

Knowledge Based Bayesian Network Construction Algorithm for Medical Data Fusion to Enhance Services and Diagnosis

Ahmed Sameh

Department of Computer and Information Systems, Prince Sultan University, Riyadh 11586, Saudi Arabia

Article history

Received: 7-05-2018

Revised: 14-12-2018

Accepted: 27-05-2019

Email: asameh@psu.edu.sa

Abstract: Traditional Bayesian networks' algorithms are treating the network construction process as an isolated and autonomous data-driven trial-and-error process and completely ignoring the domain knowledge. In this work we are proposing a new 'Semantically Aware Ontology-Based Bayesian Network construction algorithm' that is knowledge centered instead of data centered. The objective of the new algorithm is to empower patients through improving their self-diagnosis and testing by automatically constructing a set of Ontology-Based Bayesian networks using combination of domain and expert knowledge. The exciting thing about the proposed algorithm is that it uses on 'Saudi-native training data' streamed from the "Unified Medical Record" server and authenticated domain and expert knowledge extracted from the "King Abdulla Encyclopedia" server. A proof-of-concept prototype based on open-source software "Netica" and "Protégé" is implemented and tested. It demonstrates learning of probabilities, network structure and mixes discrete and continuous variables. It imports "Diabetes" patient medical record steams from the "Unified Medical Record" server to be used as training and testing datasets. It also extracts Bayesian data variables from the "King Abdullah Encyclopedia" server to aid in constructing and learning the ontology-based Bayesian networks. The prototype is implemented on an Internet server and can be accessed from medical applications on Smartphones and PDAs. It currently deals with 60 positive "Diabetes" Saudi patients and 60 negative "Diabetes" training cases. The resulting Ontology Bayesian network was tested on another 100 test cases drawn randomly from the 'Unified Medical records' server. An accuracy of diagnosis of 100% was achieved on the test data.

Keywords: Home Healthcare, Bayesian Network, Ontology, Medical Encyclopedia, Unified Medical Records, Self-Diagnosis

Introduction

A paradigm shift from traditional data mining to semantic data mining is leading to Knowledge Centered instead of Data Centered data mining techniques. Integrating Ontology and Bayesian Network is currently promoting the enrichment of our medical domain knowledge. At the moment, open interoperable ontologies are enabling medical stakeholders to communicate with no ambiguity. Unfortunately obtaining action items from such ontologies is done through deterministic reasoning which are unsuitable for the medical field. Besides, Bayesian networks offer an intuitive representation of uncertainty, which can be integrated with medical ontologies through few steps. This integration can be done by extracting the Bayesian variables (noun nodes) that are core in the specific

medical domain along with their possible attributes/values from the ontologies, mapping the relationships (verb nodes) between these variables through consulting the appropriate ontologies and then calculating the conditional probabilities required for the Bayesian nodes.

It is well known that medicine and healthcare generally lag behind in technology adoption. Thus, Kingdom of Saudi Arabia is trying to enhance its medical and healthcare sector through a number of national technology-based mega projects. Two notable projects are the Ministry of Health's "Unified Medical Record" project (MHUMR, 2016) and the National Guard's "King Abdulla Encyclopedia of Arabic Health Content" project (2018). The first project was launched by KSA's Ministry of Health in 2008 and incremental implementation has already been finished in the

areas of Riyadh and Jeddah and their surroundings. The goal of the project is to have a unified electronic medical record for every Saudi citizen and every resident in the Kingdom. The medical record for a patient includes all the necessary medical documents (diagnosis, treatment, prescriptions, medical tests, medical images such as X-

rays and MRIs, major operations, medical costs, therapy, etc.). For instance, using person's ID, the complete unified electronic medical record of the person will be collected on-the-fly from all physical data centers that are distributed over the Kingdom. Figure 1 depicts glimpses from the current unified electronic medical system.

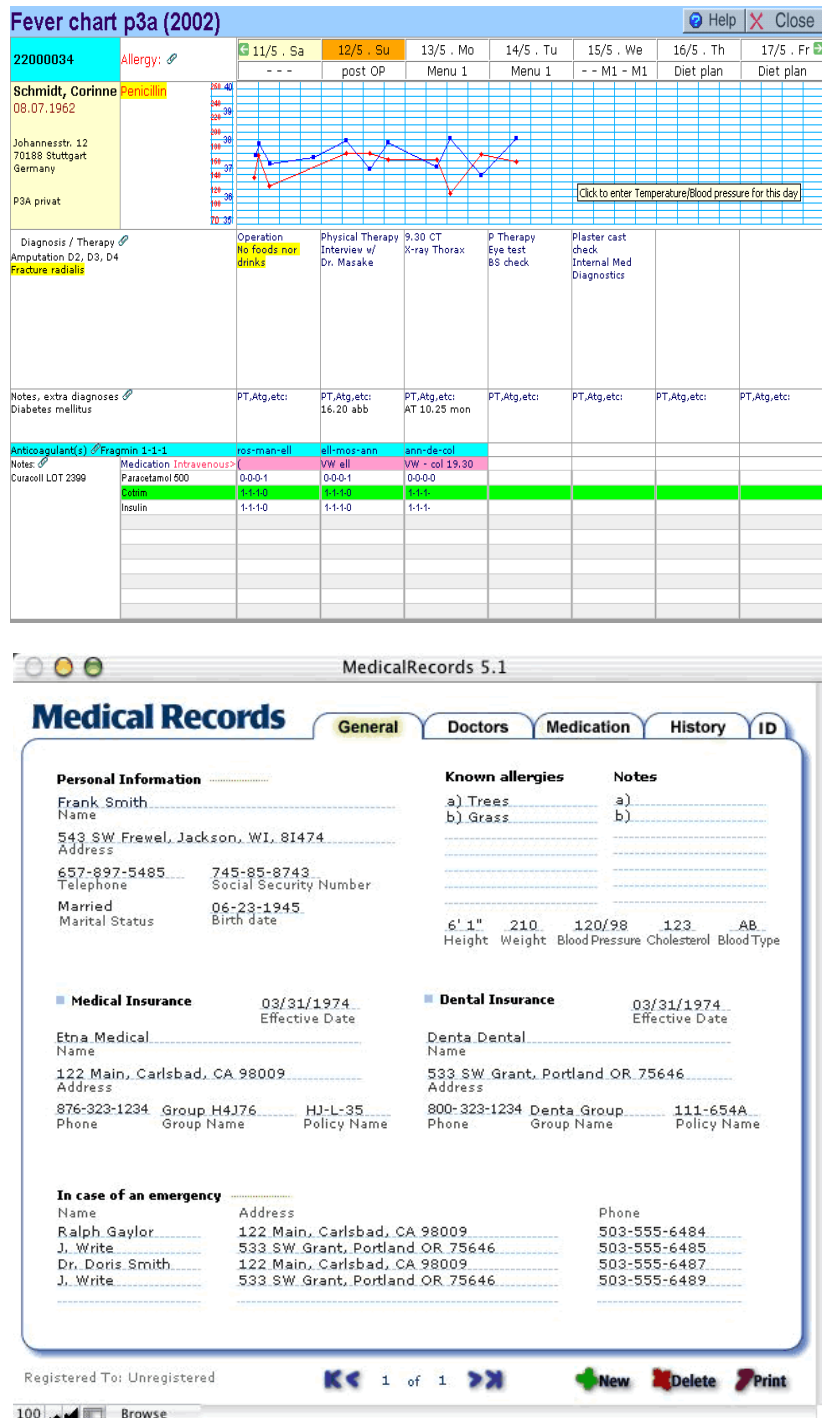


Fig. 1: KSA Ministry of Health – "Unified Medical Records" System- Screen Shots

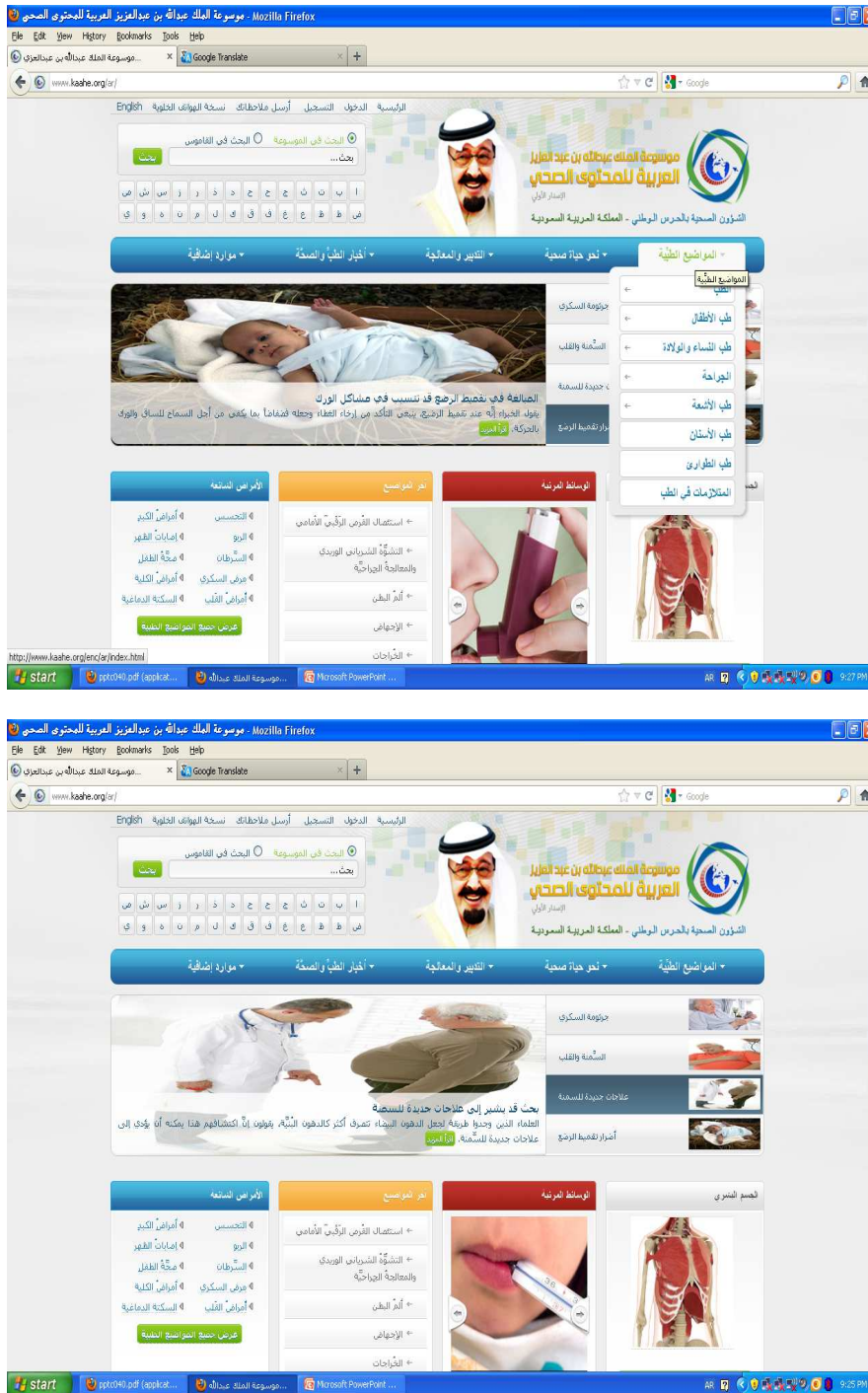


Fig. 2: Glimpses of Arabic King Abdulla National Medical Encyclopedia – Screen Shots

The second project was launched by KSA National Guards in 2009 and the first phase of the project has been launched on the Web in 2011. The goal of the project is to provide the first completed Arabic medical encyclopedia for healthcare sector. Moreover, this project takes in a medical dictionary of terms, complete descriptions of various diseases; their causes, symptoms, medications

with multimedia contents, various branches of medicine and pharmacology, therapy, history of Arabic medicine, medical library and medical news. The contents of the encyclopedia has been authenticated and verified by medical experts from the KSA National Guards. The target audiences are the general public, so although medical contents are ensured to be comprehensive but it is presented

in a straight forward simple way. The final outcome is a complete usable Arabic medical encyclopedia. Figure 2 shows glimpses from the current system.

