Journal of Medical Systems (2019) 43: 200
https://doi.org/10.1007/s10916-019-1262-0

EDUCATION & TRAINING

MCRDR Knowledge-Based 3D Dialogue Simulation in Clinical Training and Assessment

Wenli Yang[1] · David Hebert[1] · Sunwoo Kim[1] · Byeong Kang[1]

Received: 15 January 2019 / Accepted: 27 March 2019 / Published online: 23 May 2019

Abstract

Dialogue-based simulation is a real-world practice technique for medical and clinical education that provides students with an opportunity to train using hands-on experiences without putting actual patients being put at risk. In this paper, a 3D interactive dialogue-based training and assessment system that supports the detailed development of clinical trial competency for medical students in a distributed virtual environment was proposed. For clinical training, MCRDR-based natural language understanding to realize the semantic representation of written dialog from the most relevant inference results was applied, and on the basis of this, a convolutional neural network model was also used to make the generated inference more exact and reliable. For clinical assessment, the dialogue-driven competency method was used to encompass medical knowledge, communication skill as well as professionalism skill based on the collected dialogue information. Finally, the potential of the created system was demonstrated with several clinical cases. The preliminary results indicate that the system demonstrates the potential of providing efficient training and flexible assessment, while saving time, improving practical skills and making students more confident.

Keywords Dialogue simulation · MCRDR · Knowledge base · Competency assessment · Clinical training

Introduction

Clinical practice plays an important role in medical education. Traditional medical courses allow students to interact within the given clinical scenarios through group role-play. However, not all students are comfortable with role-playing, or even take the scenarios seriously and this will have an effect on their training performance. With the rapid development of information technologies, the implementation of clinical simulations has recently evolved and has been applied to several medical educational subjects. By using clinical simulation, students are provided with more opportunities to bridge the gap between in-class learning and real-world clinical experiences.

Over the past several years, clinical simulations have been adopted and have significantly influenced practices in the areas of clinical training and assessment. There are many different methods used in clinical simulations, such as task trainer, manikin-based and standardized patient simulation. Task trainer simulation is usually for testing procedural technical skills. For example, H. Owen applied task trainer to practice cricoid pressure [1]. DV Girzadas used task trainer for training and evaluating endovaginal ultrasound skills [2]. Manikin-based simulation includes different levels of fidelity simulation manikins which are used in a safe training environment, such as nursing education [3] and healthcare work [4]. Standardized patient simulation facilitates interactive training through a virtual patient [5, 6], which is trained to present patient information, physical symptoms and give investigation findings. Among these simulations, dialogue simulation is one of the most popular methods used to model patient inquiry and system recognition performance. Dialogue simulation offers a set of scenarios that include how to respond when a patient communicates with the doctor. These scenarios provide an interactive and stimulating clinical experience while providing medical students with the option of making mistakes and at the same time, learning from them. However, there are some limitations to the current dialogue simulations, which detailed as follows:

This article is part of the Topical Collection on *Education & Training*

✉ Wenli Yang
yang.wenli@utas.edu.au

[1] Discipline of Information & Communication Technology, School of Technology, Environments & Design, University of Tasmania Sandy Bay Campus, Launceston, Australia

1) These systems can only support single classifications in natural language understanding (NLU); however, most times we need to produce multiple conclusions, such as differential diagnosis, possible investigation, etc.
2) Systems suffer from a lack of modification of case-based reasoning Verification and Validation (V&V) of the quality of output generated by the system's output parser produced from NLU during dialogue simulation.
3) There is currently little or no competency assessment through dialogue simulation.

To solve these problems, in this paper, a 3D dialogue simulation system was developed, consisting of three main modules: A Multiple Classification Ripple Down Rules (MCRDR)-based NLU module, a system manager module and a dialogue-driven competency assessment module. The MCRDR-based NLU creates a semantic representation of input sentences and it produces contextual responses. The system manager manages patient information, case models, as well as the dialogue model. The dialogue-driven competency assessment model enables a collaboration between basic medical knowledge, procedural communication skills and professionalism skills during dialogue simulation.

The key objective of this research is to develop a 3D interactive system for clinical training and assessment, which not only enhances the training experience for medical students, but also efficiently achieves the automatic competency assessment. This paper is organized as follows: section 2 covers the related work; section 3 outlines the methodology used in the proposed 3D training and assessment system; a clinical case is designed and analyzed in section 4 and section 5 discusses the detailed evaluation performed together with the test results. Section 6 concludes the paper with a summary of research findings and significance.

Related work

In this section, the background related to dialogue-based clinical training and assessment methodologies is reviewed. The section provides an overview of contemporary medical training and assessment methods, followed by their limitations.

Dialogue-based theory in clinical training

For clinical training using dialogue principles, semantic representation is the major problem of focus by the natural language understand (NLU) component in the proposed system. In recent years, several NLU algorithms have been developed for semantic analysis and representation and in this section, a summarization of these approaches as applied in the medical dialogues domain, including case base, knowledge base and rule base classical logic is presented.

Case base

The case approach proposed by Stoic philosophers were used to refer to multiple concepts [7]. A case library is introduced by experts to guide the derivation of certain leading question sentences along with expected replies as in Fig. 1 where a simple NLU is achieved. This method is derived from conceptual samples rather than focusing on syntactic constructions, and the predefined sets and distance metric can be used to make referred inferences.

