driven papers* were mainly written by clinicians and their focus is on the medical application. In this type of papers, the *clinical benefit* and *medical background* are usually described in detail, while the BN *development* is usually only vaguely explained. *Hybrid papers* result from a collaboration between decision scientists and clinicians and the content of their works tended to be more equally divided between the clinical and modelling aspects of the BN. Hybrid papers were found to be the most complete of the three types of papers. The remainder of this section provides a more thorough analysis of the review findings associated with each objective, while also discussing the strengths and weaknesses of the review process

Box 2
Major findings related to the second objective - modeling approach.

1 70 of the published BNs (57%) were static models with no time element
2 A prevalence of *data-driven* over *knowledge-driven* models
3 The methodologies employed for learning *data-driven* BNs was described in greater detail than that for eliciting *knowledge-driven* BNs
4 The process of eliciting the BN structure (38 out of 40 papers) and/or parameters (16 out of 19 papers) from knowledge was rarely described sufficiently in the literature
5 Experts were a major source of knowledge for eliciting the BN structure and/or parameters
6 The literature lacked a well described overall development process
7 *Hybrid-driven* papers were found to be more generic in approach and application
8 The complete BN structure was not provided in 58 of the papers (47%) reviewed and in those where the BN structure was available, a detailed description of the variables was rarely given
9 A full list of NPTs were provided in only 7 of papers in this review

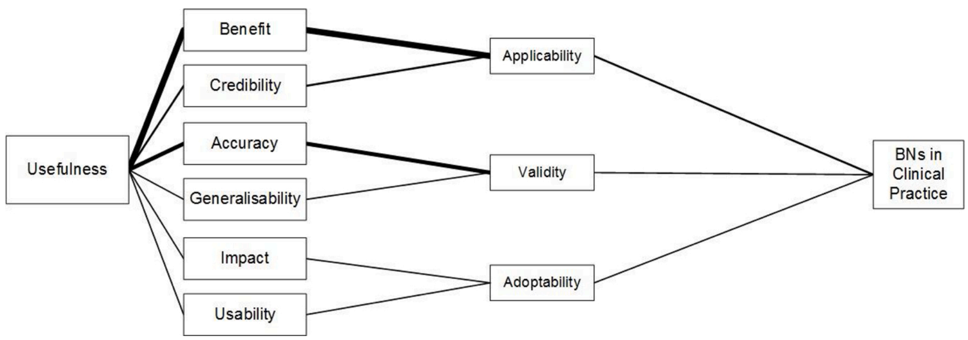

Fig. 12. Taxonomy of the third objective of the scoping review.

Box 3
Major findings related to the third objective - model usefulness.

1 The elements most frequently identified in the literature were benefit (in 101 of the papers), credibility (in 42 of the papers) and accuracy (in 94 of the papers)
2 *Knowledge-driven* BNs appeared more credible on review than *data-driven* BNs
3 A model's *generalizability* (in 6 of the papers) and *usability* (in 12 of the papers) were rarely specified
4 The clinical *impact* of developed BNs was the most underexplored element in the literature
5 A gap exists between *developing an accurate model* and *having a useful model that can be adopted in healthcare* – no indication of BN implementation

undertaken.

7.1. Discussion of findings

7.1.1. Decision support

In many of the published BNs no real reasoning was performed as the aim of most works was to identify, or learn, relationships or causal attributes from the data. In BNs where reasoning was undertaken, the vast majority sought to help clinical decision-making, such as for diagnosis or prognosis. In these instances, reasoning from evidence was most often used. *Interventional* and more especially *counterfactual reasoning* were significantly underexplored, indicating that BNs are not being used to their full potential. Despite the benefit of counterfactual reasoning in answering important questions regularly asked by clinicians, such as *Would the patient have survived, if we had operated him sooner?*, counterfactual analysis is still considered by many as a contentious research approach. There is a debate as to whether counterfactual reasoning has scientific validity, or whether it is merely a metaphysical process with no observable or testable consequences [22]. This might explain the extremely small number of BNs in which a counterfactual reasoning approach was undertaken. While most of the published BNs are static models with no or implicit time element, we know that time plays a crucial role in clinical decision-making and that BNs can model these dynamic processes yet temporal BNs remain less popular. This might be explained by the difficulty of modelling temporal BNs, as they are structurally more complex and require more computational power to perform inference. It may also be due to the perceived unavailability or incompleteness of temporal datasets from which to draw knowledge necessary to making those inferences.

7.1.2. Modeling approach

In 59 of the papers, description of the *overall BN development process* was either absent or could only be inferred from the description of the case study. In only 26 of papers was the BN development description detailed enough to reach a level of being potentially **repeatable**. The number of BNs where the structure was learned only from data was twice that of BN structures elicited from knowledge. This may be because structured learning is less time-consuming and is usually based on well-known learning algorithms, which makes the process more repeatable. The prevelance of data-driven BNs in medical applications came as a surprise if we consider the fact that they are based solely on medical data, which are not always good enough or complete enough to be used as the only source for learning a causal structure [63,111]. Many times, the quality of medical data for research purposes was described as poor, and poor data therefore means meaningless data-driven BN structures [60]. Moreover, even if the available data includes information on all necessary variables, there are typically far too many variables and far too few samples in the dataset to achieve any sensible structural learning outcome [60]. When the BN structure is learned from knowledge, the structure is more sensible as it can more correctly represent the true causal mechanisms. However, the elicitation process is time-consuming and may be prone to errors, especially when it is based only on the input of a single domain expert [4,58]. Additionally, the way a *BN structure* was elicited from knowledge was rarely described in the literature, making the process ad hoc and less able to be standardized. In other words, elicitation process descriptions were found to be ambiguous as insufficient detail was provided for the reader to understand where the BN structure came from. This makes it near impossible to duplicate the authors' process and duplicate their findings. Regardless of the origin of the BN structure, in almost half of the papers the complete BN structure was not provided. This also makes the process of verifying or using the structure impossible. In 109 of the papers the *BN parameters* were learned from data (either completely or partially). The process, either data or knowledge driven, was usually not described in detail. Parameter learning from data is more straightforward, while eliciting parameters from knowledge is less intuitive and results in subjective parameters, therefore a detailed description of the parameter elicitation should be a requirement of any BN in healthcare publication. Regardless of the origin of the parameters, the NPTs, which could be used as an important source of knowledge for other researchers, were rarely provided in the literature. More specifically, in 33 of the papers some parameters were presented, while a full list of parameters was available in only 7 papers. Finally, in spite of the advantages of mixed approaches for developing BN structures and parameters using both data and knowledge, only a small number of BNs developed by mixed approaches were identified in the literature.