In this project we are semi-automatically constructing a set of Ontology-based Bayesian networks from the above two projects that relates causes to symptoms to medications and effects using combination of domain and expert knowledge. We have automatically extracted training and testing datasets and incrementally fused Ontology-based Bayesian learning algorithms with these data. The exciting thing about the proposed system is that it is based on Saudi native training datasets streamed from the "Unified Medical Record" project, uses authenticated domain and expert knowledge extracted from the "King Abdulla Encyclopedia" project and deploys powerful incremental fusion Ontology-based Bayesian modeling that involves realistic uncertainties and conditional probabilities. The proposed system utilizes cross fertilization of Bayesian data mining techniques to provide intelligent decision/reasoning aids. Likewise, Incremental Ontology-based Bayesian propagation algorithms are used to provide adjustment and re-training of existing Bayesian networks whenever new cases and recent knowledge need to be incorporated into existing Ontology-based Bayesian models. For example in some cases, dynamic Bayesian networks are used to model disease temporal progress.

The structure of this paper is as follows: section 2 provides literature review of some similar works, section 3 provides algorithms used in integrating ontologies with Bayesian networks. In section 4 we present the overall architecture of the proposed system. We explain the technique of extracting Bayesian variables (nouns) and influencers (verbs) from the Arabic Encyclopedia and the Unified Medical records in section 5. In section 6 we show how to build Datasets from the "Unified Medical Records" databases to be used for both training and testing. In section 7 we show the ontology-based Bayesian learning algorithms digesting the training dataset. Implementation of a prototype using both "Protégé" and "Netica" is shown in section 8. Finally we conclude in section 9.

Literature Review

Linda and Akaichi (2017) proposes a new ontology: Disease-Symptom (DS-Ontology) that is utilized in an ontology-based Clinical Decision Support System to govern with certainty effective medical diagnosis. The proposed system makes use of Bayesian networks for correlating various pieces of information along with various possible conclusions extracted from the ontology.

Mobile medical system is proposed in (Eunjeong *et al.*, 2009) that retrieves patients' clinical information file for immediate diagnosis. It monitors and analyzes the context of service for various providers in many medical organizations' databases. Medical information file fusion is demonstrated by reasoning about the relationships among the various pieces of information.

Giovanni *et al.* (2016) utilizes electronic medical records that have the patient health history and relates this to the patient interaction with a healthcare facility or clinical trials. Data mining models including Bayes networks are used to automate the process of information extraction and to support health data management.

Wang *et al.* (2014) proposes a probabilistic disease progression model that uses Bayesian networks. It learns the full progression major trajectory from a set of incomplete and inaccurate records that only represents segments of the progression. Observations are incomplete and irregular. Bayesian networks helped in bridging between the hidden forward progression process and the observed medical evidence witnessed.

The paper in (Kabli *et al.*, 2008) reports on using new evolving Bayesian networks along with mixed Genetic algorithms for the problem of 'Prostate Cancer' treatment. Bayesian has been used to represent and process uncertainties inherent in the clinical practice. The proposed techniques showed more versatile and effective alternative outputs to the widely used patient tables for 'Prostate Cancer' pathology production.

The work in (Ha *et al.*, 2009) uses statistical uncertainty approach for summarizing medical Bayesian network structure/parameters from medical data records. Data is obtained from original correlation graphs and construct Bayesian networks that would crop the most matching correlation graphs possible. The proposed technique determines the network parameters with good accuracy.

The paper in (Yousefi and Dalton, 2014) proposes a Bayesian classification based on best control strategies for probabilistic Bayesian networks in successfully determining 'Cancer Prognosis' in face of inherent stochastic of cell dynamics to predict disease progression.

The work in (Hu *et al.*, 2017) proposes a new methodology that uses an ontology to supply uncertain relationships among various random variables in the Bayesian network. For example for the patients of depression, the technique by learning Bayesian structures and probability tables from Treatment Alternatives patient datasets, was able to achieve high level of diagnosis accuracy.

Integration of Bayesian Networks and Ontology

Bayesian networks are used for representing and analyzing probabilistic models of uncertainty reasoning. Bayesian networks use factored joint probability distributions (conditional probabilities) as a directed graph model (DAG). The role of DAG is for organizing information about uncertain domain variables, their relationships and a computational model for calculating the impact of specific evidence (s) on belief (s) (evidences) wherein nodes represent variables, arcs represent

probabilistic dependence among variables and conditional probabilities represent the strength of the data dependencies in the form of conditional probability tables CPT at each node with its direct parents. Their computational model is

utilized as follows: compute posterior probabilities given evidence (s) about selected nodes. It also expresses and exploits probabilistic independence for efficient and effective computations (see Fig. 3).

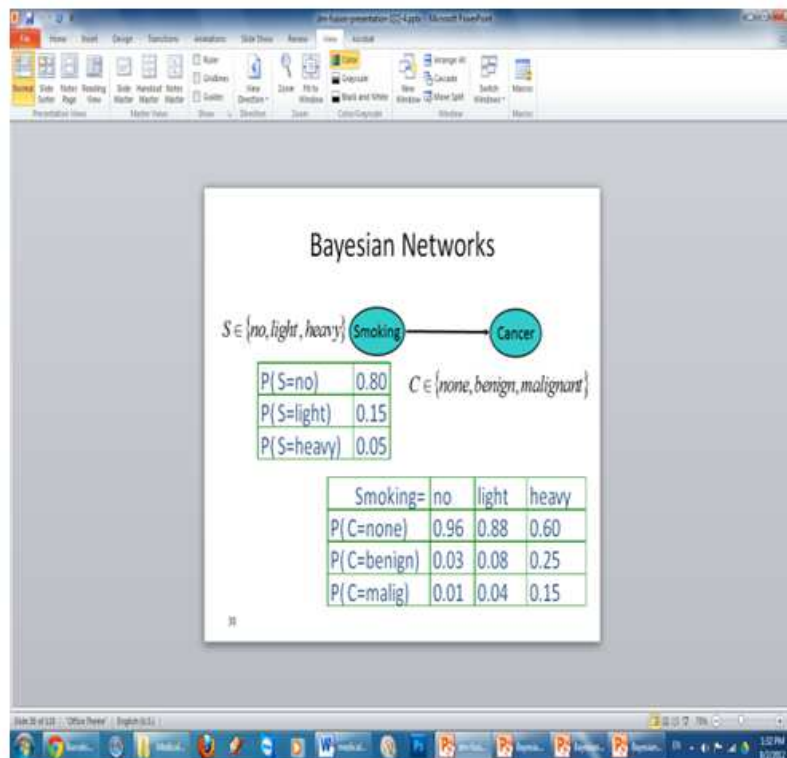
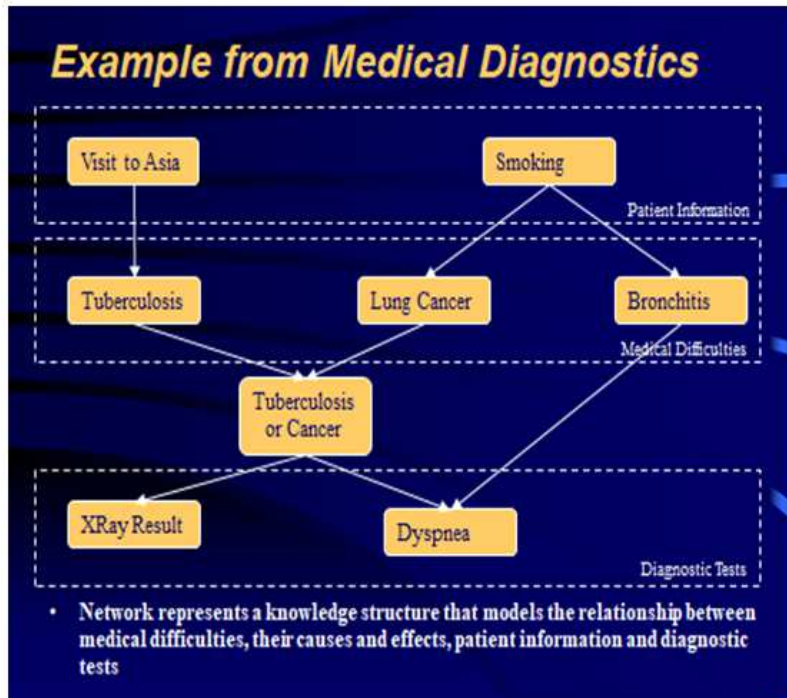


Fig. 3: Bayesian Network Node dependencies – NETICA Screen Shots

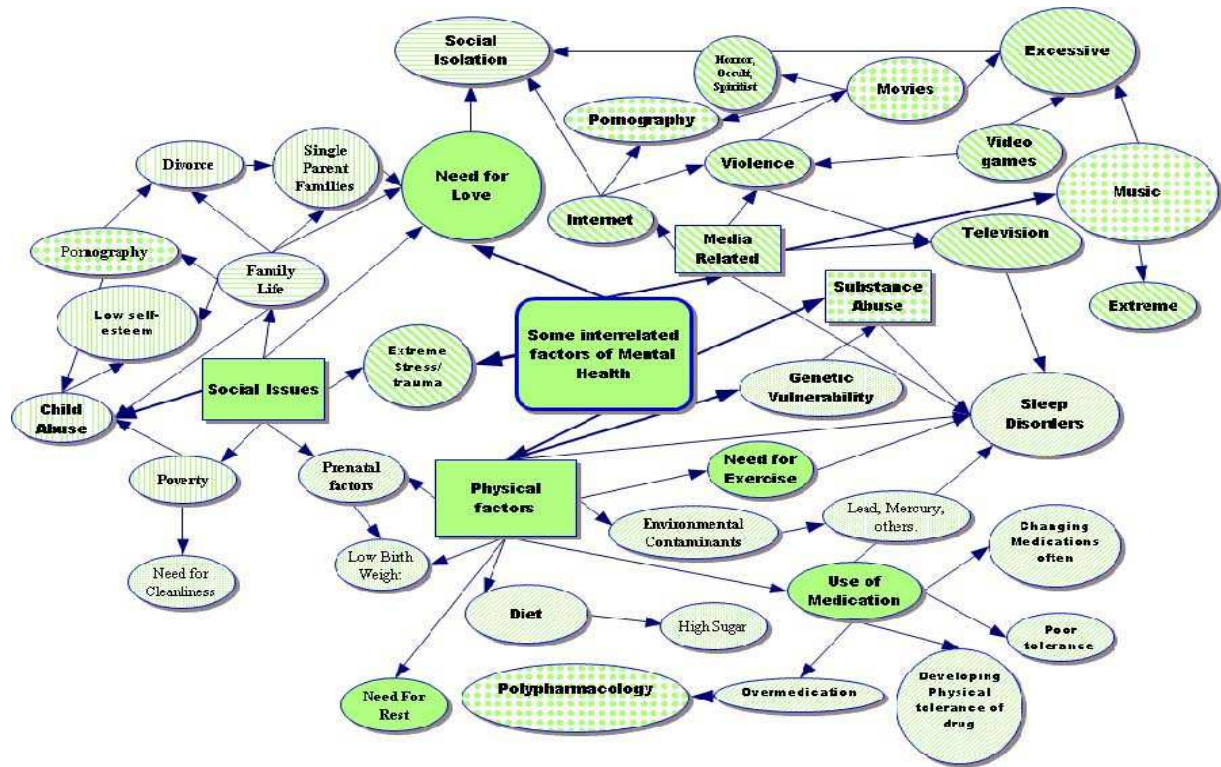


Fig. 4: Bayesian Model of a "Mental Disease" Network (Yousefi and Dalton, 2014)

As we are building medical models from “King Abdulla Encyclopedia” using Bayesian Networks, Fig. 4 shows an example of a Bayesian model of a “Mental Disease”. This model is used for making intelligent decisions/reasoning, data fusion of multiple information sources, answer queries in the form of conditional probabilities. For example: “What are the chances of developing “Depression” given that the patient has no friends?” The figure shows how the training data set are fused and summarized in the form of CPT conditional probability tables of the Bayesian network. These graphs come from mix of interactions between the artificial intelligence- AI, decision theory, reasoning and statistical communities. The reason for using them in this project comes from the development of propagation algorithms for constructing and querying them followed by a continuous availability of free commercial software tools. Currently there is growing number of creative applications that make uses Bayesian networks, we consider the current deployment to be a new creative application. The steps of constructing traditional Bayesian networks are as follows:

1. Factored the joint/conditional probability distributions as a directed graph –DAG-using:
 - Tree Data-structure for information about various uncertain variables and their dependencies
 - A computational model for calculating the impact/effect of evidence on beliefs

2. Build a Knowledge structure:

- All model's variables are depicted as graph nodes (nouns)
- Model's arcs are labeled with probabilistic dependence among variables (verbs)- CPT-conditional probabilities tables express/encode the strength of the dependency
- Various dependencies among participating nodes

3. Establish the Computational model:

- Calculates posterior probabilities given evidence about selected nodes (node and its parents)
- Exploits/expresses probabilistic independence for effective/efficient computation (delete some arcs)

Integrating Ontology and Bayesian Network is currently promoting the enrichment of our medical domain knowledge (see Fig. 5). At the moment, open, interoperable ontologies are enabling medical stakeholders to communicate with no ambiguity. Unfortunately obtaining action items from such ontologies is done through deterministic reasoning which are unsuitable for the medical field. Bayesian networks offer an intuitive representation of uncertainty.

