Bayesian Network Models of Causal Interventions in Healthcare Decision Making: Literature Review and Software Evaluation

Artem Velikzhanin, Benjie Wang and Marta Kwiatkowska

Technical Report

Department of Computer Science
University of Oxford

November 29, 2022
Abstract

This report\textsuperscript{1} summarises the outcomes of a systematic literature search to identify Bayesian network models used to support decision making in healthcare. After describing the search methodology, the selected research papers are briefly reviewed, with the view to identify publicly available models and datasets that are well suited to analysis using the causal interventional analysis software tool developed in [1]. Finally, an experimental evaluation of applying the software on a selection of models is carried out and preliminary results are reported.

Introduction

In the medical field, prediction of the risk of disease requires establishing a statistical model of risk factors and diseases, and prediction of the probability of disease according to levels of multiple risk factors. Bayesian network models are often used in healthcare settings to support decision making, such as outcome prediction and disease management. Since Bayesian network models can support causal inference and interventions, algorithms and software tools have been developed, for example [1], that can compute provable robustness guarantees and automate “what if” analysis of disease outcomes. However, in view of data often being proprietary and privacy concerns, there is a lack of publicly available models and datasets on which to evaluate the potential of these methods in healthcare settings.

This report summarises a systematic review of existing literature on Bayesian network models in medical settings undertaken as part of the FUN2MODEL project\textsuperscript{2}, with a specific focus on decision support that is based on causal interventions. A search methodology was developed to identify articles that meet a set of criteria, which were then reviewed to summarise the use cases, including the medical domain and effects of causal interventions, and insight obtained; those articles that provide publicly available models and datasets were highlighted. For a selection of models, causal Bayesian network representations were built and analysed using the IntRob software [1] \textsuperscript{3}, and preliminary results reported.

Literature Search Strategy and Analysis

This review is focused on identifying existing datasets and Bayesian network (BN) models that capture decision making in healthcare settings, and particularly those where the decision can be modelled as a causal intervention in the system. Thus, articles were selected based on the following criteria:

- availability of the (Bayesian network) model;
- availability of (categorical) data;
- appropriate targets for intervention (for example, administering a drug);
- clear intervention interpretation to determine the appropriate intervention bounds;
- an appropriate target probability (e.g. survival of patient).

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

\texttt{bayes AND (network OR model) AND (medical OR clinical) AND decision-making AND causal}

Terms such as ‘Bayesian networks’ and ‘decision making’ are included as they are widely used in the target literature. We also tried using ‘intervention’, but this proved to be less useful as intervention is also a medical term. In particular, in medicine, an intervention is a treatment, procedure, or other action taken to prevent or treat disease, or improve health in other ways [2]. As for interventions in BNs, in Woodward’s theory [3] one variable X is a direct cause of another variable Y if there exists an intervention on X such that, if all other variables are held fixed at some value, X and Y are associated. Such an account assumes a lot about the sort of intervention needed, however, and Woodward goes to great lengths to make the idea clear [4]. A preliminary search showed that there is a small amount of work in this specific area, so the term ‘causal’ was chosen as the most universal one, encompassing also the interventions that we are interested in. Due to the large number of articles, we limited the selection to the areas of Computer Science, Psychology, Medicine and Dentistry, and Neuroscience, using filters.

\textsuperscript{1}This report was funded by the ERC under the European Unions Horizon 2020 research and innovation programme (FUN2MODEL, grant agreement No.834115).

\textsuperscript{2}fun2model.org

\textsuperscript{3}https://github.com/wangben88/interventional-robustness
After extracting relevant articles based on the above criteria, we analyzed each article using natural language processing (NLP) tools to answer the following questions: in what medical areas are Bayesian network models used, and, in particular, what were the effects of causal interventions studied? For each article, we had access to the following attributes: title, abstract, authors, year, and DOI. The analysis was performed on the basis of available abstracts of articles. As of August 4, 2022, there were 1042 (983 in ScienceDirect; 24 in Scopus; 35 in PubMed) unique article abstracts available to download in ScienceDirect, Scopus and PubMed.

The entire procedure/algorithm for the analysis of the articles is shown in Appendix A in Figure A.1. In the first step, the PMIDs of the PubMed articles were converted to the BibTeX format. Then, DOI links from all three databases were converted into the same format and merged, allowing us to remove duplicate articles from the database. For the remaining articles, we extracted the article abstracts; 29 abstracts were not available.

Next, all raw abstracts were cleaned/preprocessed in order to facilitate further analysis. We used regular expressions (combinations of wildcard characters operators for searching and manipulating substrings in text) to remove all parentheses in abstracts, and tokenize the words. Tokenization allows us to decompose the text into a sequence of tokens, which can be words or sub-words.

We then removed all stop words. Stop words in natural language are the most common words which do not convey significant meaning; for example, in English these are words such as ‘the’, ‘for’, ‘at’, ‘to’, etc. English and French dictionaries were used to identify and remove the stop words. We also perform lemmatization of the tokens, which refers to converting a token to its ‘base’ or ‘dictionary’ form. This process includes, for instance, plural to singular conversion, as well as word tense normalization. Unlike stemming (another popular approach which simply removes affixes), lemmatization takes into account a broader context, including the surrounding and adjacent clauses. Finally, we employ N-grams, which are continuous sequences of tokens in a document. Bigrams (2) and trigrams (3) were used in our analysis.

For the final step, Latent Dirichlet Allocation (LDA) was used as a statistical model for the abstracts. LDA models every document as a mixture of topics, and every topic as a mixture of words. Crucially, it is possible to simultaneously estimate both parameters: finding a combination of words associated with each topic, as well as determining a combination of topics that describe each document (in this case abstracts). The LDA model was built with 30 different topics, where each topic is a combination of keywords and each keyword contributes a certain weight to the topic. After training, we obtained the following evaluation for the LDA model:

Perplexity: -15.231877825584922
Coherence Score: 0.4339549814094681

Perplexity measures how well a probability distribution predicts a sample; it thus indicates how well the topic model fits the data. Coherence, on the other hand, measures the degree of semantic similarity between words in a topic, in order to measure how ‘meaningful’ each topic is. The lower the perplexity, the better the model, and vice versa for coherence. While perplexity is an important measure of fit, we are more interested here in obtaining a conceptually coherent/consistent topic model. Therefore, in this analysis, we primarily consider the coherence score.

According to non-optimised LDA analysis, a vast proportion of the retrieved articles are in the field of research on cancer and mental disorders. This is indicated by word markers such as ‘death’, ‘cognitive’, ‘brain’, ‘emotion’ and ‘behaviour’ (see Appendix A A4-A8). On the charts, circles represent the topics. The larger the circle, the more common this theme. A good topic model should have large non-overlapping circles scattered throughout the diagram. In turn, a model with too many topics will often have many small circles that can overlap and be in the same area of the chart.

To optimize the model, hyperparameter tuning was performed. For this, the coherence score was calculated for 10 topics, for a range of values of alpha and beta hyperparameters. The maximal coherence score was obtained using 10 topics, alpha = 0.01, and beta=0.91. After the selection of hyperparameters, the accuracy of the model was:

Perplexity: -8.21063741472142
Coherence Score: 0.643855950055103

The model has become more coherent, but less sensitive. In Figure A7 in Appendix A, we see that the top-30 most salient items include tokens such as ‘brain’, ‘cognitive’, ‘death’, ‘neural’, ‘intervention’, ‘memory’ and ‘child’. According to the model, a possible interpretation is that the interventions are probabilistically relevant to 37.5% of the topics in which the study was conducted. In the distance map, the first bubble includes generic terms that can be found in almost all medical works. The behaviour token is relevant to three large clusters; it can be ignored. The second large cluster includes tokens such as ‘intervention’, ‘death’, ‘treatment’, and ‘child’. This may be due to the fact that research on interventions is mainly conducted on fatal diseases that children suffer from. The third large cluster of tokens relates to diseases associated with a decline of brain function. This can be observed if the relevance metric (λ) is 0.6. The fourth and fifth clusters are language
clusters, such as Spanish and French. Other clusters represent a small percentage of the total and are mostly devoid of particular semantics.

