

# Detecting COVID-19 Utilizing Probabilistic Graphical Models

Emad Alsuwat<sup>1</sup>, Sabah Alzahrani<sup>2</sup>  
Department of Computer Science  
College of Computers and Information Technology  
Taif University

Hatim Alsuwat<sup>3</sup>  
Department of Computer Science  
College of Computer and Information Systems  
Umm Al Qura University

**Abstract**—Probabilistic graphical models are employed in a variety of areas such as artificial intelligence and machine learning to depict causal relations among sets of random variables. In this research, we employ probabilistic graphical models in the form of Bayesian network to detect coronavirus disease 2019 (denoted as COVID-19) disease. We propose two efficient Bayesian network models that are potent in encoding causal relations among random variable, i.e., COVID-19 symptoms. The first Bayesian network model, denoted as  $BN_1$ , is built depending on the acquired knowledge from medical experts. We collect data from clinics and hospitals in Saudi Arabia for our research. We name this authentic dataset  $DS_{covid}$ . The second Bayesian network model, denoted as  $BN_2$ , is learned from the real dataset  $DS_{covid}$  depending on Chow-Liu tree approach. We also implement our proposed Bayesian network models and present our experimental results. Our results show that the proposed approaches are capable of modeling the issue of making decisions in the context of COVID-19. Moreover, our experimental results show that the two Bayesian network models we propose in this work are effective for not only extracting casual relations but also reducing uncertainty and increasing the effectiveness of causal reasoning and prediction.

**Keywords**—Coronavirus disease 2019; COVID-19; artificial intelligence; machine learning; probabilistic graphical models; causal models; Bayesian networks; detection methods

## I. INTRODUCTION

Over the last two years, the world has experienced events that changed the whole scenario and would be considered as an extremely dark phase when written as history. This is because of the outbreak of a coronavirus disease 2019 pandemic denoted as COVID-19 that results from the family of viruses and causes issues related to gastrointestinal and respiratory diseases. There are different symptoms associated with the COVID-19 including shortness of breath, cough, fever, and fatigue, etc. It is a challenge to detect COVID-19 as these symptoms are also present in other diseases such as the common cold. The testing or detection of COVID-19 depends on the past or current presence of the SARS-CoV-2. If there is no virus, then the patient cannot be declared as a COVID-19 patient. Moreover, there is a presence of fear in people that can lead to psychological responses.

Medical diagnostic errors, on the other hand, can be considered as the leading error of injury and death. The error in diagnosis of the diseases results in the wrong treatment of the patient, which delays the patient's recovery time, give time for actual diseases to spread, and form another reaction from the wrong medications. There is a need for a strong patient

safety culture to improve the quality and reducing the risks associated with the diagnosis and then the treatment of the patient. For coping with all the mentioned above challenging COVID-19 detection issues, Artificial Intelligence (AI) needs to get involved in order to make the process of detecting the virus become accurate and efficient.

Artificial intelligence and machine learning have been widely used in many real-world applications in recent years. It has been reported that machine learning techniques could be used to deliver automated tools that are able to efficiently predict and detect risks. Artificial intelligence is the way that combines large data, intelligent algorithms, and iterative processing. If artificial intelligence is utilized in the right way, it can be beneficial and powerful. In this regard, this research focuses on the detection of COVID-19 utilizing machine learning, namely using Bayesian networks which is the expanded representation of the probability distribution. Bayesian networks are now established as the indispensable tools of artificial intelligence and are vital in the fields of engineering and science.

Probability theory has provided the base for different engineering and scientific tasks. Artificial intelligence and more specifically machine learning are considered as one of the major fields that have used probability for the development of new algorithms and theorems. In terms of the probabilistic graphical models or PGM's, the Bayesian networks are used at a wide-scale [1]. Bayesian networks are the extensive range class of probabilistic models. Initially, Bayesian networks were the type of statistical model that can show the probability distribution at a complex scale. They are well-developed modeling systems that are useful for monitoring, diagnosing, and making predictions even if uncertainties are present.

In this research, we utilize the soundness of probabilistic graphical models to both build a Bayesian network model using expert's knowledge and learn the structure of the Bayesian network model from authentic data. That is, we propose two probabilistic graphical models that are effective in the context of detecting COVID-19. The proposed probabilistic models are potent in encoding causal relations among random variables, i.e., COVID-19 symptoms, and thus can extract the required knowledge in order to reduce uncertainty and increase effectiveness and accuracy of causal reasoning and prediction.

The main contributions of our research are as follows: We build a naive Bayesian network model using experts' knowledge that is potent in dealing with the uncertainty in