Melchior (2013), Fig. 6-10 show by example how to inject/encode uncertainty into the 'Aspirin Therapy for Diabetic Patients' ontology (our running example ontology). Figure 6 depicts the proposed integration methodology.

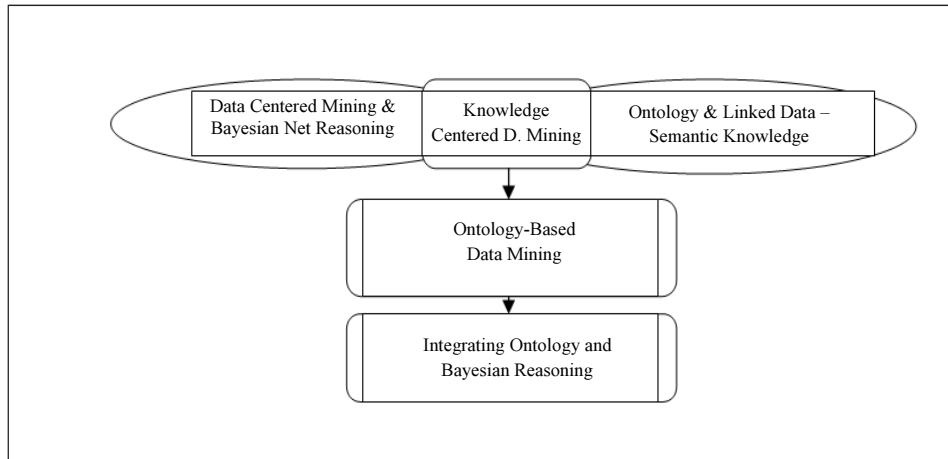


Fig. 5: Integration through Fusion, Mining and Reasoning of Medical Diagnosis Data

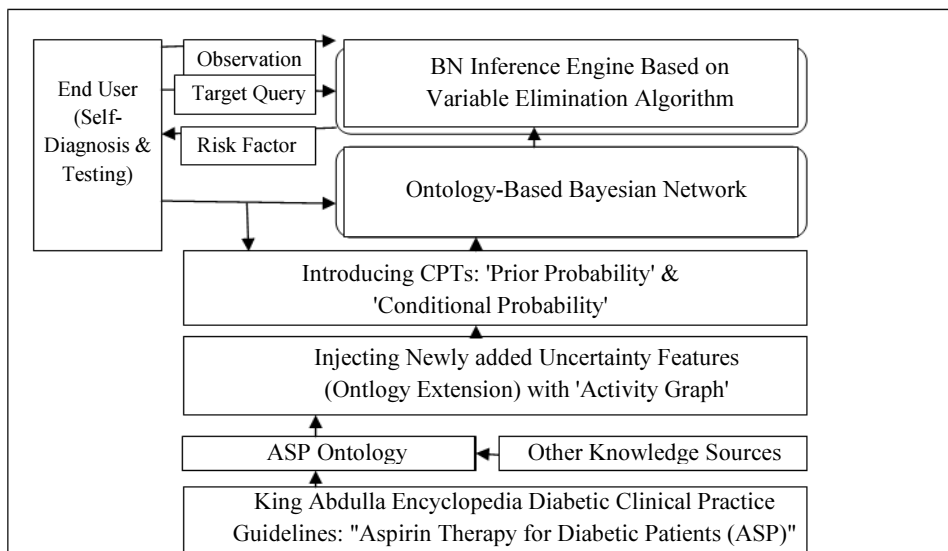


Fig. 6: Proposed Ontology Based Bayesian Network Construction Steps and User Interaction

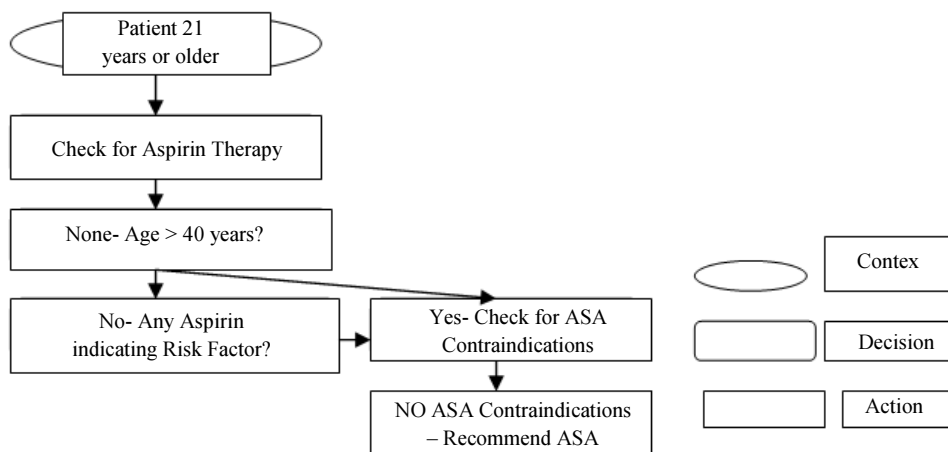


Fig. 7: Classification of portion of "Aspirin Therapy for Diabetic Patients (ASP)" ontology nodes– clinical practice guideline- activity graph variant pathways

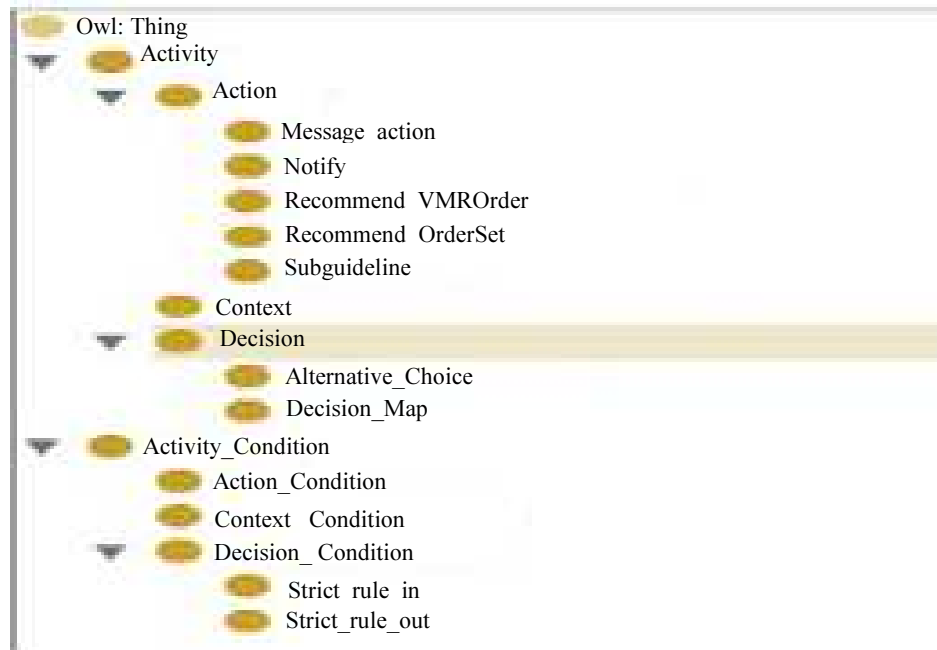


Fig. 8: Protégé class representation for the ASP ontology- screen shot

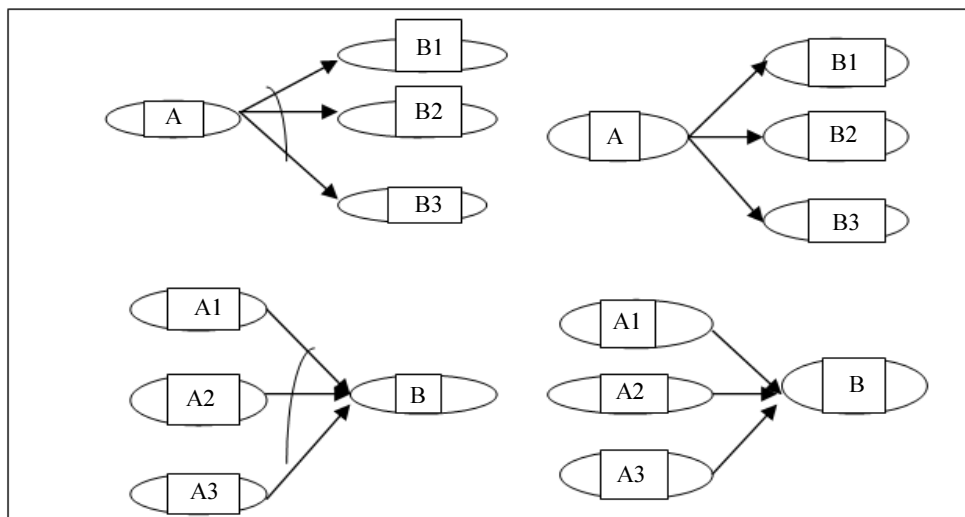


Fig. 9: Rules for the construction of the Conditional Probability Tables (CPTs) with "AND" and "OR"

First, the original 'Aspirin Therapy Encyclopedia' (extracted from King Abdulla Encyclopedia) is represented into an ontology model (using Protégé Editor) that incorporate newly added uncertainty features (encode some uncertain activities not variables). Accordingly, we introduce a kind of ontology extension by injecting uncertainty activities as nodes. We then link them to the original ontology nodes by introducing CPTs (Conditional Probabilities) as in standard BNs. As such; when the end user gives his/her observed evidences (see Fig. 6); the BN will use its inference engine (based on the famous variable elimination algorithm) to

deduct and reason out probabilities for some activities queried by the user. Accordingly the user can decide the risk of following on with his/her selected activities.

We use as our current running example the "Clinical Ontology of Aspirin Therapy for Diabetic Patients" (Fig. 7). This ontology has been extracted from King Abdulla Encyclopedia shown in Fig. 2. In it, several recommendations activities are exposed from the clinical process. They then form what is called "Activity Graph". It represents relationships among activities. We use it to demonstrate how we cope with uncertainty in the ontology.

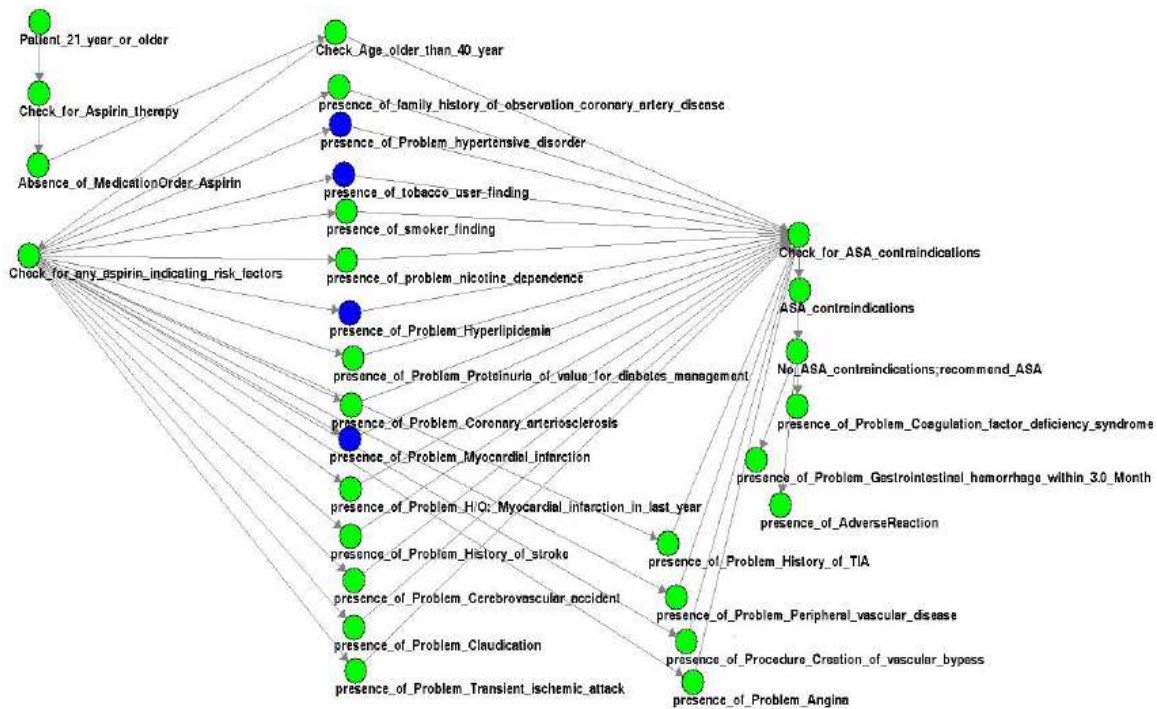


Fig. 10: An ontology-based Bayesian Network (BN) of ASP screen shot-driven according to figure 7 classification above

An activity graph has three activity variants: context, decision and action (see Fig. 7). Each part of the activity graph starts with a context node that describe the clinical context of that pathway and the conditions needed for that path to be followed. A decision node lists subsequent action activity nodes and the conditions for their execution. Finally, action nodes condense some work items that must be executed by the user.