7.1.3. Model usefulness

While 101 of published BNs addressed a clinical problem of significant importance and 94 of papers were shown to have an acceptable predictive performance, there appeared a significant gap between *having an accurate model* and *demonstrating a clinically useful model capable of delivering an actual impact on healthcare decision making*. A description of the model's credibility was observed in 42 of the papers. Particularly, papers with knowledge-driven BNs provided more convincing evidence of model credibility than those with data-driven BNs. Important elements for having a useful model, such as *generalizability*, *usability* and *impact* were either vague or overlooked in the literature. Generalizability is important for achieving BN adoption in healthcare as a model could not be used widely in practice if it has not first been proven to work on alternative populations and therefore applied to the broader health consumer population. Another important aspect considered was whether user-friendly interface and an impact analysis to study the benefit of the model and its integration into the existing workflow had been performed. In this review, no single BN adhered to all of the prescribed elements necessary for having a useful model. In addition, there was no evidence in this review to suggest that any of the BN models were in current clinical use. This does not categorically mean there are no BNs in clinical practice, only that in this review we did not find a publication describing its BN as currently engaged in clinical practice.

7.2. Strengths and limitations

The first strength of this review is that this is the largest review of BNs in healthcare targeting three objectives; (1) *decision support*, (2) *development*, and (3) *usefulness*, which, when combined, provide a comprehensive assessment of the domain. Several attributes related to each objective were identified based on a *preliminary short review* of the literature [32] followed by *a comprehensive review* process conducted by a panel of experts. The second strength is that in order to ensure a consistent review process, a set of tools were developed that include a detailed framework capturing all the necessary attributes, a systematically designed online literature review questionnaire and a manual to guide the respondents on completing the questionnaire. In the manual a detailed description on what each attribute means and a clear explanation on how to identify each attribute in a published paper were provided. An additional training event for the reviewers was conducted to make sure that all the participants were comfortable and had a robust understanding of the review process. In addition, to further guarantee that all the attributes were identified correctly, each paper was reviewed by two reviewers; one experienced and one junior. In cases of disagreement on an attribute of the review, two senior reviewers performed a final review of the paper collaboratively so that consensus could be achieved. In this review 123 papers were identified following the screening process and reviewed in order to address a significant gap in the literature. *However, this review has the following limitations*:

1 *Possible missing relevant papers*: Even if this is a representative sample of papers published in both medical and AI journals and conference proceedings, it may not reflect the entire range of the literature regarding BNs in healthcare. We looked for keywords such as *Bayesian Networks*, *probabilistic graphical models*, *medical*, *clinical* to appear in the abstract of each paper. It is possible that a small number of relevant papers were not included because they did not use the selected keywords in their abstract. This is especially true in cases where the actual name of the medical condition is described without mentioning the words *medical* or *clinical*. However, we believe that the large number of selected papers was sufficient for drawing conclusions.
2 *Limited time period*: Due to the vast number of published papers and the extensive time that would have been required for review and data extraction, we limited the study to the seven-year period 2012−2018. We believed this was a sufficient number of years to be able to draw conclusions and as such, papers published prior to that period were not reviewed. The literature search was conducted in early 2019 and data extraction and analysis occupied almost an entire year; hence, papers published between 2019 and 2020 are not included in the collection. This may be considered a limitation as it is possible that more recent publications could have less shortcomings, but we do not believe it significantly impacts the outcomes and findings of this review.
3 *Subjectivity*: *Despite* the detailed framework, the well-designed online survey and the guiding manual, we recognize that a few of the attributes were subjective and difficult to identify. For instance, to capture how repeatable was the overall development process as well as the structure and parameter learning or elicitation process, a subjective 3-scale answer was provided with the labels: Negligible, Insufficient and Sufficient detail. For some papers choosing the appropriate state was difficult and it was based on reviewer judgement. For evaluating the clinical usefulness of the BN in practice a set of elements previously described in the literature were examined. Some of these elements were again subjective and many times not clearly stated in the papers. However, to increase the correctness of our findings, each paper was read by two reviewers and a consensus was achieved in case of a disagreement.
4 *Possible missing evidence of adoption*: One of our key findings was the total absence of clinical *adoption* of any BN in the literature reviewed. However, it is possible that evidence for subsequent translation and adoption of BNs into the real world may well appear in other places, such company reports, media, general commentaries and not

necessarily in journal articles. For instance, based on limited marketing materials, there are indications that Babylon is using BNs in their non-public research and application development [145], including what is claimed as the largest BN in the world [146]. As a result, BNs that have been used in practice might have been missed. For that reason we tried to infer the adoption of the published BNs in practice either from the overall impression of the paper or based on whether the BN follows all the rules necessary for having a useful model, as a model cannot be applied in practice if it has not been proven to be useful first.

8. Recommendations for researchers

In this section, specific recommendations for improving the *reproducibility* of the modelling approach described in published papers as well as the *usefulness* of the BNs in practice are presented.

1 *The clinical benefit of the BN should be explained in depth*: As discussed in the third objective of this scoping review, the clinical benefit of a BN was usually explained in the published papers. However, there were many cases where the benefit was not provided, and therefore the aim of the BN was difficult to understand. The clinical benefit should be stated at the beginning of each paper, preferably supported with relevant literature and statistics that justify the need for a BN capable of performing better than the current practice standard. The effort of developing a BN to be used as a CDSS should be based on the need to address an important clinical problem with a significant potential benefit. Making the extra effort of thinking and describing the benefit from the beginning can help readers better understand the model's aim, and researchers better understand the future impact a model may have. One example of properly representing the benefit of the BN can be seen in the introduction of the work by Ducher et al. [147].

2 *The structure and the variables of the BN should be provided*: The second objective of this review identified that, in most papers, neither the BN structure nor the variables were clearly described. This makes it difficult to review the BN logic, and hides important knowledge that could potentially be used in other studies. As a result, the complete BN structure should be provided in a paper, and further, the variables of the BN should be fully described. Tables that include the full variable name, a description of the variable and its possible states are extremely useful and can be a significant source of knowledge that improves readers comprehension of the proposed model. An example of how variables should be described in this manner can be found in the papers published by Velikova et al. [92] and Constantinou et al. [58].

3 *The parameters of the BN should be provided*: Also, from the second objective of this review we identified that in most papers the BN parameters were not disclosed. Failure to disclose the parameters completely inhibits reproducibility of the claimed results. Additionally, complete NPTs can be a valuable source of knowledge for similar studies. To resolve this issue the marginal prior probabilities and complete NPTs should always be made available to the reader. An illustrative example on how NPTs should be presented can be found in the work of Jochems et al. [104].

4 *The BN development process should be explained in detail to the point of being repeatable*: We *identified* in the second objective of this review that 97 of published papers failed to fully describe the BN development process. Absence of a clear development process makes the research difficult to follow and impossible to duplicate. For that reason, the overall methodology should be described in a generic way to improve the readers comprehension of the approach used, and their ability to replicate the methodology used. A diagrammatic illustration of the overall methodology can be very informative, such as the one presented in Yet et al. [102].

5 *When the BN structure is elicited from knowledge, the elicitation process followed should be described in detail to the point of being repeatable*: While reviewing the modelling approach used, we observed that in cases where the BN structure was elicited from knowledge the process was rarely described. To simply say that the BN structure was elicited from experts or literature is not a sufficient description of methodology. The process of extracting variables, as well as the choice of arcs to be used, should be described clearly and open to critique by the reader. An example of a clear and concise description can be found in the works of Akhtar et al. [82] and Seixas et al. [135].

