IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Use of Artificial Intelligence and Machine Learning in Medicines with implementation of Bayesian techniques

Anubhav Kumar[1], Virat Sharma[2], Praveen Verma[3], Dr. Abhay Bhatia[4], Dr. Manish Kumar[5]

[1, 2] B.Tech CSE 4th year, Inderprastha Engineering colleges, Ghaziabad , Uttar Pradesh, India

[3, 4, 5] Dept. of CSE, Roorkee Institute of Technology, Roorkee, Uttarakhand, India

Abstract

The interest in artificial intelligence in the medical sciences has increased over the last two decades, but most studies of its applications in clinical studies have drawn criticism for having unreliable designs and poor replicability, necessitating a greater need for medical professionals to be knowledgeable about this quickly expanding field of research. The area of Bayesian artificial intelligence is briefly introduced in this article. We talk about causal inference, Bayesian networks, and their (potential) applications in clinical practice.

Keywords: medicinal, undirected graphs, machine learning, statistics

1 Introduction

The work of Alan Turing in 1950 can be credited with giving rise to artificial intelligence (AI). He developed a test, now known as the Turing Test, in which a machine must possess a specific level of intelligence to trick a human into believing he is speaking with another human (the Imitation Game). A straightforward interaction with a virtual assistant like Siri provides as a vivid illustration of how quickly AI research has advanced over the past few decades, despite the fact that such a degree of intelligence has not yet been achieved.

In the recent two decades, AI has gained attention in the medical sciences, with an increasing number of its applications receiving worldwide approval from health authorities. But the majority of doctors and medical researchers lack formal training in the fields where the field of smart medical monitoring, diagnosis, and follow-up has its roots. This is one of the key barriers to broad physician and engineer collaboration in the creation of AI software, along with historical philosophical disparities in the practise of medical and engineering professions. It might also be argued that this is a contributing factor to the ongoing replicability issue in medical AI research2, which is characterised by faulty study designs and subpar replication3.

Few works attempt to pique doctors' interest in the foundational ideas and jargon of artificial intelligence, despite the fact that there is an expanding body of research on how AI can be applied in clinical practise. They are familiar with notions from probability and statistics like correlation, regression, and confidence intervals because to the progressive shift towards quantitative methods over the past century; it is worthwhile to elaborate on these ideas and connect them to contemporary AI.

In the early 2000s, neural networks and clustering were the two AI techniques that were most frequently used in the medical literature (in the last decade). Bayesian reasoning and AI techniques are less well known, despite their potential applications in human physiology research, symptom interaction analysis, and symptom recovery research. Given that we cannot create a complete mechanistic model of diseases and the physiological pathways they affect, Bayesian models can help with reasoning in probabilistic terms when dealing with uncertainty.

This work aims is mainly just to introduce physicians to Bayesian AI through a perspective (basically clinically) onprobabilistic reasoning and Bayesian networks.

2 Probabilistic reasoning

Using information about circumstances that might be related to an event that has already occurred, Bayes' theorem estimates the probability of that occurring: It is represented using the widely used mathematical notation

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)},$$

the probability $\Pr(A \mid B)$ of an event A given prior knowledge of a condition B (that is, Bhas occurred). $\Pr(A \mid B)$ is a conditional probability, while $\Pr(A)$ and $\Pr(B)$ are known asmarginal probabilities, that is, the probabilities of observing the events A and B individually.In the context of medical diagnosis, the goal is to determine the probability $\Pr(D_i \mid C_p)$ of presence of a particular disease or disorder D_i given the clinical presentation of the patientC_p [11], the prior probabilities P_i^J of the disease in the patient's reference group, and the prior probabilities P_j^J of other diseases D_j; that is

$$\Pr(D_i \mid C_p) = \frac{P_i' \Pr(C_p \mid D_i)}{P_i' \Pr(C_p \mid D_i) + \sum_j P_j' \Pr(C_p \mid D_j)}$$

where $\Pr(C_p \mid D_j)$ is the probability of having the same clinical presentation given otherdiseases.