In Appendix A, Figure A2 shows the search results for all three databases. From ScienceDirect, in the Computer Science field 257 papers were downloaded (out of 258), in the Medicine and Dentistry fields 609 papers were downloaded (out of 632), in the Neuroscience field 342 papers were downloaded (out of 382), and in the Psychology field 271 papers were downloaded (out of 305). In total 1479 papers was downloaded and after duplicate removal 1093 papers were retained. Our overall finding is that the bulk of the analyzed work is in the areas of cancer, respiratory diseases and brain research. Interventions are likely to be found in studies that investigate disease deaths (e.g., cancer and COVID-19). A search was carried out for relevant data in the articles. Due to missing data, only 34 articles were included in this report.

Review of Relevant Articles

Due to variability in presentation of diseases, it can be difficult for doctors to prescribe optimal treatment plans. More accurate predictive models can be beneficial to provide personalised treatment and determine effective interventions. At the same time, it is important that any interventions we apply are safe; a medical error in any specific intervention could be fatal. This may be due to the individual reactions of the patient’s body to drugs or on the part of the doctor, for example, loss of concentration. From a practical perspective, rigorously testing all potential interventions/treatments for all types of patients in long clinical trials is not always feasible as valuable time may be lost, and the result may not be positive. As a result, doctors often have to rely only on their experience, and the wrong choice of intervention may aggravate the course of the disease. In this review, we are interested in applications of causal inference to help solve the ‘treatment effect’ problem. In this context, one aims to choose an intervention which has a positive causal effect on the outcome variable, which will help to determine the most effective treatment plan. The effect of an intervention can be formalized as the difference between two potential outcomes for an individual: one where they receive the intervention, and one where they do not. As a result, counterfactuals are the basis for causal conclusions, giving answers to a number of questions [5]. For example:

1. will the intervention work?
2. why did it work?
3. what combination of interventions are safe/can reduce the risks?
4. how effective is the intervention?

Unfortunately, in many articles, the concept of intervention is extremely vague. In particular, as mentioned previously, in medical articles interventions are usually understood as operational actions rather than part of a model. Also, in more than 90% of cases, data or models are not available. Thus, we selected all papers with medical data available for review. A brief summary of the relevant articles now follows.

Several papers demonstrate applications of Bayesian network analysis in cancer research. In [6], the authors build a Bayesian network model using the English Lung Cancer Database (LUCADA) and apply causal interventions. The results show that the survival-maximizing model only works correctly in 29% of cases. This is due to the lack of some of the information necessary for effective interventions; in the database, the constructed model is not able to distinguish between patients for whom surgery is indicated and those for whom it is not suitable. The model and data are not available. The authors continue their work in [7], where a clinical decision support application for lung cancer care (LCA) was developed. The goal was to help lung cancer experts make decisions about treatment choices in MDT (Multidisciplinary Team) meetings. The system is not available via the link in the article (the login page is stored in the webarchive).

The impact of obesity in early-stage breast cancer survivors on health behaviors was studied in [8]. Clinical data from 333 overweight or obese postmenopausal women who survived breast cancer were used to build a Bayesian network. The authors combine factors to determine possible interventions. For example, based on the network they developed, they conclude that older age, more sleep disturbances, and higher BMI are associated with lower physical activity. Data is available in the article, but presumably, after the anonymization of the personal data, some parameters were removed. Because of this, the dataset is incomplete. It may be possible to synthesize the missing data based on what is already known.

Factors and causal relationships influencing the choice of place of death in Swiss cancer patients were studied in [9]. One of the goals of the project was to explore the possibility of interventions provided by healthcare professionals to facilitate EOL (end of life) at home. The model was built on 116 adult patients who died from cancer between 2015 and 2016 in southern Switzerland. The model can be downloaded from the supplementary materials. Data is available on request.
Using data from the National Lung Screening Trial (NLST), a model is built in [10] to study the impact of interventions on the survival of patients with lung cancer. By interventions, the authors mean low-dose computed tomography (LDCT) irradiation to reduce patient mortality. The authors constructed their dynamic Bayesian network (DBN) using 10-fold cross-validation. The models were evaluated based on the probabilistic variable of the biopsy, taking into account all previous and current data for each of the three intervention points in the NLST study. Data is available upon request.

For the diagnosis of breast cancer, researchers used a Bayesian network [11] built using the K2 algorithm. The data is collected from The First Affiliated Hospital of Fujian Medical University, China and the Breast Cancer Wisconsin Dataset (BCWD) of the UCI machine learning repository. Naive Bayes (NB), Bayesian network (BN), ID3, J48 and NBTree classifiers are used for performance analysis. The UCI dataset is available for download.

In [12], the authors (using transcriptomic datasets) build a DBN model used for cancer classification. They used three datasets for a DBN-based approach, which is able to model time series gene expression data for classification purposes. The datasets are the pancreatic cancer dataset (GSE14426), colon cancer dataset (GSE37182), and breast cancer dataset (GSE5462). Data is available through GEOquery package.

Decisions in transplant care often involve trade-offs between potential benefit and potential harm of treatment. In [22], the authors build a model for making kidney transplant decisions. A decision-making
method based on Kidney Donor Risk Index (KDRI) was developed. The model predicts the probability of death over the next 3 years. Data is not available.

Assessing and managing the risk of violence is considered a critical component for decision making in medium secure services [23]. The researchers build a model in forensic psychiatry using a dataset referred to VoRAMSS. In order to find out the impact on the output node, sensitivity analysis is performed. The model is compared with other regression models trained on the same dataset. Data not available.

To assess and manage the risk of future violent re-offending for released prisoners with mental illness, a DSVM-P model was developed in [24]. The authors explore the issue of reducing the risk of future re-offending to an acceptable level by introducing interventions. Using sensitivity analysis applied to the interventions, the authors observe the most active symptoms. Data is not available.

A causally-based decision support tool for the management of violence risk among released prisoners using Bayesian networks was created in [25]. The risk analysis is managed and analysed through interventions on variables such as alcohol, drugs and psychiatric interventions. The model was compared with the results from HRC-20, VRAG and PCL-R risk assessment instruments. Data is not available. The process was replicated in [26]. The authors develop a risk management tool for discharged forensic prisoners using the same methodology. The model was compared with the results from HRC-20, SAPROF and PANSS risk assessment instruments. The data is not available.

Autism, or autism spectrum disorder (ASD), occurs due to developmental disorders caused by differences in the brain. For predicting autistic spectrum disorder, it is possible to use fuzzy cognitive maps (FCM) [27]. The authors used a non-linear hebbian learning for training model. In the study, experts, based on the constructed model, determine the boundaries for the acceptable probabilities of having autism. They define: definite autism, probable autism and no autism. The data is available.

A Bayesian network decision model was proposed for supporting the diagnosis of dementia, Alzheimers disease (AD), and mild cognitive impairment (MCI) [28]. In the paper the list of CDSS (clinical decision support system) was provided. The authors perform a sensitivity analysis for parameters like dementia, AD and mild cognitive impairment. The performance was tested by AUC, F1, MSE, MXE scores. The dataset based on data from the Duke University Medical Center and the Center for Alzheimer’s Disease and Related Disorders can be purchased.

Chronic obstructive pulmonary disease (COPD) is a common, preventable and treatable disease, with a worldwide prevalence of 10.1% in people aged 40 years or older [29]. Using a Bayesian model, the authors analyse the impact of critical variables on the risk of rehospitalization in patients with COPD [30]. The authors argue that it is possible to reduce these risks with the help of personalised interventions. The accuracy of the model is determined in comparison with other machine learning methods. The dataset was constructed based on St. Mary’s Hospital data.

In [31] and [32], the authors used machine learning (ML) techniques and built transplant survival models. For predicting the outcome of a heart transplant, the authors built a framework based on a BN model that can identify patient-specific survival risk scores and the interactions between the explanatory variables [31]. For testing performance of the model six evaluation metrics were used. The dataset was provided by United Network for Organ Sharing (UNOS). The dataset and software are available for download. Of the 119,873 organ transplants worldwide, 67% are kidney transplants, and 59% of those are from deceased donors, according to the World Health Organization [33]. A three-step Bayesian risk model for predicting graft survival in kidney transplantation was proposed in [32]. The UNOS data is used for building the model. The authors used several techniques, such as ANN (artificial neural network) and SVM (support vector machine), and sensitivity analysis for selecting features for the final dataset. For performance measure of the accuracy, sensitivity, specificity, f-measure, and g-mean metrics were used. Data is not available.