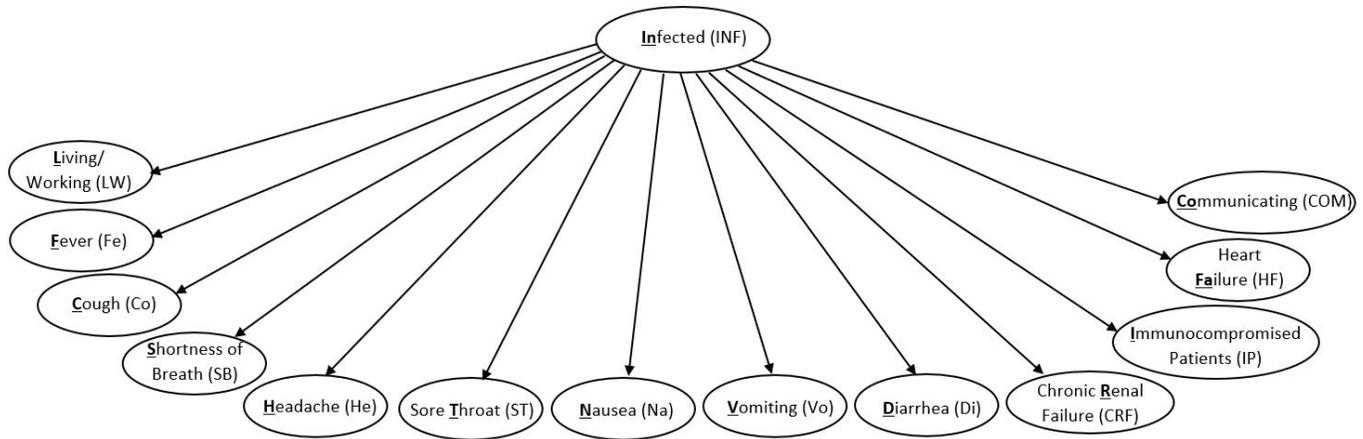


Fig. 1. A Causal Bayesian Network Model that was Built based on the Acquired knowledge from Medical Experts, Denoted as  $BN_1$ . The Direction of the Causal Relations in from the Class Variable to Attributes, i.e., top to Down. Each Node in the Diagnostic Bayesian Network Model has Two States, Either 'yes' or 'no'.

diagnosing the arriving patients in a clinic with COVID-19. We name this causal model  $BN_1$ . We collect data from clinics and hospitals in Saudi Arabia for our research. We name this authentic dataset  $DS_{covid}$ . We construct the “best possible” Bayesian network model from the dataset  $DS_{covid}$  using Chow-Liu tree approach (we refer the readers to [2], [3], [4] for complete information about Chow-Liu algorithm). We name this causal model  $BN_2$ . We have implemented our approaches and evaluate the correctness of  $BN_1$  using the authentic dataset  $DS_{covid}$ . Our results show that the naive Bayesian network model,  $BN_1$ , and the learned model using Chow-Liu tree approach,  $BN_2$ , are identical despite using different ways to recover these causal models. In addition, our experimental results indicate that the two causal models,  $BN_1$  and  $BN_2$ , are promising for reasoning under uncertainty, that is, the proposed models not only are capable to detect COVID-19 positive cases but also can be efficiently used make real-time reasoning and predictions.

The rest of this paper is structured as follows. In Section II, we present our coronavirus diagnostic naive Bayesian network model built using Expert Knowledge. In Section III, we present our coronavirus diagnostic Bayesian network model recovered from the collected authentic data. In Section IV, we present implementations of the proposed causal models and show our empirical results. In Section V, we give a brief overview of background information. In Section VI, we present the conclusion and future work of our research.

## II. CORONAVIRUS DIAGNOSTIC NAIVE BAYESIAN NETWORK MODEL FROM EXPERT KNOWLEDGE

In this section, we acquire medical knowledge from experts and then build a Bayesian network model that captures the acquired knowledge. The fundamental purpose is to use the developed Bayesian network model to diagnose the condition of arriving patients suffering from one or some of COVID-19 symptoms.

The acquired medical knowledge is as follows: Coronavirus disease (COVID-19) is an acute respiratory disease that can be detected via investigating either exposure risks or clinical signs

and symptoms. Exposure risks for coronavirus disease increase when communicating with a reported case of coronavirus in the last 14 days before onset of symptoms or living/working in a place identified to be suffering of occurrences of coronavirus in the last 14 days. Clinical signs and symptoms may include one or more of the following:

- (1) Patients suffering from fever above 37.5 degrees Celsius,
- (2) Patients suffering from cough,
- (3) Patients suffering from shortness of breath,
- (4) Patients suffering from sore throat or headache,
- (5) Patients suffering from vomiting, nausea, or diarrhea, and
- (6) Patients suffering from Heart failure, chronic renal failure, and immunocompromised patients.

This knowledge needs to be applied to the situation where a patient is presented in a clinic, and we need to assure whether the patient is infected with coronavirus disease or not. Give the set of symptoms that the patient is suffering from, the doctor would like to know the probability that the patient is infected with coronavirus disease. For instance, if the patients suffers from fever, cough, and shortness of breath, what is the chance that this patient is infected with coronavirus disease and must be quarantined immediately? When getting all information from the patient (i.e., by directly asking the patient about what symptoms the patient suffers from), the developed Bayesian network model will give the final judgement, which will significantly help the doctor and significantly reduce the diagnose time.

Therefore, the goal is to transfer this acquired knowledge into a medical network (i.e., Bayesian network model) that can be used to diagnose arriving patients in a clinic. In order to do so, we first have to determine the variables of the Bayesian network model. In the context of the above given medical knowledge from experts, the variables are either symptoms or conditions, i.e., diseases. There are, indeed, fourteen symptoms and a condition, which is the case whether the patient is infected with coronavirus disease or not. The variables in this Bayesian network model are discrete, i.e., their values are either True or False.

TABLE I. CONDITIONAL PROBABILITY TABLES FOR THE CAUSAL BAYESIAN NETWORK MODEL  $BN_1$

Node	Probability Value(s)		Type
<b>INF</b>	$P(in) = 0.01$	-	Prior Probability
<b>LW</b>	$P(l in) = 0.05$	$P(l in') = 0.01$	Conditional Probability
<b>Fe</b>	$P(f in) = 0.1$	$P(f in') = 0.01$	Conditional Probability
<b>Co</b>	$P(c in) = 0.1$	$P(c in') = 0.01$	Conditional Probability
<b>SB</b>	$P(s in) = 0.1$	$P(s in') = 0.05$	Conditional Probability
<b>He</b>	$P(h in) = 0.02$	$P(h in') = 0.01$	Conditional Probability
<b>ST</b>	$P(t in) = 0.02$	$P(t in') = 0.01$	Conditional Probability
<b>Na</b>	$P(n in) = 0.02$	$P(n in') = 0.01$	Conditional Probability
<b>Vo</b>	$P(v in) = 0.02$	$P(v in') = 0.01$	Conditional Probability
<b>Di</b>	$P(d in) = 0.02$	$P(d in') = 0.01$	Conditional Probability
<b>CRF</b>	$P(r in) = 0.02$	$P(r in') = 0.01$	Conditional Probability
<b>IP</b>	$P(i in) = 0.02$	$P(i in') = 0.01$	Conditional Probability
<b>HF</b>	$P(fa in) = 0.02$	$P(fa in') = 0.01$	Conditional Probability
<b>COM</b>	$P(co in) = 0.05$	$P(co in') = 0.01$	Conditional Probability

The structure of the acquired medical knowledge from experts is represented by the qualitative naive Bayesian network model shown in Fig. 1. There are two layers in the proposed Bayesian network model. The edges in this model go from the first layer, the condition or the variable class layer, to the second layer, the attributes or symptoms layer. One should note that give the condition or the class variable in this Bayesian network model, every two attributes are d-separated, i.e., independent.