6 *When the BN parameters are elicited from knowledge, the elicitation process followed should be described in detail to the point of being repeatable*: Another finding from the second *objective* of this review was that in many papers where the BN parameters were elicited from knowledge, the process was not described in any detail. As with the previous items discussed above, this limits the reader's ability to understand, critique and replicate the modelling process. Consequently, the process for eliciting the parameters from knowledge should be clearly described both for methodological review, and for repeatability. In cases where the parameters are extracted from experts it is necessary to explain not only how the parameters were elicited, for instance using graphs or anchors, but also the number of experts involved, time and resources consumed by the process, how the different answers among the experts have been considered, the certainty of the elicited subjective parameters and the benefits and limitations of the chosen process. Similarly, when the parameters are elicited from literature a list of the publications that were used should be provided, the process of selecting the necessary parameters should be explained, and the degree of trust for the elicited parameters. An example of a well described parameter elicitation process can be found in Yet et al. [105] and Constantinou et al. [85].

7 *When the BN structure and parameters are learned from data, the learning algorithm as well as the dataset used, and the way missing data was treated should be provided*: We identified that insufficient information was usually provided for the characteristics and limitations of the data that was used to develop data-driven BNs. This creates significant gaps in the reader's understanding of the modelling process. When learning the BN structure from data, the dataset used should be described, including presentation of elements such as the dataset origin, size and limitations, as in Lee et al. [73] and Zador et al. [131]. In addition, the learning algorithm used should be described in detail, such as in the works of Lee et al. [73] and Luo et al. [86]. A significant element that should never be overlooked is how missing data were identified and classified, for example: not missing at random or missing completely at random, and so on. How this missing data was treated, i.e. whether missing values were computed or ignored, should also be described. A clear description of the learning process makes it easier for readers to follow and potentially replicate the process. A good example of a description of data pre-processing and how to handle missing data can be found in the paper published by Sesen et al. [55].

8 *The credibility of a data-driven BN structure should be explained*: From the third objective of this review we learned that the credibility of data-driven BNs was difficult to assess as it was usually underexplored by authors. When the BN structure is learned from data, it represents the relationships that are available in the dataset, which are not always in agreement with the true causal mechanisms of a given disease. For that reason, a justification of the model's credibility is needed to ensure that the BN has a clear logic and evaluates the answer it was intended to compute. Representing true causal/ influential relationships is particularly important when the BN is intended for interventional

or counterfactual reasoning. An example of credible data-driven BN structures can be found in the works of Chao et al. [67] and Kaewprag et al. [148].

9 *A validation of the model's accuracy should be provided*: As observed in the third objective of this review, most authors provided some claim of their model's accuracy. However, their approach to assessing and validating accuracy was not always described with sufficient clarity to be easily understood and evaluated by the reader. In this review we found several ways to validate a BN. A simple approach for evaluating the model's reasoning was through the use of specific cases or scenarios. Although useful, this is not sufficient to fully validate a model's accuracy. The minimum validation process needed is internal validation, where the predictive performance of the model is assessed using resampling techniques such as bootstrapping or cross-validation. All steps followed in validation of the BN should be explained in sufficient detail to enable repeatability. In cases where low accuracy is observed, justification and potential corrections should be offered. In cases where internal validation has not been performed, this should be openly disclosed and suggested as future work, thus indicating the need for a validation study. A well-described validation and review process can be found in Yet et al. [102].

10 *An external validation to evaluate the model's generalizability should be considered*: An important finding of the third objective of this review was a lack of generalizability for most of the described models. An accurate model may not be useful in clinical practice if it is not proven to work consistently on different populations. Thus, an external validation is necessary for validating the model's performance on different datasets. Providing an external validation significantly strengthens the performance and utility of a BN. In cases where an external validation has not been performed due to data or time restrictions, it should be openly disclosed and suggested as future work as we see in the work of Park et al. [126]. An example of external validation can be found in the paper published by Luo et al. [143].

11 *The actual use of the model in practice and its impact on healthcare should be considered*: The third objective of our review identified that impact analysis was usually neglected, especially in research papers where the sole focus was on development of the BN. However, in order to bridge the gap between having accurate models and delivering impact in practice, more focus should be given to identifying issues and approaches to help clinicians adopt the decision support model. In cases where a small pilot study to investigate the potential impact of the models, such as in the work of Mcheick et al. [133], or a properly randomized control trial have not been performed, this should also be openly disclosed and identified as potential future work. Acknowledging the need for an impact study and how the model could be applied in practice to help clinical decision making provides the reader with a clear understanding of the intended use and the potential benefits that may result from use of the proposed model. A representative example acknowledging an impact analysis as a crucial future step is available in Velikova et al. [92].

9. Future directions

This scoping review revealed the following research areas that have been largely neglected, and that should receive further attention.

1 *Multi-disciplinary expert team involvement in BN development:* When developing medical BNs intended for use as CDSS there is a need for collaboration between clinicians and decision scientists. Clinical expertise is important to ensure that the clinical benefit and adoption of the model are considered, while decision scientists, computer scientists or statisticians are necessary to develop an accurate BN model. Methodologies and tools to support multidisciplinary expert involvement processes for BN development would be useful to resolving this need.

2 *Using BNs to answer hypothetical questions:* BNs should be used to perform more than just reasoning from evidence. They can also be used to answer *hypothetical questions.* More research should focus on modeling and validating medical BNs for performing *interventional* and *counterfactual reasoning.*

3 *Temporal BN models:* Greater effort is needed to identify better methods for developing temporal BNs that capture the evolving process of clinical decision making.

4 *Clinical datasets for use in harnessing ML for developing BNs:* Despite supposedly being in the era of big data and machine learning, many papers in this review identified a lack of good quality clinical datasets as a barrier to developing accurate BNs. Solving the issues of data integrity, integration and interoperability, especially in electronic health records and routine clinical datasets, will be key to ensuring that health data can be used to render credible and accurate predictions and decisions for use in the care of individual patients and their families.

5 *Mixed methodologies for developing BNs:* More work is needed to develop BNs using mixed approaches that are based on both data and knowledge.

6 *Tools and methods for eliciting expert knowledge for BN development:* Significant effort has been made for eliciting BN parameters from experts [149,150]. However, eliciting BN structure from experts remains largely *ad hoc* and research into more methodological approaches is needed in order to standardize these processes.

7 *Adoption of medical BNs in clinical practice:* While there has been significant research interest in developing accurate medical BNs, their adoption in clinical practice is less obvious. More attention should be paid to the process of translating an accurate BN to a useful and usable CDSS that has potential for benefit when applied in clinical practice. It is also important not to neglect the issue of trust. Unexplainable or black-box CDSS are not going to be trusted or adopted by clinicians. Thus, having explainable CDSS is crucial [151,152].