The corresponding OWL Protégé classes are shown in Fig. 8 for the Aspirin Therapy ontology. The above activity graph is inserted in Protégé in the 'Activity Class'. This class contains all the nodes in the activity class. Inside it, three sub-classes: Context, Decision and Action are defined. Any activity can include further internal conditions that must be fulfilled for the node to be executed. For example the "Yes; Check for ASA" activity has the following internal conditions: checking presence of family recorded history, checking existence of hypertensive disorder etc. We convert these insider conditions of activity nodes as an activity condition class in the ontology shown (see Fig. 8).

Protégé OWL adds a number of uncertainty features as property set that contains {**hascause**, **hasCondition**, **hasValue**, **hasState**, **isObserved**, **hasPriorProValue**, **hasCondiProValue**} and a number of activity instances. The property 'hascause' describes the dependency relationship, 'hascondition' attach specific probability attributes, 'hasStat' is a Boolean function, 'isObserved' is another Boolean function, a property function 'hasPriorProValue' is a Float function and a property

function 'hasCondiProValue' is a Float function. The execution of a node is triggered if the node preconditions (s) are satisfied and the execution status of its subsequent nodes is checked, as depicted in Fig. 8.

Other added uncertainty features to the Protégé OWL ontology include CPTs such as "Prior Probability" and "Conditional Probability". Both encode uncertainty levels. For example 'hasPrior- ProValue' property and 'hasCondiProValue' property are used. Figure 9 shows the rules for construction the CPTs. An activity that has no parent has only 'prior probability'. When an activity node allows the execution of a set of activities nodes simultaneously the conditional probabilities CPT are set to 1.0. As soon as an activity node allows the execution of one of a list of subsequent activities nodes, the conditional probabilities CPTs are set to 1.0/n. Once a set of activities nodes allow the execution of activity node together, the conditional probability CPT is set to 1.0; and when one of the activities can allow an activity node, the conditional probability is set to 0.0. We apply the variable elimination algorithm (Zheng and Bo-Young, 2007) to perform BN inference (a standard engine). E denotes observed evidence by the end user, Xq is the query predicate that needs to compute its posterior probability, XE is the set of observed/measured random variables. BN uses the following formula:

$$P(Xq / E) = P(Xq, E) / P(E) = \text{Sum}_X \{Xq, XE\} P(X) / \text{Sum}_X / XEP(X)$$

End users follow the designated ontology in the ontology-based BN step by step. As shown in Fig. 10, they check the 'hypertensive disorder' node, then 'tobacco user finding', then 'hyperlipidemia' node and 'myocardial infarction' node respectively to provide their input observations (Blue nodes). The system loads the relevant ASP ontology-based BN and allows the user to select a 'target' activity, to calculate its conditional probability given the provided evidence. This is the 'testing' mode, where CPTs are fixed and not update. They are used to predict the query probability of a new case (the authorized user) given the history of UMDs cases taken from the UMD database. The BN uses the famous variable elimination algorithm to perform the inference according to the above formula. The figure shows the probability of the activity node "No ASA (aspirin therapy) contraindications- recommend ASA" (Green node) to help end-user judges/decides whether or not his/her observations of aspirin risk factors are adequate and how to proceed.

The Figure also shows another query case to acquire the uncertain actual probability degree of activity instance (Green node) "presence of problem coagulation factor

deficiency syndrome". The end user will again provide his/her observed evidence as before; they check the 'hypertensive disorder' node, then 'tobacco user finding', then 'hyperlipidemia' node and 'myocardial infarction' node respectively to provide their input observations. The system again loads the relevant ASP ontology-based BN and applies the inference algorithm.

The conclusion of the above cases shows high certainties-probabilities values for the queried activities, which suggest that the user can take a decision to proceed ahead based on the fed/observed evidence.

Proposed System Architecture

We are using the Lung Cancer ontology extracted from LUCADA (Royal College of Physicians, 2015) international patient dataset, KSA Unified Medical Records (UMD), as well as King Abdulla Encyclopedia as our data sources. Figure 11 shows the client/server architecture of the proposed system.

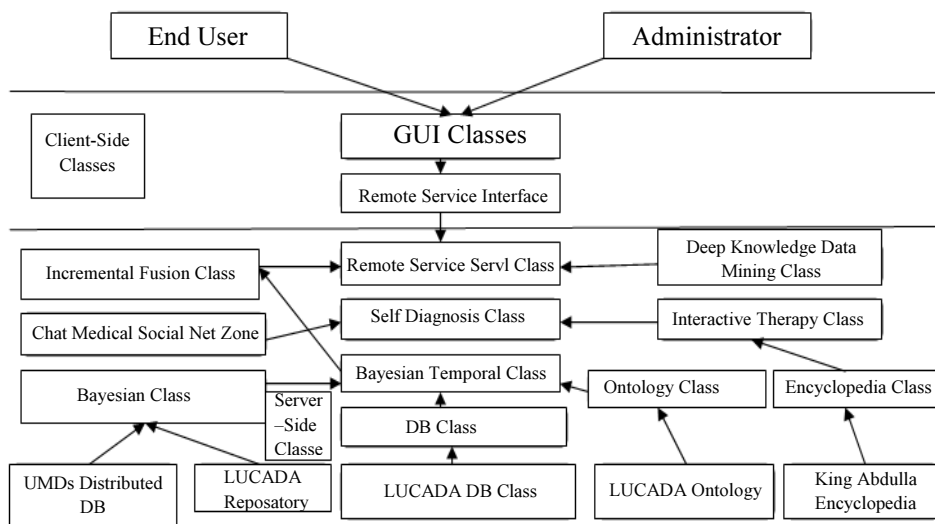


Fig. 11: Architecture of the proposed system

Table 1: Functional Modules of the proposed system

Module Name	Functions Assigned to this module	Its Role in the "Training Mode"	Its Role in the "Testing Mode"
Admin Interface- AI	Managing the Users, Encyclopedia contents, import datasets from the "Unified Medical records"	Entry Point for the Admin to train and incremental re-training of the various Ontology-Based Bayesian Networks	Observe the success rate of "reasoning" and "Diagnosis"
Remote User Interface-UI	Made up of: Diagnosis, Chat, Social Medical media, E-Learning and Incremental Enhancement Modules	Users have no role in training mode	Main users of this mode
Encyclopedia Database Class	Extract & Display relevant context-related Disease material	Both Admin and Users can access and manage the Encyclopedia Database to extract relevant datasets and edit ontologies	Use in E-Learning
Unified Records Databases Class	Extract & Display relevant context-related Diseases Medical Records	Both Admin and Users can access and manage the Unified Medical records Database to extract relevant records	Extract and compare his medical data with similar patients
Noun Extractor Class	Arabic Noun extraction using ARABIC POS-NLP algorithm	Discovering nodes for Building a Bayesian Network for a specific disease	Extracting nouns from the user queries

Table 1: Continue

Verb Extractor Class	Arabic Verb extraction using ARABIC POS- NLP algorithm	Discovering direct influence between nodes of the ontology-based Bayesian networks – this will influence the tables of probabilities	Extracting verbs from the user queries
Bayesian Structure	Builds the Inter-dependency Graph of a Bayesian Network	Before performing "training" Ontology-Based. Bayesian nets have to be constructed first Admin use both nouns and verbs extracted to do that	Use them for diagnosis and reasoning
Server Bayesian Training Class	Propagate the training dataset up/down a Bayesian Net and update the Probability tables	Admin apply this module to train various Ontology-Based Bayesian networks of various Diseases	User does not train
Server Bayesian Propagation Algorithm	Used in both training (update the probability tables) and testing (on fixed probability tables)	Used by Admin while invoking the above module (Bayesian Training)	User does not use this algorithm
Bayesian Testing Class	Propagate the testing dataset using the Bayesian Propagation Algorithm	Not used in this mode	Main user
Chat Medical Social Net Zone	From the User Interface, the User can chat with other on-line Users, Physicians, Doctors, other Users, and Admin of the system	Not used in this mode	Member of the Medical Social Network
View Training Dataset	Display any dataset extracted from "Unified Medical Records" Database- LUCADA Database	Admin and Users can activate this module to browse into selected datasets- Uses LUCADA Database	User can view them
View My Medical record	Any User can extract his/her record from the "Unified Medical record" Database and display its contents	Admin and Users can activate this module to browse into selected datasets	User can view his/her unified medical record
Server Deep Data Mining Training Class	Export extracted Unified Medical Datasets to "Microsoft Azure Machine Learning" Studio for data mining algorithms experimentations	Admin activate this module to move to Microsoft Azure Machine Learning tool to run data mining experimentations in the MS studio	User is not involved
Server Deep Data Mining Testing Class	For Data Mining Prediction, it calls Microsoft Azure Machine Learning tool and import its results	Admin activate this module to move to Microsoft Azure Machine Learning tool to run data mining experimentations in the MS studio	User is the main agent in this activity
Diseases Bayesian Network Library	Archive various Diseases Bayesian networks in a ready-to-use library so that any can be called for diagnosis- LUCADA BNs	Admin invokes this module to build the Bayesian library of various Diseases' networks- Uses LUCADA BNs	User can open the library and select any Bayesian network to deploy
Process a Query	Users can pose NLP Arabic queries to the system	Not used in this mode	User initiated
Server Increment fusion of Bayesian networks Class	Admin can re-train existing Bayesian networks by increment fusion algorithm	Admin invokes this module to re-train the Bayesian nets with any new datasets coming from the Unified Medical Records database - Uses LUCADA Database	User is not involved here
Activating E-Learning Mode Class	For medical therapy Users can combine diagnosis with Encyclopedia suggested medical therapy	Not used in this mode	User is involved here
Server Self-Diagnosis Class	Users can retrieve appropriate Bayesian networks and apply his/her data in the propagation algorithm testing mode-LUCADA Database	Not used in this mode	User is the main player here
Server Interactive Therapy/ Prognosis Class	For medical therapy Users can combine diagnosis with Encyclopedia suggested medical therapy	Not used in this mode	User is the main player here
Server Bayesian Temporal Progress Class	Users can track the progress of their disease through applying "Diagnosis" at various stages of their Diseases	Not used in this mode	User is the main player here
Server Ontology Class	Repository of various diseases ontologies for integration with various Bayesian Networks - LUCADA Ontology	Ontology-Based Bayesian algorithm uses this class to build Domain Knowledge Bayesian Networks	Same but for testing mode.

A functional table of the subsystems (Classes) and their interactions- for both the "Training" and the "Testing" Modes is shown in Table 1. There are two modes: Administrator and User. The Administrator mode composes new ontology-based Bayesian nets, train them and upload them to the system for use by Users. Users work in testing mode.

(a) Training Mode

For a particular disease (e.g., Lung Cancer Disease, or Aspirin therapy for diabetic patients (ASA), etc.), relevant individual records are retrieved from the "Unified records"

databases and form the LUCADA (Lung Disease dataset). Likewise, relevant Encyclopedia pages of the "Lung Disease" and "Aspirin therapy" are extracted from King Abdulla Encyclopedia. These pages are text processed to extract "nouns" that will be deployed as variables in the Bayesian nets. Verbs are next extracted to define various relationships among the variables. Accordingly, Bayesian network structure composes and displays to the domain expert to fill in the "Conditional Probability tables". The domain expert can overwrite/modify the dependency graph connections while filling in the probability tables. This last step can be replaced by an automated "training" mode

Bayesian Belief Network construction. The Ontology-Based Bayesian Learning algorithm of section 2 is used to construct and train the Bayesian networks. The trained Bayesian networks are now ready to answer any conditional probability questions (see the "Testing Mode" from the trained Ontology-Based Bayesian network). Questions can deal with (a) Diagnosis; (b) treatments; (c) Prevention; (d) Minimizing impact; (e) Detection; (f) Reaction; and (g) Enhance medical services.