We obtain the Conditional Probability Tables (CPTs) from medical experts depending on their subjective beliefs that reflect the practical experience they have gained. The assessment of the prior probability of the class variable and the assessment of the probability of each attribute conditional on its parents, i.e., conditional on the class variable, are give in Table I. We use the notation  $f$  to indicate a positive response on the node  $Fe$ , 'Fever' and  $f'$  to indicate a negative response on the node  $Fe$ , 'Fever'. The notation  $P(f)$  stands for  $P(Fe = f)$ . Similar notations are applied to each node in the developed Bayesian network model as follows:

- $in$  stands for a positive response on the node  $INF$ ;
- $l$  stands for a positive response on the node  $LW$ ;
- $c$  stands for a positive response on the node  $Co$ ;
- $s$  stands for a positive response on the node  $SB$ ;
- $h$  stands for a positive response on the node  $He$ ;
- $t$  stands for a positive response on the node  $ST$ ;
- $n$  stands for a positive response on the node  $Na$ ;
- $v$  stands for a positive response on the node  $Vo$ ;
- $d$  stands for a positive response on the node  $Di$ ;
- $r$  stands for a positive response on the node  $CRF$ ;
- $i$  stands for a positive response on the node  $IP$ ;
- $fa$  stands for a positive response on the node  $HF$ ; and
- $co$  stands for a positive response on the node  $COM$ .

One should note that Table I includes only positive probabilities while negative probabilities can be easily derived from this table. For instance, the probability of heart failure given that the patient has coronavirus disease is 0.02, formally can be

written as  $P(fa|in) = 0.02$ . One can derive the probability of not having heart failure given that the patient has coronavirus disease, which is  $1 - 0.02 = 0.98$ , formally can be written as  $P(fa'|in) = 0.98$ .

The proposed causal model shows that the joint probability can be written as follows:

$$P(INF, LW, Fe, Co, SB, He, ST, Na, Vo, Di, CRF, IP, HF, COM) = P(INF) P(LW|INF) P(Fe|INF) P(Co|INF) P(SB|INF) P(He|INF) P(ST|INF) P(Na|INF) P(Vo|INF) P(Di|INF) P(CRF|INF) P(IP|INF) P(HF|INF) P(COM|INF)$$

The joint probability is the product of class variable prior probability times all conditional probabilities of all other variables. On the other hand, if we want to calculate the probability that a patient has coronavirus disease given that this patient has fever and has communicated with a person who was reported to have coronavirus disease, we would do the following:

$$P(INF|Fe, COM) = P(INF, Fe, COM)/P(Fe, COM)$$

For more information about this inferential equation, we refer the readers to Bayes theorem [4]. The implementation of the proposed causal model in this section and more experiments will be presented in Section IV.

### III. CORONAVIRUS DIAGNOSTIC BAYESIAN NETWORK MODEL FROM DATA

In this section, we learn a Bayesian network model from data. We have collected data from clinics and hospitals in Saudi Arabia for our research. We name our dataset  $DS_{covid}$ , which contains the data for 60,000 patients. The dataset  $DS_{covid}$  is available at [https://emadalsuwat.github.io/resources/Covid19\\_dataset.dat](https://emadalsuwat.github.io/resources/Covid19_dataset.dat).

The dataset  $DS_{covid}$  consists of cases, a.k.a. tuples. Each case represent a unique object in the collection process. Objects or cases are patients in this research context. Each case in our dataset is characterized by fourteen variables. That is, for each tuple in our dataset, there are fourteen cells, a.k.a

data items, that are required to be filled such that each cell corresponds to a defined random variable. The dataset  $DS_{covid}$  is a discrete dataset, which means that the answer to each variables is either “yes” or “no”. For example, case #1 contains information about patient #1 in our dataset and so on. If patient #1 has fever, we shall see “yes” in the cell that corresponds to the random variable “Fever” and so on. If the value of a cell in a tuple is missing (i.e., there is no reported information), the cell is left blank.

We construct the “best possible” tree-shaped Bayesian network model from the learning dataset  $DS_{covid}$ . We then determine whether or not the best possible tree-shaped belief network learned using the collected dataset  $DS_{covid}$  is the naive Bayesian network model. If so, we say that the naive Bayesian network model is the best possible tree-shaped belief network that fits the collected dataset  $DS_{covid}$ .

We construct the “best possible” Bayesian network model from the dataset  $DS_{covid}$  using Chow-Liu tree approach, which is an approximation of the probability distribution with tree-shaped belief network. In Chow-Liu tree approach, the links are directed away from the root of the tree-shaped Bayesian network model.

Let  $DS_{covid}$  be a discrete dataset over variables  $\{V_1, V_2, \dots, V_n\}$ . Let  $BN_2 = (V, E)$  be the Bayesian network model constructed from the dataset  $DS_{covid}$  using Chow-Liu tree approach. Let  $V = \{V_1, V_2, \dots, V_n\}$  and  $E = \{(v_1, v_2) : v_1, v_2 \in V\}$  be the set of nodes and edges, respectively, in the belief network  $BN_2$ . The causal network  $BN_2$  can be learned as follows [3]:

**Step 1:** For each pair of vertices  $(V_i, V_j) \in BN_2$ , we calculate the mutual information as follows.

$$MI(V_i, V_j) = \sum_{V_i, V_j} P(V_i, V_j) \log_2 \left( \frac{P(V_i, V_j)}{P(V_i)P(V_j)} \right) \quad (1)$$

**Step 2:** Construct the mutual information weighted graph, which we denote as  $G_{MI}$ , over the set of vertices  $\{V_1, V_2, \dots, V_n\}$ , where the weight of the edge  $(V_1, V_2)$  in  $G_{MI}$  is  $MI(V_i, V_j)$ .