A review of literature listed several key features of a case-based inference as follows:

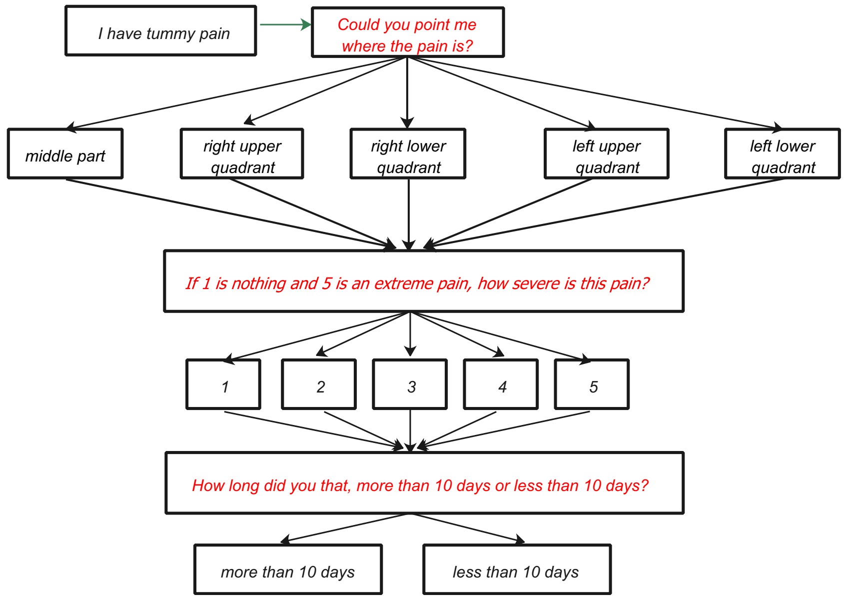
1) It presents real-world evidence, thus the case-based method can provide convincing inferences.
2) It offers independent and simple explanations of multiple problems or practices.
3) It offers opportunities to explain by a natural form which constrains the use of a case with a specific predicate.

All of these characteristics are critical points for medical training. However, for case-based construction, progressively adding cases covers more of the problem space but many cases may overlap thereby potentially hindering the search and retrieval process unnecessarily.

Knowledge base

The knowledge-based approach applies a collection of linguistic knowledge to solve typical problems in natural language understanding such as ambiguity resolution and inferencing [8]. Knowledge-based inferences are generated by matching partial meanings of input texts with stored knowledge structures and adding the resulting instantiations of the knowledge to the current interpretation of the text (see Fig. 2). Knowledge, such as that constrained in selectional conditions applies to NLU in two ways: 1) it constrains knowledge acquisition by pruning branches which are ruled out; 2) it guides matching in the remaining paths that are more relevant to the preferred understanding. Therefore, knowledge-based construction can help to reduce the search space and make natural language understanding more efficient. For instance, in the sentence "I have a tummy pain", someone trying to understand has to infer that "tummy pain" is the symptom and the relative pain degree (severe or slight), location of pain, duration of pain, etc. are used to generate appropriate matching responses. The most important issue in knowledge-based construction is error recovery, because when such an error is revealed during processing, the knowledge-based

Fig. 1 Example of case-based construction



inference must go back and resume inference after the erroneous result has been excluded.

Rule base

Rule-based refers to the knowledge representational structure consisting of an *if* component and a *then* component. Rule-based methods are frequently used in medical training and learning as each medical problem description can be represented as: Title - symptoms - diagnosis - prognosis – treatment. Consider the following example:

- *Title:*

  Instructions for treating a gastroenteritis.

- *Symptoms:*

  If you examine a man with these symptoms, including watery diarrhoea, vomiting, stomach pain, cramping, mild fever, nausea, and headache.

- *Diagnosis and prognosis:*

  Then, you will tell your patient: *"It may be gastroenteritis."*

- *Treatment:*

  If the patient has symptoms of vomit and diarrhoea, then you can treat him by oral rehydration solutions; if the patient has symptoms of fever, then you can treat by Panadol and Panadeine. Otherwise, no medical treatment is needed.

In a rule-based construction, the statements after *if* are called conditions, those after *then* may be called conclusions. Thus, a set of conditions representing some pattern that may be observed or examined can be defined, and a set of conclusions representing responses that should be taken. For natural language understanding, the *if…then…* rules are collected together in a rule base, and rule selection is always chosen from a knowledge base which is appropriate to the current circumstances. The limitation of this model is the restricted

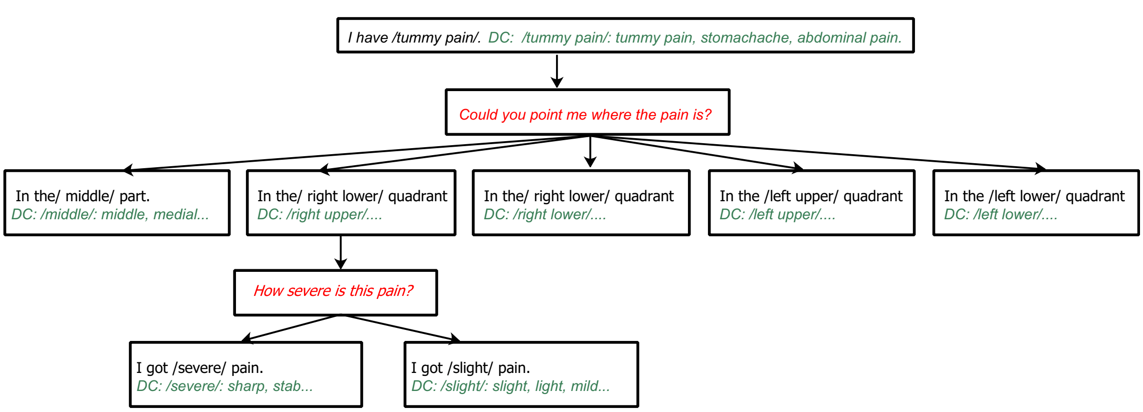


Fig. 2 Examples of knowledge-based construction

power of expression and classification, as several other branches of medical knowledge do not necessarily fit into this pattern. This means that this approach may ignore some possibly-important investigations and differential diagnoses. Table 1 lists the main features and limitations between the case base, knowledge base and rule base used for NLU.

Based on the above discussion, it is found that most NLU approaches can only be used for a single classification. However, for medical training, there is usually a need to produce multiple conclusions, such as differential diagnosis, possible investigation, etc. Multiple Classification Ripple Down Rules (MCRDR) [17] can be applied in natural language understanding but it also provides multiple classifications, which may be suitable for medical training. This research is targeted at solving the limitations exhibited by contemporary systems by utilizing the MCRDR approach for knowledge acquisition and inference. MCRDR, which is based on Ripple Down Rules (RDR) [18] focuses on an incremental knowledge acquisition technique based on classifying cases in the context in which they are presented to the system. In this way, the system can display a reasonable explanation as to how to generate and evaluate the multiple conclusions. For example, MCRDR has been used in several applications, such as web document classification [19],expert system development [20], and intelligent conversation [21]. MCRDR enables the creation and maintenance of the knowledge base without any assistance from a technical knowledge engineer. However, how to properly evaluate the generated MCRDR inference results remains an important problem to solve. In this paper, how MCRDR can be applied to dialogue-based clinical training is explored, and an evaluation method based on convolutional neural networks is proposed to verify and update generated conclusion.