Approximately 537 million adults (20-79 years) are living with diabetes. The total number of people living with diabetes is projected to rise to 643 million by 2030 and 783 million by 2045 [34]. For the prevention of complications of type 2 diabetes mellitus (T2DM), the authors propose a model that predicts six complications of T2DM [35]. The model was built using the Tabu-search and Bootstrap algorithms. Data was taken from the National Health Clinical Center for the period 1st January 2009 to 31st December 2009. The authors compare the performance between BN, BN-wopi, NB, RF and C5.0 models using AUC, 95%CI, sensitivity and specificity metrics. The article also provides a comparative performance table of models from related studies by AUC score. Dataset is available on request.

In contrast, [36] proposes to predict the complications associated with type 1 diabetes. The paper presents two types of models using two different approaches (DDO-DBN, EI-DBN) in the design of the structure of the Dynamic Bayesian Network. The difference is that the second network is built on the basis of knowledge from the medical literature. As a result of the analysis, the authors determine threshold values for continuous variables. The authors used DCCT and EDIC data sets for building the model.

In [37], the possibility of automatic construction of the Bayesian network topology and ontology from electronic medical records using the K2 algorithm was studied. The dataset was built on the basis of 10,000 anonymized patient records. The classification performance of the constructed CBN is compared with other
inference models such as Naive Bayes, Basic, Random-node-input, and Frequency-based Bayesian networks. The data is available.

Infectious diseases are the world's greatest killers, accounting for more than 13 million deaths annually among children and young adults alone [38]. Patients often self-medicate to treat infectious diseases. In such cases, a system is required that can help in the choice of antibiotics. In [39], the authors propose a system called IDDAP based on ontologies for infectious disease diagnosis and antibiotic therapy. The ontology system was built by combining existing ontologies. Performance analysis was carried out using ROC, and the system was tested on different ontologies. The ontology is available; the software has been removed from Github.

Non-communicable diseases (NCDs) kill 41 million people each year, equivalent to 71% of all deaths globally [40]. Prevention of these diseases can help in their development. In [41], the authors developed a system based on Bayesian networks. The system determines the impact of user interventions on the main risk factors. The task was to predict coronary heart disease. The desired health indications are also indicated, i.e. upper and lower limits of the normal state without complications. A number of datasets were used in this work; Pima Indians Diabetes Data Set is available on Kaggle.

While much attention has been paid to the problem of evaluating the effectiveness of therapy using observational data, little work has been done on evaluating the treatment effect of interventions [42]. In the paper, the authors propose a model that studies reducing the number of falls in elderly patients through interventions in the model. Data from 1810 patients at Lille University Hospital (France) are used to build models. Experiments were carried out with three prediction models (BN, SVM, and DT); BN structure and modalities are available.

There are a number of tasks in which it is required to make timely diagnostic and managerial decisions. One such challenge is diagnosing infections in children in the emergency department. Based on expert opinion and knowledge in the area of expertise, [43] developed a Bayesian model to predict the results of causative pathogens. The paper provides three clinical scenarios to support decision making by physicians. The model is available.

In [44], the authors identify factors associated with hypertension. To achieve this, they build a Bayesian network based on Tabu search. The study participants were selected only from cities in Shanxi province, which does not allow generalisation of the results to a wider population. For evaluation, the authors used accuracy, TPR, FPR, precision, recall and F-measure metrics using the Weka software. The model has parameters such as smoking or drinking that can be used as interventions. Dataset and model are available.

Finally, we remark that in Appendix B (see Tables 1 and 2) we summarise information on the availability of medical data. Data was collected through a systematic review of publications. In most papers, data from the UCI and Kaggle repositories are used. According to the analysis of papers, several articles were identified that met the criteria stated above. The case of interventions for fall prevention in elderly patients [42] was chosen as the best suited to further analysis.

Interventions and Disease Management: Risk Boundary Selection

In the medical field, prediction of the risk of disease requires establishing a statistical model of risk factors and diseases, and prediction of the probability of disease according to levels of multiple risk factors [45].

Usually, the reasoning about the risk boundaries is based on deviations from the normal state, as for example in [6], [19], [27], [46]. In the medical literature, there are also works which establish some limits/bounds on the observed variables that indicate the normal state of a patient, with values outside the limits indicating the presence of the disease. This section provides several examples of such works.

Some articles have outlined specific risk boundaries [14], [15],[31], [32], [41], which can be used to evaluate interventions and, therefore, to recommend more effective personal therapy.

In [32], the authors set boundaries for the response to transplantation. There are three risk groups defined: low, medium and high risk. Depending on which risk group a patient falls in, it is possible to choose personalized treatment more accurately, and, more importantly, to prevent a patient from falling into a high-risk category in advance, by determining the influence of certain factors on risk.

In [41] interval boundaries are indicated for CAD risk factors. Also in this work, the authors provide a comparative analysis of similar approaches, comparing them in terms of parameters of risk factors analysis and context awareness. Based on the model, the authors developed an application that provides recommendations on factors that can exacerbate or alleviate the course of the disease. To do this, they created several scenarios and tested the system on two datasets.

An alternative to using boundaries is finding the most optimal solution using influence diagrams, i.e., a Bayesian network augmented by utility nodes (UN).

The authors of [22] suggest using the Kidney Donor Risk Index (KDRI) to evaluate complex transplant decisions. The metric assesses the risk of transplant rejection compared to a 40-year-old healthy donor. The model was built in Netica using UN. The purpose of using utility nodes is to maximize the expected value while looking for the most optimal solution.
Regarding acceptable levels of numerical variables, [46] presents an algorithm for the treatment of diseases associated with psoriasis. For example, for the treatment of metabolic syndrome, levels are set for HDL, AHT, BP, and TG. Using the reasoning model of this work as an example, experts can build BN models, determine acceptable levels of variables, and design effective interventions.

In [47], algorithms for the diagnosis of testosterone deficiency and follow up of testosterone therapy are developed. From analysis of the literature, the authors identify the main symptoms, signs, and conditions indicative of testosterone deficiency. The levels of physical indicators are also given. The information provided in the article can be used to design intervention-based CDSS.

Interventions can also be derived from literature analysis, as was done in [49]. As a result of the analysis, the authors proposed pharmacological and non-pharmacological interventions for BPSD (dementia symptoms).

Based on the works presented in this section (and similar other works), it is possible to develop more effective systems for controlling risk levels based on interventions, which can be the next step in building effective CDSS systems.

Review Summary and Model Selection

In most of the analyzed papers, machine learning methods like support vector machines (SVM), Bootstrap Forest (BF), Bayesian belief network (BBN), and artificial neural networks (ANN) were used. When BN models are used, they may be part of a larger pipeline [31] if the data was not initially prepared/in the right format to work with BNs. AUC is most commonly used for accuracy evaluation. Bayesian network models are mostly built on the basis of expert opinions [43], but sometimes they are also built using the K2 [37] or TabuSearch [18] algorithms, then checked or supplemented by experts [42]. These authors use their own data derived from close cooperation with medical institutions (Table B1 of Appendix B) or data from repositories (Table B2) or medical databases (Table B3). Also, in some works, data from multiple sources can be combined if incomplete [41]. By interventions, the authors mostly refer to medical interventions that are already accounted for in the Bayesian network [39]. In some works, sensitivity analysis is carried out to determine the impact of various factors on the target node [16], [31] according to some machine learning classifier. Similarly, it is possible to use ‘what-if analysis’ in Bayesian networks to discern the effect of a given variable on the target by conditioning on the value of that variable [32]. This could be used as a heuristic for selecting promising interventions to study.

The main evaluation indices of a BN model are true positive rate (TPR), true negative rate (TNR), recall, and precision. Sensitivity (TPR) indicates the proportion of positive classes correctly predicted and the ability of the BN to recognize positive classes. Specificity (TNR) represents the proportion of correctly predicted negative classes and measures the ability of the BN to recognize negative classes [44].

In Table B1 (Appendix B), the found models are highlighted in green, and the proposed model for our experimental study of interventions is highlighted in blue (a shorter dataset was attached to the work). Appendix C contains the constructed models based on the available data in the papers.

Experimental Evaluation

This section of the report summarises our experiments on selected models using the software tool developed in [1]. Our workflow was primarily based upon the framework/pipeline illustrated in Appendix D for Bayesian network modelling, based on [50]. In the following experiments, since we already have access to the Bayesian network from the literature, we skipped the first few steps. For interventional modelling, we make use of the IntRob (interventional robustness) software [1], based on a recently proposed approach to providing guarantees on the effects of combinations of interventions. As we will see, this is important for providing safety guarantees with respect to different treatment (intervention) policies.

