A Novel Decision-Making Process for COVID-19 Fighting Based on Association Rules and Bayesian Methods

Salim El Khediri1,2,3,*, Adel Thaljaoui4 and Fayezy Al-Fayez4

1Department of Information Technology, College of Computer, Qassim University, Buraydah 51542, Saudi Arabia
2Department of Computer Sciences, Faculty of Sciences of Gafsa, University of Gafsa, Gafsa 2112, Tunisia
3LETI Laboratory, University of Sfax, National School of Engineers (ENIS), BP 1173, Sfax 3038, Tunisia
4Department of Computer Science and Information, College of Science at Zulfi, Majmaah University, Al-Majmaah 11952, Saudi Arabia

*Corresponding author: s.elkhediri@qu.edu.sa

Since recording the first case in Wuhan in November 2020, COVID-19 is still spreading widely and rapidly affecting the health of millions all over the globe. For fighting against this pandemic, numerous strategies have been made, where the early isolation is considered among the most effective ones. Proposing useful methods to screen and diagnose the patient’s situation for the purpose of specifying the adequate clinical management represents a significant challenge in diminishing the rates of mortality. Inspired from this current global health situation, we introduce a new autonomous process of decision-making that consists of two modules. The first module is the data analysis based on Bayesian network that is employed to indicate the coronavirus symptoms severity and then classify COVID-19 cases as severe, moderate or mild. The second module represents the decision-making based on association rules method that generates autonomously the adequate decision. To construct the model of Bayesian network, we used an effective method-oriented data for the sake of learning its structure. As a result, the algorithm accuracy in making the correct decision is 30% and in making the adequate decision is 70%. These experimental results demonstrate the importance of the suggested methods for decision-making.

Keywords: autonomous decision-making; Bayesian networks; association rules; Bayesian network’s structure learning based on data approach; COVID-19

Received 1 December 2020; Revised 10 February 2021; Editorial Decision 29 April 2021
Handling editor: Fionn Murtagh

1. INTRODUCTION

Because of the COVID-19 pandemic, the systems of the health care have been enduring crippling challenges during the past few months. According to the World Health Organization (WHO), the total deaths of this epidemic are 2 312 278 deaths at 8 February 2021 (458 544 USA) [1]. Accordingly, the coronavirus disease is among the life-threatening pandemics that threatened the lives of humans worldwide. For the purpose of limiting the pandemic spread, the quarantine procedure has played an important role. In addition, governments have encouraged researchers (in medicine, chemistry, physics, computer science, mathematics and so on) to collaborate for the goal of proposing an effective solution for this pandemic. Nevertheless, no specific and effective medication was found [2]. In [3], Wang et al. highlighted the importance of the earlier detection of the infection the process of the treatment. Thus, the earlier the symptoms are discovered the earlier the adequate
care and medication are attributed, thence the rates of mortality could be diminished [2]. For the purpose of analyzing chest radiographs, which is imperative in the detection of the coronavirus, numerous Artificial Intelligence methods have been incorporated [4]. Added to this, methods of deep learning have been employed to analyze chest computed tomography such as in [3, 5, 6] as well as [7]. For the prediction of the positive cases of COVID-19, an enhanced Neuro-Fuzzy Inference System called FPASSA-ANFIS has been utilized [8]. Furthermore, Pirouz et al. [9] presented a novel technique of classification for analyzing the relation between positive cases and parameters of the environment. Thus, owing to the unforeseen fatality rate of the COVID-19 pandemic and prompted by the benefits of former diagnosis of the severity of cases, in this study we introduce an autonomous process of decision-making founded on hybrid approaches of data analysis. In fact, we suggested combining the simpleness of the method of association rules with the efficacy of probabilistic and graphical model to analyze coronavirus characteristic symptoms for deciding the cases’ severity. To build the Bayesian network (BN) model, a novel method of structure learning was proposed as an enhancement of K2 algorithm, afterwards the parameters learning step was performed employing the algorithm of Expectation Maximization (EM). The second step allows the rules generation basing on the algorithm of frequent pattern (FP) growth. Later, through the calculation of distance separating features, the decision assigned for the closest object (possessing minimal distance) could be attributed to the novel object. The remainder of this paper is organized in the following way: Section 2 introduces the theoretic foundations employed to construct our proposal. Section 3 is devoted to depicting the process of autonomous decision-making basing on the methods of association rules and the enhancement of BNs. To construct the discrete model of BN, we suggested a new algorithm of structure learning basing on data-oriented approach. Additionally, we used the rules that were discovered to autonomously take the right decision for every single decisional situation. Afterwards, in Section 4, we provide a description of the suggested process application in the fight against the new coronavirus, known as COVID-19. In Section 5, we present a comprehensive assessment of our work in comparison with the existing works. In Section 6, we summarize the suggested contribution and we propose future works.

2. THEORETICAL BACKGROUND
A major challenge in constraining the COVID-19 horizontal spread is the early detection of the infected cases so as to apply the decision of quarantine. Enabling decision-maker to choose on their own the most suitable decision to make (isolation, medical care, hospitalization, etc.) represents the objective of the system of autonomous decision-making. Depending method accuracy, such system grants the lessening of the infected cases number, the assistance of the users in the decision-making in addition to the reduction of the high loads received by units of medical services. This section introduces the theoretic concepts needed for understanding the autonomous decision-making. The proposed process is composed of three different modules for predicting the severity of the case basing on the identified symptoms through the use of the technique of BNs, afterwards specifying the adequate decision in an autonomous way and basing on the achieved results utilizing the method of association rules.

2.1. Construction of BN
BN represents a graphical and probabilistic model. Technically, BN is signified in the form of the pair (Gr, P) in which Gr is established in the form of directed acyclic graph (DAG), whereas P denotes the distributions of conditional probability. The DAG, denominated by the pair (N, E), signifies the relationships between the N nodes. The aforementioned nodes denote the characteristic variables that are connected by a directed edges E set that designate the nodes dependencies. Every edge E betwixt two nodes Nj named children and Ni called parent characterizes directed dependence otherwise ‘causation’ relationship. The distributions of conditional probabilities included within the CPT show for every single node its conditional probabilities provided respective parents [10]. Using the subsequent formula, the distribution of the joint probability for the BN is calculated:

\[ P(Gr) = P(N_1, N_2, \ldots, N_m) = \prod_{i=1}^{m} P(N_i | Pa(N_i)) \]

In which Pa(N) denotes the node N’s parent and m signifies the overall nodes number contained within the graph.

