

Development of an artificial intelligence diagnostic model based on dynamic uncertain causality graph for the differential diagnosis of dyspnea

Yang Jiao (✉)^{1,*}, Zhan Zhang^{2,*}, Ting Zhang³, Wen Shi⁴, Yan Zhu⁵, Jie Hu⁶, Qin Zhang (✉)⁷

¹Department of General Internal Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China; ²Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China; ³Department of Pulmonary and Critical Care Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China; ⁴Department of Gastroenterology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China; ⁵Institute of Internet Industry, Tsinghua University, Beijing 100084, China; ⁶Department of Medical Administration, Suining Central Hospital, Suining 629000, China; ⁷Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

© Higher Education Press 2020

Abstract Dyspnea is one of the most common manifestations of patients with pulmonary disease, myocardial dysfunction, and neuromuscular disorder, among other conditions. Identifying the causes of dyspnea in clinical practice, especially for the general practitioner, remains a challenge. This pilot study aimed to develop a computer-aided tool for improving the efficiency of differential diagnosis. The disease set with dyspnea as the chief complaint was established on the basis of clinical experience and epidemiological data. Differential diagnosis approaches were established and optimized by clinical experts. The artificial intelligence (AI) diagnosis model was constructed according to the dynamic uncertain causality graph knowledge-based editor. Twenty-eight diseases and syndromes were included in the disease set. The model contained 132 variables of symptoms, signs, and serological and imaging parameters. Medical records from the electronic hospital records of Suining Central Hospital were randomly selected. A total of 202 discharged patients with dyspnea as the chief complaint were included for verification, in which the diagnoses of 195 cases were coincident with the record certified as correct. The overall diagnostic accuracy rate of the model was 96.5%. In conclusion, the diagnostic accuracy of the AI model is promising and may compensate for the limitation of medical experience.

Keywords knowledge representation; uncertain; causality; graphical model; artificial intelligence; diagnosis; dyspnea

Introduction

Dyspnea is a subjective respiratory discomfort that includes a variety of intensity and a heterogeneous nature of sensations [1]. It may be the main clinical manifestation of pulmonary disease, cardiovascular disease, myocardial dysfunction, anemia, and obesity as well as anxiety, anger, and other psychological feelings. Dyspnea may also exist in the same patient with a variety of factors leading to

respiratory discomfort. Although dyspnea is a common chief complaint in outpatient clinics, recent studies showed that even general practitioners in developed countries could only reach an accurate diagnosis rate of 61% in chronic dyspnea patients [2]. The shortage of medical resources in China was even more serious than the developed countries. In resource-limited areas, health care services remain at a low level, and general practitioners are very short of clinical experiences.

For the past decade, research on artificial intelligence (AI) has developed quickly. In the medical field, AI technology can assist in disease diagnosis by analyzing huge amounts of medical data. Research showed that the diagnostic accuracy based on AI reaches or even exceeds the average diagnostic accuracy based on the pathologist or

Received August 8, 2019; accepted February 13, 2020

Correspondence: Yang Jiao, peterpumch@163.com;

Qin Zhang, qinzhang@tsinghua.edu.cn

*These authors contributed equally to this work.

radiologist in the field of pathologic or imaging diagnosis [3–5]. At this stage, technological breakthroughs are mainly focused on the clinical diagnosis related to pattern recognition. However, the development of AI-aided diagnosis system is stuck for diseases whose diagnosis requires complex logical reasoning. Since the 1960s, many expert systems were developed to solve inference problems. However, how to represent and calculate various and complex uncertainties rigorously is not well solved. Dynamic uncertain causality graph (DUCG) is a recently developed AI theory based on expert causal knowledge [6,7]. It has unique advantages in the diagnosis of large-scale complex systems and has been applied to the malfunction analysis and diagnosis in complex industrial systems with a lot of uncertainties. At the same time, preliminary exploration has begun in the intelligent diagnosis of medical diseases [8,9]. The aim of the present study was to establish a disease knowledge base for patients with dyspnea as the chief complaint. With the support of a clinician expert with evidence-based medical thinking and plentiful experience in diagnosis, and combined with DUCG theory, an AI-aided diagnostic model was built and applied to compensate for the shortage of general practitioners' experience in diagnosis and shorten the training cycle of general practitioners, thus effectively saving the cost of health services for the Chinese government.

Materials and methods

Graphical representation of DUCG

The basic ideas of DUCG were presented in References [10–12], in which the virtual random functional event variables along with weights were introduced to represent and quantify the uncertain causalities among variables, resulting in that the logical operation was introduced before the probability calculation, the simplification rules were introduced to simplify the problem and scale based on the received evidence for a specific problem, the number of parameters was reduced and not so sensitive significantly, and the construction of causal knowledge base with uncertainties and inference computation became much easier and intuitive than other models, including able to deal with not only DAG (directed acyclic graph) but also DCG (directed cyclic graph), etc. DUCG is a domain knowledge-based AI system that has been successfully applied in many industry areas, such as nuclear power plants, chemical engineering systems, and aerospace systems, to resolve fault diagnoses and predictions. Similarly, DUCG can also be applied in clinical diagnoses. Different from other AI systems that are usually “black box,” DUCG is explicable. It is based on the causalities among events. For example, dyspnea (event A) can be

caused either by pneumonia (event B) or by bronchitis (event C). By giving different probability values to B causing A and C causing A, DUCG can deduce the probability of pneumonia and bronchitis if a patient acquired dyspnea. Of course, the above is the simplest case. The situation can be much more complicated. A disease can cause many abnormal states in a human body, including symptoms, physical signs, laboratory examination items, and imaging information. These abnormal states can also be caused by different diseases. A DUCG graphical representation system was developed to represent such a complex scenario, as depicted in following.