Bayes' theorem makes it possible to work with the distributions of dependent (or condi-tionally dependent) variables. However, in order to reduce the number of variables we need to observe simultaneously in probabilistic systems, it is also important to determine whethertwo variables A and B are *independent* ($A \perp B$),

$$\Pr(A \mid B) = \Pr(A),$$

or *conditionally independent* ($A \perp B \mid C$) given the value of a third variable C,

$$\Pr(A \cap B \mid C) = \Pr(A \mid C)\Pr(B \mid C).$$

Extracting conditional dependence tables is one of the building blocks upon which we can build Bayesian probabilistic reasoning. It allows to take a set of variables (say, A, B and C again) and to compute the conditional probabilities of some of them (say, $A \mid C$),

$$\Pr(A = y \mid C = x) = \frac{\sum_{B \in \{x,y\}} \Pr(C = x, B, A = y)}{\sum_{B,A \in \{x,y\}} \Pr(C = x, B, A)}.$$

Conversely, we can also take variables or set of variables that are (conditionally) independent from each other and combine them to obtain their *joint probability*. This joint probability will be structured as a larger conditional dependence table that is the product of smaller tables associated with the original variables. For instance,

$$\Pr(A = y, B = z \mid C = x) = \Pr(A = y \mid B = z)\Pr(B = z \mid C = x)$$

assuming $A \perp C$. The ability of explicitly merging and splitting set of variables to separate variables of interest we need for diagnostic purposes from redundant variables is one of the reasons that makes Bayesian reasoning easy while at the same time mathematically rigorous.

3 Machine learning and Bayesian thinking

The branch of AI known as machine learning (ML) explores the algorithms and statistical techniques that enable computers to carry out well-defined tasks without explicit instructions. Four elements are necessary for machine learning implementation.

First, we need a working world model that can be understood by computers and explains the tasks and their environment. In practice, this entails selecting a class of Bayesian models defined over the relevant exogenous variables and the variables of interest and putting those models into software. The joint probability is represented by generative models, such as Bayesian networks, which explain how variables interact with one another.

$\Pr(X_1, \ldots, X_N)$;

while discriminative models such as random forests and neural networks only focus on how a group of variables predicts a target variable by estimating

$$\Pr(X_1 | X_2, \ldots, X_N).$$

Clearly, a generative model may be a better option than a discriminative model depending on the application, or vice versa. Generative models should be chosen when describing the phenomenon we are modelling from a systems perspective. Discrimination models offer superior prediction accuracy at the expense of being less expressive if we are merely interested in forecasting some clinical occurrence, like in the case of diagnostic devices.

The ability to forecast brand-new events is typically how we gauge the model's success.

The third stage, known as learning, is encoding the knowledge about the world into the model via training data, experts, or both.

therefore, the computer system

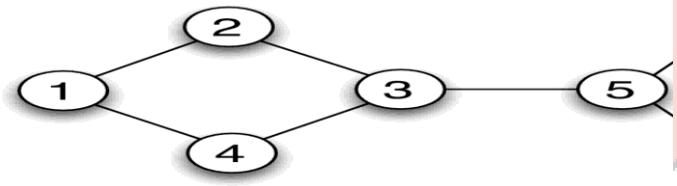

Figure 1: An undirected network of seven nodes connected through edges.

discover the model that maximizes the selected performance metric inside the required class, using either observational or experimental data or professional knowledge that is available from practitioners. Fourth, the computer decides whether and how to carry out the prescribed task while using the model as a stand-in for reality and making inferences as fresh inputs are received.

By putting machine learning applications into practice successfully is far from simple: it needs a lot of data, and choosing how to organize the model from a probabilistic and mathematical standpoint is challenging. The importance of ethical software engineering cannot be understated. Because of these factors, machine learning models should have a small number of variables to make them simple to build and understand.

In addition, unlike cutting-edge neural networks, compact models are unlikely to need a lot of computer capacity to learn.

Probability should be utilized to determine if two variables are connected given other factors because clinical settings limit our knowledge of the patients to what we can learn from them. Formally, we say two variables are connected with or probabilistically reliant on one another if the likelihood of an event occurring in one variable impacts the likelihood of an event occurring in the other. Probability-symmetric association does not distinguish between causes and effects on its own. But Bayesian network models go beyond probability theory to depict causal relationships as arcs in a graph, enabling rigorous causal inference.