For the construction of the BN model, two stages for learning its parameters and structure need to be performed. On the structure learning phase (qualitative level), the BN structure learning signifies the process of extraction of the relationships between the nodes. The relationships that were discovered denote the dependencies between the random variables that could have continuous or discrete values. On the parameters phase (quantitative level), learning the network parameters is implemented through the calculation of the distributions of the probability (CPT). The latter describes the relationships that are discovered betwixt random variables. In fact, application in a specific field requires the construction of a special BN model made specifically for this domain.

For solving the BN structure learning problem, which presents a serious challenge, numerous methods founded on the knowledge of the expert or on the data have been suggested. Numerous works like [11–13] have employed the expert-based approach for the purpose of reducing the space of the search’s complexity. Since the creation of the BN structure basing only on the aforementioned approach is really difficult, many errors could be produced [14]. Alternatively, various works such as
have used the data-driven approach for BN structure learning. Consequently, relying only on input data to learn the topology of the BN is deemed as an NP difficult problem due to the candidate graphs number, which is super exponential in the number of the variables [10, 21]. Identifying the structure, which ideally suits the data among the candidate structures, is accomplished through the use of employing the most elevated score. In this context, various functions of scoring have been presented such as the Akaike information criterion score, Bayesian Dirichlet equivalence uniform score, the K2 score and the Bayesian information criterion (BIC) score [22].

The K2 method is among the extensively used data-oriented methods for learning the BN structure. The aforementioned method uses an approach of score search to produce the final resolution. Based on the greedy approach, the method of K2 searches within the entire space to build the topology of the network. Upon the detection of the nodes’ correct order, the K2 generates better accuracy compared to similar algorithms. As shown in Fig. 11 in the appendix, the principal of this algorithm is to commence with an empty graph and then to test dependency relations between nodes to determine each nodes’ parents (the maximal parents number is provided as input). For doing that, the major steps could be recapitulated as follows:

1. Starting with node number one that cannot take parents.
2. Calculating for every node its local score while knowing the parent of the node contained in the previous nodes.
3. Choosing the parents allowing the maximal score.
4. Appending the edges betwixt detected parents and the node.

The main step of K2 method is to select the parents from the set of preceding nodes. Inspired by this idea, we describe the improved version called IK2-BN method [32].

### 2.2. Association rules method

Association rules, which was presented in 1990, is considered to be among the most significant data mining techniques [23]. The aforementioned technique enables the discovery of the association rules through the reflection of the interesting association or correlation betwixt items in a large data number [24]. Associative rules are made in the form of an implication $x \rightarrow y$ where $x$ and $y$ contain one item or an items set. The interpreting of this implication $x \rightarrow y$ is the following; ‘if there exists $x$ subsequently the item $y$ is produced’. For every rule, the parameters like the confidence and the support need to be specified. For discovering the rules of association, the algorithm of Apriori has been suggested, and it has the fourth rank among the 10 leading methods of data mining [25, 26]. This algorithm’s appellation is inspired from the frequent items’ prior knowledge utilization. In order to determine frequent items, the Apriori algorithm does the analysis of the dataset several times [27]. Furthermore, the FP-Growth algorithm [28] (shown in Fig. 12 in the appendix), which represents an enhancement of the Apriori algorithms, discovers the rules of association in a rapid way. Moreover, the FP-Growth algorithm is in fact more efficient than the Apriori algorithm [29–31]. This former algorithm is founded upon an approach of divide and rule, and it comprises these following steps:

1. The scanning of the transactional database, and the establishing of table $T$, which corresponds to the specific minimum support.
2. The scanning of transactional database again and the constructing of the FP-Tree basing on $T$.
3. The mining of the produced PF-tree in order to determine if every single subtree contains one single path: if that was the case, frequent sets of items are found from the single path. If not, it continues the search for the single path within corresponding sub-trees.

Constructing the FP-Tree is in fact the essential phase in the method of FP-Growth. Furthermore, the major feature FP-Tree in fact is addressed this algorithm’s data structures. Owing to its superior effectiveness as well as its speed for big dataset, we opted to use this algorithm in our paper for the extraction of the associative rules for the purpose of constructing the process of autonomous decision-making. In the subsequent parts of this paper, we provide more details about how to utilize the algorithm of FP-Growth and the model of BN to give help to the decision-maker.

### 3. PROPOSED DECISION-MAKING PROCESS

In Fig. 1, we present the proposed process for autonomous decision-making in order to fight against this pandemic. As shown, three modules compose our process; the first module is for the construction of the BN model, the second module is the association rules generation and the third one is the decision-making module. The build of BN model is based on data (given dataset).

The BN construction consists of structure and parameter learning phases based respectively on the improved K2 algorithm [32] and the EM algorithm. Then basing on the forward inference, the new scenario is identified. To discover the utility rules that involve the frequent set of items, FP-Growth algorithm is utilized. The aim of final module is to find the nearest object to each generated one basing on the pairwise comparison. As illustrated in Fig. 1, the generated rules are used for selecting the adequate decision. This decision is attributed to the new object that represents the new patient in our case. The patient status is determined basing on the inference method. The illustrated feedback designates an important step in our process for the future scenarios, which allows the update of the dataset by adding the new object having the adequate decision.

#### 3.1. Module 1: Bayesian model construction

To construct the probabilistic graphical model basing on the data approach, we select the variables that are pertinent for
analyzing the coronavirus infection. These variables involve the factors and the symptoms influencing the diagnosis. For this medical application, we build the Bayesian model. At the first, we begin with the phase of the structure learning, which is realized using the improved version of K2 algorithm [32]. The IK2-BN algorithm [32] applies the greedy search approach in a large space of possible graphs. The performance of the K2 algorithm depends mainly on the input nodes order, and it is the same weak point of IK2-BN. To bypass this problem, the expert gives the adequate order as an input for the used algorithm. The main idea of the ameliorated algorithm can be visualized as in Fig. 2.