We identified the Bayesian network models from [14, 15, 42] as particularly suitable for interventional analysis; in particular, we conduct experiments on the basis of the model from paper [14]. In this paper, the authors attempt to identify patients at risk for lymph node metastasis (LNM) in endometrial cancer. The Bayesian network model can be used to predict lymph node metastasis, as well as outcomes such as patient survival up to 5 years and recurrence of disease, given some (incomplete) evidence/information about the patient.

An oncologist treating a cancer patient employs the following procedure. In the first stage, a diagnosis is made and it is determined how much the tumor has spread and whether the patient is operable. In the second
stage, if the patient is operable, then they apply radiation therapy targeting the tumor and nearby tissues (this depends on many factors). After irradiation, the patient is surgically operated on, with irradiated and nearby tissue removed. Finally, chemotherapy is prescribed to prevent spread of tumour cells into lymph nodes or blood.

In their work, the authors focus on predicting lymph node metastasis, which is one of the most important prognostic factors for patient outcomes, and which can be significantly influenced by adjuvant therapy: it is thus of significant practical interest to be able to predict LNM in order to inform therapy decisions. We begin by performing feature selection for the classifier based on their impact on the LNM probability in the Bayesian network. In Table 1, we show the probability of LNM and probability of LVSI (lymph-vascular space invasion), for a range of different evidence values. The green cells are updated values and the blue one was chosen for classification.

Table 1. Probability estimates for lymph node metastasis and LVSI given different evidence [14]

| Evidence provided to the Bayesian network | Modalities | Lymph node metastasis | LVSI |
|------------------------------------------|------------|-----------------------|------|
| No evidence                              |            | 8.6                   | 16   |
| Preoperative grade                       | 1          | 4.7                   | 10.6 |
|                                          | 2          | 8.2                   | 14.6 |
|                                          | 3          | 20.7                  | 34.4 |
| Preoperative grade, L1CAM                | 1,negative | 3.9                   | 9.9  |
|                                          | 1,positive | 17.7                  | 22   |
|                                          | 2,negative | 6.6                   | 13.1 |
|                                          | 2,positive | 28.2                  | 28.1 |
|                                          | 3,negative | 17.4                  | 31.8 |
|                                          | 3,positive | 30.4                  | 41.8 |
| Preoperative grade, ***PM                | 1,favorable*| 2.8                   | 8.9  |
|                                          | 1,unfavorable**| 35.8              | 45   |
|                                          | 2,unfavorable | 5.1                  | 11.6 |
|                                          | 3,favorable | 16.9                  | 29.3 |
|                                          | 3,unfavorable| 37.1                 | 45.4 |
| Preoperative grade, Ca-125              | 1,normal   | 1.5                   | 8.9  |
|                                          | 1,elevated | 22.8                  | 20.3 |
|                                          | 2,normal   | 2.8                   | 11.8 |
|                                          | 2,elevated | 24.7                  | 28.2 |
|                                          | 3,normal   | 7.7                   | 28.5 |
|                                          | 3,elevated | 60.9                  | 52.2 |

*Favorable: all IHC stainings were normal (ER, PR positive, L1CAM negative, p53 wildtype).  
*Unfavorable (ER, PR negative, L1CAM positive, p53 mutant).  
Ca-125, cancer antigen 125; ER, estrogen receptor; L1CAM, L1 cell adhesion molecule; PR, progesterone receptor.  
**PM - molecular profile.

In Table 2, we show a classification of risk groups for LNM.

Table 2. LNM risk table [14]

| Predicted probability | Risk group   |
|-----------------------|--------------|
| < 1%                  | Very low     |
| 1%-5%                 | Low          |
| 6%-15%                | Intermediate |
| 16%-25%               | High-intermediate |
| > 25%                 | High         |

According to Table 1, the following variables were selected as features for the classifier: ER, PR, L1CAM, p53, preoperative grade, CA125. We used a threshold of 0.05 for the prediction of LNM, in accordance with
the risk profiles in Table 2. We use the BNC_SDD software to obtain a logical representation of the resulting Bayesian network classifier (BNC), which we use for further analysis in the IntRob software.

![Bayesian network diagram](image)

Fig. 1 The chosen Bayesian network. The target node (LNM) is shown in pink. Yellow nodes were chosen as features for classification, while green nodes represent intervention targets.

```
{
"id": "1",
"name": "Cd",
"fleltype": "net",
"vars": 6,
"root": "LNM",
"leaves": ["5R","PrimaryTumor","PR","L1CAM","p53","CA125"],
"threshold": 0.65,
"input_filepath": "networks/",
"output_filepath": "output/"
}
```

Fig. 2 JSON file for LNM classifier

For our first experiment, we considered the probability of LNM under interventions on adjuvant therapy. Such an intervention represents a change in treatment policy; it can be seen, in particular, that adjuvant therapy has a causal influence on LNM from the Bayesian network graph. In particular, we asked the following question: what are the potential consequences of instituting a bad treatment policy? To analyze this, we used the IntRob software to compute an upper bound on the probability of lymph node metastasis, under all possible interventions (treatment policies). In Figure 5, we see that the worst-case policy may result in 60% of patients suffering metastasis, up from 9% under the current policy. This shows that it is vital to choose the right adjuvant therapy for each patient depending on their characteristics, and failing to do so can be unsafe. In Figure 6, we conducted another test to see how different interventions on adjuvant therapy affect the false negative rate of the trained classifier; that is, whether this might result in the classifier mispredicting patients as not at risk of LNM when they actually are. Interestingly, the classifier is surprisingly robust, with a false negative probability of only 0.001. This can possibly be attributed to the fact that the classifier uses indicators causally downstream from LNM (p53, L1CAM, etc.), which provide sufficient information to correctly predict LNM, even with the shift observed above in the LNM distribution.

The results of the working of IntRob software can be seen in Figures 3-5.
For the second experiment, we study the probability of recurrence, again under interventions to the adjuvant therapy. Applying the IntRob software to derive the worst-case probability of recurrence over all possible interventions, we find in Figure 8 that the probability of (regional-distant) recurrence can reach as high as 38%, once again demonstrating the risks of inappropriate treatment.
Finally, for the last experiment, we considered 5-year survival (‘Survival 5yr’). We performed a sensitivity analysis in Figure 9 according the framework in Appendix D, to find relevant features to use for the classifier. Based on this, we trained a classifier using the following nodes/features: PrimaryTumor, CA125, Therapy, Recurrence. The input file for the ODD compiler is shown on Figure 10.

We now consider how 5-year survival can be improved using interventions. In Figure 13, we use the IntRob software to assess the maximal probability of 5-year survival, under all possible interventions to recurrence and adjuvant therapy. We find in Figure 13 that the upper bound on this probability is 1 (up to floating point error), indicating that there may be significant room for improvement of patient outcomes by improving treatment policies and reducing recurrence.
Fig. 9 Sensitivity analysis

```
{
   "id": "1",
   "name": "C4",
   "filetype": "net",
   "vars": 4,
   "root": "Survival5yr",
   "leaves": ["PrimaryTumor", "CA125", "Therapy", "Recurrence"],
   "threshold": 0.5,
   "input_filepath": "networks/",
   "output_filepath": "output/"
}
```

Fig. 10 JSON config file for 5-year survival classifier

The output file after compiling in BNC_SDD software is shown in Figure 11.

```
[PrimaryTumor, Therapy, CA125, Recurrence]
3 3 S1 S1 S1
2 2 3 3
1 1 2 2 2 2
0 0 1 1 1
```

Fig. 11 ODD file for 5-year survival classifier
Fig. 12 IntRob software for the final experiment

Fig. 13 Upper bound on 5-year survival probability under best-case intervention
In conclusion, our analysis has revealed a number of insights into the Bayesian network model. The obtained BN bounds can be utilised to develop subsequent treatments strategies, as well as used as input into models that improve QoL or QoL [8] or facilitate EOL at home [9]. It is also possible to use medical protocols that reduce the likelihood of falling into a certain group.

Acknowledgements
This project was funded by the ERC under the European Unions Horizon 2020 research and innovation programme (FUN2MODEL, grant agreement No.834115).
References

[1] B. Wang, C. Lyle, and M. Kwiatkowska, “Provable guarantees on the robustness of decision rules to causal interventions,” 2021.