(b) Testing Mode

Figure 19 shows the modified "King Abdulla Encyclopedia"'s Client page. Patient ID is added so that the "Unified Records" relevant data for this particular patient will be retrieved from all distributed databases of the "Unified Records" database servers. The learnt Ontology-Based Bayesian networks for various diseases (see the above section) will be stored in an indexed library. Relevant ones will be recalled by the user to run the ontology BNs integration algorithm- in testing mode using his/her own data on the retrieved Bayesian models and answer his/her own queries/questions.

Text mining techniques are used to extract variables and their dependencies structures of the Bayesian networks- see next section. Text from both the Encyclopedia and the unified servers will be analyzed to discover relevant variables Bayesian nodes (nouns) and their dependencies (verbs). Keyword verbs such as: affect, raise, lower, increase, decrease, develop, etc. are searched to provide basis for such discoveries. End user then can compose his queries in natural language and it will be converted to "Conditional Probability" queries to the Bayesian networks.

As such, the proposed system will empower a patient user by allowing him/her not only to browse through the Encyclopedia for general knowledge about his/her diseases, but also to retrieve relevant Ontology-Based Bayesian networks and apply his/her own data values and get personalized self diagnosis, treatments, preventive recommendations, minimizing the impact of his/her disease, reaction to his/her situation and suggestions for enhancing medical services regarding his/her case. The patient can also compare various treatments as well.

Figure 12 shows as another example the result of using the Lung Cancer ontology extracted from LUCADA (Royal College of Physicians, 2015) international patient dataset. In the figure classes are represented by circles; object properties are represented by edges. We use OWL 2 version in our proposed system. We enrich this ontology by adding the capability of uncertainty through a set of construction rules (see section 2) that transform the ontology into a Bayesian

network that preserves the semantic of the original ontology. Lung Cancer specific variables are selected from the LUCADA ontology. One advantage of the produced Bayesian network is its 'adaptively'; where the probability statistical distributions inspiring the Bayesian network can be mechanically updated to include newly added patients cases. This leads to an evolving ontology-based Bayesian network. Our experimental results indicate that manually constructed 'traditional' Bayesian models don't represent the best models of either actual predictive performance or to fit the data.

Extracting Bayesian Variables from King Abdulla Encyclopedia and the "Unified Medical Records"

In addition to King Abdulla Arabic Medical Encyclopedia, other knowledge sources such as treatment database, unified medical records and clinical drugs can be used to extract more variables into the ontology-based Bayesian net. For example, some data cases can be added as instances to existing classes in the ontology. This can be considered part of what is called "Ontology Evolution" (Royal College of Physicians, 2015).

Free-text fields in the UMDs such as 'Discharge summary', 'Progress notes', 'Complaints of the patient', 'history of the patient' and 'Communication log' are scanned and manually annotated by a domain expert. 992 UMD records are utilized in the implementation phase. Fields such as 'symptoms', 'Disease' and 'indication' are main field to be extracted. These fields are used along with other structured fields in the UMDs. Collected data are put in a table form of all files. Figure 13 shows how this table is converted to a knowledge graph. UMD instances increases the weights gradually to represent the reliability of a connection.

Figure 13, two colors of nodes according to the type of entity, the red nodes represent 'symptom' entities, while the green nodes represent 'Disease' nodes. This graph contains 109 kinds of 'Disease' and 577 different kinds of 'Symptoms'. As a whole, 1030 pieces of knowledge of knowledge in the graph.

King Abdulla Encyclopedia related pages (e.g., to Lung disease) are extracted. Then NLP Arabic processing- POS (Part of Speech) to isolate "Nouns" and "Verbs" separately. Nouns are represented as nodes (variables) in the Bayesian network. Verbs are the labels on the links between the nodes. Captions on images and figures and may be multi-media descriptions are also used and processed. No further processing like stemming or morphology is needed at this stage. Figure 14 shows sample of extracted nouns and verbs and the Bayesian graph produced out of them.

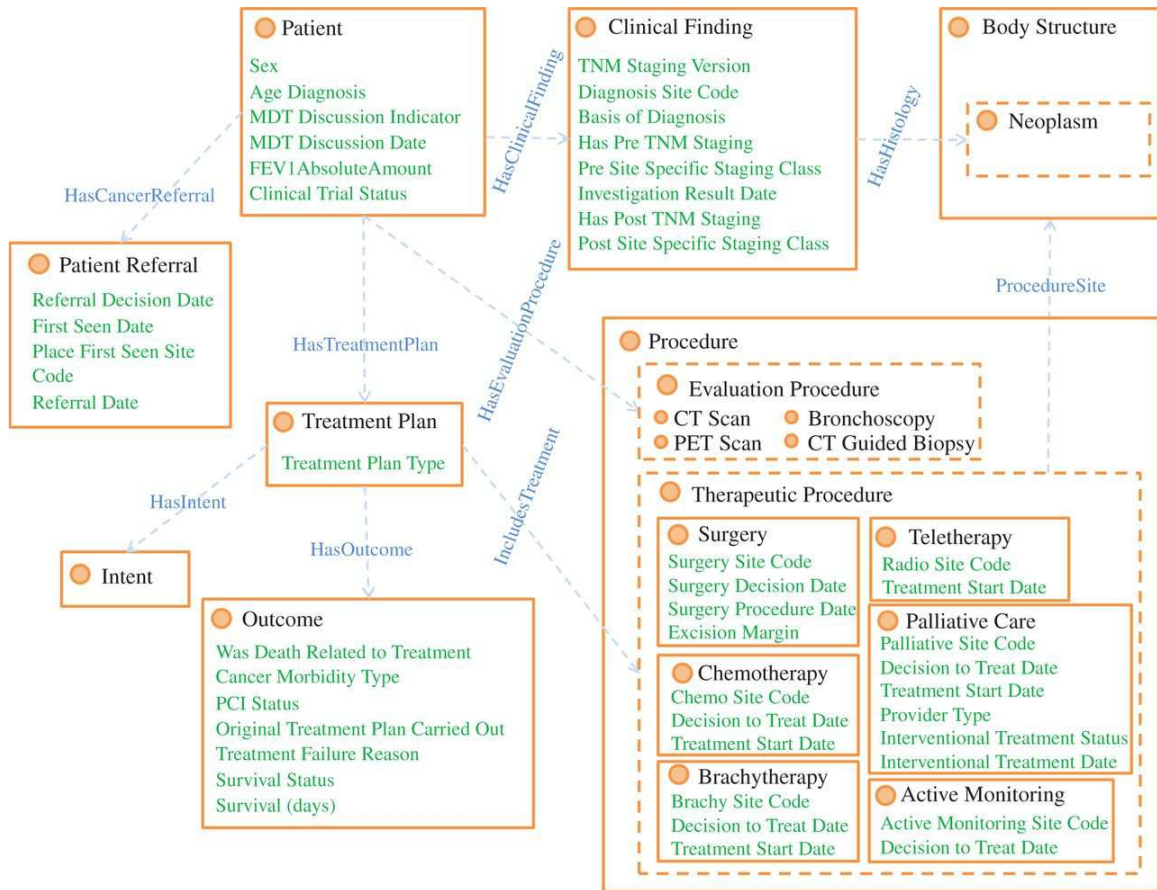


Fig. 12: Lung cancer ontology extracted from the large international lung cancer patient dataset, namely LUCADA (Royal College of Physicians, 2015) with the help of the King Abdulla's Encyclopedia- a screen shot

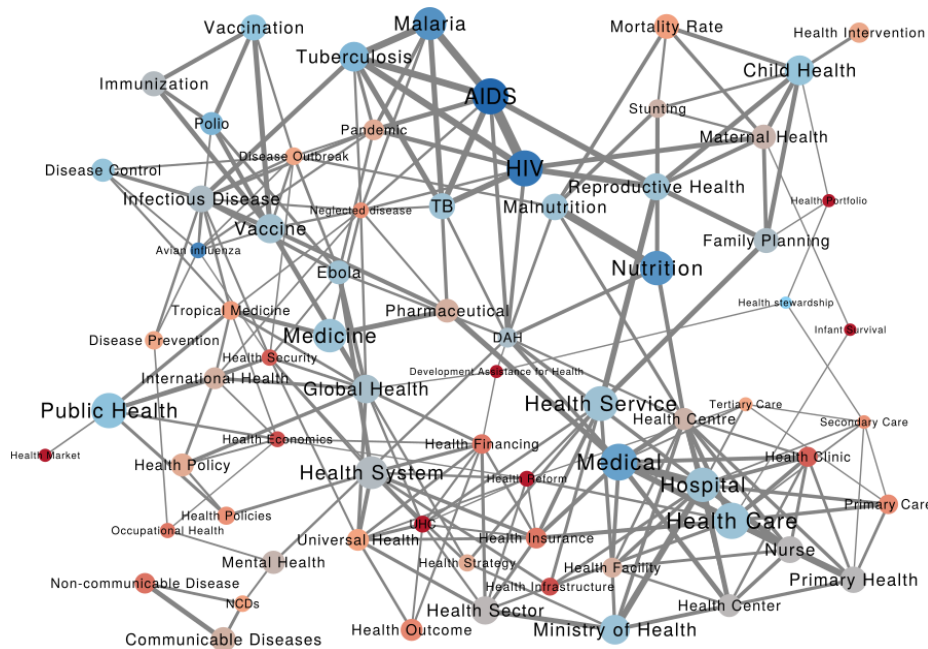


Fig. 13: Topology of the knowledge graph. A screen shot

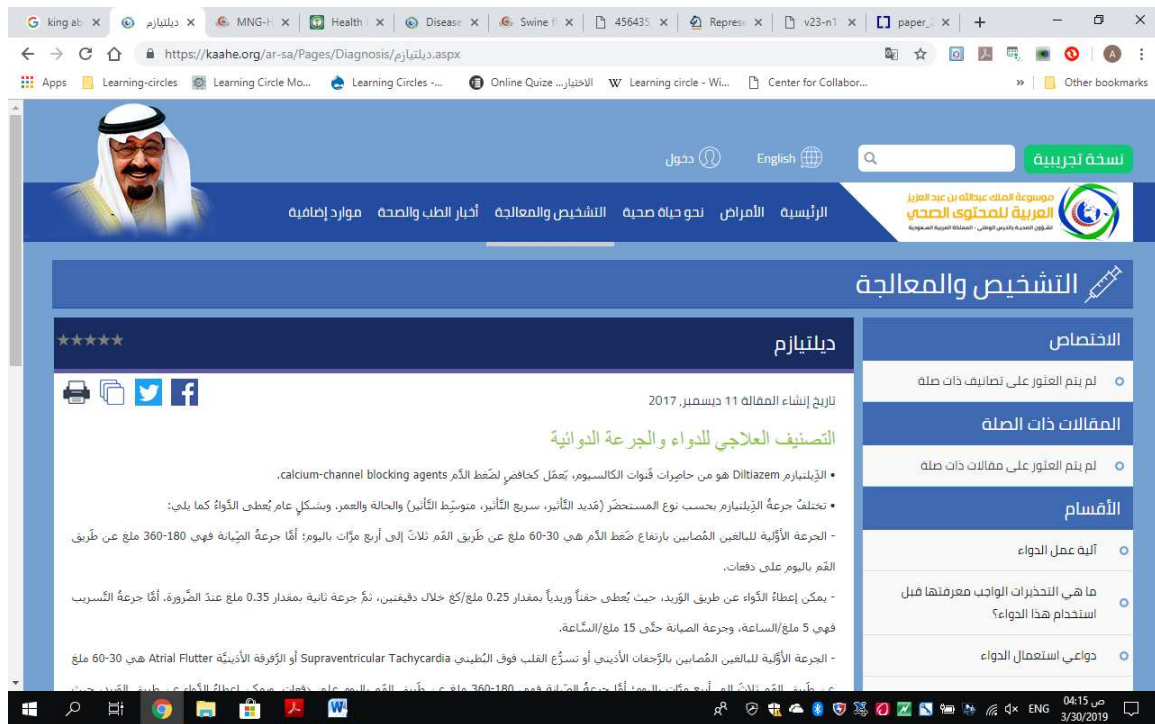


Fig. 14: Extracting "Nouns" and "Verbs" from King Abdulla Encyclopedia- a screen shot

On-the-Fly Construction of Training Datasets from the "Unified Medical Records" (UMDs) Databases

The "Unified Medical Records" (UMDs) are stored as a distributed database. Centralized server/database is primitively expensive since Patients can perform various medical activities at various places in the Kingdom and it is easier then and more efficient to have distributed solution. Centralized solution would require high data traffic. Records can be retrieved by "Patient Name", "Saudi ID", or "Iqama ID" for individuals. Group search for "Disease" type is also possible. For our case of "Lung Disease", we can search with keywords extracted from the Encyclopedia like "Repository Disease", or "Lung Disease". All patients' records with this type of disease are retrieved and extracted from the distributed database.