As illustrated in Fig. 2, the expert gives the node order as the input order for the IK2-BN [32]. Then, the search space is partitioned into two search subspaces. For instance, for the fifth node (8), its parents can be found in its predecessors (4, 2, 3 and 7) and its children can be determined in its successors (6, 1, 5 and 9). According to [32], the IK2-BN algorithm includes two steps: the first step allows to searching the parents in the first subspace (parents’ subspace) and to searching for children (in the subspace of children). The IK2-BN algorithm can be explained as follows:

For each node:
1. Calculate the local score of this node.
2. The parents set is initialized as empty set.
3. For searching its parents: In the predecessors set, the opted node that increases further the score is inserted in the parents set.
4. The execution of the previous step carries on until the specified maximum number of parents is obtained or there is no node allowing the increase of the score.
5. The children set is initialized as empty.
6. For searching its children: In the successors set, the node that decreases further the score is inserted in the children set.
7. The execution of the previous step carries on until the specified maximum number of children is obtained or there is no node allowing the decrease of the score.
8. Finally, the determined node can be assigned as child if the corresponding score is lower than the maximum score of the parent (using the absolute value).

The previous description of the IK2-BN algorithm [32] can be formulated by Fig. 3.
FIGURE 3. The IK2-BN algorithm steps [32].

As a resume, we can note that the IK2-BN algorithm [32] includes two main parts: (i) the search of the parents (colored in green) and (ii) the search of the children (colored in blue) for each node. This extension of K2 algorithm aims to produce the closest structure to the original one of the BN. After the generation of the learnt structure, the EM algorithm allows the learning of the BN parameters. This step permits the quantification of the dependencies between the nodes composing the BN topology.

3.2. Module 2: generation of the association rules

This module aims to mine the data for discovering the utile rules. For this aim, we employed the algorithm FP-Growth. This algorithm is characterized by the scalability and the efficiency in mining the FPs. The functioning of this method starts by the database compression with the creation the FP-Tree that embodies the most frequent items. The second step is to divide the compressed database into conditional databases where the association between each one and the FP is performed. Finally, each database is separately mined.

The core of this algorithm is the building of the FP-Tree that enables it to extract directly the frequent items (itemset). This latter is described in Fig. 4.

The frequent itemsets having the minimum support are utilized to produce rules with high confidence (where confidence ≥ minimum confidence). As the following section illustrates, relying on the generated rules, the adequate decision is attributed to the new patient.

3.3. Module 3: decision-making

To autonomously make the adequate decision based on the approach variable, we based our process on the exploitation of the generated rules. The proposed approach is founding on the calculation of the distance measure. Thence, the distance between each pair of objects allows the specification of the corresponding decision. The features of these objects have considerable values, which are utilized for calculating the distance.

To autonomously attribute the proper decision, the algorithm allows the following:
hybrid algorithms were applied in the medical field, especially can be used to update the initial database for future use. The rules methods to autonomously make the proper decision that allows the generation of the BN structure shown in the above-indicated variables. As shown in Fig. 5, to learn the discrete BN using the proposed method, the used variables were discretized. The used dataset involves 24 variables where 21 are the clinical symptoms and we added the variables age, proposed decision and disease classes. In this work, our aim is to analyze the symptoms using the probabilistic graphical representation, to predict the class of the new patient(s) using the BN model, and then the adequate decision is autonomously specified. To learn the BN model, the training dataset includes 616 patients and the test dataset involves 203 samples. In the following subsection, we will describe the BN construction phase using the specified approach.

4. APPLICATION IN MEDICAL CONTEXT

4.1. Used dataset

The used dataset [33] includes 819 patients. This dataset was pre-processed before the exploration by cleaning the missing data and inconsistent data relying on the previous researches like [2, 34, 35], [36, 37]. The objective of this work is to discover the COVID-19’s symptoms and to explore their importance for specifying the severity of the case that is useful for making the adequate decision. Generally, the coronavirus disease-19 has clinical manifestations including fever, cough, fatigue, muscle pain, diarrhea, respiratory distress, pneumonia, etc. [2]. According to the three works [35–37], the major symptoms of coronavirus are fever, shortness of breath, myalgia, cough, diarrhea, sputum production, chest pain, hemoptysis and sore throat. These clinical symptoms are detected at least in 1 and at most in 98 patients among 100 patients [35–37]. Additionally, we consider the age factor and we add the severity classes and the attributed decisions to classify the cases as mild, moderate and severe. In Fig. 5, we visualize the discretized variables that are useful for the classification of the disease (mild or moderate or severe).

The patients can be classified into three classes named disease classes, which are mild, moderate and severe basing on the above-indicated variables. As shown in Fig. 5, to learn the discrete BN using the proposed method, the used variables were discretized. The used dataset involves 24 variables where 21 are the clinical symptoms and we added the variables age, proposed decision and disease classes. In this work, our aim is to analyze the symptoms using the probabilistic graphical representation, to predict the class of the new patient(s) using the BN model, and then the adequate decision is autonomously specified. To learn the BN model, the training dataset includes 616 patients and the test dataset involves 203 samples. In the following subsection, we will describe the BN construction phase using the specified approach.

4.2. BN model construction

To diagnosis the COVID-19 cases, the BN model is built using the above-described dataset and by employing and ameliorating the Toolbox FullBNT-1.0.3 (matlab project). For this, the learning phases of the BN structure and the BN parameter have to be performed. Finally, the application of the IK2-BN algorithm allows the generation of the BN structure shown in Fig. 6. This topology represents the dependency relationships between the given variables (in Table 1). As specified in Fig. 1, the expert gives the nodes order in order to apply the IK2-BN algorithm that is sensitive to this order (indicated in Table 1).