[2] “Nci dictionary of cancer terms.” https://www.cancer.gov/publications/dictionaries/cancer-terms/def/intervention. Accessed: 2022-10-11.

[3] J. Woodward, Making Things Happen: A Theory of Causal Explanation. Oxford University Press, 2003.

[4] F. Eberhardt and R. Scheines, “Interventions and causal inference,” Philosophy of Science, vol. 74, pp. 981–995, Dec. 2007.

[5] S. Chaieb, A. B. Mrad, and B. Hnich, “Interventions recommendation system for preventing future falls in older adults,” Procedia Computer Science, vol. 192, pp. 192–201, 2021.

[6] M. B. Sesen, A. E. Nicholson, R. Banares-Alcantara, T. Kadir, and M. Brady, “Bayesian networks for clinical decision support in lung cancer care,” PloS One, vol. 8, no. 12, p. e82349, 2013.

[7] M. B. Sesen, M. D. Peake, R. Banares-Alcantara, D. Tse, T. Kadir, R. Stanley, F. Gleeson, and M. Brady, “Lung Cancer Assistant: a hybrid clinical decision support application for lung cancer care,” Journal of the Royal Society, Interface, vol. 11, p. 20140534, Sept. 2014.

[8] S. Xu, W. Thompson, J. Kerr, S. Godbole, D. D. Sears, R. Patterson, and L. Natarajan, “Modeling interrelationships between health behaviors in overweight breast cancer survivors: Applying Bayesian networks,” PloS One, vol. 13, no. 9, p. e0202923, 2018.

[9] H. Kern, G. Corani, and D. Huber, “Impact on place of death in cancer patients: a causal exploration in southern Switzerland,” BMC Palliative Care | Full Text.

[10] P. Petousis, S. X. Han, D. Aberle, and A. A. T. Bui, “Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: A dynamic Bayesian network,” Artificial Intelligence in Medicine, vol. 72, pp. 42–55, Sept. 2016.

[11] S. Liu, J. Zeng, H. Gong, H. Yang, J. Zhai, Y. Cao, J. Liu, Y. Luo, Y. Li, L. Maguire, and X. Ding, “Quantitative analysis of breast cancer diagnosis using a probabilistic modelling approach,” Computers in Biology and Medicine, vol. 92, pp. 168–175, Jan. 2018.

[12] K. Kourou, G. Rigas, C. Papaloukas, M. Mitsis, and D. I. Fotiadis, “Cancer classification from time series microarray data through regulatory Dynamic Bayesian Networks,” Computers in Biology and Medicine, vol. 116, p. 103577, Jan. 2020.

[13] K.-M. Wang, K.-J. Wang, and B. Makond, “Survivability modelling using Bayesian network for patients with first and secondary primary cancers,” Computer Methods and Programs in Biomedicine, vol. 196, p. 105686, Nov. 2020.

[14] C. Reijnen, E. Gogou, N. C. M. Visser, H. Engerud, J. Ramjith, L. J. M. v. d. Putten, K. v. d. Vijver, M. Santacana, P. Bronsert, J. Bulten, M. Hirschfeld, E. Colas, A. Gil-Moreno, A. Reques, G. Mancebo, C. Krakstad, J. Trovik, I. S. Haldorsen, J. Huvila, M. Koskas, V. Weinberger, M. Bednarikova, J. Hausnerova, A. A. M. v. d. Wurff, X. Matias-Guiu, F. Amant, E. Consortium, L. F. A. G. Massuger, M. P. L. M. Snijders, H. V. N. Ksters-Vandevelde, P. J. F. Lucas, and J. M. A. Pijnengbog, “Preoperative risk stratification in endometrial cancer (ENDORISK) by a Bayesian network model: A development and validation study,” PLOS Medicine, vol. 17, no. 5, p. e1003111, 2020. Publisher: Public Library of Science.

[15] X. Jiang, A. Wells, A. Brufsky, and R. Neapolitan, “A clinical decision support system learned from data to personalize treatment recommendations towards preventing breast cancer metastasis,” PloS One, vol. 14, no. 3, p. e0213292, 2019.

[16] N. E. Fenton, S. McLachlan, P. Lucas, K. Dube, G. A. Hitman, M. Osman, E. Kyrimi, and M. Neil, “A privacy-preserving Bayesian network model for personalised COVID19 risk assessment and contact tracing,” July 2020. Pages: 2020.07.15.20154286.

[17] B. A. Akinnuwe, S. G. Fashoto, E. Mbunge, A. Odumabo, A. S. Metfula, P. Mashwama, F.-M. Uzola, O. Owolabi, M. Okpeku, and O. O. Amusa, “Application of intelligence-based computational techniques for classification and early differential diagnosis of COVID-19 disease,” Data Science and Management, vol. 4, pp. 10–18, Dec. 2021.
[18] X. Wang, J. Pan, Z. Ren, M. Zhai, Z. Zhang, H. Ren, W. Song, Y. He, C. Li, X. Yang, M. Li, D. Quan, L. Chen, and L. Qiu, “Application of a novel hybrid algorithm of Bayesian network in the study of hyperlipidemia related factors: a cross-sectional study,” *BMC Public Health*, vol. 21, p. 1375, July 2021.

[19] Z. Zhang, J. Zhang, Z. Wei, H. Ren, W. Song, J. Pan, J. Liu, Y. Zhang, and L. Qiu, “Application of tabu search-based Bayesian networks in exploring related factors of liver cirrhosis complicated with hepatic encephalopathy and disease identification,” *Scientific Reports*, vol. 9, p. 6251, Apr. 2019. Number: 1 Publisher: Nature Publishing Group.

[20] N. A. Loghmanpour, M. K. Kanwar, M. J. Druzdzel, R. L. Benza, S. Murali, and J. F. Antaki, “A new Bayesian network-based risk stratification model for prediction of short-term and long-term LVAD mortality,” *ASAIO journal (American Society for Artificial Internal Organs: 1992)*, vol. 61, pp. 313–323, June 2015.

[21] Y. Wu, C. McLeod, C. Blyth, A. Bowen, A. Martin, A. Nicholson, S. Mascaro, and T. Snelling, “Predicting the causative pathogen among children with osteomyelitis using Bayesian networks improving antibiotic selection in clinical practice,” *Artificial Intelligence in Medicine*, vol. 107, p. 101895, July 2020.

[22] R. Neapolitan, X. Jiang, D. P. Ladner, and B. Kaplan, “A Primer on Bayesian Decision Analysis With an Application to a Kidney Transplant Decision,” *Transplantation*, vol. 100, pp. 489–496, Mar. 2016.

[23] A. C. Constantinou, M. Freestone, W. Marsh, and J. Coid, “Causal inference for violence risk management and decision support in forensic psychiatry,” *Decision Support Systems*, vol. 80, pp. 42–55, Dec. 2015.

[24] A. C. Constantinou, M. Freestone, W. Marsh, N. Fenton, and J. Coid, “Risk assessment and risk management of violent reoffending among prisoners,” *Expert Systems with Applications*, vol. 42, pp. 7511–7529, Nov. 2015.

[25] J. W. Coid, S. Ullrich, C. Kallis, M. Freestone, R. Gonzalez, L. Bui, A. Igoumenou, A. Constantinou, N. Fenton, W. Marsh, M. Yang, B. DeStavola, J. Hu, J. Shaw, M. Doyle, L. Archer-Power, M. Davoren, B. Osumili, P. McCrone, K. Barrett, D. Hindle, and P. Bebbington, *Development of a Bayesian network for the risk management of violent prisoners*. NIHR Journals Library, Nov. 2016.

[26] J. W. Coid, S. Ullrich, C. Kallis, M. Freestone, R. Gonzalez, L. Bui, A. Igoumenou, A. Constantinou, N. Fenton, W. Marsh, M. Yang, B. DeStavola, J. Hu, J. Shaw, M. Doyle, L. Archer-Power, M. Davoren, B. Osumili, P. McCrone, K. Barrett, D. Hindle, and P. Bebbington, *Development of a Bayesian network for risk management of patients discharged from forensic mental health services*. NIHR Journals Library, Nov. 2016.

[27] A. Kannappan, A. Tamilarasi, and E. I. Papageorgiou, “Analyzing the performance of fuzzy cognitive maps with non-linear hebbian learning algorithm in predicting autistic disorder,” *Expert Systems with Applications*, vol. 38, pp. 1282–1292, Mar. 2011.