Variable description

In DUCG, the event variables $V \in \{B, X, D, G, BX, SX, RG, C\}$ are used to represent clinical information and diseases as follows:

B_i variable drawn as a square  represents the basic or root cause event variables. There is no input or cause for B_i variable but at least one output or consequence. Normally, B_i represents diseases.

X_i variable drawn as a circle  represents the consequence or intermediate event variable. X_i variable has at least one input. Normally, it represents symptoms, physical signs, laboratory examination items, imaging information, risk factor, and so forth.

D_i variable drawn as a pentagon  represents the default cause variable of X_i or SX_i . When there is no input for an X_i or SX_i variable (the description is given later), a D_i variable represents the default input indicating the unknown cause.

G_i variable drawn as a logic gate  represents the logic gate variable. Its states represent the state combinations of the input variables.

SG_i variable drawn as a double line logic gate  represents a special logic gate variable whose states indicate the meaningful state combinations of risk factors of a B_i variable. The difference between G_i and SG_i is that only one state of SG_i can be true, given the observed states of risk factors represented by X_i variables.

The black directed arc  is used between a logic gate variable (G_i or SG_i) and its input variables. It does not have a value but only indicates the input variables of logic gates.

BX_i variable drawn as a double line circle  represents the output of the SG_i variable, which is the corresponding B_i variable with revised state probability distribution by SG_i . The change rates from B_i to BX_i is in the SA_i variable between them.

SA_i variable drawn as a double line directed arc 

from an SG_i to a BX_i represents the probability distribution change variable that changes the state probability distribution from B_i to BX_i according to the state of SG_i .

SX_i variable drawn as a hexagon  denotes the special X_i variable representing a clinical golden standard. Its abnormal state directly points to a disease with confidence $0 < \theta_{ij} \leq 1$, $j \neq 0$ (0 indexes the normal state).

RG_i variable drawn as a directed shield  represents the reversal logic gate variable. Its state depends on the state combination of its output variables. Given the states of the output variables, the state of RG_i becomes known and can be treated as a special evidence.

F_i variable drawn as a single line directed arc  represents a weighted causal functional variable from parent variable V_i to child variable X_i or SX_i . $F_{n,i} \equiv (r_{n,i}/r_n)A_{n,i}$, where n indexes the child variable and i indexes the parent variable. $r_{n,i} > 0$ represents the causal relationship intensity between V_i and X_n or SX_n . $r_n \equiv \sum_i r_{n,i} \cdot A_{n,i}$ is an event matrix representing the uncertain functional causality between V_i and X_n or SX_n . $A_{nk,ij}$ is a virtual event and a member of the event matrix $A_{n,i}$, indicating that V_{ij} causes X_{nk}/SX_{nk} , where j indexes the state of V_i and k indexes the state of X_n/SX_n . Correspondingly, $F_{nk,ij}$ is a member of the weighted event matrix $F_{n,i}$. All the variables along with graphical symbols and functional descriptions are summarized in Table 1.

Conditional weighted functional event variables

Similar to the above/ordinary F_i variable, the conditional weighted functional variable is drawn as a dashed directed arc , which is used in the same way as an ordinary F_i , but with an observable validation condition $Z_{n,i}$. $Z_{n,i}$ is an event or a group of events. When $Z_{n,i}$ is observed as true, the dashed directed arc becomes the ordinary directed arc; otherwise, it is eliminated.

Similar to the above/ordinary double line directed arc, the conditional dashed double line directed arc  is used in the same way as the ordinary double line directed arc but with a validation condition $Z_{n,i}$. $Z_{n,i}$ is an event or a group of events. When $Z_{n,i}$ is observed as true, the dashed double line directed arc becomes the ordinary double line directed arc; otherwise, it is eliminated. They are also included in Table 1.

DUCG example and inference method

In the DUCG example shown in Fig. 1, there are three B -type variables, eight X -type variables, three D -type variables, one G -type variable, three SG -type variables, three BX -type variables, one SX -type variable, and one RG -

type variable. In this example, B_1 , B_2 , and B_3 represent three disease variables, and X_9 , X_{10} , and X_{11} are the three risk factors of the three diseases, respectively. The state probability distributions of BX_1 , BX_2 , and BX_3 can be calculated according to the state observation of X_9 , X_{10} , and X_{11} , either known or unknown. The logics are specified in SG -type variables.

Usually, X_i , SX_i , and RG_i are the evidence variables whose states are to be observed. The observed states are the observed evidence E . The inference or diagnosis is to calculate $\Pr\{BX_{kj}|E\}$, given that the state probability distribution of BX_k has been calculated according to the states of the risk factors of B_k as described above. Then, given $E = X_{1,1}X_{2,0}X_{3,1}SX_{4,2}X_{5,0}RG_{6,0}RG_{7,2}$, the suspected occurrence probability of BX_{kj} ($j \neq 0$) is calculated using the equation below:

$$h_{kj}^s = \Pr\{BX_{kj}|E\} = \frac{\Pr\{BX_{kj}E\}}{\Pr\{E\}}. \quad (1)$$

In DUCG, state 0 usually indexes the normal state.

$BX_{kj}E$ and E in Eq. (1) are expanded as the sum-of-products composed of only BX_i and A_i events with r parameters. During the expansion, the logic operations such as inclusions and exclusions are applied. The weights $(r_{n,i}/r_n)$ are also calculated. Given the parameters $b_{ij} \equiv \Pr\{B_{ij}\}$, $a_{nk,ij} \equiv \Pr\{A_{nk,ij}\}$, and $r_{n,i}$, the two probabilities of the two sum-of-products of $BX_{kj}E$ and E in Eq. (1) can be calculated, respectively; thus, h_{kj}^s can be calculated. The rank of BX_{kj} according to h_{kj}^s is the inference or diagnostic results.