**Step 3:** Given the  $G_{MI}$  graph, build a maximum weight spanning tree, which we denote as  $G_{SPT}$ .

**Step 4:** Direct the resulting maximum spanning tree  $G_{SPT}$ . This can be accomplished by selecting any node in the resulting graph as the root of the tree and then set the direction of the remaining edges to point outward from it. We denote the resulting tree as  $BN_2$ .

In this section, we have shown that it is feasible to estimate the Bayesian network model from medical recorders of patients, i.e., from the collected data. That is, the result of the formal steps presented above is the Bayesian network directed acyclic graph. An important point to be mentioned is that one should use Expectation–Maximization algorithm (a.k.a EM algorithm) to estimate the conditional probability distributions of the constructed causal model,  $BN_2$ , from data. For more information about the EM algorithm, we refer the readers to [5]. The implementation of learning the structure of the causal model  $BN_2$  from the collected data using Chow-Liu tree approach as proposed in this section and more experiments will be presented in Section IV.

#### IV. IMPLEMENTATIONS OF CAUSAL MODELS AND EXPERIMENTAL RESULTS

In this section, we implement our proposed two Bayesian network models,  $BN_1$  and  $BN_2$ , using *Hugin<sup>TM</sup> Researcher* 8.4. We then present some experimental results to show the practical capability of the proposed causal models.

##### A. Implementation of the Causal Model $BN_1$

In this subsection, expert knowledge presented in section II can be modeled by the implemented Bayesian network  $BN_2$  shown in Figure 2. The Bayesian network  $BN_2$  consists of fourteen nodes: Infected (INF), Living/Working (LW), Fever (Fe), and so on. Each node has two states: “Yes” or “No”. For instance, if the presented patient has Fever, we choose “Yes” as the answer for this symptom and vice versa.

Fig. 2 shows the running mode of the Bayesian network model  $BN_1$ . The figure is split into two parts: Node List Pane on the left hand side and Network Pane on the right hand side. It is important to point out that we have expanded all nodes getting ready to enter evidence or facts. For instance, if a patients has Nausea, we then just click on the state “Yes” under the node Nausea in the Node List Pane and so on.

One should note that any two variables in the Bayesian network model  $BN_2$  are independent given the state of the hypothesis variable. That is, if the state of the hypothesis variable Infect is given, then any two variables of the symptoms nodes are independent. Another important note is that causality is obvious and clear in the implemented model  $BN_2$ . That is, the class variable, Infect in this case, is the cause of the other variables, i.e., symptoms. Entering information about the state symptoms (either “Yes” or “No”) will increase/decrease the joint probability that the patient has COVID-19 disease.

##### B. Implementation of the Causal Model $BN_2$

In this subsection, we use the four steps presented in Section III to learn the structure of the Bayesian network model from data. That is, we use the Chow-Liu algorithm to recover the structure of  $BN_2$  from the collected dataset  $DS_{covid}$ . We present the first few lines of the authentic dataset  $DS_{covid}$  in Fig. 3. The data file of the dataset  $DS_{covid}$  we use in this implementation consists of 60,000 cases, i.e., tuples.

We use *Hugin<sup>TM</sup> Researcher* 8.4 to learn the Bayesian network model  $BN_2$  from the dataset  $DS_{covid}$ . We use *Hugin<sup>TM</sup>* implementation of Chow-Liu algorithm with the default significance level at 0.05 [6]. It is important to keep the level of significance at the default setting and not to increase this value in order to reduce the running time of the Bayesian network structure learning algorithm to the best possible running time.

The recovered Bayesian network model from the dataset  $DS_{covid}$  is as shown in Fig. 4. When comparing the result of Chow-Liu algorithm,  $BN_2$ , with the result of the original model we built using experts’ knowledge,  $BN_1$ , we observe the following: 1) All nodes have been recovered correctly, 2) All links have been recovered correctly, and 3) There is no missing or misoriented links, i.e., the original and the learned Bayesian network models indicate the same causal