[28] F. L. Seixas, B. Zadrozny, J. Laks, A. Conci, and D. C. Muchahut Saade, “A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer’s disease and mild cognitive impairment,” *Computers in Biology and Medicine*, vol. 51, pp. 140–158, Aug. 2014.

[29] P. Jarhyan, A. Hutchinson, D. Khaw, D. Prabhakaran, and S. Mohan, “Prevalence of chronic obstructive pulmonary disease and chronic bronchitis in eight countries: a systematic review and meta-analysis,” *Bulletin of the World Health Organization*, vol. 100, pp. 216–230, Mar. 2022.

[30] S. Lee, S. Wang, P. A. Bain, C. Baker, T. Kundinger, C. Sommers, and J. Li, “Reducing COPD Readmissions: A Causal Bayesian Network Model,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4046–4053, 2018. Conference Name: IEEE Robotics and Automation Letters.

[31] A. Dag, K. Topuz, A. Oztekin, S. Bulur, and F. M. Megahed, “A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival,” *Decision Support Systems*, vol. 86, pp. 1–12, June 2016.

[32] K. Topuz, F. D. Zengul, A. Dag, A. Almehni, and M. B. Yildirim, “Predicting graft survival among kidney transplant recipients: A Bayesian decision support model,” *Decision Support Systems*, vol. 106, pp. 97–109, Feb. 2018.

[33] “Global Observatory on Donation and Transplantation.” [http://www.transplant-observatory.org/](http://www.transplant-observatory.org/). Accessed: 2022-10-11.

[34] “IDF diabetes atlas 2021.” [https://diabetesatlas.org/atlas/tenth-edition/](https://diabetesatlas.org/atlas/tenth-edition/). Accessed: 2022-10-11.
[35] S. Liu, R. Zhang, X. Shang, and W. Li, “Analysis for warning factors of type 2 diabetes mellitus complications with Markov blanket based on a Bayesian network model,” *Computer Methods and Programs in Biomedicine*, vol. 188, p. 105302, May 2020.

[36] S. Marini, E. Trifoglio, N. Barbarini, F. Sambo, B. Di Camillo, A. Malovini, M. Manfrini, C. Cobelli, and R. Bellazzi, “A Dynamic Bayesian Network model for long-term simulation of clinical complications in type 1 diabetes,” *Journal of Biomedical Informatics*, vol. 57, pp. 369–376, Oct. 2015.

[37] Y. Shen, L. Zhang, J. Zhang, M. Yang, B. Tang, Y. Li, and K. Lei, “CBN: Constructing a clinical Bayesian network based on data from the electronic medical record,” *Journal of Biomedical Informatics*, vol. 88, pp. 1–10, Dec. 2018.

[38] I. of Medicine, *Informing the Future: Critical Issues in Health*. National Academies Press, 2003.

[39] Y. Shen, K. Yuan, D. Chen, J. Colloc, M. Yang, Y. Li, and K. Lei, “An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription,” *Artificial Intelligence in Medicine*, vol. 86, pp. 20–32, Mar. 2018.

[40] “The Global Alliance for Chronic Diseases (GACD).” [https://www.gacd.org](https://www.gacd.org). Accessed: 2022-10-11.

[41] F. Pittoli, H. D. Vianna, J. L. Victria Barbosa, E. Butzen, M. Gaedke, J. S. Dias da Costa, and R. B. Scherer dos Santos, “An intelligent system for prognosis of noncommunicable diseases risk factors,” *Telematics and Informatics*, vol. 35, pp. 1222–1236, Aug. 2018.

[42] S. Chaieb, A. Ben Mrad, and B. Hnich, “From Personal Observations to Recommendation of Tailored Interventions based on Causal Reasoning: a case study of Falls Prevention in Elderly Patients,” *International Journal of Medical Informatics*, vol. 163, p. 104765, July 2022.

[43] J. A. Ramsay, S. Mascaro, A. J. Campbell, D. A. Foley, A. O. Mace, P. Ingram, M. L. Borland, C. C. Blyth, N. G. Larkins, T. Robertson, P. C. M. Williams, T. L. Snelling, and Y. Wu, “Urinary tract infections in children: building a causal model-based decision support tool for diagnosis with domain knowledge and prospective data,” *BMC Medical Research Methodology*, vol. 22, p. 218, Aug. 2022.

[44] J. Pan, H. Rao, X. Zhang, W. Li, Z. Wei, Z. Zhang, H. Ren, W. Song, Y. Hou, and L. Qiu, “Application of a Tabu search-based Bayesian network in identifying factors related to hypertension,” *Medicine*, vol. 98, p. e16058, June 2019.

[45] X. Zhang, Z. Yuan, J. Ji, H. Li, and F. Xue, “Network or regression-based methods for disease discrimination: a comparison study,” *BMC Medical Research Methodology*, vol. 16, Aug. 2016.

[46] E. Daudén, S. Castañeda, C. Suárez, J. García-Campayo, A. Blasco, M. Aguilar, C. Ferrándiz, L. Puig, and J. Sánchez-Carazo, “Abordaje integral de la comorbilidad del paciente con psoriasis,” *Actas Dermosifiliográficas*, vol. 103, pp. 1–64, Jan. 2012.

[47] J. Buvat, M. Maggi, A. Guay, and L. O. Torres, “Testosterone deficiency in men: Systematic review and standard operating procedures for diagnosis and treatment,” *The Journal of Sexual Medicine*, vol. 10, pp. 245–284, Jan. 2013.

[48] K. D. I. G. O. K. H. C. W. Group, “KDIGO 2018 clinical practice guideline for the prevention, diagnosis, evaluation, and treatment of hepatitis c in chronic kidney disease,” *Kidney International Supplements*, vol. 8, pp. 91–165, Oct. 2018.

[49] J. Watt, Z. Goodarzi, A. C. Tricco, A.-A. Veroniki, and S. E. Straus, “Comparative safety and efficacy of pharmacological and non-pharmacological interventions for the behavioral and psychological symptoms of dementia: protocol for a systematic review and network meta-analysis,” *Systematic Reviews*, vol. 6, Sept. 2017.

[50] E. Kyrimi, S. McLachlan, K. Dube, M. R. Neves, A. Fahmi, and N. Fenton, “A comprehensive scoping review of bayesian networks in healthcare: Past, present and future,” *Artificial Intelligence in Medicine*, vol. 117, p. 102108, July 2021.
Appendix A

Fig. A1 Research algorithm

Fig. A2 Articles identified
Fig. A3 Bag of worlds output
Fig. A4 LDA
Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 1 (26.1% of tokens)

- study
- research
- health
- review
- intervention
- development
- care
- current
- literature
- potential
- process
- method
- impact
- support
- analysis
- new
- impending
- including
- outcome
- future
- related
- assessment
- specific
- provide
- reported
- identify
- clinical
- system
- result
- practice

Overall term frequency
Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_{t} p(t | w) * log(p(t | w)p(b)) for topics t see Chuang et al (2012)]
2. relevance(topic w | topic t) = \lambda * p(w | t) + (1 - \lambda) * p(w | b); see Sievert & Shirley (2014)

Fig. A4 LDA
Fig. A5 LDA
Fig. A6 LDA

Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 3 (11.5% of tokens)

1. saliency(term, w) = frequency(w) * [sum_t ponder(t, w) * log(tfidf(w, t)) for topics t, see Chuang et al (2012)]
2. relevance(term, w | topic t) = λ * p(w | t) + (1 - λ) * p(w) / |t|, see Sievert & Shirley (2014)
Fig. A7 LDA
Fig. A7 LDA

Intertopic Distance Map (via multidimensional scaling)

Marginal topic distribution

Top-30 Most Relevant Terms for Topic 1 (45.7% of tokens)

1. saliency(term w) = frequency(w) \times \log(1 + \frac{1}{term frequency within the selected topic})

2. relevance(term w | topic t) = \lambda \times \log(w | t) + (1 - \lambda) \times \log(w | p(t))