Construction of the DUCG knowledge base with dyspnea as the chief complaint

A DUCG knowledge base uses the event variables defined in the above section to represent causalities with uncertainties among the variables in the domain. The result is a directed graph called a DUCG. A common disease spectrum that may cause dyspnea is determined by our professional clinician team and then a disease pool including 28 diseases is formed, as shown in Table 2.

Subgraphs were constructed using the DUCG knowledge base editor of the DUCG Cloud Platform. Each subgraph contains only one B -type variable (representing a disease), combining the disease knowledge, epidemiological data, and clinical diagnostic experience, across disciplines. Fig. 2 is an example of the subgraph of chronic obstructive pulmonary disease (COPD) contained in the DUCG knowledge base editor. All the 28 constructed subgraphs were then combined by fusing the same variables automatically through the DUCG platform editor. The result is the DUCG knowledge base about

Table 1 Functions of variables in DUCG

Type	Symbol	Medical meaning
B_i		The B -type variable indicates a disease
X_i		The X -type variable can represent (1) risk factors, such as smoking and alcoholism; and (2) clinical indications, such as symptoms, physical signs, laboratory examination, and imaging information
BX_i		The BX -type variable stands for a disease affected by risk factors
SX_i		The SX -type variable indicates some laboratory examinations or imaging information through which the disease can be diagnosed directly
G_i		The G -type variable is used to express the logical relationship of clinical indications
RG_i		The RG -type variable is used to express the effect of combinations of clinical indications on diseases
SG_i		The SG -type variable is a special logic gate. It is only used to express the logical relationships between risk factors and diseases
D_i		The D -type variable represents the default input indicating the unknown cause
F_i		The F -type variable indicates the causal relationship intensity between two clinical indications
Conditional F_i		The function of the conditional F -type variable is the same as the F -type variable but with an observable validation condition; if the condition is true, then the variable functions, and vice versa
SA_i		The SA -type variable is used to represent the influence of risk factors on the probability of disease onset
Conditional SA_i		The function of conditional SA -type variable is the same as the SA -type variable but with an observable validation condition

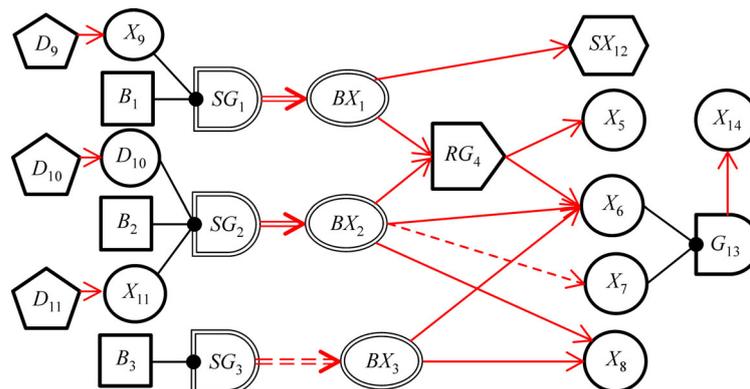


Fig. 1 Example of a dynamic uncertain causality graph.

Table 2 Diseases causing dyspnea in the DUCG knowledge base

<i>B</i> -type variable	Description	<i>B</i> -type variable	Description
<i>B</i> ₁	Carbon monoxide poisoning	<i>B</i> ₁₅	Pericardial effusion
<i>B</i> ₂	Metabolic acidosis	<i>B</i> ₁₆	Hemochromatosis
<i>B</i> ₃	HCM	<i>B</i> ₁₇	End-stage tumor
<i>B</i> ₄	Pulmonary infection	<i>B</i> ₁₈	COPD
<i>B</i> ₅	PAH	<i>B</i> ₁₉	Laryngospasm
<i>B</i> ₆	Interstitial lung disease	<i>B</i> ₂₀	Foreign body in air passage
<i>B</i> ₇	Pulmonary alveolar proteinosis	<i>B</i> ₂₁	Obesity
<i>B</i> ₈	PE	<i>B</i> ₂₂	Scoliosis
<i>B</i> ₉	Heart failure	<i>B</i> ₂₃	Pleural effusion
<i>B</i> ₁₀	HPS	<i>B</i> ₂₄	Asthma
<i>B</i> ₁₁	DCM	<i>B</i> ₂₅	Bronchitis
<i>B</i> ₁₂	Anaemia	<i>B</i> ₂₆	Guillain-Barre syndrome
<i>B</i> ₁₃	Renal failure	<i>B</i> ₂₇	Myasthenia gravis
<i>B</i> ₁₄	Constrictive pericarditis	<i>B</i> ₂₈	Psychology

Abbreviation: HCM, hypertrophic cardiomyopathy; PAH, pulmonary artery hypertension; PE, pulmonary embolism; DCM, dilated cardiomyopathy; HPS, hepatopulmonary syndrome; COPD, chronic obstructive pulmonary disease.

dyspnea (Fig. 3), which is used to diagnose the possible disease of any patient with dyspnea as his/her chief complaint.

The DUCG Cloud Platform was developed by Beijing Tsingrui Intelligence Technology Co., Ltd. to implement the DUCG theory applied in disease diagnoses. Any clinician who is interested in and qualified to test this platform can contact the second or last author to obtain an account and password.