The learnt DAG represents the discovered relations between the discrete variables. Each candidate graph has a BIC score as explained in Section 2.1. According to this score, we can determine how the learnt topology is closest to the original structure of the BN. The algorithm IK2-BN has as input the order entered by the expert and then produces the DAG shown in Fig. 6. This latter illustrates the dependency relationships between the 24 variables where the resulted BIC score is—311 387. The nodes corresponding to these variables are directly or indirectly connected to the node ‘disease class’, which demonstrates how the selected symptoms are important for detecting the case severity. In the network topology, the used variables are represented as a child node and/or a parent node to specify
the dependency relations. From this graph, for each node, the first child node is defined as primary symptom and the second child node (child of the child) is defined as secondary symptom. At the quantitative level, we employed the EM method for estimating the probability distributions of the BN. The generated BN model is outlined in Fig. 7 and Fig. 8. To explain the dependency relations, we specify some values in order to indicate the characteristics of each disease class.

**TABLE 1.** The order entered by the expert.

| Node name         | Node number | Node name      | Node number |
|-------------------|-------------|----------------|-------------|
| Disease class     | 1           | Headache       | 13          |
| Age               | 2           | Joint pain     | 14          |
| Fever             | 3           | Throat pain    | 15          |
| Cough             | 4           | Runny nose     | 16          |
| Breathing         | 5           | Abdominal pain | 17          |
| Pneumonia         | 6           | Diarrhea       | 18          |
| Malaise           | 7           | Loss appetite  | 19          |
| Fatigue           | 8           | Muscle pain    | 20          |
| Chest pain        | 9           | Sputum         | 21          |
| Flu symptoms      | 10          | Vomiting       | 22          |
| Thirst            | 11          | Sneeze         | 23          |
| Chills            | 12          | Nausea         | 24          |
As Fig. 7 shows, the definition of no-symptoms means that the 0 value is checked in each node. Consequently, as highlighted using red circle, the probability of mild class is equal to 97.42%. However, for the class moderate, the probability is 0.65% and for the class severe is 1.93%. To explain these two values that are not expected, we emphasize the impact of the age factor. This latter has a direct effect on the identification of the severity of the case. Furthermore, the marginal probability distributions prove that the secondary symptoms (directly linked to the node ‘Malaise’) were not detected during the diagnosis (probability of 0 is higher than 99%).

Figure 8 represents the second scenario, which demonstrates the role of the pneumonia symptom in classifying the case as severe. As illustrated, if the pneumonia is present (see the node circled in red), the probability of severe class is 99.74%. This factor is notably detected in the patients having age between 51 and 60 years. Basing on the probability distribution, we highlight that the probabilities of high fever, mild fever, fever and no fever are respectively 0.07, 7.5, 0.06 and 92.32%. The learnt DAG demonstrates that if the patient has pneumonia, the fever has no importance (the probability of no fever is 92.32%) in specifying the disease class.

To predict the class of the disease for the new patients (testing the learnt model), we used the method of variable elimination for the inference phase. Then, the extracted rules using the association rules method is the main step for autonomously making the adequate decision as described in the following section.

4.3. Autonomous decision-making

Based on the proposed modules, we aim to integrate them to ensure the autonomous process of decision-making. In this module, the data mining is realized using the RapidMiner tool. In fact, for discovering the association between the given symptoms, the disease classes and the assigned decisions, the FP-Growth algorithm is utilized for analyzing the decisional situations. Basing on the examined factors including the clinical features, the patient age and the disease class (severity of the case), the used method of association rules mining easily produces clear description of the situation. Determining the adequate decision has crucial role in the control and the progress of the patient status. There are multiple decisions that are assigned to the COVID-19 cases such as the admission for the treatment at hospital, the seek care, the quarantine, the sought treatment at hospital, the medical attention and so on. Among the above-defined symptoms, we select the patient age, the disease’s class and the attributed decisions for analyzing the association rules. In the experimentation, we used 54 patients as samples for training the system and we reserved 270 patients to test the learned model. The opted samples involve a diverse symptoms added to the age factor and the assigned decisions for providing a potent foundation to achieve the test phase.

As observed, a specific decision can be attributed for different patients. Accordingly, the algorithm was performed multiple times in order to produce relevant association rules that involve different decisions. To examine these decisions, we assigned to the minimum support (min-sup) and minimum confidence (min-conf) the values 0.01 and 0.4. Indeed, several rules were generated; for instance, we presented in Fig. 9 four rules including 10 items (or more) for minimizing the rate of matching of input object and saved objects.

As illustrated in Fig. 9, the produced rules is formulated as ‘IF F1= V1, F2= V2 . . . Fn= Vn, . . . THEN D=decision’. Where F represents the Feature, V designates the Value of each feature and D is the decision given to the patient according to the values of these features. As shown, each rule corresponds to the object characterized by n features (the number of features is at least 9) and a proper decision. The output of the BN model represents an important factor in the process of decision-making. For this reason, all selected rules involve the value of the case’s severity (predicted class) for making the adequate decision. The analysis of 258 rules is done to define the occurrence of factors (or a combination of many factors) that are among the most frequent factors (illustrated in Table 2).

To automatically make the adequate decision, we applied the approach based on variable for an example of object having the following characteristics: ‘Joint pain = 0, Fatigue = 0, Malaise = 0, Runny nose = 0, Breathing = 0, Chills = 0, Pneumonia = 0, Cough = 0, Disease class = 1, age = 3’. This object has no symptoms, and the discrete value of age is 3 (means that the age is between 21 and 30) and classified as mild case. As indicated, this object is characterized by the values attributed to the 10 features. Then, to define the decision that is adequate for this object, we performed the following steps:

1. The objects having the class mild (class disease 1) are ordered using the value of the confidence.
2. The distances between the given object and the set of ordered objects are calculated.
3. The distance attributed to the input is the decision of the object having the maximum confidence and the minimum distance. For the given example and as shown in Fig. 10, the distance between it and the selected object (framed in red) is zero. Consequently, the decision made for the new object or patient is 1.
5. EVALUATION AND DISCUSSION

The integrated methods used for autonomous decision-making consist in the Bayesian model construction and the mining of the association rules.