10. Summary and conclusions

BNs can be successfully applied to model complex medical problems requiring reasoning under uncertainty. An immense interest is demonstrated by the volume of literature proposing medical BNs and the research effort expended globally in developing BNs to support clinical decision making. This scoping review, drawing on 123 papers that propose BNs in healthcare, evaluated the decision support, development and usefulness of BN models proposed for use in clinical practice. We found that the majority of published BNs focus on supporting simple decisions where reasoning from evidence is required, with authors neglecting more complex reasoning approaches that BNs may be applied to, and the dynamic nature of clinical decision making. There is a preference for data-driven BNs even though medical datasets usually lack integrity, presenting with poor quality and incomplete. There is also an important lack of repeatable modelling approaches and a significant gap between development of a claimed accurate model, and demonstration of a clinically useful model capable of delivering actual impact on healthcare. Consequently, we have proposed several future research directions and recommendations for improving reproducibility of modelling approaches described in the literature, and for improving the overall usefulness of medical BNs. With these future directions we hope the quality and repeatability of published papers on BNs in healthcare will improve and that all researchers in this domain can work towards closing the chasm between research interest and clinical adoption.

Declaration of Competing Interest

The authors report no declarations of Interest.

Acknowledgments

EK, SM, MRN, AF and NF acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC) under project EP/P009964/1: PAMBAYESIAN: Patient Managed decision-support using Bayes Networks. KD acknowledges financial support from the School of Fundamental Sciences, Massey University, New Zealand, for his study sabbatical and visits with the PamBayesian team.

References

[1] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers Inc.; 1988.
[2] Cooper GF. The computational complexity of probabilistic inference using Bayesian belief networks. Artif Intell 1990;42(2–3):393–405. https://doi.org/10.1016/0004-3702(90)90060-D.
[3] Heckerman D. A tutorial on learning with Bayesian networks. 1995 [Online]. Available: /home/obada/Desktop/MAS1-ISW/Robot Learning/A tutorial on learning with Bayesian networks.pdf.
[4] Bielza C, Larranaga P. Bayesian networks in neuroscience: a survey. Front Comput Neurosci 2014;8:1–23. https://doi.org/10.3389/fncom.2014.00131.
[5] Tennant PWG, et al. Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. medRxiv 2019. https://doi.org/10.1101/2019.12.20.19015511.
[6] McLachlan S, Dube K, Hitman GA, Fenton NE, Kyrimi E. Bayesian networks in healthcare: distribution by medical condition. Artif Intell Med 2020;107 (January):101912. https://doi.org/10.1016/j.artmed.2020.101912.
[7] Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? BMJ 1995;311(December):1539–41. https://doi.org/10.1136/bmj.311.7019.1539.
[8] Blackmore CC. Clinical prediction rules in trauma imaging: Who, how, and why? Radiology 2005;235:371–4. https://doi.org/10.1148/radiol.2352040307.
[9] Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 2006;144:201–9 [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-33644855006&partnerID=40&md5=903fb2ad9598fcbdc1984db9a9af8784.
[10] Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338(January):1487–90. https://doi.org/10.1136/bmj.b606.
[11] Adams ST, Leveson SH. Clinical prediction rules. BMJ 2012;344:d8312. https://doi.org/10.1136/bmj.d8312.
[12] Pearl J. Graphical models. Causality Intervention 1993;8.
[13] Fenton N, Neil M. Risk assessment and decision analysis with bayesian networks. second edi. New York: Chapman and Hall/CRC; 2019.
[14] Bayes T. An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc London 1763;53:370–418. https://doi.org/10.1098/rstl.1763.0053.
[15] Dranca L, et al. Using Kinect to classify Parkinson's disease stages related to severity of gait impairment. BMC Bioinformatics 2018;19(1):1–15. https://doi.org/10.1186/s12859-018-2488-4.
[16] Philipp P, Beyerer J, Fischer Y. Expert-based probabilistic modeling of workflows in context of surgical interventions. 2017. https://doi.org/10.1109/COGSIMA.2017.7929589.
[17] Flores MJ, Nicholson AE, Brunskill A, Korb KB, Mascaro S. Incorporating expert knowledge when learning Bayesian network structure: a medical case study. Artif Intell Med 2011;53(3):181–204. https://doi.org/10.1016/j.artmed.2011.08.004.
[18] Woodward J. Making things happen. A theory of causal explanation. 2003.
[19] Pearl J. Causality: models, reasoning, and inference. 2nd edition 2009.
[20] Shpitser I, Pearl J. Complete identification methods for the causal hierarchy. J Mach Learn Res 2008;9:1941–79. doi: 10.1.1.210.7793.
[21] Hagmayer Y, Sloman SA, Lagnado DA, Waldmann MR. Causal reasoning through intervention. Oxford: Oxford University Press; 2007.
[22] Dawid AP. Causal inference without counterfactuals. J Am Stat Assoc 2000;95 (450):407–24.
[23] Sloman SA. Counterfactuals and causal models: introduction to the special issue. Cogn Sci 2013;37(6):969–76. https://doi.org/10.1111/cogs.12064.
[24] Tiffen J, Corbridge SJ, Slimmer L. Enhancing clinical decision making: development of a contiguous definition and conceptual framework. J Prof Nurs 2014;30(5):399–405. https://doi.org/10.1016/j.profnurs.2014.01.006.
[25] Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. J Eval Clin Pract 2001;7(May 2): 97–107. https://doi.org/10.1046/j.1365-2753.2001.00284.x.
[26] Druzdzel MJ, Flynn RR. Decision support systems. Encyclopedia of library and information science. second edition 2002.
[27] Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Aust J Physiother 2006;52(3):157–63. https://doi.org/10.1016/S0004-9514(06)70024-1.
[28] Patel VL, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009;46(1):5–17. https://doi.org/10.1016/j.artmed.2008.07.017.
[29] Shortliffe EH, Cimino JJ. Biomedical informatics: computer applications in health care and biomedicine. fourth edition 2013.
[30] Lucas P. Bayesian networks in medicine: a model-based approach to medical decision making. 2001.
[31] Lucas P, van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. Artif Intell Med 2004;30(March):201–14. https://doi.org/10.1016/j.artmed.2003.11.001.
[32] Kyrimi E, McLachlan S, Dube K, Fenton N. Bayesian networks in healthcare: the chasm between research enthusiasm and clinical adoption. medRxiv Prepr 2020. https://doi.org/10.1101/2020.06.04.20122911.
[33] Kyrimi E, Dube K, Fenton N, Fahmi A, Neves MR, Marsh W, et al. Bayesian Networks in Healthcare: What is preventing their adoption? Artif Intell Med 2021;116:102079. https://doi.org/10.1016/j.artmed.2021.102079 [Online].
[34] Dube K, McLachlan S, Daley B, Hitman GA, Fenton N, Kyrimi E. Survey tool approach for collaborative literature reviews (STaR): case studies in artificial intelligence in healthcare. Manuscr Prep 2021.
[35] Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338(jun04 2):b606. https://doi.org/10.1136/bmj.b606.
[36] Altman GD, Vergouwe Y, Royston P, Karel KGM, Moons GM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:605.
[37] Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: A review. J Clin Epidemiol 2008;61(November): 1085–94. https://doi.org/10.1016/j.jclinepi.2008.04.008.
[38] Gilchrist A. Thesauri, taxonomies and ontologies - an etymological note. J Doc 2003;59(1):7–18. https://doi.org/10.1108/00220410310457984.
[39] Kakabadse NK, Kakabadse A, Kouzmin A. Reviewing the knowledge management literature: towards a taxonomy. J Knowl Manag 2003;7(4):75–91. https://doi.org/10.1108/13673270310492967.
[40] Bowman JP, Nichols DS. Aequorivita gen. nov., a member of the family Flavobacteriaceae isolated from terrestrial and marine Antarctic habitats. Int J Syst Evol Microbiol 2002;52(5):1533–41. https://doi.org/10.1099/ijs.0.01976-0.
[41] Triplett J. Phylogeny and taxonomy of the genus Arundinaria (Poaceae: bambusoideae). Sourced from:. 2018. https://linearpopulationmodel.blogspot.com/2018/09/phylogeny-and-taxonomy-of-genus.html.
[42] Atkinson SP. Graduate competencies, employability and educational taxonomies: critique of intended learning outcomes. Pract Evid Scholarsh Teach Learn High Educ 2015;10(2):154–77.
[43] Van Der Maaten LJP, Postma EO, Van Den Herik HJ. Dimensionality reduction: a comparative review. J Mach Learn Res 2009;10:1–41. https://doi.org/10.1080/13506280444000102.
[44] Jordan D. First proposal for a two-way radio article taxonomy. Sourced from:. 2007. https://en.wikipedia.org/wiki/File:Two-way_Radio_taxonomy_diagram.png.
[45] Gianaroli L, Magli MC, Gambardella L, Giusti A, Grugnetti C, Corani G. Objective way to support embryo transfer: a probabilistic decision. Hum Reprod 2013;28 (5):1210–20. https://doi.org/10.1093/humrep/det030.
[46] Qian X, Wang Y, Jiang X. Parallel bayesian network modelling for pervasive health monitoring system. the International Parallel and Distributed Processing Symposium, IPDPS 2014:1631–7. https://doi.org/10.1109/IPDPSW.2014.182.
[47] Wang Z, et al. 3-D stent detection in intravascular OCT using a Bayesian network and graph search. IEEE Trans Med Imaging 2015;34(7):1549–61. https://doi.org/10.1109/TMI.2015.2405341.
[48] Agarwal P, Verma R, Mallik A. Ontology based disease diagnosis system with probabilistic inference. 2016. https://doi.org/10.1109/IICIP.2016.7975383.
[49] Ong MS, et al. Learning a comorbidity-driven taxonomy of pediatric pulmonary hypertension. Circ Res 2017;121(4):341–53. https://doi.org/10.1161/CIRCRESAHA.117.310804.
[50] Karaca Y, Sayici G, Mathematics T, Methods Q. Bayesian networks for sub-groups of multiple sclerosis. Int J Electron Mech Mechatronics Eng 2013;3(1):455–62.
[51] Khademi M, Nedialkov NS. Probabilistic graphical models and deep belief networks for prognosis of breast cancer. EEE 14th International Conference on Machine Learning and Applications, ICMLA 2015:727–32. https://doi.org/10.1109/ICMLA.2015.196.
[52] Refai A, Merouani HF, Aouras H. Maintenance of a Bayesian network: application using medical diagnosis. Evol Syst 2016;7(3):187–96. https://doi.org/10.1007/s12530-016-9146-8.
[53] Kuang D, et al. Depression recognition according to heart rate variability using Bayesian Networks. J Psychiatr Res 2017;95:282–7. https://doi.org/10.1016/j.jpsychires.2017.09.012.
[54] Liu S, et al. Quantitative analysis of breast cancer diagnosis using a probabilistic modelling approach. Comput Biol Med 2018;92(May 2017):168–75. https://doi.org/10.1016/j.compbiomed.2017.11.014.
[55] Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. Bayesian networks for clinical decision support in lung cancer care. PLoS One 2013;8(12): 1–13. https://doi.org/10.1371/journal.pone.0082349.
[56] Sesen MB, et al. Lung cancer assistant: a hybrid clinical decision support application for lung cancer care. J R Soc Interface 2014;11:20140534. https://doi.org/10.1098/rsif.2014.0534.
[57] Liu R, et al. Pathway-finder: an interactive recommender system for supporting personalized care pathways. IEEE International Conference on Data Mining Workshops, ICDMW 2015:1219–22. https://doi.org/10.1109/ICDMW.2014.37. no. January.
[58] Constantinou AC, Freestone M, Marsh W, Fenton N, Coid J. Risk assessment and risk management of violent reoffending among prisoners. Expert Syst Appl 2015; 42(21):7511–29. https://doi.org/10.1016/j.eswa.2015.05.025.