4 Bayesian Networks

We have introduced in this section about the fundamental notions of Bayesian networks. Specialized textbooks have given more elaborated review on this subject[9], and the software to implement them is completely available from the **bnlearn** package[14] for statistical environment the R.

4.1 The use of graphs to represent interactions among entities

The study of graphs is a topic in mathematics known as graph theory. These structures are used in the medical sciences to express interactions between elements such as symptoms, signs, or biological markers.

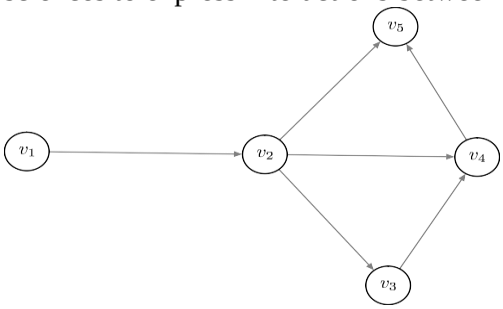

Figure 2: A Bayesian network, or any Directed Acyclic Graph consisting of five nodes. These Edges are directed from one node to another node.

A graph G is understood as a set V of *nodes* (also known as *vertices*) representing variables (or other feature of the data) that are connected through a set A of *edges* (also known as *arcs*). Let us consider a network of five nodes, such as that shown in Figure 1. In this case, the set of nodes comprises and the set of edges

$$\mathbf{V} = \{v_1, v_2, v_3, v_4, v_5\},$$

$$\mathbf{A} = \{a_{12}, a_{14}, a_{23}, a_{24}, a_{25}, a_{34}\},$$

where a_{12} represents the edges between node v_1 and node v_2. The network represented in Figure 1 is termed as an *undirected* network, as the edges are not directed in any of the particular direction, so therefore for the above said graph $(v_i, v_j) = (v_j, v_i)$. Undirected networks are commonly used for representing the pairwise interactions among some psychopathological symptoms[6]. The edges are unweighted, so that any of the edge can either be present by $a_{ij} = 1$ or if absent $a_{ij} = 0$ between any of the two nodes; or they can be weighted, so that some edges can be stronger than the others in any considered network, and can have either a positive or negative sign. For such instance, a weighted edge can represent a partial correlation estimate[15] so that conveying the existence of a conditional association relationship between any two variables can be defined.

4.2 Directed Acyclic Graphs

Contrarily, directed acyclic graphs serve as the foundation for Bayesian networks (DAGs). Due to the fact that a DAG only has directed edges, (vi, vj) /= (vj, vi), as the former is vi vj and the latter is vj vi. Figure 2 displays a case in point.

The head of the arrow represents the consequence, and the tail node represents the cause, in causal links represented by these arcs. Loops and cycles—the result of a node acting on itself—cannot exist in Bayesian networks (for any of the instance, A goes to B, B goes to C, and C goes to A).

A Bayesian network's main objective is where to express the conditional independence that is set of relationships between variables (that is, variables that do not predict each other). In addition to the DAG, Bayesian networks are well defined by the global probability for distribution of $\mathbf{X}$ (with X_i being the variable that corresponds to the node v_i in the network) with parameters $\mathbf{\Theta}$,

$$\Pr(\mathbf{X}, \mathbf{\Theta}) = \prod_{i=1}^{N} \Pr(X_i \mid \Pi_{X_i}; \Theta_{X_i})$$

where Π_{X_i} represent the parent nodes of X_i. This factorization derives from the *Markov property* of Bayesian networks, that is, every variable X_i depends on its parents Π_X [10].

The three most of the common probability distributions for Bayesian networks are *Gaussian*, *discrete* and *conditional linear Gaussian* [9]. Discrete Bayesian networks have, for instance, that have been used in expert systems for differentiating between diseases likewise tuberculosis and lung cancer[16]. Gaussian Bayesian

networks are common is system biology and genetics and are responsible for reconstructing of direct and indirect gene effects [17], and at the last with both conditional Gaussian Bayesian networks they havebeen used to study various clinical treatments and conditions [7;18].

5 Structure learning of Bayesian networks

In this among section we will have introduced and will learn the concepts of graphical separation and probabilistic independence.