5.1. Evaluation of the BN model

To construct the Bayesian model, we employed an improved method based on data approach for BN structure learning. The IK2-BN algorithm [32] extends the K2 algorithm by the subdivision of the search space for searching the children and the parents. As above-described, for making the proper decision, we introduced novel variable-based approach by computing the distance between used objects and the input one. The efficiency evaluation allows us to demonstrate the advantages of the proposed method compared to the previous works.

As above-described, we proposed to use the improved K2 method for structure learning named IK2-BN algorithm. It consists of the parents’ search phase and the children’ search phase using the score [32]. To test the performance of IK2-BN algorithm for structure learning, we employed the benchmark networks described in Table 3.

For evaluating the performance of the IK2-BN algorithm, the structure difference method is used, which is based on
determining the number of edges that are correct, added, deleted and reversed. The difference is calculated by detecting the errors in the learnt DAG compared to the original network topology. The number of errors is defined by summing the number of added edges, of reversed edges and of deleted edges. These metrics are represented using, respectively, the terms AE, RE and DE. In addition, we used the terms Structural Difference (SD) and Correct Edges (CE). To evaluate the efficiency of IK2-BN algorithm, it was executed thousands of times using the databases ASIA-1000 (Network-Number of cases), ASIA-3000, ALARM-1000 and ALARM-3000. In Table 4, we exhibit the obtained experimental results that represent containing the best deviation of the evaluation metrics.
As illustrated in Table 4, we represent for each network with specific number of samples the values of the metrics indicating the difference between the learned network topology and the original one. By analyzing the number of CE, we notice that the used algorithm allows the generation of 100% of the edges of the initial structure of ASIA network for 1000 and 2000 cases. In addition, it is able to learn the correct topology of CANCER network for the cases 1000, 2000, 5000 and 10 000. However, for ASIA (10 000 and 5000), the number of DE and AE are one edge. The used IK2-BN algorithm allows to determine seven correct edges in all the executions. Accordingly, we can note the efficacy of this algorithm for the small networks. In addition, for ALARM network (1000 and 2000 samples), it can correctly learn 42 edges among 46 edges with 4 wrongly deleted edges, no edge is reversed and 4 edges are incorrectly added. For ALARM-1000 and ALARM-10000, only one edge is added and four edges are deleted comparing to the original structure of ALARM network.

Moreover, we introduce a comparison of the results produced by the used algorithm with other works employed for solving the problem of BN structure learning. Thence, we compare the experimental results of the improved K2 algorithm with Ko and Kim algorithm [38], Tabar *et al.* algorithm [17] and Ai [18]. Each algorithm is performed 100 times using the well-known networks, which are ASIA and ALARM (with 1000, 2000, 5000 and 10 000 samples). As illustrated in the Table 5, the higher values are indicated in bold and the second best values are represented using starred values to exhibit the efficacy of employed algorithm number of samples.

Table 5 outlines the structure difference for each network using the indicated algorithms. Asia network is a small network, and its topology is composed of eight nodes and eight edges. The used algorithm allows to determine correctly 87.5% of the initial structure. As proved by the values in Table 5, the IK2-BN algorithm gives important results compared to the other methods for ASIA network. In addition, as marked in bold, it generates the lowest values, which are 0 or 1 incorrectly AE, 0 DE and 0 RE. The number of the edges that are incorrectly deleted is 1, which is the second best value. Furthermore, to demonstrate the algorithm efficacy, we introduce a comparative study of the experimental results obtained for ALARM network with those of the other algorithms.

As illustrated using bold values for ALARM network, the number of CE learned by the improved version of K2 algorithm is higher than the others. Additionally, the IK2-BN method allows to produce the lower number of RE and AE. For ALARM network, the number of errors generated by the used algorithm is five, eight, five and five, respectively, for 10 000, 5000, 2000 and 1000 cases. In addition, this algorithm allows the generation of the lowest number of accidentally AE and RE. The IK2-BN returns the highest number of DE compared

---

**TABLE 2.** Frequent factors in the generated rules.

| Factor                      | Occurrence |
|-----------------------------|------------|
| Disease class               | 100%       |
| Joint pain                  | 93.02%     |
| Chest pain                  | 93.02%     |
| Disease class, age          | 84.49%     |
| Joint pain, chest pain      | 76.74%     |
| Joint pain, chest pain, disease class | 65.11%     |
| Joint pain, chest pain, fatigue | 62.01%     |
| Joint pain, disease class   | 60.46%     |

---

**FIGURE 10.** Selecting the adequate decision.
### TABLE 3. Used networks.

| Benchmark network | Samples                  | Nodes | Edges |
|-------------------|--------------------------|-------|-------|
| ALARM             | 250-500-1000-2500-5000-10000 | 37    | 46    |
| ASIA              | 250-500-1000-2500-5000-10000 | 8     | 8     |
| CANCER            | 250-500-1000-2500-5000-10000 | 5     | 4     |

### TABLE 4. Experimental results.

|          | AE | DE | RE | CE |
|----------|----|----|----|----|
| ASIA     |    |    |    |    |
| 1000     | 0  | 0  | 0  | 8  |
| 2000     | 0  | 0  | 0  | 8  |
| 5000     | 1  | 1  | 0  | 7  |
| 10000    | 1  | 1  | 0  | 7  |
| CANCER   |    |    |    |    |
| 1000     | 0  | 0  | 0  | 4  |
| 2000     | 0  | 0  | 0  | 4  |
| 5000     | 0  | 0  | 0  | 4  |
| 10000    | 0  | 0  | 0  | 4  |
| ALARM    |    |    |    |    |
| 1000     | 4  | 4  | 0  | 42 |
| 2000     | 4  | 4  | 0  | 42 |
| 5000     | 1  | 4  | 0  | 42 |
| 10000    | 1  | 4  | 0  | 42 |