[59] Zarikas V, Papageorgiou E, Regner P. Bayesian network construction using a fuzzy rule based approach for medical decision support. Expert Syst 2015;32(3): 344–69. https://doi.org/10.1111/exsy.12089.

[60] Constantinou AC, Fenton N, Marsh W, Radlinski L. From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. Artif Intell Med 2016;67:75–93. https://doi.org/10.1016/j.artmed.2016.01.002.

[61] Xu S, et al. Modeling interrelationships between health behaviors in overweight breast cancer survivors: applying Bayesian networks. PLoS One 2018;13(9):1–13. https://doi.org/10.1371/journal.pone.0202923.

[62] Neapolitan R, Jiang X, Ladner DP, Kaplan B. A primer on Bayesian decision analysis with an application to a kidney transplant decision. Transplantation 2016;100(3):489–96. https://doi.org/10.1097/TP.0000000000001145.

[63] Hernandez-Leal P, et al. Discovering human immunodeficiency virus mutational pathways using temporal Bayesian networks. Artif Intell Med 2013;57(3):185–95. https://doi.org/10.1016/j.artmed.2013.01.005.

[64] Cai Z-Q, et al. Analysis of prognostic factors for survival after hepatectomy for hepatocellular carcinoma based on a bayesian network. PLoS One 2015;10(3): e0120805. https://doi.org/10.1371/journal.pone.0120805.

[65] Fuster-Parra P, Tauler P, Bennasar-Veny M, Ligeza A, López-González AA, Aguiló A. Bayesian network modeling: a case study of an epidemiologic system analysis of cardiovascular risk. Comput Methods Programs Biomed 2016;126: 128–42. https://doi.org/10.1016/j.cmpb.2015.12.010.

[66] Solomon NP, Dietsch AM, Dietrich-Burns KE, Styrmisdottir EL, Armao CS. Dysphagia management and research in an acute-care military treatment facility: the role of applied informatics. Mil Med 2016;181(5S):138–44. https://doi.org/10.7205/milmed-d-15-00170.

[67] Chao YS, et al. A network perspective on patient experiences and health status: the medical expenditure panel survey 2004 to 2011. BMC Health Serv Res 2017; 17(1):1–12. https://doi.org/10.1186/s12913-017-2496-5.

[68] Constantinou AC, Yet B, Fenton N, Neil M, Marsh W. Value of Information analysis for interventional and counterfactual Bayesian networks in forensic medical sciences. Artif Intell Med 2015. https://doi.org/10.1016/j.artmed.2015.09.002.