Assessment in clinical simulation

With the development of simulated clinical training, simulated assessment tools are now widely used to evaluate the performance of medical students on virtual patient during clinical practice. The current assessment methods involve many techniques using objective structural clinical examination, knowledge base modelling, etc.

Objective structured clinical examination

Objective structured clinical examination [22] provides a procedural assessment of clinical skills in a given clinical case, selected by experts according to the following criteria: patient information, detailed description of symptoms, related vital signs, and so on. Based on the given information, students are asked to make investigations and proffer possible diagnoses of the patient's problems. The scope of structured clinical examination is essentially patient information, a demonstration of physical examinations, and an assessment of a narrow range of technical skills. The structured clinical examination takes a short time to execute assessment over a range of instructions. This approach has been used successfully in some applications including basic medical education [23] and patient management [24]. However, it is difficult to organize a wide variety of materials into a structured case.

Knowledge base modelling

This approach relies on linking a students' conceptual understanding to a knowledge base [25], intended to confirm the knowledge acquisition. The knowledge base attempts to represent the information about a word, sentence or topic, and the knowledge inference represents logical conditions and referred conclusions. This approach can work in certain cases, but there is still a need to make sure explicit understanding is achieved and correct choices are given. Compared with objective structured clinical examinations, the positive side of knowledge base is that it allows for a dynamic conversation rather than a static approach based on standardized cases. However, validation of the inference results is necessary, and there is often difficulty in combining different knowledge domains.

Table 1 Comparisons between Case base, Knowledge base and Rule base

	Case base	Knowledge base	Rule base
Derivation	conceptual samples	syntactic constructions	conceptual samples syntactic constructions
Feature	Provides convincing and independent inference, and it is simple to explain	Matches the meanings of input text with its knowledge, which can reduce the search space and make natural language understanding more efficient.	Use a set of conditions representing some pattern that may be observed or examined, and a set of conclusions representing responses that should be taken.
Limitation	Sentences must match exactly; covers more of the search space and cases sometimes overlap.	When some errors are revealed during processing, the knowledge base search should be conducted again.	Restricted power of expression and classification, since there are some other possible conclusions based on the same conditions.
Typical application	[9, 10]	[11–13]	[14–16]

Objective structured clinical examination is applied in a collection of cases for the examination and evaluation standardized tasks, while the knowledge base approach is used for measuring the understanding of the vocabulary. However, there is a lack of evidence with regard to the effectiveness of using the assessment of competency. In this paper, a dialogue-driven competency assessment method is proposed, which can not only infer the knowledge base assessment, but also encompass communication skills in a collection of dialogue analysis, as well as clinical skills in the given cases.

Methodology

An interactive 3D dialogue system was designed and developed with a constrained natural language understanding capability to provide medical students with the ability to repeat real-world clinical practices individually. This section describes the overall proposed system architecture and the detailed training and assessment methodologies used for natural language understanding, virtual patient and the assessment module.

System architecture

The functional architecture of the proposed system can be utilized by medical teachers and students running on an individual computer, without other software or plug-in requirements. The key functional modules consist of an MCRDR-based natural language understand (NLU) component, the system manager and competence-based assessment module as shown in Fig. 3.

The proposed system is based on an application-independent software architecture, which builds a semantic representation of input text for practicing clinical skills with case-based learning. Each functional module is described as follows;

1) MCRDR-based NLU module

The task of the NLU module is to infer an appropriate patient response to the utterances of the student. Student utterances are parsed using MCRDR-based representation, which contains *Text transformation*, *Case-based classification*, *Context variable identification*, *Dictionary identification* as well as *MCRDR inference*. Based on real-world clinical trials, text transformation is designed to transform extracted sentences into index terms or features. A case-based classification model is then built which is used to select the most relevant context and queries to represent specific utterance cases. After that, the selected contextual variables are considered as keywords or key phrases pattern-matched from utterances by simple meta-rules referring to dictionary terms. These terms are defined in the dictionary identification