### TABLE 5. Comparison of the gained results with the results of three algorithms.

| For ASIA database | Edge | IK2-BN Ko and Kim [38] Tabar et al. [17] Ai [18] | For ALARM database | Edge | IK2-BN Ko and Kim [38] Tabar et al. [17] Ai [18] |
|-------------------|------|---------------------------------|-------------------|------|---------------------------------|
| ASIA 1000         | CE   | 7                               | 5                 | 4    | 4                               |
|                    | DE   | 1*                              | 0                 | 0    | 1                               |
|                    | RE   | 0                               | 3                 | 4    | 2                               |
|                    | AE   | 0                               | 1                 | 0    | 3                               |
|                    | SD   | 1                               | 4                 | 3    | 6                               |
| ASIA 2000         | CE   | 7                               | 5                 | 5    | 4                               |
|                    | DE   | 1*                              | 0                 | 0    | 1                               |
|                    | RE   | 0                               | 3                 | 3    | 2                               |
|                    | AE   | 0                               | 1                 | 0    | 3                               |
|                    | SD   | 1                               | 4                 | 3    | 6                               |
| ASIA 5000         | CE   | 7                               | 5                 | 5    | 4                               |
|                    | DE   | 1*                              | 0                 | 0    | 1                               |
|                    | RE   | 0                               | 3                 | 3    | 1                               |
|                    | AE   | 1                               | 1                 | 1    | 3                               |
|                    | SD   | 2                               | 4                 | 4    | 6                               |
| ASIA 10000        | CE   | 7                               | 5                 | 6    | 5                               |
|                    | DE   | 1*                              | 0                 | 0    | 1                               |
|                    | RE   | 0                               | 3                 | 3    | 1                               |
|                    | AE   | 1                               | 1                 | 1    | 3                               |
|                    | SD   | 2                               | 4                 | 4    | 5                               |

Bold value indicates the best result compared to the cited methods.
TABLE 6. Comparative study of the three algorithms (IK2-BN, K2 and Evo-BN).

|                | IK2-BN algorithm | Evo-BN algorithm | K2 algorithm |
|----------------|------------------|------------------|-------------|
| ASIA 1000      |                  |                  |             |
| AE             | 0                | 1                | 2.95        |
| DE             | 1                | 2                | 2.05        |
| RE             | 0                | 0.85             | 2.95        |
| CE             | 7                | 6                | 5.95        |
| ALARM 1000     |                  |                  |             |
| AE             | 4*               | 1.1              | 16.6        |
| DE             | 4*               | 2.4              | 15.75       |
| RE             | 0                | 1.2              | 16          |
| CE             | 42               | 39.6             | 26.25       |
| ALARM 5000     |                  |                  |             |
| AE             | 1                | 2.3              | 9.96        |
| DE             | 4                | 4.1              | 9.45        |
| RE             | 0                | 2.1              | 9.6         |
| CE             | 42               | 37.9             | 32.55       |

Bold value indicates the best result compared to the cited methods.

TABLE 7. Experimental results comparison to five algorithms.

|                | IK2-BN algorithm | NDPSO-BN algorithm | BNC-PSO algorithm | ABC-B algorithm | SCA algorithm | MMHC algorithm |
|----------------|------------------|---------------------|-------------------|-----------------|---------------|----------------|
| ASIA 3000      |                  |                     |                   |                 |               |                |
| AE             | 0                | 0                   | 0                 | 0               | 0             | 0              |
| DE             | 0                | 1                   | 1                 | 1               | 1             | 1              |
| RE             | 0                | 0                   | 0                 | 0               | 1             | 1              |
| SD             | 0                | 1                   | 1                 | 1               | 2             | 2              |
| ALARM3000      |                  |                     |                   |                 |               |                |
| AE             | 2*               | 1                   | 3                 | 1               | 17            | 2              |
| DE             | 4*               | 1                   | 1                 | 1               | 28            | 1              |
| RE             | 0                | 0                   | 1                 | 2               | 4             | 9              |
| SD             | 6                | 3                   | 6                 | 4               | 49            | 12             |
| ALARM5000      |                  |                     |                   |                 |               |                |
| AE             | 1*               | 1                   | 1                 | 0               | 24            | 1              |
| DE             | 4                | 1                   | 1                 | 1               | 27            | 2              |
| RE             | 0                | 0                   | 0                 | 0               | 6             | 6              |
| SD             | 5                | 2                   | 3                 | 1               | 57            | 9              |

Bold value indicates the best result compared to the cited methods. As illustrated in the Table 5, the higher values are indicated in bold and the second best values.

to the other methods (Tabar et al. algorithm [17], Ko and Kim algorithm [38] and Ai algorithm [18]).

As shown in Table 5, the used IK2-BN algorithm outstrips the other improved versions of K2 algorithm. As demonstrated, the used algorithm is efficient for solving the BN structure learning problem as proved using the terms of CE and SD. Furthermore, we describe in Tables 6 and 7 new comparative study using other proposed methods. Firstly, we represent in Table 6 the comparison of IK2-BN algorithm, original K2 algorithm and Evo-BN algorithm [39].

For evaluating the algorithm performance, we used ASIA network for 1000 samples and ALARM network for 1000 and 5000 samples, and then IK2-BN algorithm was executed 30 times for learning the BN structures. Table 6 describes the experimental results produced by the IK2-BN and the Evo-BN and K2 algorithms studied in [39]. Our aim is to compare the performances of the used algorithm, the Evo-BN algorithm and the K2 algorithm.

The results shown in Table 6 demonstrate that the used algorithm outstrips the two algorithms for ASIA-1000. It allows to obtain 7 CE, 0 RE, 0 AE and 1 DE. For ALARM for 1000 cases, it produces 0 RE that represents the lower value, 42 CE and 4 AE and the same number of DE that is the second best result. Additionally, our method produces the lowest number of added, deleted and reversed edges and the highest correct edges’ number for ALARM-5000. Finally, we can prove that the improved version of K2 algorithm named IK2-BN algorithm remarkably outperforms its original version.