[69] Kim M, Cheeti A, Yoo C, Choo M, Paick JS, Oh SJ. Non-invasive clinical parameters for the prediction of urodynamic bladder outlet obstruction: analysis using causal Bayesian networks. PLoS One 2014;9(11):9–14. https://doi.org/10.1371/journal.pone.0113131.

[70] Fröhlich H, Bahamondez G, Götschel F, Korf U. Dynamic Bayesian network modeling of the interplay between EGFR and hedgehog signaling. PLoS One 2015;10(11):1–14. https://doi.org/10.1371/journal.pone.0142646.

[71] Jin Y, Su Y, Zhou XH, Huang S. Heterogeneous multimodal biomarkers analysis for Alzheimer's disease via Bayesian network. Eurasip J Bioinform Syst Biol 2016; 2016(1):4–11. https://doi.org/10.1186/s13637-016-0046-9.

[72] Zeng Z, Jiang X, Neapolitan R. Discovering causal interactions using Bayesian network scoring and information gain. BMC Bioinformatics 2016;17(1):1–14. https://doi.org/10.1186/s12859-016-1084-8.

[73] Lee S, Jiang X. Modeling miRNA-mRNA interactions that cause phenotypic abnormality in breast cancer patients. PLoS One 2017;12(8):e0182666. https://doi.org/10.1371/journal.pone.0182666.

[74] Noyes N, Cho KC, Ravel J, Forney LJ, Abdo Z. Associations between sexual habits, menstrual hygiene practices, demographics and the vaginal microbiome as revealed by Bayesian network analysis. PLoS One 2018;13(1):1–25. https://doi.org/10.1371/journal.pone.0191625.

[75] Vila-Francés J, et al. Expert system for predicting unstable angina based on Bayesian networks. Expert Syst Appl 2013;40(12):5004–10. https://doi.org/10.1016/j.eswa.2013.03.029.

[76] Farmer N. An update and further testing of a knowledge-based diagnostic clinical decision support system for musculoskeletal disorders of the shoulder for use in a primary care setting. J Eval Clin Pract 2014;20(5):589–95. https://doi.org/10.1111/jep.12153.

[77] Moreira LB, Namen AA. A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. Comput Methods Programs Biomed 2018;165: 139–49. https://doi.org/10.1016/j.cmpb.2018.08.016.

[78] Zhang Y, et al. Identification of reciprocal causality between non-alcoholic fatty liver disease and metabolic syndrome by a simplified Bayesian network in a Chinese population. BMJ Open 2015;5(9). https://doi.org/10.1136/bmjopen-2015-008204.

[79] Cardin F, Minicuci N, Andreotti A, Granziera E, Militello C. Factors influencing challenging colonoscopies during anesthesiologist-assisted deep sedation. Saudi J Gastroenterol 2016;22(1):64–8. https://doi.org/10.4103/1319-3767.173761.

[80] Soto-Ferrari M, Prieto D, Munene G. A Bayesian network and heuristic approach for systematic characterization of radiotherapy receipt after breast-conservation surgery. BMC Med Inform Decis Mak 2017;17(1):1–16. https://doi.org/10.1186/s12911-017-0479-4.

[81] Gross TJ, Araújo RB, Vale FAC, Bessani M, Maciel CD. Dependence between cognitive impairment and metabolic syndrome applied to a Brazilian elderly dataset. Artif Intell Med 2018;90(January):53–60. https://doi.org/10.1016/j.artmed.2018.07.003.

[82] Akhtar MJ, Utne IB. Human fatigue's effect on the risk of maritime groundings - A Bayesian Network modeling approach. Saf Sci 2014;62:427–40. https://doi.org/10.1016/j.ssci.2013.10.002.

[83] Lee J, Lee W, Park IS, Kim HS, Lee H, Jun CH. Risk assessment for hypertension and hypertension complications incidences using a Bayesian network. IIE Trans Healthc Syst Eng 2016;6(4):246–59. https://doi.org/10.1080/19488300.2016.1232767.

[84] Takenaka S, Aono H. Prediction of postoperative clinical recovery of drop foot attributable to Lumbar degenerative diseases, via a Bayesian network. Clin Orthop Relat Res 2017;475(3):872–80. https://doi.org/10.1007/s11999-016-5180-x.

[85] Constantinou AC, Freestone M, Marsh W, Coid J. Causal inference for violence risk management and decision support in forensic psychiatry. Decis Support Syst 2015;80:42–55. https://doi.org/10.1016/j.dss.2015.09.006.

[86] Luo Y, et al. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis. Radiother Oncol 2017;123 (1):85–92. https://doi.org/10.1016/j.radonc.2017.02.004.

[87] Sandri M, Berchialla P, Baldi I, Gregori D, De Blasi RA. Dynamic Bayesian networks to predict sequences of organ failures in patients admitted to ICU. J Biomed Inform 2014;48:106–13. https://doi.org/10.1016/j.jbi.2013.12.008.

[88] Orphanou K, Stassopoulou A, Keravnou E. DBN-extended: a dynamic Bayesian network model extended with temporal abstractions for coronary heart disease prognosis. IEEE J Biomed Health Informatics 2016;20(3):944–52. https://doi.org/10.1109/JBHI.2015.2420534.

[89] Yousefi L, Tucker A, Al-Luhaybi M, Saachi L, Bellazzi R, Chiovato L. Predicting disease complications using a stepwise hidden variable approach for learning dynamic Bayesian networks. IEEE 31st Symposium on Computer-Based Medical Systems 2018:106–11. https://doi.org/10.1109/CBMS.2018.00026. vol. June.

[90] Exarchos KP, et al. Prediction of coronary atherosclerosis progression using dynamic Bayesian networks. the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2013:3889–92. https://doi.org/10.1109/EMBC.2013.6610394.

[91] Exarchos KP, et al. A multiscale approach for modeling atherosclerosis progression. IEEE J Biomed Health Informatics 2015;19(2):709–19. https://doi.org/10.1109/JBHI.2014.2323935.

[92] Velikova M, Van Scheltinga JT, Lucas PJF, Spaanderman M. Exploiting causal functional relationships in Bayesian network modelling for personalised healthcare. Int J Approx Reason 2014;55:59–73. https://doi.org/10.1016/j.ijar.2013.03.016.

[93] Orphanou K, Stassopoulou A, Keravnou E. Temporal abstraction and temporal Bayesian networks in clinical domains: a survey. Artif Intell Med 2014;60(3): 133–49. https://doi.org/10.1016/j.artmed.2013.12.007.

[94] Haddad T, Himes A, Campbell M. Fracture prediction of cardiac lead medical devices using Bayesian networks. Reliab Eng Syst Saf 2014;123:145–57. https://doi.org/10.1016/j.ress.2013.11.005.

[95] Chang YS, Fan CT, Lo WT, Hung WC, Yuan SM. Mobile cloud-based depression diagnosis using an ontology and a Bayesian network. Future Gener Comput Syst 2015;43–44:87–98. https://doi.org/10.1016/j.future.2014.05.004.

[96] Vemulapalli V, et al. Non-obvious correlations to disease management unraveled by Bayesian artificial intelligence analyses of CMS data. Artif Intell Med 2016;74: 1–8. https://doi.org/10.1016/j.artmed.2016.11.001.