Inference and calculation methods

Once the patient’s clinical information (evidence *E*) is collected, the DUCG Inference Engine begins the calculation process. First, the DUCG is simplified based on *E* according to the simplification rules 1–10 and 16 [11,13]. Through general graph simplification, the variables and causal relationships unrelated to the evidence and hypothesis are deleted to narrow the range of hypotheses for determination and reduce the amount of inference calculation. According to the commonly used assumption of a single disease diagnosis, the simplified DUCG is further decomposed into a group of sub-DUCGs for every *B*_{*kj*} (*j* ≠ 0, representing an abnormal state), with each subgraph retaining only one *B*-type variable. The subgraph can be further simplified by applying the above rules. During the simplification process, the parts that are not related to the evidence *E* and the variables for evaluation will be simplified according to the simplification rules. Some sub-DUCGs along with their corresponding *B*_{*k*} may be eliminated during the simplification. The remaining sub-DUCGs correspond to *B*_{*kj*}, *j* ≠ 0, and all these *B*_{*kj*} compose the hypothesis space. If the hypothesis space has only *B*_{*kj*}, the inference is completed and the disease is diagnosed as *B*_{*kj*}. If there are multiple hypotheses, then the probability for each hypothesis must be calculated and sorted. First, we calculate Pr{*E*|sub-DUCG_{*kj*}}. Second, we calculate

$$h_{kj}^s \equiv \text{Pr}\{E|\text{sub-DUCG}_{kj}\} / \sum_{k,j \neq 0} \text{Pr}\{E|\text{sub-DUCG}_{kj}\}.$$

Finally, we sort *h*_{*kj*}^{*s*} and get the diagnosis results.

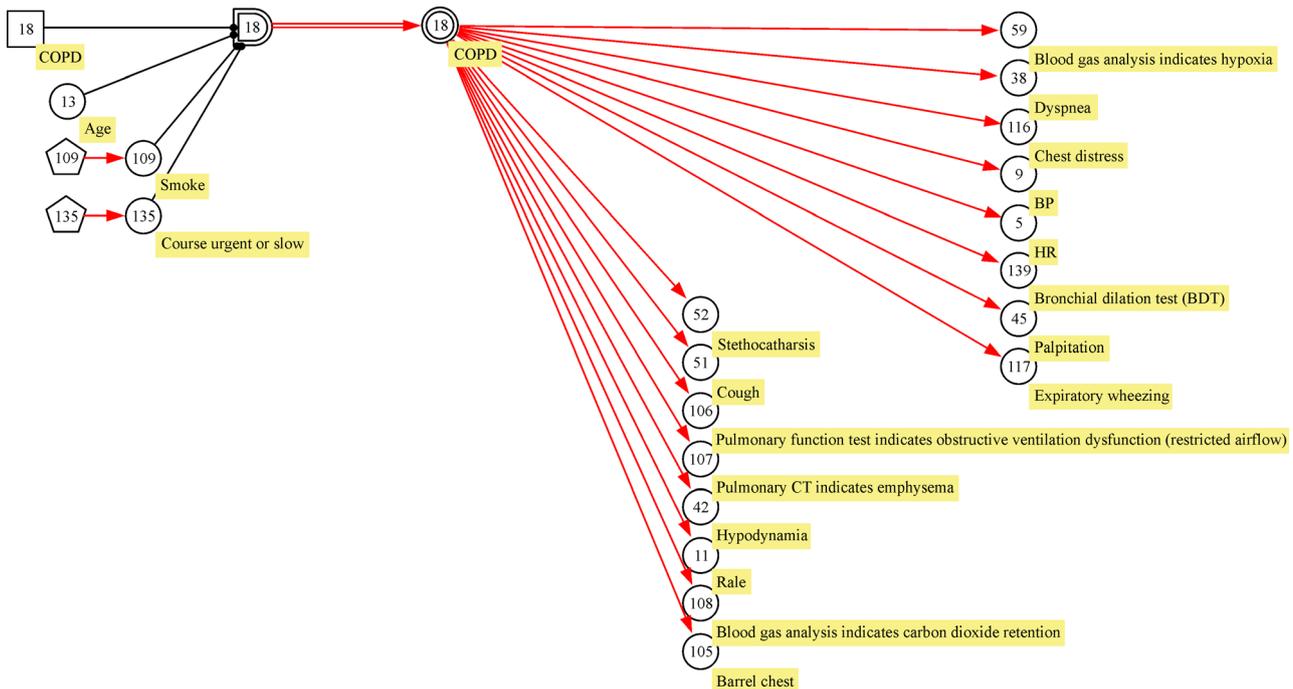


Fig. 2 DUCG subgraph example of chronic obstructive pulmonary disease.