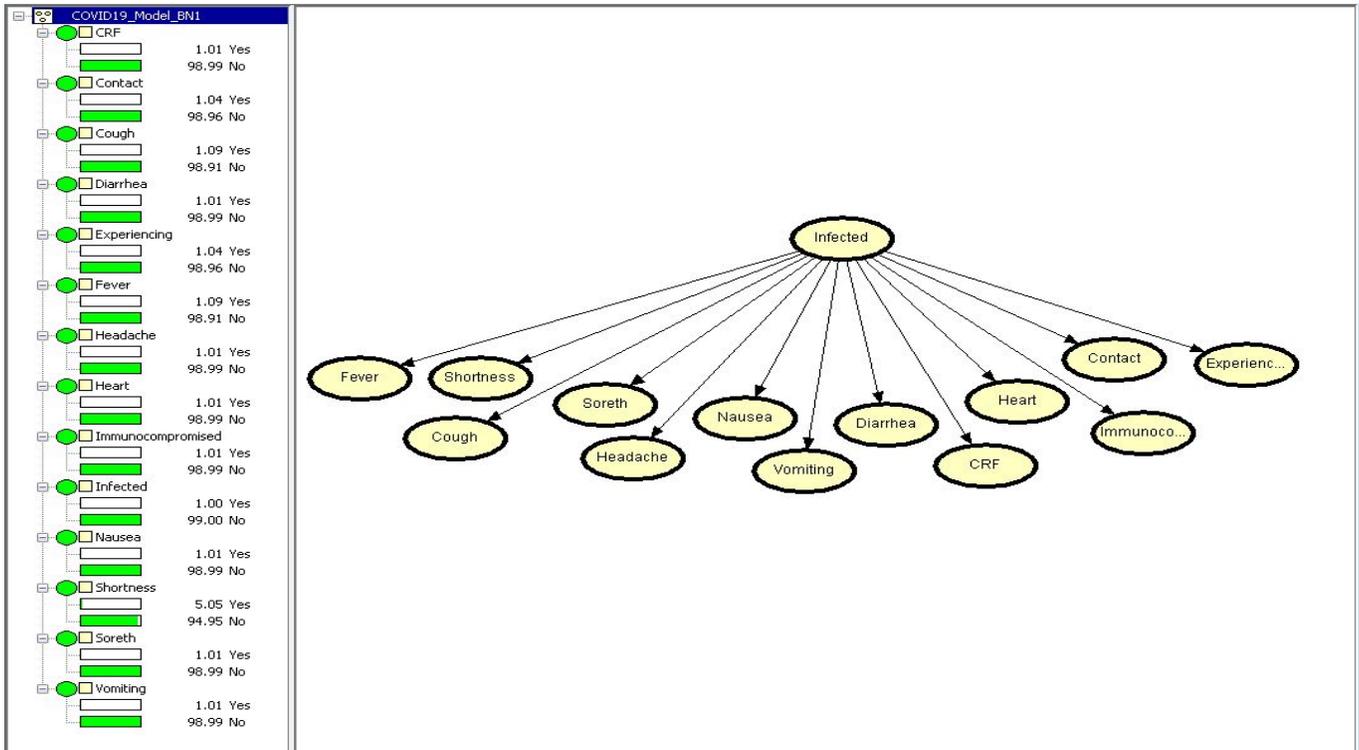


Fig. 2. The Implemented Bayesian Network Model  $BN_1$ .

#	Fever	Headache	Soreth	Nausea	Vomiting	CRF	Heart	Infected	Immunoco...	Diarrhea	Shortness	Cough	Experiencing	Contact
0	No	No	No	No	No	No	No	No	No	No	No	No	No	No
1	No	No	No	No	No	No	No	No	No	No	No	No	No	No
2	No	No	No	No	No	No	No	No	No	No	No	No	No	No
3	No	No	No	No	No	No	No	No	No	No	No	No	No	No
4	No	No	No	No	No	No	No	No	No	No	No	No	No	No
5	No	No	No	No	No	No	No	No	No	No	No	No	No	No
6	No	No	No	No	No	No	No	No	No	No	No	No	No	No
7	No	No	No	No	No	No	No	No	No	No	No	No	No	No
8	No	Yes	No	No	No	No	No	No	No	No	No	No	No	No
9	No	No	No	No	No	No	No	No	No	No	No	No	No	No
10	No	No	No	No	No	No	No	No	No	No	No	No	No	No
11	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No
12	No	No	No	No	No	No	No	No	No	No	No	No	No	No
13	No	No	No	No	No	No	No	No	No	No	No	No	No	No
14	No	No	No	No	No	No	No	No	No	No	No	No	No	No
15	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No

Fig. 3. The First Few Cases (Tuples) of the Data File of the Authentic Dataset  $DS_{covid}$ .

relationships. We thus conclude that Chow-Liu structure learning algorithm was able to perfectly recover the structure of the Bayesian network model  $BN_2$  from the collected dataset  $DS_{covid}$ .

### C. Reasoning using the Proposed Models

In this subsection, we employ the implemented Bayesian network models,  $BN_1$  and  $BN_2$ , for reasoning under uncertainty. That is, given some symptoms that a presented patient in a clinic suffers from, the medical practitioner can then utilize one of these models to find out the probability that this patients is infected with COVID-19.

**Using  $BN_1$  for Reasoning under Uncertainty:** Assume that a patient presented in a clinic where this patient has been in contact with a person who has COVID-19 and suffers from Fever and Sore Throat. The medical practitioner employs the

Bayesian network model  $BN_1$  to calculate the inferential (conditional) probability  $P(INF = "Yes" | Fe = "Yes", ST = "Yes", COM = "Yes")$ , which can be calculated by dividing the joint probability  $P(INF = "Yes", Fe = "Yes", ST = "Yes", COM = "Yes")$  by the probability of evidence  $P(Fe = "Yes", ST = "Yes", COM = "Yes")$ .

The medical practitioner can simply enter the evidence on the nodes "Fever", "Soreth" and "Contact". That is, the medical practitioner chooses the state "Yes" on these three nodes as shown in Fig. 5.

We eventually propagate these pieces of evidence in order to calculate the conditional probability. The conditional probability is calculated by dividing the joint probability by the probability of the evidence.  $BN_1$  can be used to obtain the joint probability as follows:  $P(INF = "Yes", Fe = "Yes", ST = "Yes", COM = "Yes") = 8.5916e - 7$ .