5.1 The *Markov* property

In Bayesian networks, if any of the two nodes are unconnected (i.e, no edge is shared), thatmeans that they are also conditionally independent found: this is termed as the *Markov* property[10]. The graphical separation is used to imply the probabilistic independence among them,

$$A \perp_G B \,|\, C \Longrightarrow A \perp_P B \,|\, C.$$

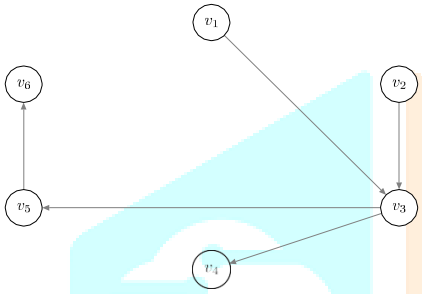

Figure 3: A Bayesian network consist of six nodes for illustration of graphical separation (that means that two nodes are not connected in any of the particular network). For any instance, 1 is separated from 4 and 5 is through 3; 2 is separated from 4 and 5 through 3, and 3 is separated from 6 through 5.

Making of such network by itself it is very clear representing the conditional independence relation-ships between the nodes. For this such reason, the DAG is termed as an *independence map* of the various variables.

$$\Pr(\mathbf{X}, \boldsymbol{\Theta}) = \prod_{i=1}^{N} \Pr\left(X_i \,|\, \Pi_{X\boldsymbol{i}} ; \Theta_{X\boldsymbol{i}}\right),$$

The Markov property makes it possible to write decomposing the larger model Pr($\mathbf{X}$, $\boldsymbol{\Theta}$) into a set of smaller models Pr $(X_i \,|\, \Pi_{X\boldsymbol{i}} ; \Theta_{X\boldsymbol{i}})$ that are easier to understand. The possibility of such decompositions is only when the absence of loops and cycles is there in that graphs. Figure 3 represents a Bayesian network with six nodes. Two nodes, say v_1 and v_4, are graphically separated by node v_3, and are therefore conditionally independent given node v_3:

$$v_1 \perp_G v_4 \,|\, v_3 \Longrightarrow \Pr(v_1, v_4 \,|\, v_3) = \Pr(v_1 \,|\, v_3) \Pr(v_4 \,|\, v_3).$$

Figure 3 also shows a specific kind of relationship in a Bayesian network, that is the one among nodes v_1, v_2 and v_3. Both v_1 and v_2 have an edge pointing to v_3 and the two shareno connection. This type of motif, sometimes referred to as a v-structure or a collider, is frequently regarded as one of the fundamental components of Bayesian networks. Contrary to popular belief, the two causes in a collider are known to be negatively linked. As a result, conditioning on the common effect in the collider (that is, examining relationships while modifying the effect) yields different estimates than examining the two causes separately. This phenomenon, often referred to as Berkson's bias or collider bias, is a significant source of prejudice in the medical sciences[22].

5.2 Blankets by Markov

The probability distribution of the variables Xi is thought of as being represented by a DAG, which may be retrieved by determining which nodes are conditionally (in)dependent. To assess whether two nodes in a network are (in)dependent or conditionally (in)dependent using an algorithm, d- separation is a valuable tool. If conditioning on all members of S prevents all pathways (sequences of nodes and edges with A as the starting node and B as the ending node) between A and B, then two nodes A and B are d-separated by S. A collider is known to block all overlapping paths. The Markov blanket of node A in the graph G is known as the set S. The Markov blanket is by definition made up of a node's parents (nodes with edges directed towards A), children (nodes that get edges directed from A), and spouses, or the other parents of the children. All nodes outside of the Markov blanket are independent of the node of interest, making it possible to investigate a target node of interest while ignoring the rest of the Bayesian network.

5.3 Bayesian networks and Causality

The links between variables in Bayesian networks can be simply understood as causal interactions because they are built on DAGs. However, before considering an edge as a causal impact, three presumptions should be made. First, given its direct causes, each variable (node) must be conditionally independent of both its indirect and direct non-effects (this is the causal translation of the Markov property).

Second, a DAG must accurately represent the probability distribution of X such that the only dependencies in the probability distribution are those resulting from d-separations in the DAG. There must be no latent variables that serve as confounding factors, which is the third premise that follows from the first two (therefore developing causal effects on one or several nodes in the network without the DAG reporting such relationships).