Fig. A7 LDA
Fig. A9 LDA
| Title                                                                 | First Author                  | Year | Field   | Int | Mod | Dataset | Data Description                                                                 | Links  | Comment                                                                 |
|----------------------------------------------------------------------|------------------------------|------|---------|-----|-----|---------|---------------------------------------------------------------------------------|--------|--------------------------------------------------------------------------|
| Bayesian Networks for Clinical Decision Support in Lung Cancer Care  | Senes M. Berkan              | 2013 | cancer  | Yes | Yes | NA      | LUCADA database                                                                  | DB     | used the Junction Tree algorithm                                           |
| Lung Cancer Assistant: a hybrid clinical decision support application for lung cancer care. | Senes M. Berkan              | 2014 | cancer  | Yes | Yes | NA      | LUCADA database                                                                  | Software| based on prev. paper                                                       |
| Modeling interrelationships between health behaviors in overweight breast cancer survivors: Applying Bayesian networks | Xu S                         | 2018 | cancer  | Yes | NA  | Yes     | Dataset is not full. 333 postmenopausal overweight or obese breast cancer survivors participating | Dataset|                                                                       |
| Impact on place of death in cancer patients: a causal exploration in southern Switzerland | Kern H.                     | 2020 | cancer  | Yes (?) | Yes by request | 116 adult patients who died from cancer between 2015 and 2016 in southern Switzerland | Software|                                                                        |
| Title                                                                 | First Author | Year | Field   | Int | Mod | Dataset   | Data Description | Links                                                                 | Comment                                          |
|----------------------------------------------------------------------|--------------|------|---------|-----|-----|-----------|------------------|----------------------------------------------------------------------|--------------------------------------------------|
| Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: A dynamic Bayesian network | Petousis P.  | 2016 | cancer  | (?) | Yes | by request | National Lung Screening Trial | Models available in supplementary files | Shorter dataset                                   |
| Quantitative analysis of breast cancer diagnosis using a probabilistic modelling approach | Liu S.       | 2018 | cancer  | No  | Yes | Yes        | Breast Cancer Wisconsin Dataset (UCI) | Dataset                                                                 | Used two datasets. Second one is not available   |
| Cancer classification from time series microarray data through regulatory Dynamic Bayesian Networks | Kourou K.    | 2020 | cancer  | No  | No  | Yes        | GSE14426, GSE37182, GSE5462 | ?                                                                 | Pancreatic cancer dataset, colon cancer dataset, breast cancer dataset |
| Survivability modelling using Bayesian network for patients with first and secondary primary cancers | Wang K       | 2020 | cancer  | No  | Yes | Partially | National Health Insurance (NHI) database of Taiwan | Resulting conditional probability tables in the paper |                                                                |
| Preoperative risk stratification in endometrial cancer (ENDORISK) by a Bayesian network model: A development and validation study | Reijnen, C.  | 2020 | cancer  | No  | Yes | by request | International Federation of Gynecology and Obstetrics (FIGO) or ENITEC | Model                                                                 | Risk interpretation (Bounds) available          |
| Title                                                                 | First Author         | Year | Field          | Int | Mod | Dataset       | Data Description                                                                 | Links          | Comment                               |
|-----------------------------------------------------------------------|----------------------|------|----------------|-----|-----|---------------|----------------------------------------------------------------------------------|----------------|---------------------------------------|
| A clinical decision support system learned from data to personalize treatment recommendations towards preventing breast cancer metastasis | Jiang, X.            | 2020 | breast cancer  | No  | Yes | Yes           | LSDB-5YDM                                                                        | Dataset        |                                       |
| A privacy-preserving Bayesian network model for personalised COVID19 risk assessment and contact tracing. | Fenton, N.           | 2020 | COVID-19       | No  | Yes | Yes           | Data in the model                                                                | Age-naRisk model | Made a sensitivity analysis           |
| Application of intelligence-based computational techniques for classification and early differential diagnosis of COVID-19 disease | Akin-nuwesi, B.      | 2021 | COVID-19       | No  | No  | Yes           |                                                                                  | Software & data |                                       |
| Title                                                                 | First Author | Year | Field            | Int | Mod | Dataset | Data Description                                                                 | Links                                      | Comment                                                                 |
|----------------------------------------------------------------------|--------------|------|------------------|-----|-----|---------|-----------------------------------------------------------------------------------|--------------------------------------------|-------------------------------------------------------------------------|
| Application of a novel hybrid algorithm of Bayesian network in the study of hyperlipidemia related factors: a cross-sectional study. | Wang, X.     | 2021 | Hyperlipidemia    | No  | Yes | by request | Hyperlipidemia in Shanxi province; A total of 4567 complete data were left, including 2236 males, accounting for 49.0%, and 2331 females, accounting for 51.0% | R code available in supplementary files | Setted inter.iamb-Tabu hybrid algorithm to establish the BNs structure; Used a Netica software; Bounds available in file sup.1 |
| Application of tabu search-based Bayesian networks in exploring related factors of liver cirrhosis complicated with hepatic encephalopathy and disease identification | Zhang, Z.    | 2019 | cirrhosis of the liver | No  | Yes | Yes       | Information about patients with cirrhosis who were hospitalized in the Department of Gastroenterology, First Hospital of Shanxi Medical University from January 2006 to December 2015 and who had complete medical records | Data in supplementary files | Used Netica |
| Title                                                                 | First Author | Year | Field       | Int | Mod | Dataset | Data Description                                                                                                                                                                                                 | Links                  | Comment      |
|---------------------------------------------------------------------|--------------|------|-------------|-----|-----|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|--------------|
| A new Bayesian network-based risk stratification model for prediction of short-term and long-term LVAD mortality | Loghmanpour NA | 2015 | heart failure | Yes | Yes | by request | Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS) database                                                                                                                                  | Model in sup.files      | GeNie        |
| Predicting the causative pathogen among children with osteomyelitis using Bayesian networks - improving antibiotic selection in clinical practice | Wu Y         | 2020 | osteomyelitis | Yes (?) | Yes | NA | From a retrospective cohort (n=295) of children admitted to Princess Margaret Hospital for Children (PMH), Perth, WA, Australia from 2002 to 2007                                                                 | Model in sup.files      | Netica       |
| A Primer on Bayesian Decision Analysis With an Application to a Kidney Transplant Decision. | Neapolitan R  | 2016 | Kidney Transplant | Yes (?) | Yes | NA | Model in paper                                                                                           | KDRI index for bounds  |              |
| Title                                                                 | First Author | Year | Field       | Int | Mod | Dataset                      | Data Description                                                                                                                                                                                                 | Links                  | Comment                                      |
|----------------------------------------------------------------------|--------------|------|-------------|-----|-----|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|---------------------------------------------|
| Causal inference for violence risk management and decision support in forensic psychiatry | Constanti- nou, A | 2015 | psychiatry  | Yes | Yes | NA                           | VoRAMSS dataset The dataset consists of questionnaire, interviewing and assessment data from 386 patients, out of whom 343 are males and 43 are females. Interviews were performed at 6 and 12 months post-discharge | Model in paper         | Made a sensitivity analysis                 |
| Risk assessment and risk management of violent reoffending among prisoners | Constanti- nou, A | 2015 | psychiatry  | Yes | Yes | NA                           | Prisoner Cohort Study (PCS) dataset                                                                                                           | DSVM-P model (NA)      | Made a sensitivity analysis                 |
| Development of a Bayesian network for the risk management of violent prisoners | Coid JW       | 2016 | psychiatry  | Yes | Yes | NA                           | The sample included 953 prisoners (778 men and 175 women) assessed after release into the community.                                                                                           | Model in paper         | Used AgenaRisk. Violence risk management with interventions. Compare results with HRC-20, SAPROF, PANSS scales. |
| Title                                                                 | First Author | Year | Field         | Int | Mod | Dataset | Data Description                                                                                                                                                                                                 | Links | Comment                  |
|----------------------------------------------------------------------|--------------|------|---------------|-----|-----|---------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|--------------------------|
| Development of a Bayesian network for risk management of patients discharged from forensic mental health services | Coid JW      | 2016 | psychiatry    | Yes | Yes | NA      | VoRAMSS dataset; The sample used for parameterisation and learning of the network included 386 patients discharged from MSSs in the UK (343 men, 43 women) over a 12-month period.                                | Model in paper | Used AgenaRisk. Compare results with HRC-20, SAPROF, PANSS scales.                                                                                                           |
| Analyzing the performance of fuzzy cognitive maps with non-linear hebbian learning algorithm in predicting autistic disorder | Kannappan, A | 2011 | autistic disorder | No  | No  | Yes     | In paper                                                                                                                                            | No model | No model                                                              |
| A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer’s disease and mild cognitive impairment | Seixas, F.   | 2014 | dementia      | No  | Yes | Paid    | Duke University Medical Center (Washington, USA) and the Center for Alzheimer’s Disease and Related Disorders                                        | Data   | Model in paper                                                       |
| Reducing COPD Readmissions: A Causal Bayesian Network Model          | S. Lee       | 2018 | COPD          | Yes | Yes | NA      | Dataset from St. Mary’s Hospital                                                                                                                     | Model in paper | Model in paper                                                      |
| Title | First Author | Year | Field | Int | Mod | Dataset | Data Description | Links | Comment |
|-------|--------------|------|-------|-----|-----|---------|-----------------|-------|---------|
| A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival | Dag A | 2016 | heart transplant survival | No | No | Yes | UNOS dataset | Data and software available to download | Sensitivity analysis |
| Predicting graft survival among kidney transplant recipients: A Bayesian decision support model | Topuz, K. | 2018 | kidney transplant | Yes(?) | No | NA | The UNOS dataset included information on all kidney waiting-list registrations and transplants that had been recorded in the U.S and reported between June 30, 2004, and March 31, 2015 | Model in paper. United Network for Organ Sharing (UNOS) data | Bounds available. Sensitivity analysis |
| Analysis for warning factors of type 2 diabetes mellitus complications with Markov blanket based on a Bayesian network model | Liu S. | 2020 | type 2 diabetes mellitus | No(?) | Yes | by request | dataset was collected from National Health Clinical Center between 1st January 2009 and 31st December 2009 | Model in paper Datasets |
| Title                                                                 | First Author | Year | Field          | Int  | Mod   | Dataset | Data Description                                                                 | Links                  | Comment                                                                 |
|----------------------------------------------------------------------|--------------|------|----------------|------|-------|---------|----------------------------------------------------------------------------------|------------------------|------------------------------------------------------------------------|
| A Dynamic Bayesian Network model for long-term simulation of clinical | Marini S.    | 2015 | type 1 diabetes| Yes  | Partially | Yes     | Diabetes Control and Complications Trial / Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) | Dataset                | Used ODDS ratio. Automatically construct a Bayesian topology by K2 algorithm |
| complications in type 1 diabetes                                     |              |      |                |      |       |         |                                                                                  |                        |                                                                        |
| CBN: Constructing a clinical Bayesian network based on data from the | Shen Y.      | 2018 | EHR            | No   | No    | Yes     | ?                                                                                   | Sql-scenario Software  |                                                                        |
| electronic medical record                                            |              |      |                |      |       |         |                                                                                  |                        |                                                                        |
| An ontology-driven clinical decision support system (IDDAP) for      | Shen, Y.     | 2018 | disease diagnosis and antibiotic using | No   | No    | Yes     | OWL file; based on the IDO, NCBI, HPO, DrugBank, and DO databases OWL available. | Software was deleted from git | OWL file structure NA                                                  |
| infectious disease diagnosis and antibiotic prescription              |              |      |                |      |       |         |                                                                                  |                        |                                                                        |
| An intelligent system for prognosis of noncommunicable diseases’ risk | Pittoli F.   | 2018 | CAD            | No   | No    | Partially | Pima Indians Diabetes Data Set                                                    | Model in paper         | CAD risk factors and their values intervals.                           |
| Title                                                                 | FirstAuthor | Year | Field               | Int | Mod | Dataset | Data Description                                                                                                                                                                                                 | Links               | Comment |
|----------------------------------------------------------------------|-------------|------|---------------------|-----|-----|---------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|---------|
| From Personal Observations to Recommendation of Tailored Interventions based on Causal Reasoning: a case study of Falls Prevention in Elderly Patients | Chaieb, S. | 2022 | Falls Prevention in Elderly Patients | Yes | Yes | Yes     | 1810 elderly patients from the multidisciplinary falls consultations, between January 2005 and December 2018, at Lille University Hospital,                                                                                     | Model in paper      |         |
| Urinary tract infections in children: building a causal model-based decision support tool for diagnosis with domain knowledge and prospective data | Ramsay J.  | 2022 | Urinary tract infections | No  | Yes | Yes     | From May 2019 to November 2020, 391 children were enrolled in the prospective cohort study                                                                                                                         | Models & data       |         |
| Application of a Tabu search-based Bayesian network in identifying factors related to hypertension | Jinhua Pan  | 2019 | Hypertension        | No  | Yes | Yes     | In total, 39 neighborhood committees and villages in Shanxi Province were selected as survey sites                                                                                                                | Dataset Model in paper | Used Netica |
| Dataset                          | Context                                                                 | Sample size | No. of features | Link     |
|---------------------------------|-------------------------------------------------------------------------|-------------|----------------|----------|
| post-op                         | determine where patients in a postoperative recovery area should be sent to next | 90          | 8              | UCI      |
| contraceptive                   | contraceptive method choice                                            | 1473        | 9              | UCI      |
| lymph                           | lymphography                                                            | 148         | 18             | UCI      |
| tumor                           |                                                                         | 339         | 17             | UCI      |
| nurs                            |                                                                         | 12960       | 8              | UCI      |
| Medical Insurance dataset       | make predictions of the insurance cost people will have to pay           | 3630        | 6              | Kaggle   |
| lung                            | lung cancer data set                                                    | 32          | 56             | UCI      |
| Breast Cancer Prediction        | predict from the dataset whether a person has Breast Cancer: Benign or Malignant | 683         | 11             | Kaggle   |
| Erbil Heart Disease Dataset     | common factors or characteristics contribute to the cardiovascular disease | 333         | 21             | Kaggle   |
| Pima Indians Diabetes           | Diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset | 768         | 9              | Kaggle   |
| Coronary Heart Disease          | Contains cases of coronary heart disease (CHD) and variables associated with the patient’s condition | 462         | 10             | Kaggle   |
| Haberman’s Survival             | Survival of patients who had undergone surgery for breast cancer        | 306         | 4              | Kaggle   |
| Breast Cancer Wisconsin         | Predict whether the cancer is benign or malignant                       | 569         | 32             | Kaggle   |
Table 3. Databases