Patient records are structured and more information can be extracted from each record. All attributes (nouns) values can be extracted and form a complete dataset for both training and testing of the Bayesian network with spread sheet like tables of attributes as columns and patient values as rows. . Usually datasets are divided into 80/20% for training and testing for the Bayesian network processing. The extracted datasets with the variables can form excel formatted dataset that can be used to train known data mining models such as "Classification, Clustering, Association Rules, Regression, etc." This is the subject of another paper (Sameh, 2019) (Fig. 14).

Adopting Integrated Ontology/Bayesian Algorithms

Our approach is to extend the ontology structure by adding probability-specific properties as in (Zheng *et al.*, 2008). In this section we compare our approach to others who have adapted similar ontology/Bayesian integration to expose our advantages.

A newly proposed knowledge engineering methodology that is used to construct BNs from ontologies is presented in (Zheng *et al.*, 2008). A sequence of structured steps of graphical format is used with unambiguous management of the modeling process. Domain experts participate in making modeling decisions. The proposed technique is applied successfully to the domain of 'Lung Cancer'.

Automatic construction of Bayesian Networks from telecommunication ontologies is described in (Ha *et al.*, 2009). Concepts related to BNs are integrated within the ontologies. To create the BN, an instance of each leaf class that inherits from the <root> concept is created. Properties and relations of nodes are embarks on using the constraints offered by ontology axioms.

Ontology for Clinical Practice Guidelines in (Zheng *et al.*, 2008) has been manually developed with impeded uncertainty variables that allow the assembly of conditional probability tables (CPTs). Automatic creation of the corresponding BNs given a domain specific ontology is shown through a number of examples. Successful testing of the proposed system has been demonstrated.

Gregory and Herskovits (1991; Royal College of Physicians, 2015) uses the same approach of constructing BNs directly from existing domain ontologies. They provide algorithms for creating CPTs externally using domain training instances. They provide BNs from the ontologies. User intervention is a necessity especially if the ontology does not completely cover the domain knowledge. Bayesian extensions are provided to existing ontologies. Extensions to the ontology language OWL are provided to inject probabilities into ontology descriptions and to add uncertainties in the field of ontology description. The new language is called 'PR-OWL' (Wang *et al.*, 2014) can provide set of structural transformations to add extensions to legacy ontologies. The transformations are both syntactical as well as semantic. Probabilistic markup extensions are introduced. Several cases are described as examples. The work in (Hu *et al.*, 2017) describes a system for medical diagnosis that integrates ontology with Bayesian networks. Used ontologies are extended by domain experts to add uncertainty reasoning to them. New pathologies are discovered by end users along with their probabilities. A reasoning module validates the consistency of any newly introduced evidence for diagnosis. The OWL ontologies are the bases of the proposed system. The system was tested by domain pathologists. Traditional Chinese Medicine integrated diagnosis system has been implemented in (Ha *et al.*, 2009). Biological ontologies are used to compose Bayesian Networks from such domain ontologies. OWL semantic techniques are used to query and reason with the system. Good performance was established.

Other similar systems include OntoBayes (Wang *et al.*, 2014), Z. Ding work in (Wang *et al.*, 2014) and BayesOWL in (Lori, 2017). They all have used 'PR-OWL' -a probabilistic ontology. They all provide sound formalism for representing uncertainties about entries and relationships that exist in the medical domain.

Experimental System

Now that we have described the theoretical foundation and looked at the current available solutions of integrating ontology and Bayesian networks, it is time to translate this into an actual prototype. A work diagram (experimental Studio) of the implemented prototype can be found in Fig. 15 (my own figure that shows the interaction of tools used in building the prototype). Several tools are used to experiment with the prototype. For our implementation we choose to use OWL to represent our ontology as it is the current standard for storing ontologies. With Protégé editor we can parse, edit and store ontologies from different types of OWL files and other ontology formats. Protégé also offers the option to add our own plug-in written in JAVA. The Protégé plug-in BNGen (Bayesian Network Generator) enabled us to click a Bayesian network together and save it as a Bayesian network 'xdsl file' using JSMILE6 (Gregory and Herskovits, 1991).

We make use of two existing tools in knowledge structuring and visualization; Bayesian Networks using *NETICA* reasoning and ontologies using *Protégé Editor*. While theoretically speaking, both techniques represent knowledge through network associations between nodes/edges, the information that motivates these representations is vastly different. Bayesian Networks in general capture relationships using statistical probabilities, whereas ontologies represent structured parameterized formalization of relationships.

The hierarchical structure between the Bayesian network nodes can be automatically constructed by the program code, which is based on the hierarchical relationship between the entities in the medical ontology. All of the Bayesian network nodes are created in parallel in memory and the parent-child relationship between nodes is created by using synchronization techniques between executing threads, so the child node will wait until all of its parent nodes are created. After the construction algorithm is completed, the conditional probability table will be filled with the default value (this operation is done automatically by *NETICA*).

The structure of a conditional probability table depends on the number of parent nodes. At the time of construction, we use the Noisy-Or algorithm (Sameh *et al.*, 2018) to reduce the number of probability parameters. Noisy-Or can greatly simplify the conditional probability table. And in practical applications it is easier to estimate the conditional probability between any two nodes than to estimate the joint conditional probability of multiple nodes for a node. *NETICA* provides a variety of programming language APIs that can manipulate the Bayesian network and provide a software product for visualizing the Bayesian network.

As in (Ha *et al.*, 2009) we use Protégé editor with BNGen Plug-in that is inspired by BN tab that uses *NETICA* Bayes modeler to convert ontologies into Bayesian networks by taking an original OWL ontology and adding probabilistic information, like conditional probabilities, for the construction of Bayesian networks by annotating this information with additional markups.

BNTab2 (-Bayesian Network Tab) is a Protege3 plugin that translates an OWL ontology into a Bayesian network that can be used in NORSYS NETICA4, a program to work with Bayesian networks. BNTab is used as a result of our practical need to be able to model ontologies and convert them into Bayesian networks.

The BNGen Protégé plug-in is partly inspired by BNTab (Sameh *et al.*, 2018), but only on the visual aspects of the plug-in (Fig. 16). Where BNTab uses *NETICA* as Bayesian modeler we have also chosen to use JSMILE, the JAVA API of SMILE (Zheng *et al.*, 2008).

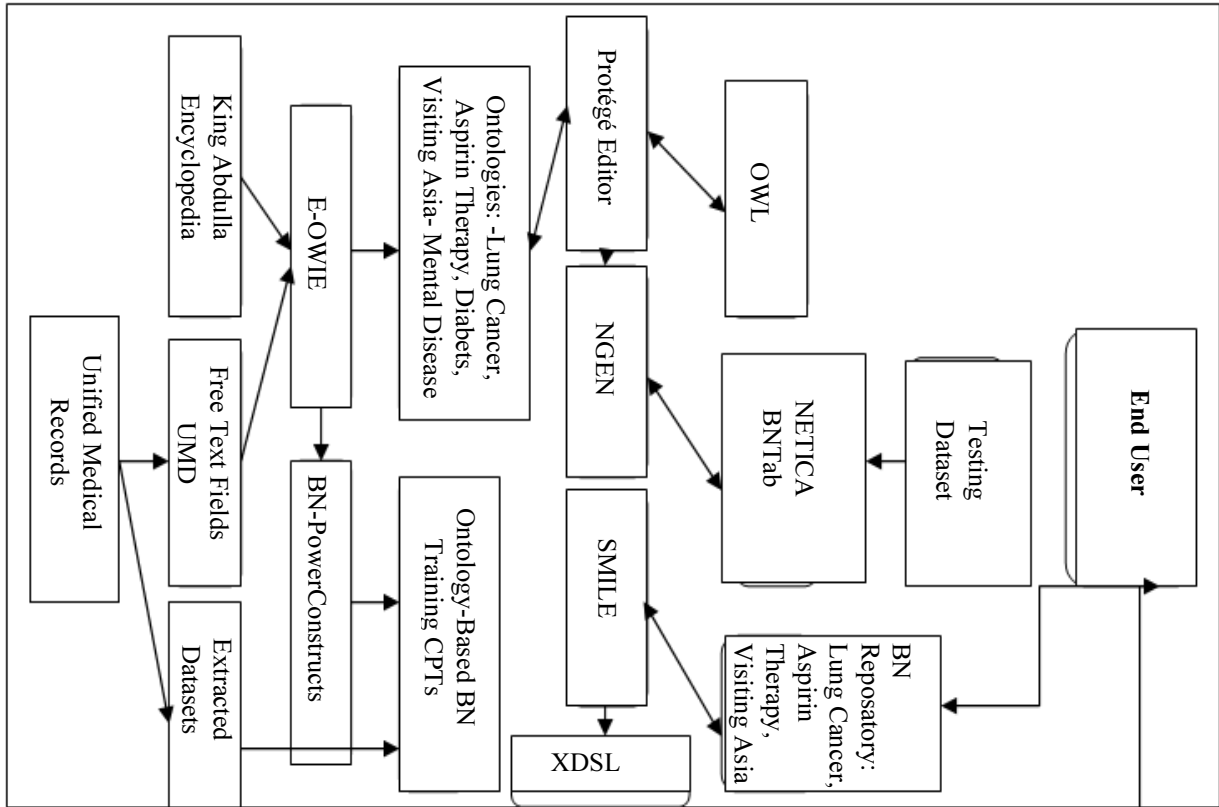


Fig. 15: Work diagram- tools used to build the prototype

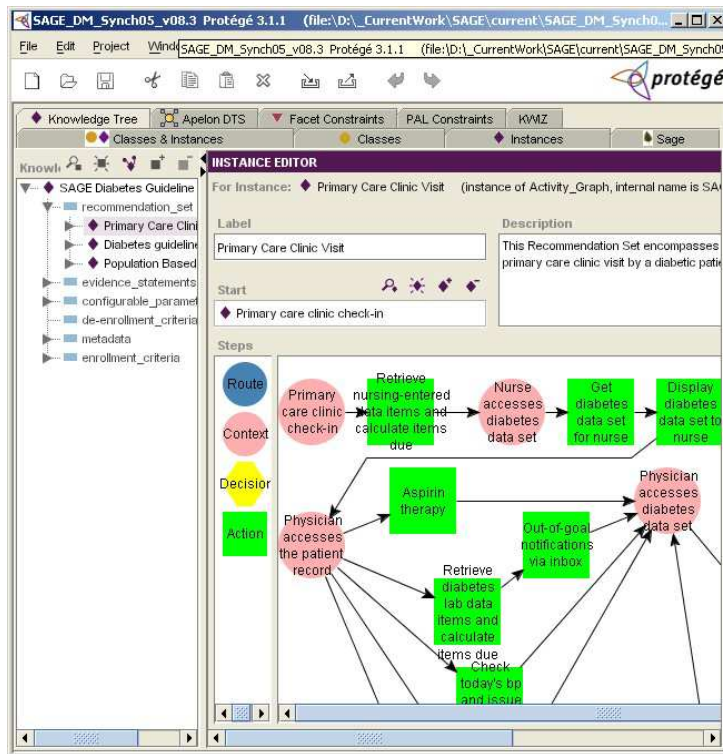


Fig. 16: A screen shot of BNGen using the ASP ontology

SMILE is a free C++ framework and supports multiple types of Bayesian networks, multiple file formats (including the NETICA file format) and has a standalone GUI (Genie 8) (Zheng *et al.*, 2008). Once the user has selected an entity he/she wants to allow BNGen helps the user by only showing relevant parts of the ontology to the user based on the entity hierarchy (taxonomy) and object properties (including part of property).

Compared to BNTab; we limit the amount of information that is provided to the user by not showing irrelevant information. When a variable and a dependency are selected in the interface the user can add the relationship to the Bayesian network by pressing the 'add relationship' button, BNGen will add the relationship to the selected entities list on the right where the variables and their relationships for the Bayesian network are displayed.

The plug-in also provides the option to create reachable by parts in the Bayesian network. These provide a way to link two entities together using an object property of the classifying node. Finally BNGen provides the option to select a state space for the added nodes. If no state space is selected the node will have two default states that SMILE initially provides for each node. Once the user has created the Bayesian network it can be saved to the file system and viewed or edited by the user using editors like Genie (Zheng *et al.*, 2008). Currently BNGen lacks support for data properties and the option to set the probability of the selected states. As such they have to be set in a Bayesian network editor after generating the Bayesian network.