The third presumption is crucial in clinical settings because it states that any confounding variables should be eliminated from the experimental design in order to safely interpret a directed connection as a causal effect.

Randomization is a frequent method to accomplish this since it breaks any existing causal link between the randomised variables and any external effects. The probabilistic and causal interpretations of Bayesian networks must be distinguished from one another. From a causal standpoint, the asymmetry between cause and effect—if we act on the cause, we may influence the consequence, but if we act on the effect, the cause is left unaffected—is what distinguishes arcs from other shapes. Given the reversibility of Bayes' theorem, this is false from a probabilistic standpoint. For instance, if we consider again the DAG in Figure 2 we can write

$$\Pr(v_1, v_2, v_3, v_4, v_5) = \Pr(v_1)\Pr(v_2 \mid v_1)\Pr(v_3 \mid v_2)\Pr(v_4 \mid v_2, v_3)\Pr(v_5 \mid v_2, v_4)$$

where each node has a distribution conditional on its parents. However, for the nodes v_1 and v_2 we have that $\Pr(v_1)\Pr(v_2 \mid v_1) = \Pr(v_2)\Pr(v_1 \mid v_2)$. This implies that the DAG in which the arc $v_1 \rightarrow v_2$ is reversed into $v_2 \rightarrow v_1$ encodes the same probability distribution as that in Figure 2, despite having different arcs. Only arcs that are a part of a collider, or those that would form a new collider or initiate a cycle if their orientation were reversed, can be modified in this fashion, making them uniquely identifiable even without applying causal assumptions.

Another result of the duality between the probabilistic and causal interpretation of Bayesian networks is that we may compute the conditional probability of any pair of variables regardless of how we construct the DAG. A diagnostic DAG with arcs going from symptoms to diseases or a prognostic DAG with arcs pointing from diseases to symptoms may make more sense depending on the application. For each diagnostic DAG that displays the same probability distribution, a prognosis DAG exists, and vice versa. From a purely statistical perspective, this is significant to note. Given that the DAG is easier to read, it follows that one will be easier to grasp than the other. However, any conditional probability that we may want to determine will be the same.

6 Discussion: applications and limitations of Bayesian Artificial Intelligence in Medicine

Doctors who practice in a number of medical disciplines can benefit from using Bayesian AI and networks. Applications of Bayesian AI in medicine can be categorized into four main categories: diagnostic reasoning, which entails giving a target patient a diagnosis based on clinical evidence; prognostic reasoning; making predictions about the future; treatment selection; and studying functional interactions between clinical evidence, such as Several examples from the four main domains described above illustrate the Bayesian AI's enormous potential.

The first step in drawing conclusions is to gather clinical information from the electronic medical record. Systems to build clinical Bayesian networks using electronic medical records have been developed [26]. Additionally, prognostic Bayesian networks are utilized to forecast patient mortality [27].

Third, in challenging circumstances [28], Bayesian networks are also utilized for clinical decision assistance and treatment selection.

Fourth, studying functional interactions between symptoms in the field of psychiatry holds great promise. The classification of mental disorders is undergoing a paradigm shift, and the new perspective of mental disorders as networks of mutually influencing components [29] offers a promising environment for the use of Bayesian reasoning. There have already been attempts to depict the interplay of symptoms for illnesses such Identifying depression as DAGs in cross-sectional data and recovering potential causal connections between them [6]. Future research in this field might, for instance, incorporate various variables in networks (other than symptoms). The presumptions necessary to accurately learn and perform inference on the structure of Bayesian networks restrict their applicability. Researcher-designed studies must take this into consideration: removing confounding variables is by far the most challenging undertaking in this regard[6].

Finally, Bayesian artificial intelligence captures uncertainty in medical reasoning through the promising Bayesian network model. They can be automatically learned from data and rigorously combine graphs and probabilities, with algorithms that automate reasoning and use the graphical part of the model to guide a computer system in computing probabilities and predicting events of interest.

References

[1] Turing AM. Computing machinery and intelligence. Mind. 1950;59(236):433.

[2] Briganti G, Moine OL. Artificial intelligence in medicine: Today and tomorrow. Frontiers in Medicine. 2020;7:27.

[3] Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. The Lancet Digital Health. 2019;1(6):e271–e297.

[4] Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minimally Invasive Therapy & Allied Technologies. 2019;28(2):73–81.