To further demonstrate the efficacy of the IK2-BN algorithm, we propose a comparative including more algorithms. This latter allows the demonstration of its relevance in learning the BN structure. Table 7 shows the experimental results
decision is the same decision of the object having the lowest distance and the highest confidence in the objects list. Finally, for evaluating the efficacy of the decision-making method, we propose to determine its precision of attributing the proper decision. Therefore, we define two principal categories of decisions, which are (i) isolation category including quarantined at home with medical care, quarantined, no medical attention and medical attention; (ii) hospitalization category involving admitted at hospital for treatment, admitted for isolation in hospital, hospitalized, sought treatment at hospital, seek care, quarantined at hospital, immediately taken to hospital for isolation, hospitalized and admitted for isolation and treatment.

For testing the variable approach-based method, 10 patients have been randomly selected from the rest of the patients (270 cases). Then, we apply the algorithm used for determining the adequate decision basing on calculating the lowest distance. For the selected patients, we indicate if the attributed decision is correct, adequate or incorrect for each patient. In Table 9, we illustrate the generated results.

As shown in Table 9, we exhibit the evaluation of the decisions made for the 10 patients using the proposed decision-making process. If the attributed decision is the same decision of the object, it is considered as correct decision. The decision is an adequate decision if it is not the correct one, but it is considered as adequate for the new patient. However, the decision is mentioned as incorrect if it is not proper for the given patient. As illustrated, the accuracy of specifying the proper decision is 100%, where 30% is for correctly attributing the decision.

6. CONCLUSION

Given the unexpected effects that the coronavirus had on human existence in a short amount of time, we used in this work the methods of machine learning to analyze the major factors to predict this disease’s severity and to propose the adequate decision. For analyzing the primary factors of the coronavirus disease, we have proposed a new BN model that is based on enhancing the algorithm of structure learning. Subsequently, through the use of the method of association mining, our proposal made the most suitable decision for COVID-19.

---

**TABLE 8.** Confusion matrix.

| Classifier results | Class 1 | Class 2 | Class 3 | Classification overall | Precision |
|--------------------|---------|---------|---------|-------------------------|-----------|
| Class 1            | 99      | 11      | 5       | 115                     | 86.087%   |
| Class 2            | 11      | 7       | 5       | 23                      | 30.435%   |
| Class 3            | 0       | 0       | 65      | 65                      | 100%      |
| Actual overall     | 110     | 18      | 75      | 203                     | 86.667%   |
| Recall             | 90%     | 38.889% | 86.667% |                         |           |

---

generated by the algorithms IK2-BN, NDPSO-BN [19], ABC-B [40], BNC-PSO [20], SCA [41] and MMHC [42] for each database (repeated 500 times). This table illustrates the difference between the original network topology and the learned structure.

As shown in the table, the ability of used algorithm in learning the correct structure surpasses the other algorithms for ASIA network for 3000. In addition, for ALARM network (for 3000 cases), the IK2-BN algorithm generates two added edges and four deleted edges representing the second lower values. The used algorithm produces 0 RE that represents the best result as marked in bold. For ALARM network (for 5000 samples), the ABC-B algorithm allows to discover the closest network topology to the initial one with 0 AE, 1 DE, 0 RE and 1 SD. The other algorithms (NDPSO-BN, BNC-PSO and IK2-BN) generate less important results. The improved K2 algorithm allows to obtain 1AE that represents the second best result, 4 DE and 0 RE.

As proved by the comparative of the experimental results, the used IK2-BN algorithm generates the closest structure to the original one for each network. The used algorithm is enable to produce at least 7 correct edges among 8 for ASIA network and 42 correct edges are learned among 46 for the well-known networks. Finally, for evaluating the efficacy of the BN model in the prediction of the correct class, the used dataset was partitioned into two datasets: for the training that includes 616 cases and for the test that includes 203 patients.

As shown in the confusion matrix (Table 8), the method's accuracy is 0.84, the precision average is 0.72 and the recall average is 0.71. Accordingly, the results generated by the BN model are acceptable for the aim of predicting the severity of the case (disease class: mild, moderate and severe).

5.2. Evaluation of the decision-making system

As above-described in Section 3, the module of autonomous decision-making is highly reliant: at the first hand on the class that is generated by the Bayesian model and at the second hand on the rules quality. For making the proper decision for the input patient, it is important to review all the generated rules that have the same disease class. Then, the assigned
infected cases in an autonomous way. First, in order to find the hidden relations between variables, we decided to introduce a novel K2 algorithm version namely (IK2-BN) through the extension of the space of search for parents as well as children. Afterwards, basing on the algorithm of FP-Growth producing the useful rules, we determined the most suitable decision. In order to do this, our algorithm does the calculation of the distance separating features to identify the closest object. As cited above, the proposed decision-making module have been tested using 10 patients and it gives either the exact decision (30%) or the adequate decision for the case (70%). Lastly, we provided the results of the experiments, which appear to be acceptable as primary proposition which needs concrete application in order to be approved. Moreover, the expert’s intervention is essential in considering all the factors concerning the progression of the disease. Our proposed process of autonomous decision-making relies primarily on the precision of the method of classification. For the future work, the used decisions should be ameliorated through the use of specific treatments. Furthermore, we suggest to expand the size of the dataset to improve the classification algorithm’s accuracy.

SUPPLEMENTARY MATERIAL

Supplementary material is available at www.comjnl.oxfordjournals.org.

DATA AVAILABILITY STATEMENT

Data available on request. The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

[1] Coronavirus disease (COVID-19) pandemic: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed 8 February 2021).

[2] Tan, L., Wang, Q., Zhang, D., Ding, J., Huang, Q., Tang, Y.-Q., Wang, Q. and Miao, H. (2020) Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct. Target. Ther., 5, 1–3.

[3] Wang, S. et al. (2020) A deep learning algorithm using CT images to screen for coronavirus disease (COVID-19). medRxiv.

[4] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z. and Xia, L. (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology, 200642.

[5] Li, L. et al. (2020) Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology, 296, E65–E71.

[6] Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P.D., Zhang, H., Ji, W., Bernheim, A. and Siegel, E. (2020) Rapid AI development cycle for the coronavirus (covid-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv preprint, arXiv:2003.05037.

[7] Wang, L. and Wong, A. (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. arXiv preprint, arXiv:2003.09871.