Our classifier is also written in JAVA and supports shape file datasets and the xdsl Bayesian networks as input. The classifier takes care of the connection with the Bayesian networks and deals with the user wishes to specify how and which data is fed to the Bayesian networks. The output is freely specifiable by the user. For testing we are using one of the fastest and most effective rule engine shell (Jess) that calls the NETICA API. It is used to construct the BN tree and apply the up/down propagation algorithm.

Figures 17 and 18 show a part of the ontology-based produced BN that is used to reason under uncertainty about ASP using NETICA API with UMDs records for testing. The data set used is composed of 155 patients UMDs. Test cases are fed manually through the GUI of the system and performance figures such as: the Area Under Curve (AUC), the Receiver Operating perceived Characteristics (ROC), True positive, False negative and other associated measures are all measured as in Table 2. AUC represents the computed percentage of correct diagnosis cases and its value is between 1 and 0. According to tested data-queries, AUC value of was 0.7664.

Table 2: Results of Experiment 4

Performance	Measure Value
TP, FP Positive prediction	0.8659
TN, FN, Negative prediction	0.8533
Sensitivity	0.7313
Specificity	0.7664
Accuracy	0.8761

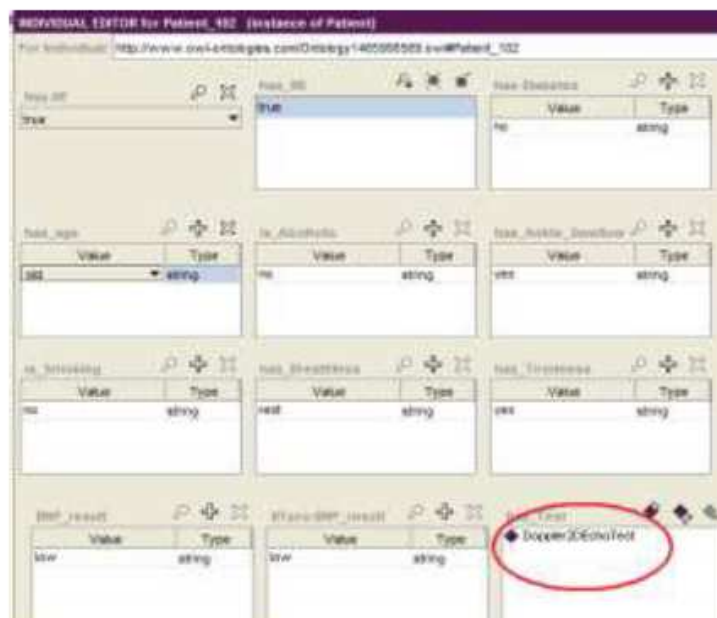
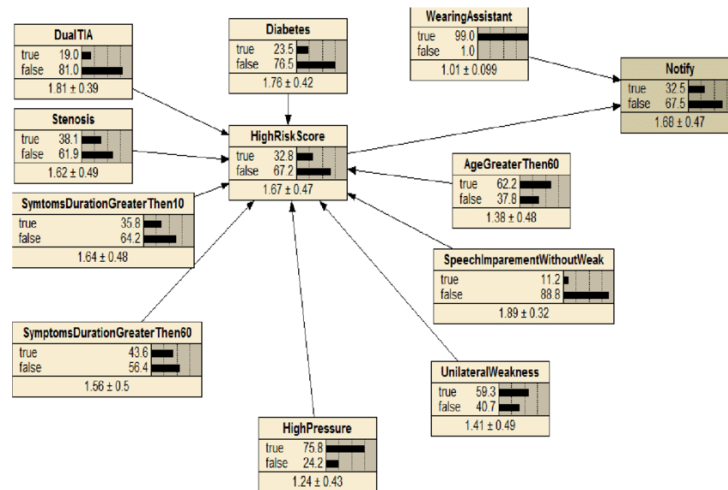


Fig. 17: Shows a screen shot that shows how the inference rules are applied to one of the patients: the shown red circle in the figure represents the recommendation that the patient should make 'Dopper-2D Echocardiography' to test if he/she has 'HF' or not



BNP	TroBNP	Breathless	LiverChinosis	PreviousMI	Doppler2DEcho	age	yes	no
hight	hight	rest	yes	yes	normal	old	55	45
hight	hight	rest	yes	yes	normal	young	45	55
hight	hight	rest	yes	yes	abnormal	old	98	2
hight	hight	rest	yes	yes	abnormal	young	95	5
hight	hight	rest	yes	no	normal	old	40	60
hight	hight	rest	yes	no	normal	young	38	62
hight	hight	rest	yes	no	abnormal	old	98	2
hight	hight	rest	yes	no	abnormal	young	95	5
hight	hight	rest	no	yes	normal	old	55	45
hight	hight	rest	no	yes	normal	young	45	55
hight	hight	rest	no	yes	abnormal	old	98	2
hight	hight	rest	no	yes	abnormal	young	95	5
hight	hight	rest	no	no	normal	old	40	60
hight	hight	rest	no	no	normal	young	38	62
hight	hight	rest	no	no	abnormal	old	98	2

Fig. 18: Part of the BN to reason under uncertainty and part of CPT creating using NETICA API- screen shots

Other performance measures such as TP, FP, TN and FN as reported. Sensitivity is the ability to correctly/effectively diagnose a disease, Specificity is the ability to correctly decide that the patient don't have the disease, Accuracy measure the system statistical bias for both having and not having a disease, PPV is defined as the probability of a patient truly diagnosed to have the disease, NPV is defined as the probability/certainty of a patient not to have the disease. The ratio of these values shows that around 75% of the test dataset were correctly diagnosed.

We have used E-OWIE, which is an enhanced version of OWIE (Zheng *et al.*, 2008). It is used to capture information from unstructured free-text. King Abdulla Encyclopedia and some fields in the UMDs are also text.

We use E-OWIE to extract relevant information from both. Context information are stored in the 'comment section' associated with its original concept. Axioms of the ontology are extracted too. The integration of OWIE resolves many types of conflicts in the extracted data.

E-OWIE uses BN-PowerConstructor to compute the probabilities among the attributes and the structure of the BN. BN-PowerConstructor learns both the structure and the parameters of the BN in addition to watching the constraints. Its graphical interface is shown in Fig. 3. 300+ UMDs are shown to the BN-PowerConstructor during its training phase.

The above tools (see newly created Fig. 6) have been used to implement and demonstrate the proposed prototype. It implements learning of probabilities, network structure and mix of discrete and continuous variables. As another

demo, we demonstrate another case with 60 “Diabetes” – variation diseases and 100 findings. It imports “Diabetes” patient medical record steams from the “Unified Medical Record” server to be used as training set. It also extracts Bayesian data from the “King Abdullah Encyclopedia” server to aid in constructing and learning the Bayesian Diagnosis networks. The prototype is implemented on an Internet server and can be accessed from medical applications on Smartphones and PDAs.

'Asia Case' is another case used for testing the proposed system. Figure 19 shows the BN produced from integrating King Abdulla's Encyclopedia for 'Lung Cancer' segment. Each node represents some condition of the patient. Dependent nodes are attached with CPTs to each other. The CTPs are adjusted through both forward and backward propagations. Thus, node 'smoking' increases the chances of getting node 'Lung Cancer' and getting node 'Bronchitis' and can cause 'Abnormal Lung X-Ray'.

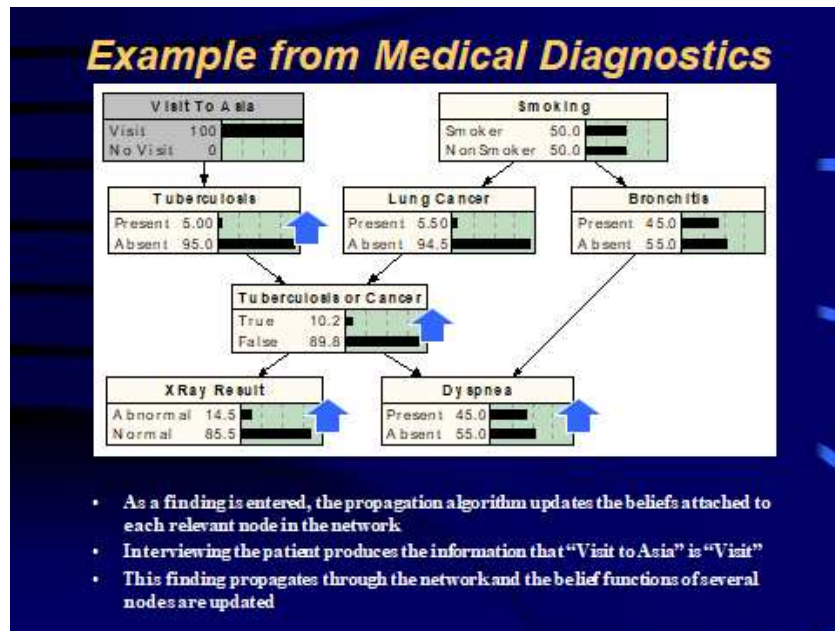


Fig. 19: NETICA screenshot DURING bayesian propagation training, Screen shots

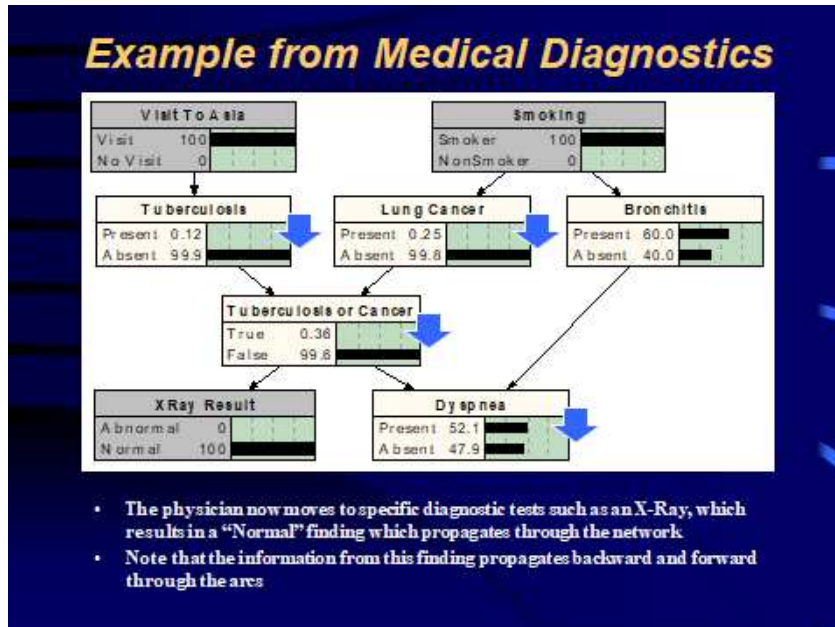


Fig. 20: Bayesian network in testing mode answering a query, screen shots

Figure 19 shows the conditional probabilities tables between the nodes. Figure 18 the training mode propagation algorithm updated the "probability tables". As shown, Relationship type knowledge is modeled by set of certain functions and conditional probability distributions for uncertain functions. The BN algorithm processes the two types of relationship information to provide an unconditional probability distribution for each node random variable in the BN.

Figure 20 demonstrates the training mode. The posed inference problem is to compute the probability of an event given a set of evidence (input values). This can be performed with the repeated sequential solicitation of Bayes theorem to answer the query. The user enters his/her ID in the front page (Fig. 20). His/her medical information gets displayed on the screen (retrieved from the unified medical record). The background of the web site as we see in the figure are "King Abdulla Encpobedia". The user determines that he/she is suffering 'breathing diffulty', 'Dyspnea', 'Xray results', etc. The user then conclues that the evidences indictes 'Bronchiti' and not 'Lung Cancer' nor 'Tuberculosis'.

As more findings and observations are provided by the user such as assigning "true" to node "Visiting Asia". This information propagates through the Bayesian network up and down to propagate this new evidence in the CPTs. The new evidence about "smoking" is set to "true" is followed. Further data collection are provided by the user, then the user starts investing diagnosis test results such as "X-Ray" – set to "normal", "Difficult breathing", "Dyspnea", etc. Variable elimination inference algorithm

then concludes that the user has bronchitis and does not have tuberculosis or lung cancer.