[5] Lucas PJF, Van der Gaag LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. Artificial intelligence in medicine. 2004;30(3):201–214.

[6] Briganti G, Scutari M, Linkowski P. Network structures of symptoms from the Zung Depression Scale. Psychological Reports. 2020:0033294120942116.

[7] Liew BXW, Scutari M, Peolsson A, Peterson G, Ludvigsson ML, Falla D. Investigating the Causal Mechanisms of Symptom Recovery in Chronic Whiplash Associated Disorders using Bayesian Networks. The Clinical Journal of Pain. 2019;35(8):647–655.

[8] Pearl J, Russell S. Bayesian networks. UCLA Cognitive Systems Laboratory; 2011. R-277.

[9] Scutari M, Denis JB. Bayesian Networks: With Examples in R. Chapman & Hall; 2015.

[10] Korb KB, Nicholson AE. Bayesian artificial intelligence. 2nd ed. CRC press; 2010.

[11] Miettinen OS, Caro JJ. Foundations of medical diagnosis: what actually are the param- eters involved in Bayes' theorem? Statistics in medicine. 1994;13(3):201–209.

[12] Murphy KP. Machine Learning: A Probabilistic Perspective. The MIT Press; 2012.

[13] Beam AL, Manrai AK, Ghassemi M. Challenges to the Reproducibility of Ma- chine Learning Models in Heath Care. Journal of the American Medical Association. 2020;323(4):305–306.

[14] Scutari M. Learning Bayesian Networks with the bnlearn R package. Journal of Statis- tical Software. 2010;35(3):1–22.

[15] Briganti G, Chantal C, Braun S, Fried EI, Linkowski P. Network analysis of empathy items from the interpersonal reactivity index in 1973 young adults. Psychiatry Research. 2018;265:87–92.

[16] Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society: Series B (Methodological). 1988;50(2):157–194.

[17] Kruijer W, Behrouzi P, Bustos-Korts D, Rodríguez-Álvarez MX, Mahmoudi SM, Yandell B, et al. Reconstruction of networks with direct and indirect genetic effects. Genetics. 2020;214(4):781–807.

[18] Dao MC, Everard A, Aron-Wisnewsky J, Sokolovska N, Prifti E, Verger EO, et al. Akkermansia

muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. Gut. 2016;65(3):426– 436.

[19] Correa M, Bielza C, Pamies-Teixeira J. Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. Expert systems with applications. 2009;36(3):7270–7279.

[20] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery. 2019;9:e1312.

[21] Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biometrics Bulletin. 1946;2(3):47–53.

[22] Geiger D, Verma T, Pearl J. d-separation: From theorems to algorithms. In: Machine Intelligence and Pattern Recognition. vol. 10. Elsevier; 1990. p. 139–148.

[23] Pearl J, Verma T. A theory of inferred causation. In: Studies in Logic and the Founda- tions of Mathematics. vol. 134. Elsevier; 1995. p. 789–811.

[24] Spirtes P, Glymour C, Scheines R. Discovery Algorithms for Causally Sufficient Struc- tures. In: Spirtes P, Glymour C, Scheines R, editors. Causation, Prediction, and Search. New York, NY: Springer; 1993. p. 103–162.

[25] Shen Y, Zhang L, Zhang J, Yang M, Tang B, Li Y, et al. CBN: Constructing a clinical Bayesian network based on data from the electronic medical record. Journal of biomedicalinformatics. 2018;88:1–10.

[26] Verduijn M, Peek N, Rosseel PMJ, de Jonge E, de Mol BAJM. Prognostic bayesian networks: I: Rationale, learning procedure, and clinical use. Journal of Biomedical Informatics. 2007;40(6):609–618.

[27] Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. Bayesian networks for clinical decision support in lung cancer care. PloS ONE. 2013;8(12):e82349.

[28] A Bhatia, A Kumar, I Khan, V Kumar. Analysis of Pattern Recognition (text mining) with Web Crawler.- International Transactions in Applied Sciences, 2011.

[29] A Bhatia "Comparative Study of Pattern Matching Using Text Mining" is published in International Journal of Research Review in Engineering, Science & Technology ISSN: 2278-6643 (Online) Volume-1, Issue-1 (June, 2012 Edition).