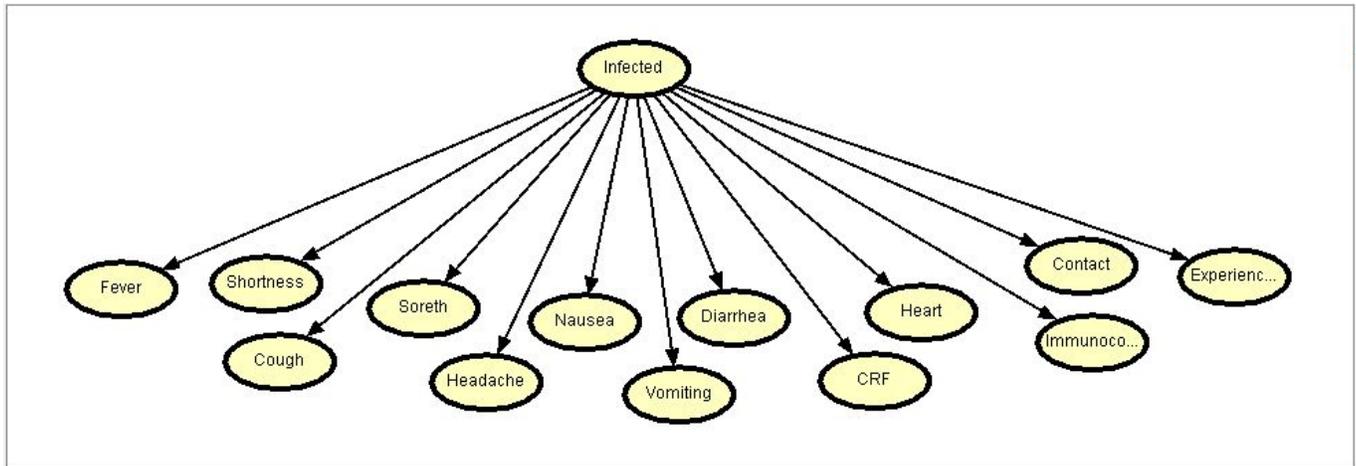


Fig. 4. The Result of using Chow-Liu Algorithm to Learn the Bayesian Network Model  $BN_2$  from the Authentic Dataset  $DS_{covid}$ .

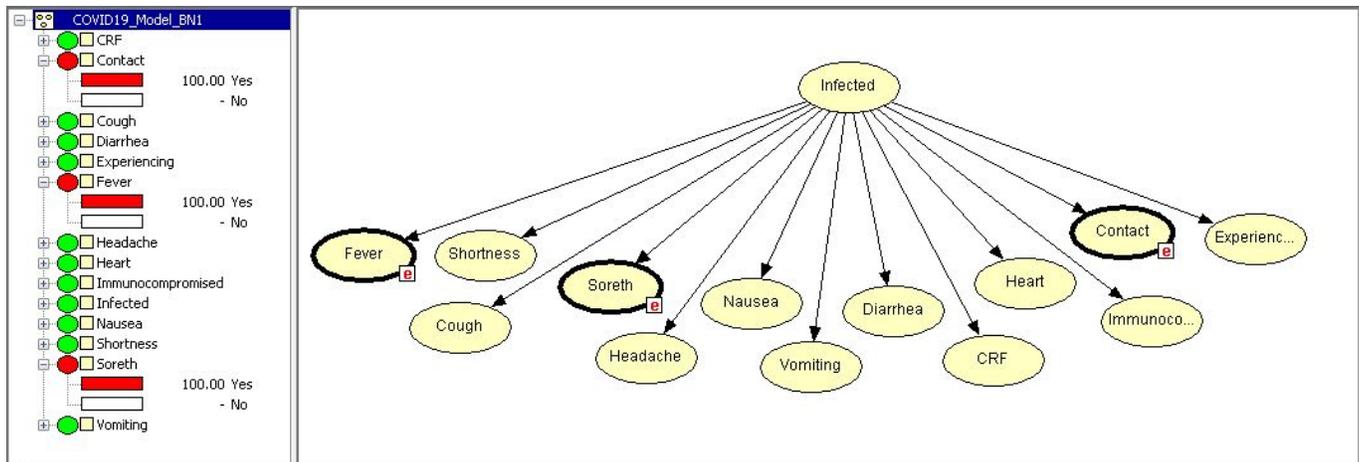


Fig. 5. The Bayesian Network Model  $BN_1$  After Entering the Evidence on the Three Nodes: Fever, Soreth and Contact.

$BN_1$  also can be used to obtain the probability of the evidence as follows:  $P(Fe = \text{“Yes”}, ST = \text{“Yes”}, COM = \text{“Yes”}) = 1.99e - 6$ . Thus, the conditional probability can be calculated as follows:  $P(INF = \text{“Yes”} | Fe = \text{“Yes”}, ST = \text{“Yes”}, COM = \text{“Yes”}) = \frac{8.5916e-7}{1.99e-6} = 0.432$ .

**Using  $BN_2$  for Reasoning under Uncertainty:** Assume that an arriving patient presented at a clinic suffers from the following symptoms: Fever and Shortness of Breath, and Cough. In addition, the patient has reported that there was a communication with a reported case of coronavirus in the last 14 days before onset of symptoms. The medical practitioner thus employs the Bayesian network model  $BN_2$  to calculate the inferential probability  $P(INF = \text{“Yes”} | Fe = \text{“Yes”}, SB = \text{“Yes”}, Co = \text{“Yes”}, COM = \text{“Yes”})$ , which can be calculated by dividing the joint probability  $P(INF = \text{“Yes”}, Fe = \text{“Yes”}, SB = \text{“Yes”}, Co = \text{“Yes”}, COM = \text{“Yes”})$  by the probability of evidence  $P(Fe = \text{“Yes”}, SB = \text{“Yes”}, Co = \text{“Yes”}, COM = \text{“Yes”})$ .

The medical practitioner can simply provide the evidence on the nodes “Fever”, “Shortness”, “Cough”, and “Contact”. That is, the medical practitioner selects the state “Yes” on these

four nodes as shown in Fig. 6.

These pieces of evidence are then propagated with the ultimate goal of calculating the conditional probability. The conditional probability is calculated as follows: 1) Employ the Bayesian network model  $BN_2$  to obtain the joint probability. Using the model  $BN_2$ , the joint probability  $P(INF = \text{“Yes”}, Fe = \text{“Yes”}, SB = \text{“Yes”}, Co = \text{“Yes”}, COM = \text{“Yes”}) = 4.0411e - 7$ . 2) Employ the Bayesian model  $BN_2$  to find the probability of the evidence. Using  $BN_2$ , the probability of the evidence  $P(Fe = \text{“Yes”}, SB = \text{“Yes”}, Co = \text{“Yes”}, COM = \text{“Yes”}) = 5.495e - 7$ . 3) Calculate the conditional probability as follows:  $P(INF = \text{“Yes”} | Fe = \text{“Yes”}, SB = \text{“Yes”}, Co = \text{“Yes”}, COM = \text{“Yes”}) = \frac{4.0411e-7}{5.495e-7} = 0.735$ .

we have shown that our implemented causal models are efficient in detecting patients with COVID-19 disease given the set of symptoms they suffer from. It is important to point out that our causal models are an outstanding technique to control the disease. Moreover, these models are great as they enable medical practitioners to take decisions under uncertainty.