| Database                          | Description                                                                 | Link          |
|-----------------------------------|-----------------------------------------------------------------------------|---------------|
| WorldBank Data catalog            |                                                                             | link          |
| OptumLabs Data Warehouse          | 160 million de-identified records across claims and clinical information to conduct investigations on populations | link          |
| NSCH Datasets                    | National Survey of Children’s Health                                         | link          |
| NLST Datasets                    | Cancer Institute                                                            | link          |
| Medical Information Mart for Intensive Care | Freely available medical data for research                                    | link          |
| Osteosarcoma Database            | Contains 911 protein-coding genes and 81 microRNAs associated with osteosarcoma | link          |
| Hospital Episode Statistics (HES)| Containing details of all admissions, outpatient appointments and A and E attendances at NHS hospitals in England | link          |
| MIMIC-III Clinical Database      | Health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012 | link          |
| CENTER-TBI                       | Contains prospectively collected data of more than 4,500 patients with TBI in Europe | link          |
| ISCoS                            | Spinal Interventions and Surgical Procedures Basic Data Set                  | link          |
| BioGPS                            | Gene and protein function                                                    | link          |
| GDC portal                       | Genomic Data Commons Data Portal                                            | link          |
| PPMI                             | Parkinson’s disease                                                         | link          |
| Open Data UK                     | data published by central government, local authorities and public bodies   | link          |
| National Lung Screening Trial    | Lung Cancer                                                                  | link          |
| STS Intermacs Database           | North American registry for the clinical outcomes of patients who receive an FDA-approved mechanical circulatory support (MCS) device to treat advanced heart failure | link          |
| BayesFusion                      | Interactive Model Repository. Models available to download                  | link          |
Section C

Fig. C1: Application of a Tabu search-based Bayesian network in identifying factors related to hypertension
Fig. C2: From Application of tabu search based Bayesian networks in exploring related factors of liver cirrhosis complicated with hepatic encephalopathy and disease identification
Fig. C3: From Hip Fracture in the Elderly A Re-Analysis of the EPIDOS Study with Causal Bayesian Networks
Fig. C4: From Preoperative risk stratification in endometrial cancer (ENDORISK) by a Bayesian network model: A development and validation study
Fig. C5: Paper "From Personal Observations to Recommendation of Tailored Interventions based on Causal Reasoning: a case study of Falls Prevention in Elderly Patients" (A proposed network for interventions)
Appendix D

Fig. D1: Framework
Fig.D2: simplify