Figure 21 shows how the system empowers patients through: Improving his/her self-diagnosis and testing, interaction with other similar-context patients/specialists, creation of medical social networks and participation in interactive therapy/prognosis e-learning. A "Chat" Zone is in the lower right corner of the above user interface. Users who open accounts to use this system form a "Medical Social Network". These can be "Patients", "Medical Doctors", "Medical staff", "Technicians", etc. Medical social networks a trend that currently is picking up big momentum (Sameh, 2019). The proposed system utilizes cross fertilization of Bayesian data mining techniques to provide intelligent decision aids, data fusion, Arabic text features extraction, automatic free text understanding and the synergy between these. Incremental Bayesian propagation algorithms are also used to provide adjustment and re-training of existing Ontology-based Bayesian networks whenever new cases and recent knowledge need to be incorporated into existing ontology-based Bayesian models. In some cases, dynamic Bayesian networks are used to model disease temporal progress.

Another investigated case study deals with 60 "Diabetes" –variation diseases and 100 findings. It imports "Diabetes" patient medical record steams from the "Unified Medical Record" server to be used as training and testing datasets. It also extracts ontology-based Bayesian data from the "King Abdullah Encyclopedia" server to aid in constructing and learning the ontology-based Bayesian networks. The prototype is implemented on an Internet server and can be accessed from medical applications on Smartphones and PDAs.

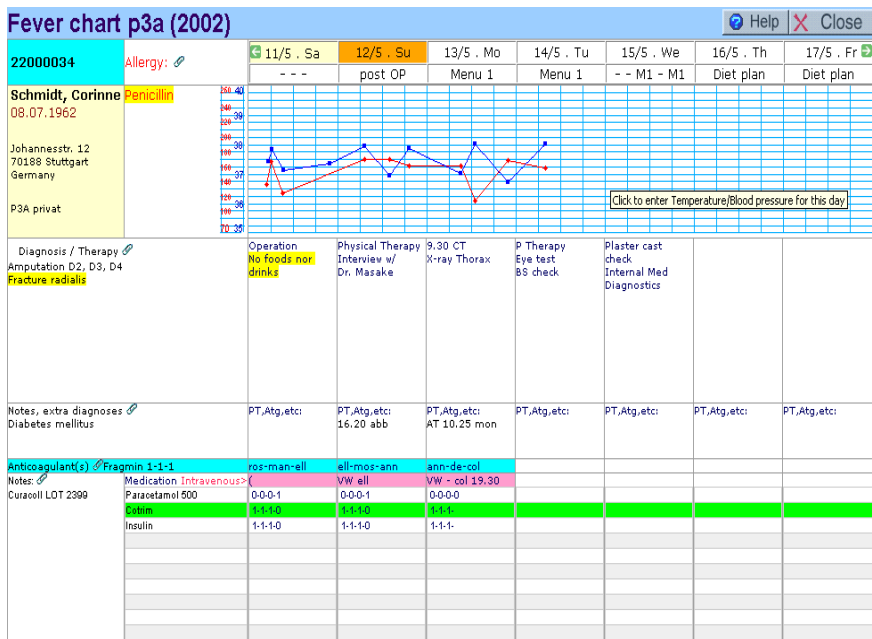


Fig. 21: The end-user login using his/her "National ID" to load its "Medical Record" from the ministry of health on top of the "King Abdulla Encyclopedia". Now he/she can load any ontology-based bayesian diagnosis network and apply his/her case-screen

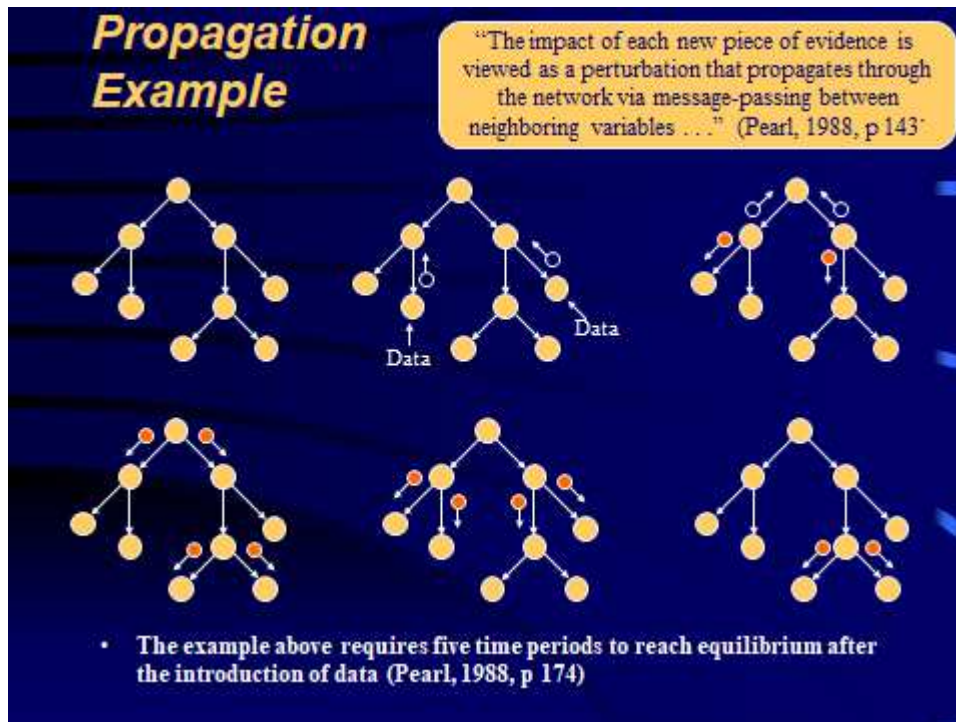


Fig. 22: Bayesian network dynamics to check disease progress over time, a screen shot

Figure 22 shows the propagation of values during the testing of the “Lung Cancer” case Bayesian network. Values are propagated from evidence to sub-conclusions to conclusions. Probability tables are used to compute propagated values. Figure 21 shows how to model temporal relationships in a dynamic Bayesian network. This can be used to model a progress of a disease. In some cases, dynamic Bayesian networks are used to model disease temporal progress. The extracted datasets with the variables can form excel formatted dataset that can be used to train known data mining models such as "Classification, Clustering, Association Rules, Regression, etc." This is the subject of another paper (Sameh, 2019).

In this particular case study, Bayesian network variables and structure are extracted from both the “Unified Medical Records” and the “King Abdulla Encyclopedia” servers. Unconditional and conditional probability tables are computed from the data stream of related cases of “Lunge Cancer” patients coming from the “Unified” server. Both continuous and discrete variables are handled following standard Bayesian rules. Data are divided into both training and verification data. Verification data sets are used to determine when to stop the training phase. As we can see from the figures, domain expert can intervene and modify both the structure and the probability tables computed. The figures show data extracted from both the “Encyclopedia” and the “Unified Records” servers that allowed the construction of the ontology-based Bayesian network. These data describes the dependencies, structure of the network. The figures show

sample extracted training data from the “unified” server that are used to train the ontology-based Bayesian networks.

Conclusion

Traditional Bayesian networks development involves specialized domain experts to establish relevant links and relationships as well as accurate conditional probabilities. Such a process is time consuming and costly. Moreover, produced networks are inconsistent and difficult to update. Although medical decision making makes good use of Bayesian networks, the aforementioned issues still on. This research alleviates this posed challenges surrounding BN modeling by integrating ontology of knowledge in disease which also shows the dependency arcs similar to BN links and relations; implementing a software system that operates on ontological nodes/edges using uncertain reasoning to create Ontology-Based BN topologies. We present that network topologies built using the proposed software exploits and represents medical domain structures to allow compact representations of complex medical models for various diseases that are query-able. Although various approaches could have been adopted (see section 6) to integrate domain knowledge into BN; but the new algorithms in sections 2-5 show how to systematically automate the integration process in such a flexible way that preserves the advantages of both ontology and Bayesian nets. Protégé, Netica,

OWIE, BNGen, BN PowerConstructor and BENTab are some of the tools used to implement the proposed system. The work in this paper is part of a larger research project "RAHA" (Sameh *et al.*, 2018). It's an I-Home healthcare system for KSA Patients that can perform self-diagnosis at home, E-learn and chat with similar patients through a private medical social network. In this paper we have exposed and explained the self-diagnosis part using Ontology-based Bayesian networks. We showed how to extract training data from the "Unified Medical Records- UMDs" distributed databases. We also showed how to build, train and maintain the Bayesian diagnosis networks using both propagation and incremental fusion algorithms. Finally, a small scale prototype was tested and demonstrated. Our ultimate goal is to setup and deploy "RAHA" hoping it will give positive boost to current Saudi Healthcare system.

Acknowledgement

The author would like to recognize RIC and PSU for their support. The Ministry of health Medical Records (UMRs) used in this project were provided with permission to the University. We would like to thank the reviewers for their detailed reviews and insightful comments, which have helped to improve the quality of this paper.

Ethics

The author testifies that this article is original and contains unpublished material. All ethical standards that ensure scholarly integrity have been followed.

References

- Eunjeong, P., D. Seungwook, H. Shin and H.S. Nam, 2009. An adaptive streaming technique for interactive medical systems in mobile environment. Proceedings of the 6th International Conference on Mobile Technology, Application and Systems. Sep. 02-04, IEEE Xplore press, New York, NY, USA. DOI: 10.1145/1710035.1710066
- Gregory, F.C. and E. Herskovits, 1991. A bayesian method for constructing bayesian belief networks from databases. Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence, Jul. 13-15, ACM, Los Angeles, CA, USA, pp: 86-94.
- Giovanni, C., S. Qiulings, H.G. Pietro, T. Giuseppe and Z. Aidong *et al.*, 2016. Feature selection model for diagnosis, electronic medical records and geographical data correlation. Proceedings of the 7th International Conference on Bioinformatics, Computational Biology and Health Informatics, DOI: 10.1145/2975167.2985847
- Ha, S., S. Kim, M. Suh, H. Seong and K.M. Jeong *et al.*, 2009. A statistical technique for structure learning of Bayesian networks from data. ICHIT '09: Proceedings of the 2009 International Conference on Hybrid Information Technology.
- Hu, H. A.H. Elrafey and L. Kerschberg, 2017. Using Modular Ontologies to Capture Causal Knowledge contained in Bayesian Networks. ASONAM '17: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, Aug. 3-5, IEEE Xplore Press, Sydney, Australia, pp: 904-907. DOI: 10.1145/3110025.3110136
- Kabli, R., J. McCall, F. Herrmann and E. Ong, 2008. Evolved bayesian networks as a versatile alternative to partin tables for prostate cancer management. GECCO '08: Proceedings of the 10th annual conference on Genetic and evolutionary computation, Jul. 12-16, IEEE Xplore Press, Atlanta, GA, USA, pp: 1547-1554. DOI: 10.1145/1389095.1389392
- Linda, M. and J. Akaichi, 2017. DS-ontology: A disease-symptom ontology for general diagnosis enhancement. ICISDM '17: Proceedings of the 2017 International Conference on Information System and Data Mining, Apr. 01-03, IEEE Xplore press, New York, NY, USA, 99-102. DOI: 10.1145/3077584.3077586
- Lori, H., 2017. Mental disease bayesian repository. Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017
- Melchior, A., 2013. Data Enrichment of Spatial Database Using Ontologies and Bayesian Networks. Master Thesis from University Utrecht.
- MHUMR, 2016. Unified Electronic Medical Record, Ministry of Health's, Unified Medical Record. Saudi Arabia.
- Royal College of Physicians, 2015. National Lung Cancer Audit. <https://www.rcplondon.ac.uk/projects/national-lung-cancer-audit>
- Sameh, A., J. Jamai and A. Zarrad, 2018. Empowering Saudi's Healthcare System with an Integrated On-Line Web-Based I-Home Healthcare Platform (RAHA). A research project proposal submitted to KACST.
- Sameh, A., 2019. Medical Social Network, A working research paper
- The National Guard's, 2018. King Abdullah Encyclopedia of Arabic Health Content. <https://www.kaahe.org/en>
- Wang, X. D. Sontag and F. Wang, 2014. Unsupervised learning of disease progression models. KDD '14: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, Aug. 24-27, IEEE Xplore Press, New York, New York, USA, pp: 85-94. DOI: 10.1145/2623330.2623754

Yousefi, M.M.R. and L.A. Dalton, 2014. Optimal cancer prognosis under network uncertainty. Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology and Health Informatics, (BCB '14), Sep. 20-23, IEEE Xplore Press, Newport Beach, California, pp: 682-683.
DOI: 10.1145/2649387.2660827

Zheng, H.T., B.Y. Kang and H.G. Kim, 2008. An Ontology-Based Bayesian Network Approach for Representing Uncertainty in Clinical Practice Guidelines. In: Uncertainty Reasoning for the Semantic Web I. URSW 2006, da Costa P.C.G. (Ed.), Springer, Berlin, Heidelberg.