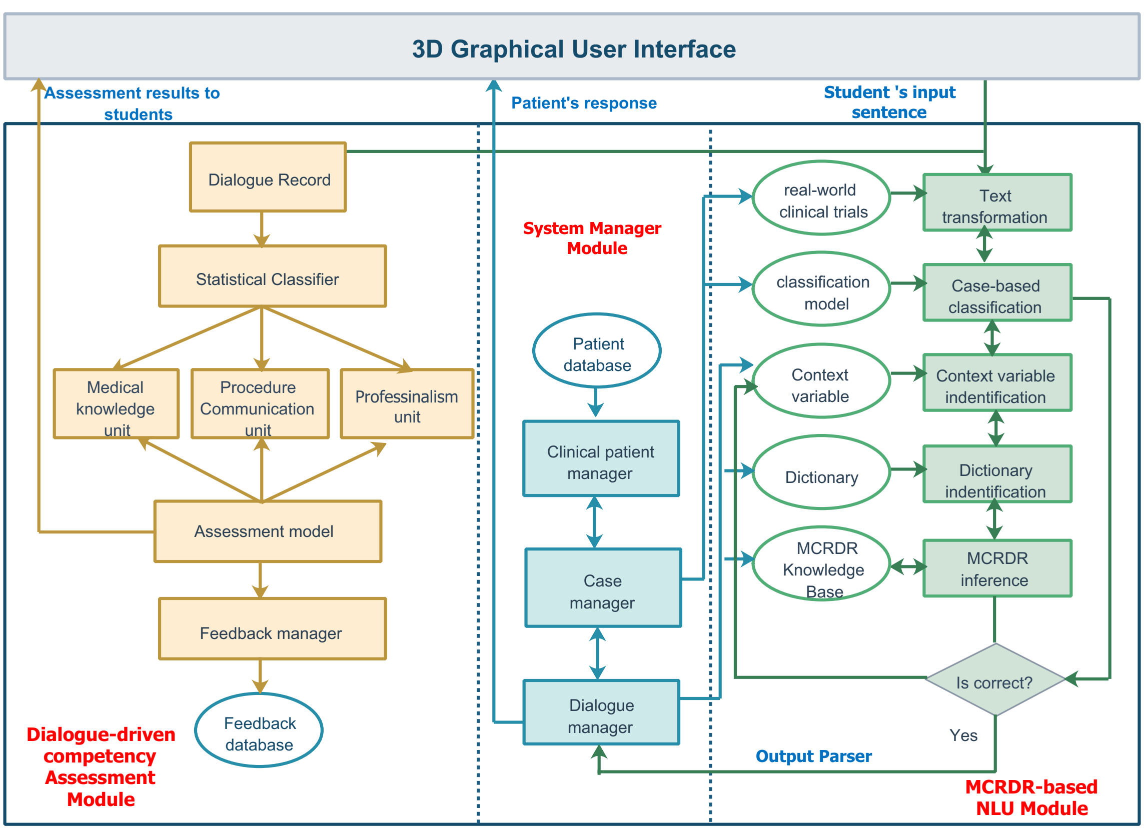


Fig. 3 System Architecture of 3D dialogue system for medical training and assessment

module by a set of registered synonyms. In this way, the semantic representation of the student's input will be constructed. The represented input case is then inferenced to the MCRDR knowledge-base (that contains the pooled knowledge in the form of rules associated with the medical case-based practices), and a post-inference evaluation loop chooses the appropriate responses returned from the inference.

2) System manager module

The task of the system manager module is to generate and update the medical knowledge needed to obtain clinical dialogue descriptions. This module comprises three components: *Patient manager* for managing and selecting traditional patient inquiry interactions against a database concerning specific virtual patient models, *Case manager* for managing real-world clinical trials and a case-based classification model, and *Dialogue manager*, necessary for building dialogue models capable of adapting to the requirements of each user. The *Dialogue manager* requires the *Clinical patient manager* and the *Case manager* model datasets in order to determine which rules need to be added to the knowledge-base.

3) Dialogue-driven competency assessment module

The basic goal of this module is to analyze and measure the doctor-patient clinical skills in terms of application of both medical knowledge and practice skills to apply relevant knowledge. All dialogue records acquired by the user interface will be stored as a list of *evidence*. The statistical classifier is then used to split the data into multiple inferences of competencies including the medical knowledge unit, as well as the procedural communication and the professionalism unit, based on the classification of collected evidence. For each unit, the assessment model's encapsulated medical-technological knowledge is used to model a correct or reasonable interaction of each case that is then applied to evaluate the training results for medical students.

MCRDR-based NLU module

This section describes the procedure for developing the NLU module by the application of MCRDR as an underlying knowledge-base representation, inference and knowledge-acquisition methodology. As was discussed in section 2, to achieve more natural and contextual responses, the MCRDR knowledge-base is represented as a decision tree comprising of nodes defining classification rules (that also refer to meta-data such as context variables and dictionary terms). For each dialog session, MCRDR-based NLU creates a pattern-matched semantic representation of a sentence, as a result of inference. The six steps used to achieve intelligent NLU are as follows;

1) To identify key phrases, text transformation is used to select the most important tokens representing the target sentences from real-world clinical cases. This transformation has three sub-steps consisting of tokenization (part of speech (POS) tagging), stopping and stemming as shown in Fig. 4.
2) Case-based key phrase or word classifications are used to label the most relevant information for the specific clinical case, which is based on the CNN (Convolutional Neural Network) structure [26]. Each case can be represented as a n × k non-static vector with *n* target sentences and *k* tagged terms output from step 1, which are then used as input into a deep convolutional network. To identify the design of the network structure, we used greedy layer-wise pre-training [27] to guide how many layers for convolutional layers. Firstly, we set the original network structure as two convolutional layers. The main steps are as follows:

i) The labelled sentences are divided into X_{train}, X_{test} . Then trains the X_{train} and obtains the feature map 1, then inputs into fully-connected layers and uses X_{test} to evaluate the classification accuracy.
ii) Add one more convolutional layer as well as pooling layer and continue forward-propagates feature map 1 into convolutional layer 3 and obtains feature map 2, then inputs into fully-connected layers and calculates the classification accuracy.
iii) Loop the above the operation until we got the following evaluation results shown in Table 2

Based on the above evaluation results, the network structure with five convolutional layers followed by max pooling, and two fully-connected (FC) layers was proposed as shown in Fig. 5.

As an Example, referring to Fig. 5, it is assumed that the output label case (*gastroenteritis*) is phased by five important sentences, with each sentence consisting of six terms. Thus, each sentence $X_{1:5}$ can be represented by a 5 × 6 matrix as network input shown in Eq. (1).

$$X_{1:5} = x_1 \oplus x_2 \oplus x_3 \oplus x_4 \oplus x_5 \quad (1)$$

The convolutional layer aims to extract feature $c_i$with different kernel size $h \times h$ ($h < 5$), and then output feature map c with all extracted features c_i.

$$c_i = \mathrm{f}(\mathrm{S} \cdot X_{i:i+h-1} + b) \quad (2)$$

$$\mathrm{c} = \left[c_{1,}c_2, \ldots c_{5-h+1}\right] \quad (3)$$

In Eq. (2), *f* is the active function – . The window consisted of *h* terms represented as $X_{i:i+h-1}$, and a bias unit *b* appended to the start of the input and each convolutional layer. The