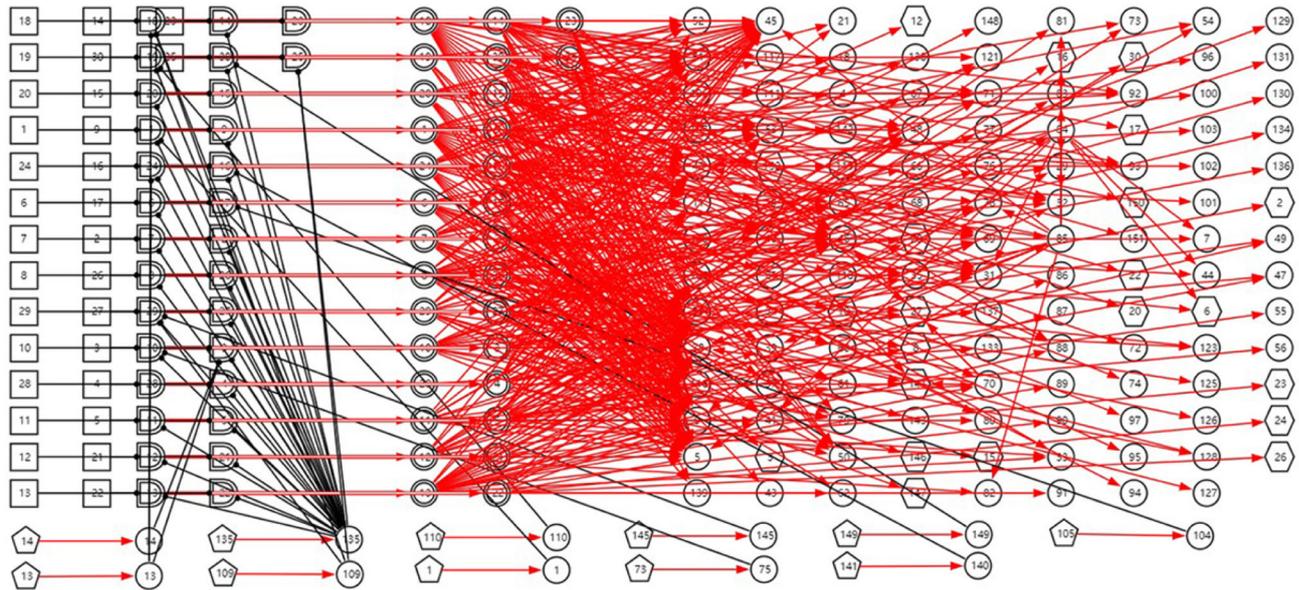


Fig. 3 Knowledge graph of the AI diagnostic model of dyspnea database.

Verification of diagnostic model of dyspnea

Electronic hospital records (EHRs) without any identity characteristics were screened retrospectively from Suining Central Hospital, a tertiary hospital in Sichuan Province, China, as a third party to verify the above model. The EHR system was initiated at the end of 2012, and so the medical records from January 2013 to December 2018 were covered. These records were sorted according to the types of diseases included in the dyspnea disease knowledge base. In each type of medical record diagnosed as the correlative disease, no more than 10 medical records with dyspnea as a chief complaint were randomly selected as test samples (for the same disease, 10 randomly selected cases were enough to verify the correctness). For relatively rare diseases less than 10 cases, all were included in the test to increase the proportion of rare diseases. Note that the low-quality case records (insufficient information or incorrect diagnoses) were eliminated from the selected cases according to the agreement between our expert team and the third-party test team. The DUCG test platform was used to verify the diagnostic accuracy of the model by feeding in the patient’s basic information, medical history, physical signs, and manual inspections in the case records by the test team. The DUCG diagnosis results were compared with the medical records. The coincident diagnoses were considered as correct; otherwise, it was considered as incorrect.

The so-called correct means: (1) the ranking first diagnosis by DUCG is in the record, which includes 158 cases out of the total 195 correct cases; and (2) more than one diagnosis were suggested by DUCG and all were listed

in the record or agreed on by the test team and the DUCG knowledge base constructor. The total includes 37 cases out of 195 correct diagnoses. The other $202 - 195 = 7$ cases were incorrect. Note that some diseases listed in the record but not in the diagnoses of DUCG could be true because they were not the causes of the main complaint of dyspnea, a condition that was discussed and agreed on by the test team and the DUCG knowledge base constructor.

Results

The AI diagnostic model based on DUCG for the differential diagnosis of dyspnea was constructed as described above. The disease set with dyspnea as the chief complaint was established on the basis of expert knowledge and epidemiological data. It contained 28 diseases and syndromes. Differential diagnosis approaches were established and optimized by clinical experts. The constructed DUCG contained 132 symptoms, physical signs, and serological and imaging results (Fig. 3). It was the combination of 28 subgraphs, as illustrated in Fig. 2, that fused the same variables in different subgraphs. The DUCG of Fig. 3 was the knowledge base actually used to diagnose the diseases causing dyspnea. In other words, the inference and calculation for any diagnosis of actual cases with dyspnea were based on the DUCG in Fig. 3.

A total of 25 959 records containing the dyspnea symptom were found in the EHR system, and 202 medical records with diagnoses confirmed by specialist for tests were randomly selected according to the above verification method. After the risk factors, clinical symptoms, signs, and available examination results were manually entered

by clicking the state of the corresponding variable shown on the screen, the DUCG-based AI diagnosis was executed. The output results were sorted according to the calculated h_{kj}^s . Among the cases, 195 were diagnosed correctly. The diagnosis accuracy rate was $195/202 = 96.5\%$. The results are listed in Table 3. The table reveals that 9 diseases had only 10 or less records and 5 did not have any record, because they were too rare under this main complaint and could not be found in the EHR system.

As an example of a test, Fig. 4 shows the sorted results for the tested case, in which $B_{9,1}$ (heart failure) ranked first and $B_{15,1}$ (pericardial effusion) ranked the second. Fig. 5 is the causality diagram explaining why disease $B_{9,1}$ ranked first, that is, because it perfectly caused all the abnormal evidence, and Fig. 6 illustrates why $B_{15,1}$ (pericardial effusion) ranked lower, that is, because two abnormal symptoms $X_{33,1}$ and $X_{73,1}$ could not be explained by $B_{15,1}$.

Discussion

Whether for general practitioners or specialists in respiratory or cardiology, dyspnea is one of the most common chief complaints in practice. However, due to the serious uneven distribution of health care resources in China, patients with dyspnea cannot be identified effectively in resource-limited regions. Misdiagnosis and misuse of treatments are very common. It will take a long time and high investment to improve this situation by relying solely on the training of general and specialized personnel.