[8] Al-Qaness, M.A., Ewees, A.A., Fan, H. and Abd El Aziz, M. (2020) Optimization method for forecasting confirmed cases of COVID-19 in China. J. Clin. Med., 9, 674.

[9] Pirouz, B., Shafiee Haghshenas, S., Shafiee Haghshenas, S. and Piro, P. (2020) Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. Sustainability, 12, 2427.

[10] Huang, L., Cai, G., Yuan, H. and Chen, J. (2019) A hybrid approach for identifying the structure of a Bayesian network model. Expert Syst. Appl., 131, 308–320.

[11] Cano, A., Masegosa, A.R. and Moral, S. (2011) A method for integrating expert knowledge when learning Bayesian networks from data. IEEE Trans. Syst. Man Cybern. B Cybern., 41, 1382–1394.
[12] Yet, B., Perkins, Z., Fenton, N., Tai, N. and Marsh, W. (2014) Not just data: a method for improving prediction with knowledge. J. Biomed. Inform., 48, 28–37.

[13] De Campos, L.M. and Castellano, J.G. (2007) Bayesian network learning algorithms using structural restrictions. Int. J. Approx. Reason., 45, 233–254.

[14] Amirkhani, H., Rahmati, M., Lucas, P.J. and Hommersom, A. (2016) Exploiting experts’ knowledge for structure learning of Bayesian networks. IEEE Trans. Pattern Anal. Mach. Intell., 39, 2154–2170.

[15] Tang, C., Yi, Y., Yang, Z. and Sun, J. (2016) Risk analysis of emergent water pollution accidents based on a Bayesian network. J. Environ. Manag., 165, 199–205.

[16] Wang, J. and Liu, S. (2018) Novel binary encoding water cycle algorithm for solving Bayesian network structures learning problem. Knowl.-Based Syst., 150, 95–110.

[17] Tabar, V.R., Eskandari, F., Salimi, S. and Zareifard, H. (2018) Finding a set of candidate parents using dependency criterion for the K2 algorithm. Pattern Recogn. Lett., 111, 23–29.

[18] Ai, X. (2017) Node importance ranking of complex networks with entropy variation. Entropy, 19, 303.

[19] Wang, J. and Liu, S. (2019) A novel discrete particle swarm optimization algorithm for solving Bayesian network structures learning problem. Int. J. Comput. Math., 96, 2423–2440.

[20] Gheisari, S. and Meybodi, M.R. (2016) Bnc-psso: structure learning of bayesian networks by particle swarm optimization. Inf. Sci., 348, 272–289.

[21] Chickering, D.M., Heckerman, D. and Meek, C. (2004) Large-sample learning of Bayesian networks is NP-hard. J. Mach. Learn. Res., 5, 1287–1330.

[22] Jiang, Y., Liang, Z., Gao, H., Guo, Y., Zhong, Z., Yang, C. and Liu, J. (2018) An improved constraint-based Bayesian network learning method using Gaussian kernel probability density estimator. Expert Syst. Appl., 113, 544–554.

[23] Jiang, Y., Zhao, M., Hu, C., He, L., Bai, H. and Wang, J. (2019) A parallel FP-Growth algorithm on World Ocean Atlas data with multi-core CPU. J. Supercomput., 75, 732–745.

[24] Lin, K.W., Chung, S.-H. and Lin, C.-C. (2016) A fast and distributed algorithm for mining frequent patterns in congested networks. Computing, 98, 235–256.

[25] Agrawal, R. and Srikant, R. (1994) Fast algorithms for mining association rules. In Bocca, J.B., Jarke, M., Zaniolo, C. (eds) Proceedings of 20th International Conference on very large data bases, VLDB (Vol. 1215), pp. 487–499. Morgan Kaufmann, San Francisco.

[26] Fünfkranz, J. and Kliegr, T. (2015) A brief overview of rule learning. In Bassiliades, N., Gottlob, G., Sadri, F., Paschke, A., Roman, D. (eds) 9th International Symposium on Foundations, Tools, and Application, RuleML 2015LNCS (Vol. 9202), pp. 56–69. Springer, Heidelberg.

[27] Saabith, A.S., Sundararajan, E. and Bakar, A.A. (2016) Parallel implementation of apriori algorithms on the Hadoop-MapReduce platform-an evaluation of literature. J. Theor. Appl. Inf. Technol., 85, 321.
Appendix

K2 Algorithm

Input: A set of n nodes, an ordering on the nodes, an upper bound u on the number of parents a node may have, and a database D containing m cases.
Output: For each node, a printout of the parents of the node.

for i = 1 \to n 
\pi_i = \emptyset;
P_{odd} = g(i, \pi_i);
OKT0Proceed := true;
while OKT0Proceed and | \pi_i | < u do
let z be the node in Pred(x_i) - \pi_i that maximizes g(i, \pi_i \cup \{z\});
P_{new} := g(i, \pi_i \cup \{z\});
if P_{new} > P_{odd} then
P_{odd} := P_{new};
\pi_i := \pi_i \cup \{z\}
else
OKT0Proceed := false
end if
end while
write(Node: x_i, Parents of x_i: \pi_i)
end for

FIGURE 11. Pseudo code of K2 algorithm.

| FP-Growth pseudo code. |
|-------------------------|
| Input: A dataset containing sequences of places S, minimum support sup\_min |
| Output: Frequent sets of places FS |
| 1  F \leftarrow \emptyset; |
| 2  I \leftarrow \emptyset; |
| 3  foreach x_i \in S do |
| 4  foreach l_i \in x_i do |
| 5  | F[l_i] += 1; |
| 6  | I \leftarrow I \cup \{l_i\}; |
| 7  SortAndRemove(F, sup\_min); |
| 8  root \leftarrow i; |
| 9  foreach x_i \in S do |
| 10  | SortByFrequency(x_i, F); |
| 11  | BuildTree(x_i, root, sup\_min); |
| 12  foreach i \in I do |
| 13  | Growth(root, i, sup\_min, FS); |
| 14  return FS |

FIGURE 12. Pseudo code of FP-Growth algorithm.