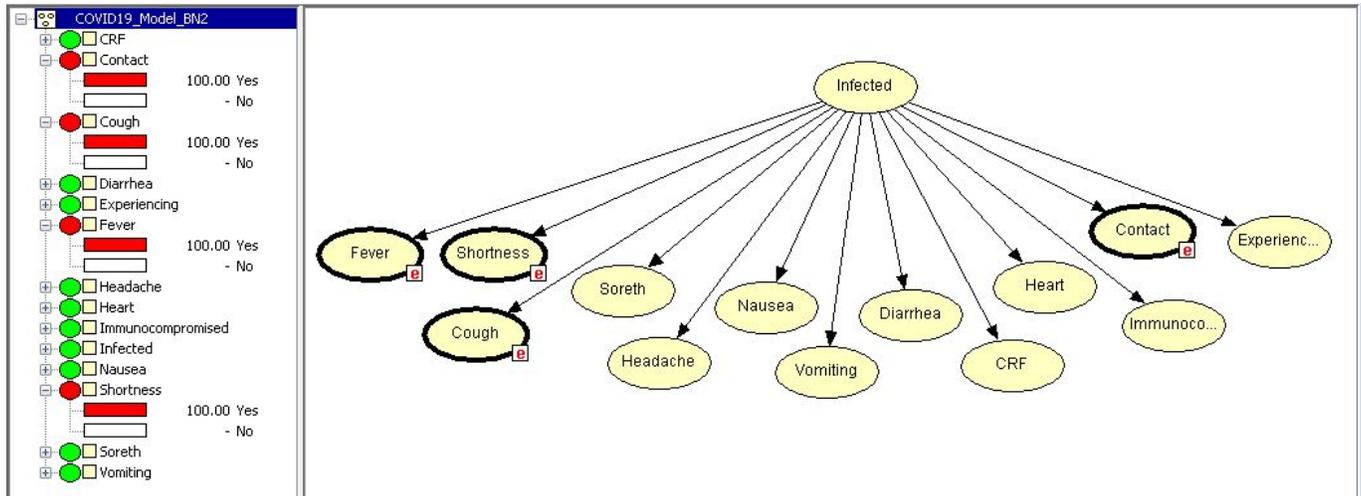


Fig. 6. The Bayesian Network Model  $BN_2$  After Entering the Evidence on the Three Nodes: Fever, Shortness, Cough, and Contact.

## V. BACKGROUND INFORMATION

In this section, we will give a brief overview of the use of artificial intelligence, namely Bayesian network models, a favorable tool for decision making. More detailed information about probabilistic graphical models and Bayesian network can be found in [4], [7], [8], [9].

Artificial intelligence is a technology that became prevalent in all different kinds of businesses. However, its use in healthcare has only started in recent years. This increased use is because the technologies have the transforming potential in different aspects of healthcare [10], [11], [12]. Machine learning is a statistical technique for learning the training models with data and for fitting models to data. It is one of the common forms of artificial intelligence. It can be evident from the fact that 63% of corporates that use artificial intelligence were employing machine learning in their corporates and businesses [10]. There are different applications of machine learning in healthcare. For instance, precision medicine- predicting the best treatment method on the basis of the patient attributes and treatment context [10], [13].

The use of probabilistic graphical models has started in 1985 when Pearl (1985) combines the probabilities theories and graphs for obtaining the comprehensible representation of the probability distribution. Further improvements in the specific models can be observed by the research of Lauritzen and Speigelhalter (1988) by Cooper and Herskovits (1992) [1], [14], [4]. The need to overcome difficult and complex problems that exist in the real world has increased the demand for using effective meta-heuristic algorithms that are able to get solutions by performing a momentous search for getting possible solutions.

The probabilistic graphical models have traditionally been used for different purposes. For instance, research indicates the use of probabilistic graphical models for dealing with age estimation of living people [15]. The importance of age estimation of the living individual has been raised because of the civil, criminal, and administrative judicial situations, and it has a significant role in the medicolegal and forensic services all around the globe. Other researchers used the Bayesian

approach for determination of clavicular epiphysis [15]. Moreover, it has been suggested to use probabilistic graphical models for evolutionary algorithm and solving complex problems. The evolutionary computation was performed perfectly using the probabilistic graphical models [15].

The diagnostic is considered as one of the major practices in artificial intelligence. Research shows several probabilistic diagnostic methods that can be implemented in the upcoming medicine research. We refer the readers to [16] for a complete list of possible uses of artificial intelligence in medicine. Another widely use of Bayesian networks other than medical research is detecting spam (a.k.a. spam filtering) [17]. Bayesian networks has also been employed and showed outstanding results in intrusion detection. Recent survey on intrusion detection using machine learning can be found in [18], [19].

Another significant application of Bayesian networks is predicting the type of hematological malignancies [20]. Bayesian networks have helped scientists to model Acute Myeloid Leukemia (AML), which is a subtype of blood cancer. Thus, research has shown that Bayesian network was not only able to model and predict many subtypes of cancer but also help physicians choose the most effective treatment. Bayesian network performance in prediction and reasoning under uncertainty was compared in the literature with other machine learning types. Indeed, Bayesian networks was proved to be a very efficient predictive model in finding probabilistic graphical relationships among random variables. Recently, researchers have employed Bayesian networks to improve screening of COVID-19 cases [21]. The above literature has given model examples that suggests the use of Bayesian networks in healthcare sector for detecting and differentiating between different and complex medical conditions. Previous work for detecting COVID-19 has focused on different machine learning tools except Bayesian networks even though Bayesian networks are very powerful in reasoning under uncertainty [22], [23], [24]. In this work, we explore the feasibility of using the probabilistic graphical models in the form of Bayesian networks to build an effective diagnosis system for controlling and detecting COVID-19 disease. The proposed

Bayesian models are useful in current situation for increasing the speed of diagnosis, which can be helpful in treating and coping with this disease efficaciously.