With the rapid development of AI research in recent years, AI application has broadened in the medical field. AI is gradually driving the deep reformation of the medical system and forming a new technological revolution. AI diagnosis and treatment systems can simulate the thinking and reasoning process of physicians and are superior to

Table 3 Example of a COPD patient's test result

Variable	Disease	Total cases in hospital	Randomly selected and tested cases	Correct diagnoses	Diagnostic accuracy
B_1	Carbon monoxide poisoning	58	10	10	100%
B_2	Metabolic acidosis	9	9	9	100%
B_3	HCM	10	10	9	90%
B_4	Pulmonary infection	10	10	9	90%
B_5	PAH	559	10	10	100%
B_6	Interstitial lung disease	296	10	10	100%
B_7	Pulmonary alveolar proteinosis	1	1	1	100%
B_8	PE	101	10	10	100%
B_9	Heart failure	429	10	9	90%
B_{10}	HPS	0	0	0	0
B_{11}	DCM	151	10	9	90%
B_{12}	Anemia	3871	10	10	100%
B_{13}	Renal failure	1099	10	10	100%
B_{14}	Constrictive pericarditis	7	7	7	100%
B_{15}	Pericardial effusion	300	10	10	100%
B_{16}	Hemochromatosis	0	0	0	0
B_{17}	End-stage tumor	9	9	9	100%
B_{18}	COPD	13 900	10	10	100%
B_{19}	Laryngospasm	0	0	0	0
B_{20}	Foreign body in air passage	40	10	10	100%
B_{21}	Obesity	5	5	5	100%
B_{22}	Scoliosis	9	9	9	100%
B_{23}	Pleural effusion	1469	10	10	100%
B_{24}	Asthma	2294	10	8	80%
B_{25}	Bronchitis	1330	10	9	90%
B_{26}	Guillain-Barre syndrome	0	0	0	0
B_{27}	Myasthenia gravis	2	2	2	100%
B_{28}	Psychology	0	0	0	0
	Total	25 959	202	195	96.53%

Abbreviation: COPD, chronic obstructive pulmonary disease; HCM, hypertrophic cardiomyopathy; PAH, pulmonary artery hypertension; PE, pulmonary embolism; DCM, dilated cardiomyopathy; HPS, hepatopulmonary syndrome.

Disease ID	Disease name	State probability	Operation
$B_{9,1}$	Heart failure	43.059%	↕ ⓧ
$B_{15,1}$	Pericardial effusion	0.157%	↕ ⓧ
$B_{14,1}$	Constrictive pericarditis	0.032%	↕ ⓧ
$B_{23,1}$	Pleural effusion	0.029%	↕ ⓧ
$B_{8,1}$	PE	0.028%	↕ ⓧ
$B_{24,1}$	Asthma	0.019%	↕ ⓧ
$B_{29,1}$	Aerothorax	0.001%	↕ ⓧ
$B_{2,1}$	Metabolic acidosis	<0.001%	↕ ⓧ
$B_{5,1}$	PAH	<0.001%	↕ ⓧ
$B_{18,1}$	COPD	<0.001%	↕ ⓧ
$B_{20,1}$	Foreign body in air passage	<0.001%	↕ ⓧ
$B_{26,1}$	Guillain-Barre syndrome	<0.001%	↕ ⓧ
$B_{12,1}$	Anaemia	<0.001%	↕ ⓧ
$B_{4,1}$	Pulmonary infection	<0.001%	↕ ⓧ
$B_{27,1}$	Myasthenia gravis	<0.001%	↕ ⓧ
$B_{13,1}$	Renal failure	<0.001%	↕ ⓧ
$B_{17,1}$	End-stage tumor	<0.001%	↕ ⓧ
$B_{1,1}$	HCM	<0.001%	↕ ⓧ

Fig. 4 Sorted diagnoses by DUCG for a test case, according to h_{ij}^k . Abbreviation: PE, pulmonary embolism; PAH, pulmonary artery hypertension; COPD, chronic obstructive pulmonary disease; HCM, hypertrophic cardiomyopathy.

human beings in memory, calculation speed, and accuracy. Therefore, these systems have natural advantages in decision support. In recent years, relevant theoretical research results have emerged in the world, including fuzzy logic, artificial neural network, Bayesian network, decision tree, and case-based reasoning [14,15]. In the past

decade, AI technology in the field of intelligent medical diagnosis has developed mainly along the path of machine learning based on big data. However, this technology kept the complex diseases requiring logical reasoning diagnosis an unsolved challenge in the development of a computer-aided diagnosis system. Several main bottlenecks exist in the development of machine learning technology. First, existing machine-learning theoretical methods have poor interpretability themselves. Second, such mainstream technology only relies on big data and pattern recognition in structure building and parameter learning, but it is short of effective knowledge representation and precise diagnostic reasoning. In complex situations such as blurred symptoms, false symptoms, multiple causes, incomplete knowledge, and causal cycles existing, having no specialized knowledge cannot solve the problem. Third, the accuracy of the big data machine-learning model is high only when the test case is similar to training data set. Once the application scenario deviates, the accuracy will be much lower than expected. For example, the model trained by the medical record data from hospital A may lead to significant errors when validated by the medical record data from hospital B, unless the quality and format of the medical record data of hospitals A and B are basically the same. However, this is usually unrealistic, and significant diversity exists in the clinical application in different levels of hospitals in China. DUCG theory can express uncertain knowledge concisely in the form of a causality graph and provide concise expression and precise reasoning method. It can also express complex logical thinking processes and uncertain causality between disease and clinical manifestation. The results are highly interpretable and the theory can