[97] Alobaidi ATS, Mahmood NT. Modified full Bayesian networks classifiers for medical diagnosis. International Conference on Advanced Computer Science Applications and Technologies, ACSAT 2013 2013:5–12. https://doi.org/10.1109/ACSAT.2013.10.

[98] Loghmanpour NA, Druzdzel MJ, Antaki JF. Cardiac health risk stratification system (CHRiSS): A Bayesian-based decision support system for left ventricular assist device (LVAD) therapy. PLoS One 2014;9(11):1–10. https://doi.org/10.1371/journal.pone.0111264.

[99] Ojeme B, Mbogho A. Selecting learning algorithms for simultaneous identification of depression and comorbid disorders. Procedia Comput Sci 2016;96:1294–303. https://doi.org/10.1016/j.procs.2016.08.174.

[100] Liu M, Hommersom A, van der Heijden M, Lucas PJF. Hybrid time Bayesian networks. Int J Approx Reason 2017;80:460–74. https://doi.org/10.1016/j.ijar.2016.02.009.

[101] Yetton BD, McDevitt EA, Cellini N, Shelton C, Mednick SC. Quantifying sleep architecture dynamics and individual differences using big data and Bayesian networks. PLoS One 2018;13(4).

[102] Yet B, Perkins Z, Fenton N, Tai N, Marsh W. Not just data: a method for improving prediction with knowledge. J Biomed Inform 2014;48(April):28–37. https://doi.org/10.1016/j.jbi.2013.10.012 [doi].

[103] Kalet AM, Gennari JH, Ford EC, Phillips MH. Bayesian network models for error detection in radiotherapy plans. Phys Med Biol 2015;60(7):2735–49. https://doi.org/10.1088/0031-9155/60/7/2735.

[104] Jochems A, et al. Distributed learning: developing a predictive model based on data from multiple hospitals without data leaving the hospital – a real life proof of concept. Radiother Oncol 2016;121(3):459–67. https://doi.org/10.1016/j.radonc.2016.10.002.

[105] Yet B, Bastani K, Raharjo H, Lifvergren S, Marsh W, Bergman B. Decision support system for Warfarin therapy management using Bayesian networks. Decis Support Syst 2013;55(2):488–98. https://doi.org/10.1016/j.dss.2012.10.007.

[106] Yet B, Perkins ZB, Rasmussen TE, Tai NRM, Marsh DWR. Combining data and meta-analysis to build Bayesian networks for clinical decision support. J Biomed Inform 2014;(August). https://doi.org/10.1016/j.jbi.2014.07.018.

[107] Bandyopadhyay S, et al. Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data. Data Min Knowl Discov 2015;29(4):1033–69. https://doi.org/10.1007/s10618-014-0386-6.

[108] Caillet P, Klemm S, Ducher M, Aussem A, Schott AM. Hip fracture in the elderly: a re-analysis of the EPIDOS study with causal Bayesian networks. PLoS One 2015; 10(3):1–12. https://doi.org/10.1371/journal.pone.0120125.

[109] Yang X, Li L, Wang J, Huang J, Lu S. Cardiovascular mortality associated with low and high temperatures: determinants of inter-region vulnerability in China. Int J