## VI. CONCLUSION AND FUTURE WORK

In this research, we demonstrated the applicability of utilizing probabilistic graphical models in the form of Bayesian networks to detect patients with COVID-19 disease. We proposed two diagnostic Bayesian network models:  $BN_1$  and  $BN_2$ . We implemented the proposed Bayesian network models and conducted our experimental results. Our results indicate this research succeeded in designing and building two effective Bayesian network models to tackle the issue of dealing with arriving patients who may have COVID-19 disease. Our experimental results also show that our proposed Bayesian network models are effective in not only extracting the casual relations but also reducing uncertainty and increasing the effectiveness of causal reasoning and prediction.

In our future work, we aim to extend Bayesian network models we built in this research in order to address more diseases beside COVID-19. In particular, we aim to extend our Bayesian network models to handle rare subtypes of blood cancer as such diseases are very unlikely to get funded or find interested researchers.

## ACKNOWLEDGMENT

This project was supported by the Deanship of Scientific Research at Taif University. The authors would like to thank Taif University for providing financial support for this research project (Project number: 1-441-64)

## REFERENCES

- [1] P. Larrañaga, H. Karshenas, C. Bielza, and R. Santana, "A review on probabilistic graphical models in evolutionary computation," *Journal of Heuristics*, vol. 18, no. 5, pp. 795–819, 2012.
- [2] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.
- [3] T. D. Nielsen and F. V. Jensen, *Bayesian networks and decision graphs*. Springer Science & Business Media, 2009.
- [4] R. E. Neapolitan, "Learning bayesian networks, vol. 38 pearson prentice hall," *Upper Saddle River, NJ*, 2004.
- [5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [6] A. L. Madsen, F. Jensen, U. B. Kjaerulff, and M. Lang, "The hugin tool for probabilistic graphical models," *International Journal on Artificial Intelligence Tools*, vol. 14, no. 03, pp. 507–543, 2005.
- [7] A. Dawid, R. Cowell, S. Lauritzen, and D. Spiegelhalter, "Probabilistic networks and expert systems." Springer-Verlag, 1999.
- [8] P. Congdon, *Bayesian statistical modelling*. John Wiley & Sons, 2007, vol. 704.
- [9] R. G. Cowell, P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer Science & Business Media, 2006.
- [10] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future healthcare journal*, vol. 6, no. 2, p. 94, 2019.
- [11] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, "A survey on ambient intelligence in healthcare," *Proceedings of the IEEE*, vol. 101, no. 12, pp. 2470–2494, 2013.
- [12] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: past, present and future," *Stroke and vascular neurology*, vol. 2, no. 4, 2017.
- [13] R. Hamamoto, K. Suvarna, M. Yamada, K. Kobayashi, N. Shinkai, M. Miyake, M. Takahashi, S. Jinnai, R. Shimoyama, A. Sakai *et al.*, "Application of artificial intelligence technology in oncology: Towards the establishment of precision medicine," *Cancers*, vol. 12, no. 12, p. 3532, 2020.
- [14] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, prediction, and search*. MIT press, 2000.
- [15] E. Sironi, M. Gallidabino, C. Weyermann, and F. Taroni, "Probabilistic graphical models to deal with age estimation of living persons," *International journal of legal medicine*, vol. 130, no. 2, pp. 475–488, 2016.
- [16] P. Szolovits, *Artificial intelligence in medicine*. Routledge, 2019.
- [17] B. Nelson, M. Barreno, F. J. Chi, A. D. Joseph, B. I. Rubinstein, U. Saini, C. Sutton, J. Tygar, and K. Xia, "Misleading learners: Co-opting your spam filter," in *Machine learning in cyber trust*. Springer, 2009, pp. 17–51.
- [18] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: A review," *expert systems with applications*, vol. 36, no. 10, pp. 11 994–12 000, 2009.
- [19] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications surveys & tutorials*, vol. 18, no. 2, pp. 1153–1176, 2015.
- [20] R. Agrahari, A. Foroushani, T. R. Docking, L. Chang, G. Duns, M. Hudoba, A. Karsan, and H. Zare, "Applications of bayesian network models in predicting types of hematological malignancies," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.
- [21] S. Eyheramendy, P. A. Saa, E. A. Undurraga, C. Valencia, C. Lopez, L. Mendez, J. Pizarro-Berdichevsky, A. Finkelstein-Kulka, S. Solari, N. Salas *et al.*, "Improved screening of covid-19 cases through a bayesian network symptoms model and psychophysical olfactory test," *medRxiv*, 2021.
- [22] I. Arpaci, M. Al-Emran, G. Marques, M. Al-Kabi *et al.*, "A survey of using machine learning algorithms during the covid-19 pandemic," *Emerging Technologies During the Era of COVID-19 Pandemic*, vol. 348, pp. 1–8, 2021.
- [23] S. Abir, S. N. Islam, A. Anwar, A. N. Mahmood, and A. M. T. Oo, "Building resilience against covid-19 pandemic using artificial intelligence, machine learning, and iot: A survey of recent progress," *IoT*, vol. 1, no. 2, pp. 506–528, 2020.
- [24] J. Nayak, B. Naik, P. Dinesh, K. Vakula, B. K. Rao, W. Ding, and D. Pelusi, "Intelligent system for covid-19 prognosis: A state-of-the-art survey," *Applied Intelligence*, vol. 51, no. 5, pp. 2908–2938, 2021.