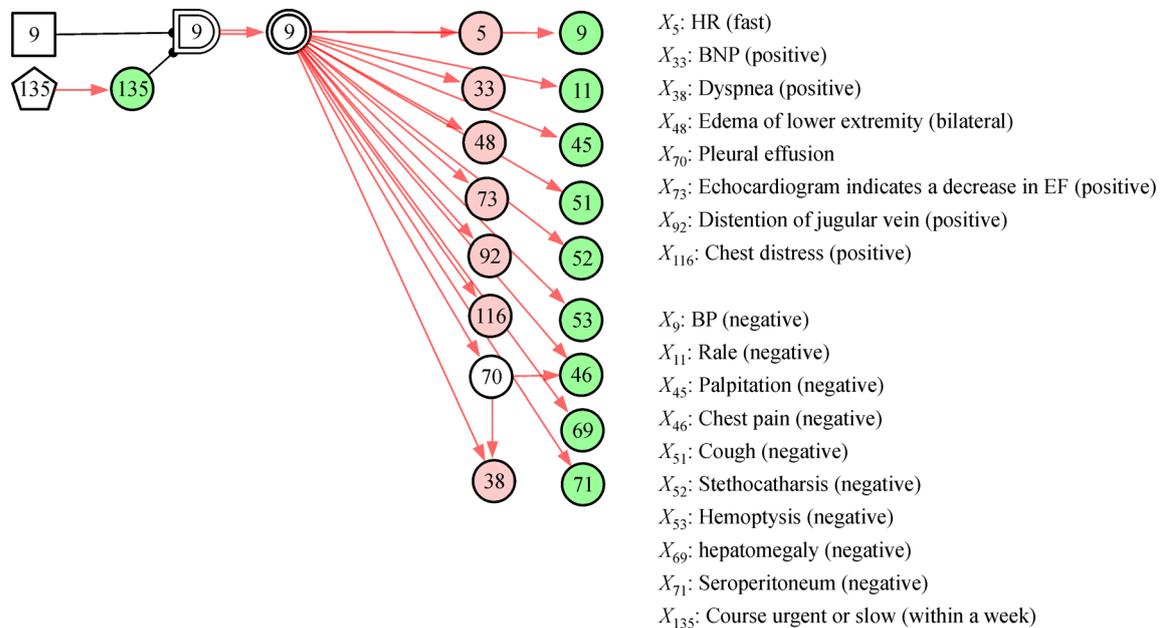


Fig. 5 Graph of the causal relationship between the diagnosis heart failure (ranking the first in Fig. 4) and its evidence.

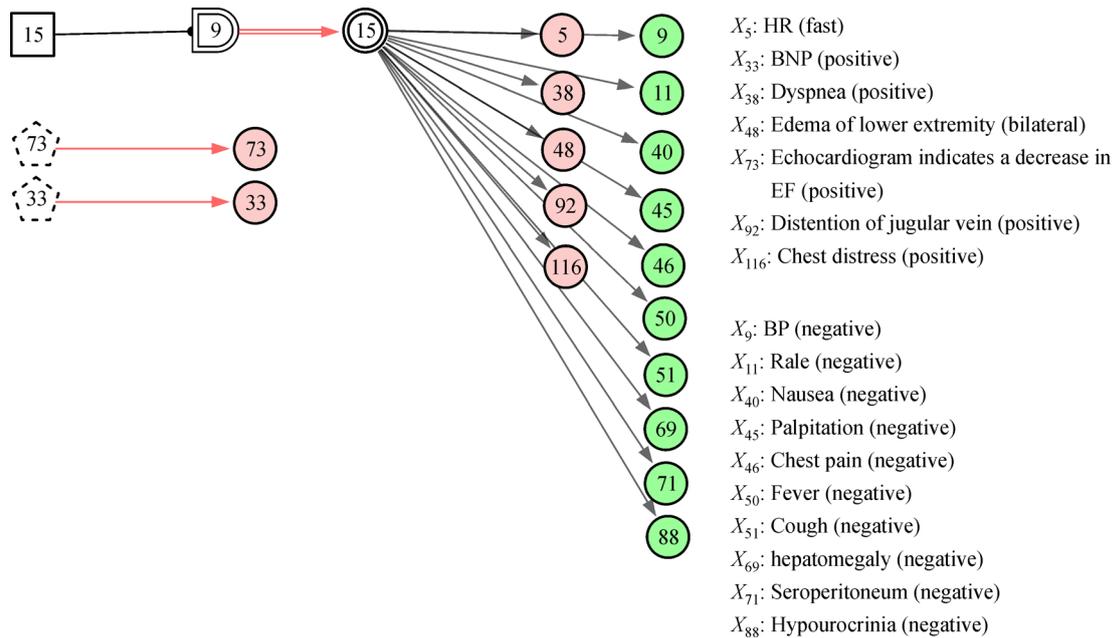


Fig. 6 Graph of the causal relationship between the disease pericardial effusion (ranking second in Fig. 4) and its evidence, in which two isolated abnormal symptoms cannot be explained by the disease.

effectively solve the contradiction between the requirement of standardization, accuracy, and depth labeling workload of data samples and insufficient data samples in clinical practice. DUCG theory has been proven to be effective in fault monitoring, prediction, and diagnosis in large complex industrial systems, such as nuclear power plants, spacecraft fault diagnosis, and chemical process fault diagnosis [6], without relying on big data.

This study is a pilot study in the application of DUCG theory based on clinical expert knowledge in AI auxiliary diagnosis. The construction of the model does not depend on data learning but chooses the chief complaint as the entry point to cover related diseases. It fits into the scenario of primary care clinic or first visit in outpatient clinic and follows the actual inquiring mode of the doctor. The future model can be optimized and expanded according to the epidemiological characteristics of patients in different regions and hospitals at different levels. At the same time, the goal of covering all common diseases in general practice can be achieved by constructing knowledge bases for different chief complaints. Through the validation conducted in a third-party tertiary hospital, the accuracy rate of the DUCG in this study reached 96.5% for patients complaining of dyspnea. Achieving such satisfactory diagnosis accuracy is probably due to the main advantages of DUCG theory as follows. First, compared with Bayesian network and other reasoning methods, DUCG does not depend mainly on complex and conditional probability, and the results can be interpreted by clinical language. Based on the interpretability, users can judge whether the diagnostic hypothesis is reasonable according to their own

clinical knowledge and experience. Second, traditional diagnostic models depend heavily on more accurate probabilistic parameters, but it is difficult to obtain invariable accurate probabilistic parameters in clinical practice. The establishment of diagnostic models based on DUCG theory does not strictly depend on these parameters, because only their relative values are meaningful. Therefore, the DUCG system is robust and easy to construct.