Environ Res Public Health 2015;12(6):5918–33. https://doi.org/10.3390/ijerph120605918.
[110] Tylman W, et al. Real-time prediction of acute cardiovascular events using hardware-implemented Bayesian networks. Comput Biol Med 2016;69:245–53. https://doi.org/10.1016/j.compbiomed.2015.08.015.
[111] Magrini A, Luciani D, Stefanini FM. A probabilistic network for the diagnosis of acute cardiopulmonary diseases. Biom J 2018;60(1):174–95. https://doi.org/10.1002/bimj.201600206.
[112] Marvasti NB, Yoruk E, Acar B. Computer-aided medical image annotation: preliminary results with liver lesions in CT. IEEE J Biomed Health Informatics 2018;22(5):1561–70. https://doi.org/10.1109/JBHI.2017.2771211.
[113] Cypko MA, et al. Validation workflow for a clinical Bayesian network model in multidisciplinary decision making in head and neck oncology treatment. Int J Comput Assist Radiol Surg 2017;12(11):1959–70. https://doi.org/10.1007/s11548-017-1531-7.
[114] Sa-ngamuang C, Haddawy P, Luvira V, Piyaphanee W, Iamsirithaworn S, Lawpoolsri S. Accuracy of dengue clinical diagnosis with and without NS1 antigen rapid test: comparison between human and Bayesian network model decision. PLoS Negl Trop Dis 2018;12(6):1–14. https://doi.org/10.1371/journal.pntd.0006573.
[115] Jiao Y, Wang XH, Chen R, Tang TY, Zhu XQ, Teng GJ. Predictive models of minimal hepatic encephalopathy for cirrhotic patients based on large-scale brain intrinsic connectivity networks. Sci Rep 2017;7(1):1–12. https://doi.org/10.1038/s41598-017-11196-y.
[116] Aushev A, et al. Feature selection for the accurate prediction of septic and cardiogenic shock ICU mortality in the acute phase. PLoS One 2018;13(11):1–18. https://doi.org/10.1371/journal.pone.0199089.
[117] Zarringhalam K, Enayetallah A, Reddy P, Ziemek D. Robust clinical outcome prediction based on Bayesian analysis of transcriptional profiles and prior causal networks. Bioinformatics 2014;30(12):69–77. https://doi.org/10.1093/bioinformatics/btu272.
[118] Gámez-Pozo A, et al. Functional proteomics outlines the complexity of breast cancer molecular subtypes. Sci Rep 2017;7(1):1–13. https://doi.org/10.1038/s41598-017-10493-w.
[119] Martin N, Fenton N, Nielsen L. Building large-scale Bayesian networks. Knowl Eng Rev 2000;15(3):257–84. https://doi.org/10.1017/S0269888900003039.
[120] Shen Y, et al. CBN: Constructing a clinical Bayesian network based on data from the electronic medical record. J Biomed Inform 2018;88(September):1–10. https://doi.org/10.1016/j.jbi.2018.10.007.
[121] Nandra R, Parry M, Forsberg J, Grimer R. Can a bayesian belief network Be used to estimate 1-year survival in patients with bone sarcomas? Clin Orthop Relat Res 2017;475(6):1681–9. https://doi.org/10.1007/s11999-017-5346-1.
[122] Lou Y, Tu SW, Nyulas C, Tudorache T, Chalmers RJG, Musen MA. Use of ontology structure and Bayesian models to aid the crowdsourcing of ICD-11 sanctioning rules. J Biomed Inform 2017;68(2017):20–34. https://doi.org/10.1016/j.jbi.2017.02.004.
[123] Moreira MWL, Rodrigues JJPC, Oliveira AMB, Ramos RF, Saleem K. A preeclampsia diagnosis approach using Bayesian networks. the IEEE International Conference on Communications (ICC) 2016:1–5. https://doi.org/10.1109/ICC.2016.7510893.
[124] Fenton N, Neil M, Lagnado D, Marsh W, Yet B, Constantinou A. How to model mutually exclusive events based on independent causal pathways in Bayesian network models. Knowledge-Based Syst 2016;113:39–50. https://doi.org/10.1016/j.knosys.2016.09.012.
[125] Berchialla P, et al. Predicting severity of pathological scarring due to burn injuries: a clinical decision making tool using Bayesian networks. Int Wound J 2012;11(3):246–52. https://doi.org/10.1111/j.1742-481X.2012.01080.x.
[126] Park E, Chang HJ, Nam HS. A Bayesian network model for predicting post-stroke outcomes with available risk factors. Front Neurol 2018;9(September):1–11. https://doi.org/10.3389/fneur.2018.00699.
[127] Lappenschaar M, Hommersom A, Lucas PJF, Lagro J, Visscher S. Multilevel Bayesian networks for the analysis of hierarchical health care data. Artif Intell Med 2013;57(3):171–83. https://doi.org/10.1016/j.artmed.2012.12.007.
[128] Turhan M, Bostanci A, Bozkurt S. Estimation of cardiovascular disease from polysomnographic parameters in sleep-disordered breathing. Eur Arch Oto-Rhino-Laryngology 2016;273(12):4585–93. https://doi.org/10.1007/s00405-016-4176-1.
[129] Bukhanov N, Balakhontceva M, Kovalchuk S, Zvartau N, Konradi A. Multiscale modeling of comorbidity relations in hypertensive outpatients. Procedia Comput Sci 2017;121:446–50. https://doi.org/10.1016/j.procs.2017.11.060.
[130] Seixas AA, et al. Sleep duration and physical activity profiles associated with self-reported stroke in the United States: application of Bayesian belief network modeling techniques. Front Neurol 2018;9(July):1–13. https://doi.org/10.3389/fneur.2018.00534.
[131] Zador Z, Sperrin M, King AT. Predictors of outcome in traumatic brain injury: new insight using receiver operating curve indices and Bayesian network analysis. PLoS One 2016;11(7):1–18. https://doi.org/10.1371/journal.pone.0158762.
[132] Ji Z, Wang B. Identifying potential clinical syndromes of hepatocellular carcinoma using PSO-based hierarchical feature selection algorithm. Biomed Res Int 2014;2014(1). https://doi.org/10.1155/2014/127572.
[133] Mcheick H, Saleh L, Ajami H, Mili H. Context relevant prediction model for COPD domain using Bayesian belief network. Sensors (Switzerland) 2017;17(7):1–24. https://doi.org/10.3390/s17071486.
[134] Hu H, Kerschberg L. Evolving medical ontologies based on causal inference. EEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018:954–7. https://doi.org/10.1109/ASONAM.2018.8508259.
[135] Seixas FL, Zadrozny B, Laks J, Conci A, Muchaluat Saade DC. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. Comput Biol Med 2014;51:140–58. https://doi.org/10.1016/j.compbiomed.2014.04.010.
[136] Ducher M, et al. Comparison of a Bayesian network with a logistic regression model to forecast IgA nephropathy. Biomed Res Int 2013;2013. https://doi.org/10.1155/2013/686150.
[137] Van der Heijden M, Velikova M, Lucas PJF. Learning Bayesian networks for clinical time series analysis. J Biomed Inform 2014;48:94–105. https://doi.org/10.1016/j.jbi.2013.12.007.
[138] Merli M, Moscatelli M, Mariotti G, Pagliaro U, Bernardelli F, Nieri M. A minimally invasive technique for lateral maxillary sinus floor elevation: a Bayesian network study. Clin Oral Implants Res 2016;27(3):273–81. https://doi.org/10.1111/clr.12524.
[139] McNally RJ, Mair P, Mugno BL, Riemann BC. Co-morbid obsessive-compulsive disorder and depression: a Bayesian network approach. Psychol Med 2017;47(7):1204–14. https://doi.org/10.1017/S0033291716003287.
[140] Li P, et al. Large-scale exploration and analysis of drug combinations. Bioinformatics 2015;31(12):2007–16. https://doi.org/10.1093/bioinformatics/btv080.
[141] Velikova M, Lucas PJF, Samulski M, Karssemeijer N. On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks. Artif Intell Med 2013;57(1):73–86. https://doi.org/10.1016/j.artmed.2012.12.004.
[142] Yin Z, Zhao Y, Lu X, Duan H. A hybrid intelligent diagnosis approach for quick screening of Alzheimer's disease based on multiple neuropsychological rating scales. Comput Math Methods Med 2015;2015. https://doi.org/10.1155/2015/258761.
[143] Luo Y, et al. A multiobjective Bayesian networks approach for joint prediction of tumor local control and radiation pneumonitis in nonsmall-cell lung cancer (NSCLC) for response-adapted radiotherapy. Med Phys 2018;45(8):3980–95. https://doi.org/10.1002/mp.13029.
[144] Lappenschaar M, et al. Multilevel temporal Bayesian networks can model longitudinal change in multimorbidity. J Clin Epidemiol 2013;66(12):1405–16. https://doi.org/10.1016/j.jclinepi.2013.06.018.
[145] Armstrong S. The apps attempting to transfer NHS 111 online. BMJ 2018;360 (January 2018):1–5. https://doi.org/10.1136/bmj.k156.
[146] Ventures MMC. Beyond the hype: artificial intelligence podcast Interview - Babylon health. 2017 [Online]. Available: https://www.healthcare.digital/single-post/2018/03/29/Healing-Healthcare-with-Artificial-Intelligence-"Beyond-The-Hype"-with-Dr-Ali-Parsa-Founder-and-CEO-of-Babylon-Health#!.
[147] Ducher M, Mounier-Véhier C, Lantelme P, Vaisse B, Baguet JP, Fauvel JP. Reliability of a Bayesian network to predict an elevated aldosterone-to-renin ratio. Arch Cardiovasc Dis 2015;108(5):293–9. https://doi.org/10.1016/j.acvd.2014.09.011.
[148] Kaewprag P, Newton C, Vermillion B, Hyun S, Huang K, Machiraju R. Predictive models for pressure ulcers from intensive care unit electronic health records using Bayesian networks. BMC Med Inform Decis Mak 2017;17(Suppl 2). https://doi.org/10.1186/s12911-017-0471-z.
[149] How to elicit many probabilities. the Fifteenth Conference on Uncertainty in Artificial Intelligence 1999:647–54 [Online]. Available: http://dl.acm.org/citation.cfm?id=2073869.
[150] van der Gaag LC, Renooij S, Witteveen CLM, Aleman BMP, Taal BG. Probabilities for a probabilistic network: a case-study in oesophageal cancer. Artif Intell Med 2002;25(2):123–48.
[151] Lacave C, Díez FJ. A review of explanation methods for Bayesian networks. Knowl Eng Rev 2002;17(April 2):107–27.
[152] Kyrimi E, Mossadegh S, Tai N, Marsh W. An incremental explanation of inference in Bayesian networks for increasing model trustworthiness and supporting clinical decision making. Artif Intell Med 2020;103(August 2019):101812. https://doi.org/10.1016/j.artmed.2020.101812.