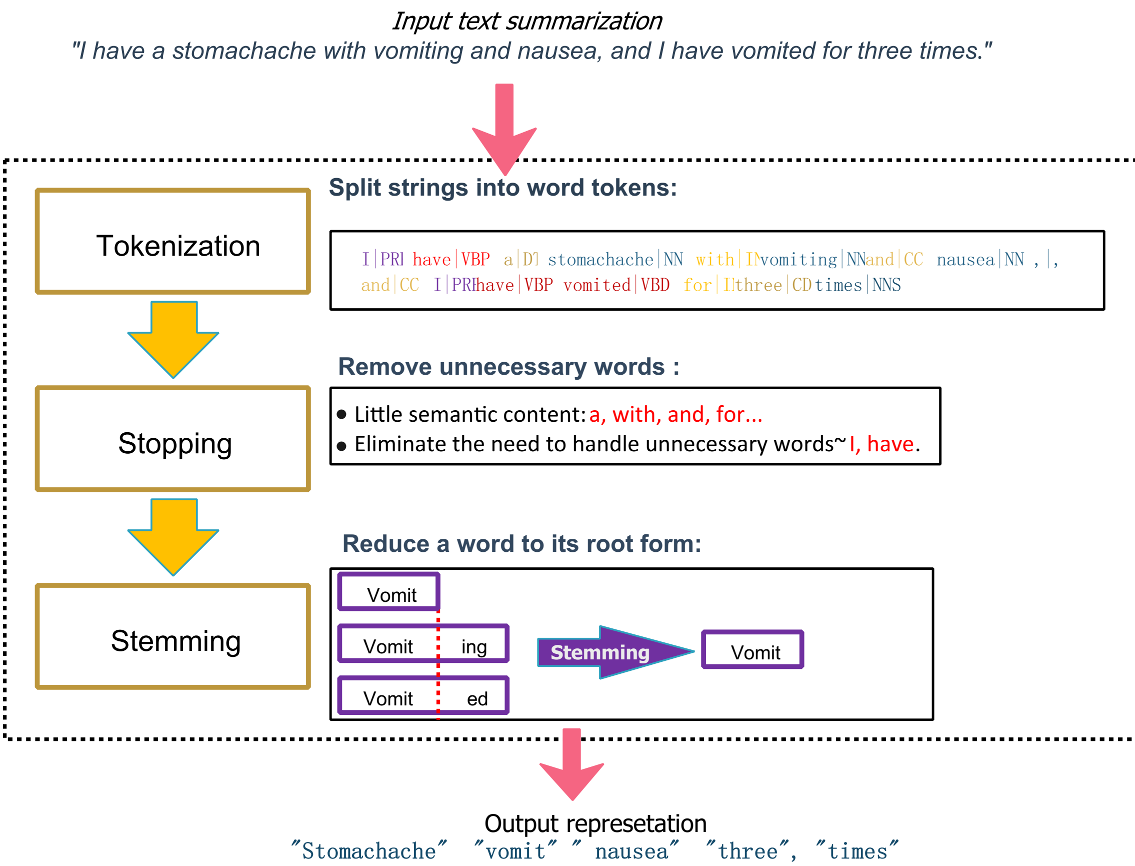


Fig. 4 The process of text transformation

feature maps extracted from the convolutional layers are used as input in max pooling layers, and then the output is $\hat{c} = \max(c)$. Finally, two fully-connected layers are used to get the classified feature vector y as defined in Eq. (4).

$$y = W \cdot z + b \qquad (4)$$

$z = [\hat{c_1}, \hat{c_2}, \ldots \hat{c_m}]$, m is the number of convolutional layers. W is the weights of two FC layers.

3) Based on the above classification model, several context variables can be selected to represent the specific case for training. As a case-based dialog progresses, the system can retain the matched data across inference requests by the use of context variables, and then reference a context variable value from the MCRDR rule conditions attribute to show different responses depending on a value provided by the user. Prior to inference, the dialog parser or the application environment defines context variable values to pass information with the current dialog (which is in effect a maintenance of context). Figure 6 shows examples of these two passing context methods.
4) The dictionary is a collection of terms with one or more synonyms, used as criteria to match context variables and create inference requests for case-based acquisition. The dictionary has a large array of medical terms, including conditions, diseases, treatments, signs and symptoms. When a user starts an inquiry with any context variable term present, the dictionary is scanned to find a matching entry which replaces the specific synonym with a registered representative term of that same variable. Table 3 gives some examples of dictionary term definitions.
5) MCRDR inference is used to maintain the domain knowledge in the clinical training system. For each parsed utterance, MCRDR inference provides a justification for a conclusion in a particular context from a range of different cases. The structure of MCRDR inference can be described as an n-ary tree with each node representing a rule. Figure 6 provides such an example of MCRDR inference for a particular clinical case.

Table 2 The comparisons of the classification accuracy using different convolutional layers

Layers number	Test loss	Test accuracy
2 Conv layers+2 pooling layers	5.473407745361328	0.33
3 Conv layers+3 pooling layers	4.625582695007324	0.33
4 Conv layers+4 pooling layers	3.5451345443725586	0.5
5 Conv layers+5 pooling layers	1.0478006601333618	0.85
6 Conv layers+6 pooling layers	1.6827958822250366	0.63

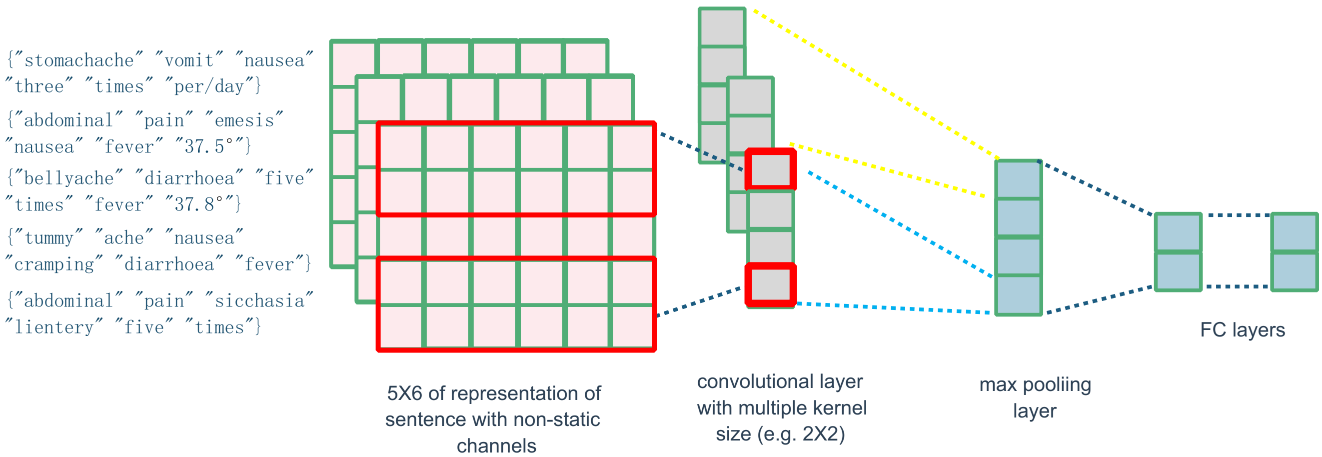


Fig. 5 Deep neural network architecture used in case-based classification

The red boxes shown in Fig. 6 represent MCRDR dialogues that are satisfied for the clinical case: appendicitis{/stomachache/, /severe pain/, /lower right/} as below:

- *Patient: "I have a **stomachache**".*
- *Student: How **severe** is this pain?*
- *Patient: **severe pain** in the **lower right** abdominal pain.*
- *Student: In general, it may be caused by **appendicitis**.*

6) Pattern evaluation is used to evaluate the input dialog to determine satisfied rules in the MCRDR knowledge-base during inference. According to the conditions for the generated inference, we identify the six most important key words using text transformation or use zero-padding if it is less than six words. Based on the selected conditions, then we input them into deep learning model generated in Step 2 to check if the classification is the same or similar as the created inference results from step 5. If the output results are consistence between MCRDR inference and case classification model, then we can finalize the generated inference and save it into our knowledge base. Otherwise, it is necessary to fall back to step 3 and update the dictionary term, as well as context variables until they agree with the case-based classification model.

In order to test the efficiency for such an update loop, we select three cases *gastroenteritis*, *gall-stone* and *laryngitis*, then ask medical experts to get the ground truth to evaluate the overlapping ratio between with pattern evaluation and without pattern evaluation.

Table 4 showed the comparisons with the ground truth, the average overlapping radio with pattern evaluation is about 79.6%, while the radio without pattern evaluation is about 51.4%. From these results, we can see that although standard MCRDR includes rigorous and robust verification and

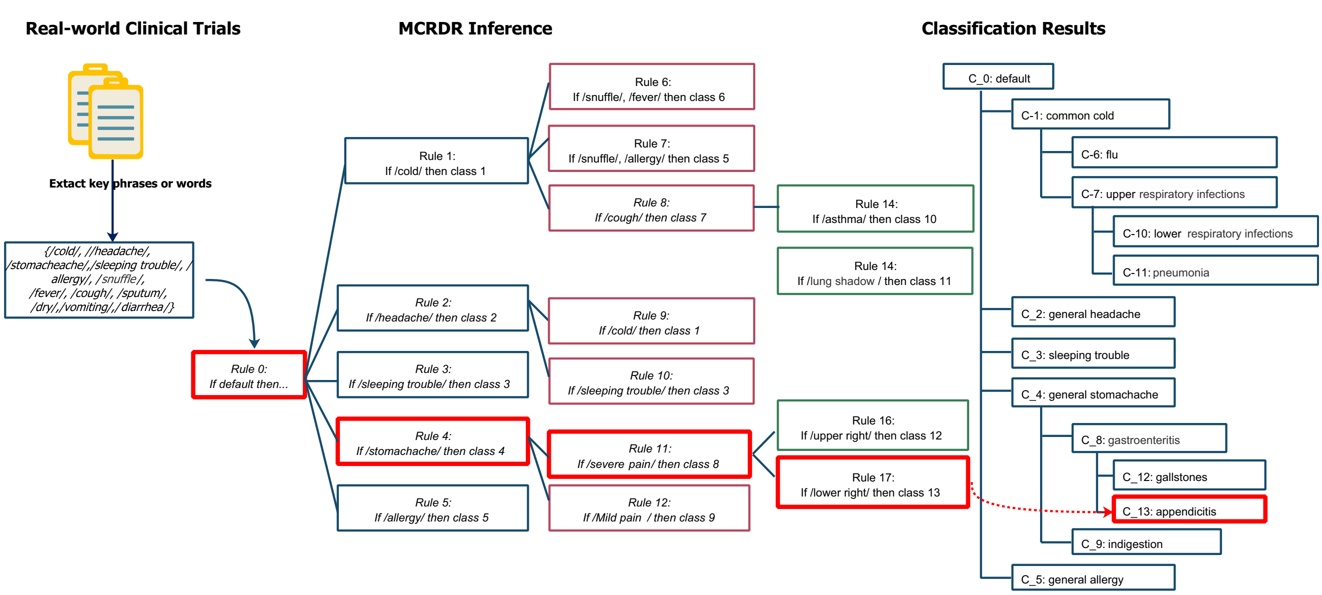


Fig. 6 An example of inference for MCRDR

Table 3 Examples of dictionary term definition

Types	Representative term	Synonyms
symptoms	/vomit/	vomit nausea emesis sicchasia
	/cold/	cold flu influenza
Treatments	/chemotherapy/	chemo chemotherapy
	/radiotherapy/	radiation radiotherapy
Medical examination	/CT/	CT computed tomography
	/FBC/	FBC CBC Full Blood Count Full blood Test Complete Blood Count

validation, the system relaxes this process when using dialog-based input. As a result of this, the standard approach fails to evaluate inference results with illogical conclusions or the addition of rules which contain identical antecedents to existing rules, while the proposed MCRDR integrated with pattern evaluation can help us to select the conditions more accurately, based on the selected conditions, we can improve the interaction efficiency with the generated MCRDR inferences.

System manager module

The proposed 3D dialogue system is an interactive as well as distributed platform, allowing medical students to repeat access and practice in case-based virtual training. In this system, the medical tutors use relational domain patient databases to gather basic and historical information regarding patients for virtual patient management. The real-world clinical training samples such as case summary, examination findings, investigation findings and differential diagnosis, etc. are collected in order to generate case-based classification models. In addition, the dialogue manager is designed to create, read, update and delete the MCRDR knowledge-base. The detailed flow of the system manager module is described in Fig. 7.

In the system, attributes such as patient name, age, gender, symptoms, past/ongoing medical condition or hospitalization, family history, etc. are covered. All of this data is obtained from traditional approaches to obtain a patient's basic and medical history information.