Note that the disease base in this paper covered only common diseases with dyspnea as the chief complaint, and other related rare diseases with very low incidence in the population were not included. Therefore, further evaluation or optimization and improvement of the model are necessary. In spite of this, the result of this pilot study proved that the technology based on DUCG theory is applicable in practice, which can compensate for the lack of experience of general practitioners in the differential diagnosis of dyspnea and provide cost-effective medical solutions for improving the quality of health services in China.

Acknowledgements

This research was funded by the research project entitled “DUCG theory and application of medical aided diagnosis-algorithm of introducing classification variables in DUCG” by the Institute of Internet Industry, Tsinghua University. We appreciate the assistance by the staff of Suining Central Hospital, Sichuan Province, China, in validating the artificial intelligence diagnostic model presented in this paper.

Compliance with ethics guidelines

Yang Jiao, Zhan Zhang, Ting Zhang, Wen Shi, Yan Zhu, Jie Hu, and Qin Zhang declare no conflicts of interests. This study protocol has been approved by the Ethics Committee of Suining Central Hospital, Sichuan Province, China. The Ethics Committee waived the requirement for informed consent because anonymous data were analyzed retrospectively.

References

1. Parshall MB, Schwartzstein RM, Adams L, Banzett RB, Manning HL, Bourbeau J, Calverley PM, Gift AG, Harver A, Lareau SC, Mahler DA, Meek PM, O'Donnell DE; American Thoracic Society Committee on Dyspnea. An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med* 2012; 185(4): 435–452
2. Pesola GR, Terla V, Malik N, Ahsan H. Chronic dyspnoea: finding the cause to reduce mortality. *J Thorac Dis* 2018; 10(Suppl 33): S4057–S4060
3. Elhteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak JAWM; the CAMELYON16 Consortium, Hermsen M, Manson QF, Balkenhol M, Geessink O, Stathonikos N, van Dijk MC, Bult P, Beca F, Beck AH, Wang D, Khosla A, Gargeya R, Irshad H, Zhong A, Dou Q, Li Q, Chen H, Lin HJ, Heng PA, Haß C, Bruni E, Wong Q, Halici U, Öner MÜ, Cetin-Atalay R, Berseth M, Khvatkov V, Vylegzhanin A, Kraus O, Shaban M, Rajpoot N, Awan R, Sirinukunwattana K, Qaiser T, Tsang YW, Tellez D, Annuscheit J, Hufnagl P, Valkonen M, Kartasalo K, Latonen L, Ruusuvoori P, Liimatainen K, Albarqouni S, Mungal B, George A, Demirci S, Navab N, Watanabe S, Seno S, Takenaka Y, Matsuda H, Ahmady Phoulady H, Kovalev V, Kalinovsky A, Liauchuk V, Bueno G, Fernandez-Carrobles MM, Serrano I, Deniz O, Racoceanu D, Venâncio R. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; 318(22): 2199–2210
4. Daliri MR. A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. *J Med Syst* 2012; 36(2): 1001–1005
5. Salas-Gonzalez D, Górriz JM, Ramírez J, López M, Alvarez I, Segovia F, Chaves R, Puntonet CG. Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees. *Phys Med Biol* 2010; 55(10): 2807–2817
6. Zhang Q. Dynamic uncertain causality graph for knowledge representation and probabilistic reasoning: directed cyclic graph and joint probability distribution. *IEEE Trans Neural Netw Learn Syst* 2015; 26(7): 1503–1517
7. Zhang Q, Yao Q. Dynamic uncertain causality graph for knowledge representation and reasoning: utilization of statistical data and domain knowledge in complex cases. *IEEE Trans Neural Netw Learn Syst* 2018; 29(5): 1637–1651
8. Dong C, Wang Y, Zhang Q, Wang N. The methodology of dynamic uncertain causality graph for intelligent diagnosis of vertigo. *Comput Methods Programs Biomed* 2014; 113(1): 162–174
9. Hao SR, Geng SC, Fan LX, Chen JJ, Zhang Q, Li LJ. Intelligent diagnosis of jaundice with dynamic uncertain causality graph model. *J Zhejiang Univ Sci B* 2017; 18(5): 393–401
10. Zhang Q. Dynamic uncertain causality graph for knowledge representation and reasoning: discrete DAG cases. *J Comput Sci Technol* 2012; 27(1): 1–23
11. Zhang Q, Dong CL, Cui Y, Yang ZH. Dynamic uncertain causality graph for knowledge representation and probabilistic reasoning: statistics base, matrix and fault diagnosis. *IEEE Trans Neural Netw Learn Syst* 2014; 25(4): 645–663
12. Zhang Q. Dynamic uncertain causality graph for knowledge representation and probabilistic reasoning: directed cyclic graph and joint probability distribution. *IEEE Trans Neural Netw Learn Syst* 2015; 26(7): 1503–1517
13. Dong C, Zhao Y, Zhang Q. Assessing the influence of an individual event in complex fault spreading network based on dynamic uncertain causality graph. *IEEE Trans Neural Netw Learn Syst* 2016; 27(8): 1615–1630
14. Ceccon S, Garway-Heath DF, Crabb DP, Tucker A. Exploring early glaucoma and the visual field test: classification and clustering using Bayesian networks. *IEEE J Biomed Health Inform* 2014; 18(3): 1008–1014
15. Lin RH, Chuang CL. A hybrid diagnosis model for determining the types of the liver disease. *Comput Biol Med* 2010; 40(7): 665–670