Secondly, real-world clinical training samples are processed in this step. As mentioned in traditional training practice, case summary, physical exam finding, investigations ordered, differential diagnosis, final diagnosis, etc. are integrated as the components of the clinical case base. This base can serve medical tutors when determining how to build a case-based classification model and how to more accurately evaluate MCRDR inference.

In addition, reference ranges of context variables, as defined in Table 5, for each case with a specific patient, are bound with one case class to assist with the design of the MCRDR knowledge-base.

Finally, in comparing patient information with the generated case model, the system will generate the context variables from each specific case class. The medical tutor can review them and refine these auto-generated context variables as components of MCRDR knowledge-base.

Dialogue-driven competency assessment module

The dialogue-driven competency assessment module provides an ongoing process of continually building multiple units such as medical knowledge, procedural communication skills, and professionalism practice. Three main steps are involved for competency classification and assessment and these are shown in Fig. 8.

1) The assessment module compiles input dialogue extracted from a medical student's input text stored in an information record base. The output of this step is a list of evidential items which are used for verification of the student's clinical skills.
2) The statistical classifier is the key step in the assessment module, which is based on a competency classification model to classify various evidential items into multiple competencies including basic medical knowledge, clinical inquiry as well as diagnosis process. Each input solution s is represented as a *n-vector*: $V(s) = (T_1, \omega_1(s); T_1, \omega_2(s); \ldots; T_n, \omega_n(s)$, where T_i is each term extracted from solution s, and $\omega_i(s)$ is the weighting of term T_i to identify how important a term is to a solution in a collection (Eq. (5)).

$$Weight_{T,s} = TF(T_i) \times \log\left(\frac{N}{DF(T_i)}\right) \tag{5}$$

Where.

$TF(T_i)$ (Term Frequency)	the number of times that T_i occurs in s.
$DF(T_i)$ (Document Frequency)	the number of solutions containing term T_i
N	total number of solutions.

We then apply a DBN-DNN (Deep Belief network-Deep Neural Network) with three hidden layers, that consists of three RBM (Restricted Boltzmann Machine) stacked together to be a DNN structure. The output of each former RBM layer is considered as the input of next RBM layer, which are

Table 4 The comparisons between generated MCRDR inferences with pattern evaluation and without pattern evaluation

Case 1: gastroenteritis

• The ground truth: /Severe/ /abdominal pain/ in your /upper or centre abdomen/; pain that /spreads to your right shoulder or back/; /nausea/; /vomiting/; /diarrhoea/, /fever/.

Types	Responses (with pattern evaluation)	Responses (without pattern evaluation)
Conditions	/abdominal pain/, /heartburn/, /5–6/, /Upper abdomen/, /vomit/, /trips/.	/abdominal pain/, /4–5/, /Around the belly button/, /vomit/.
Interaction examples	• I have a strong abdominal pain and heartburn. • How strong is the pain? Let's say in a 1 to 10 scale, how would you describe the intensity of the pain? • Between 5 and 6. • Can you point the pain location? • Upper abdomen. • Have you vomited? • Yes, I have vomited three times and made several trips. • You might have gastroenteritis	• I have a strong abdominal pain. • How strong is the pain? Let's say in a 1 to 10 scale, how would you describe the intensity of the pain? • Between 4 and 5. • Can you point the pain location? • Around the belly button. • Have you vomited? • Yes, I have vomited three times. • You might have gastroenteritis or appendicitis, you need to have FBC and LFT tests.
Overlap ratio with ground truth	75%	37.5%

Case 2: gall-stone:

The ground truth: /Severe/ /sudden/ /abdominal pain/ in the /upper right/ portion of your abdomen, /Back pain between your shoulder blades/, /Nausea/ or /vomiting/, /Yellowing of your skin/, /High fever/

Types	Responses (with pattern evaluation)	Responses (without pattern evaluation)
Conditions	/sudden/, /abdominal pain/, / nausea/, /7–8/, /Upper right/, /vomit/, /back pain between shoulder blades/, /high fever/.	/abdominal pain/, /7–8/, /right/, /vomit/, /back pain between shoulder blades/, /low fever/.
Interaction examples	• I have a sudden abdominal pain and nausea. • How strong is the pain? Let's say in a 1 to 10 scale, how would you describe t he intensity of the pain? • Between 7 and 8. • Can you point the pain location? • Upper right of the abdomen. • Do you have the back pain between your shoulder blades? • Yes, last several hours. • Do you have fever? • Yes, high fever about 38.8 C • You might have cholesterol gallstone.	• I have a abdominal pain and also vomited several times. • How strong is the pain? Let's say in a 1 to 10 scale, how would you describe the intensity of the pain? • About 7–8. • Can you point the pain location? • right of the abdomen. • Do you have the back pain between your shoulder blades? • Yes, last several hours. • Do you have fever? • Yes, fever about 37.5 C • You might have appendicitis or cholesterol gallstone, you need to have FBC and LFT tests.
Overlap ratio with ground truth	88.9%	66.7%

Case 3: Laryngitis:

The ground truth: /Hoarseness/, /Weak voice/ or /voice loss/, /dry cough/, /sore throat/, /fever/, /trouble swallowing/, /cough up blood/.

Types	Responses (with pattern evaluation)	Responses (without pattern evaluation)
Conditions	/cough/, /dry/, /blood/, /sore throat/, /hoarseness/, /fever/	/cough/, /phlegm/, /blood/, /sore throat/, /asthma/, /fever/
Interaction examples	• I have a severe cough • Do you bring up phlegm or is it a dry cough? • Dry cough. • Has there ever been blood in it? • Yes, sometimes I noticed some blood in it. • Any other symptoms?. • I also have sore throat and hoarseness. • Do you have fever? • Yes, low fever about 37.8 C • You might have laryngitis.	• I have a severe cough • Do you bring up phlegm or is it a dry cough? • I bring up a bit of phlegm. • Has there ever been blood in it? • Yes, sometimes I noticed some blood in it. • Any other symptoms?. • I also have sore throat and asthma. • Do you have fever? • Yes, high fever about 39 C • You might have a flu or laryngitis.
Overlap ratio with ground truth	75%	50%

stacked in order to be the basic framework of a DBN. The network structure of DBN-DNN is shown in Fig. 9.

For the classification model training, a CD (contrastive divergence) algorithm was applied, which is based on supervised learning in layers. As shown in Fig. 9, an input vector X was used, as well as a binding label vector to initial visual layer v_1, and v_1 and h_1 which are considered as the first RBM to train the necessary parameters such as individual weights representing the strength of connections between v_{i1} and h_{i1}, individual offset vector for v_{i1} and h_{i1}, etc. Then, h_1 will be regarded as visual layer v_2, and the procedure is repeated for the training of the second RBM (v_2 and h_2) and third RBM (v_3 and h_2) respectively. The detailed training algorithm is described as below (Algorithm 1).

3) After each solution is labelled as a defined class of competence, the integrated assessment combines each competency unit to build the model with both skills and knowledge in order to verify that a student is performing the training more accurately. For basic medical knowledge assessment, an interactive quiz is used to test a student's knowledge. For assessment clinical inquiries, based on the dialogue recordings between student and virtual patient, focus was placed on the investigation of any ignored aspects during clinical training. During diagnosis processing assessment, the similarity between a student solution and the correct solution was calculated. Finally, a student is able to view the final assessment and submit feedback to the medical tutor.

Algorithm 1 Competency Supervised Training

Input: $CST=\{X, L, n, \epsilon, \omega, b, c, S\}$

X: input vector distribution;
L: input label distribution;
n: number of layers to train;
ϵ: learning rate for random gradient descent in CD algorithm;
ω: RBM weight matrix(number of hidden layers, number of input vector);
b: RBM offset vector for input units;
c: RBM offset vector for hidden units;
S: pre-defined criterion;
for each $i \in [1, n]$ **do**
    initialize $\omega_i = 0, b_i = 0, c_i = 0$;
    initialize visual layer v_0 from X with L;
    while not satisfied with S **do**
        for each hidden unit $j \in [1, 3]$ **do**
            assign $v_{i,j}(v_{i,j-1}|h_{i,j})$ as visual layer in $RBM_{i,j}$
            calculate parameter $Q_{i,j}(c_{i,j}, \omega_{i,j})$ of $RBM_{i,j}$
            update $v_{i,j} = c_{i,j} + \sum \omega_{i,j} v_{i,j-1}$
        end for
        for each visual unit $k \in [1, 3]$ **do**
            assign $h_{i,k}$ as hidden layer in $RBM_{i,k}$
            calculate parameter $P_{i,k}(b_{i,k}, \omega_{i,k})$ of $RBM_{i,k}$
            update $h_{i,k} = b_{i,k} + \sum \omega_{i,jk} v_{i,k-1}$
        end for
        $\omega_i \leftarrow \omega + \epsilon(h_{i,2} v_{i,2}{\prime} - h_{i,3} v_{i,3}{\prime})$
        $b_i \leftarrow b_i + \epsilon(v_{i,2} - v_{i,3})$
        $c_i \leftarrow c_i + \epsilon(h_{i,2} - h_{i,3})$
    end while
end for

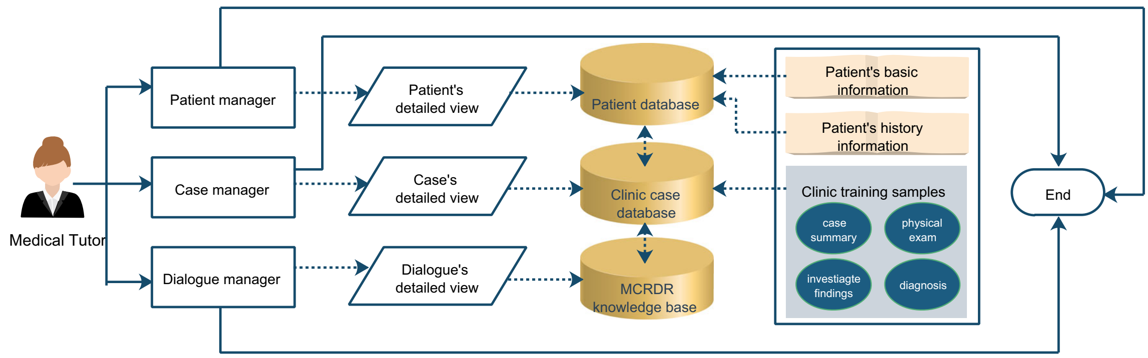


Fig. 7 System manager module flow diagram

Dialogue cases

The design of a 3D dialogue system is based on real-world clinical trials in order to automatically realize medical training and assessment. In this section, a *Cholecystitis* dialogue simulation was used to explore the realities of the proposed methodology and explain the system's value within this domain. Medical tutors confirmed the patient information, clinical

Table 5 Reference ranges of context variables examples

Case Class		Pancreatitis	Cholecystitis
Virtual patient	Name	Nina Tyler	Jane Smith
	Age	21	35
	Gender	Female	male
	Presenting Symptom	Abdominal pain, nausea, vomiting, loss of appetite, worse after eating, radiating pain to the back, fever	Abdominal pain, nausea, vomiting, loss of appetite, radiating pain, worse after fatty food, fever
	Duration	Few days	Few days
	Past Medical History	Appendectomy	Vasectomy
	Medication	Multi-vitamin supplements	Nil
	Family History	Hypertension	N/A
	Social History	Lives with friends, family lives nearby, currently a Uni student, never smoked, drinks around 8 standards on "night out"	Lives with his wife, no children, currently work as a plumber, has 5 pack-year of smoking, enjoys 1–2 cans of beer every night, no recreational drug history
Case model	Summary	Pancreatitis is a sudden inflammation of pancreas.	Cholecystitis is a swelling and irritation of gallbladder.
	Examinations	Temp 37.8C BP 110/78 PR 110 and regular RR 16	Temp 37.5C BP 120/95 PR 100 and regular RR 18
	Investigations	FBC: Elevated WCC LFT: Possible increase AST & GGT Elevated C-reactive protein Serum amylase & lipase: Elevated CT imaging may be done to determine the severity	FBC: Elevated WCC Elevated C-reactive protein LFT: Mildly elevated bilirubin Abdominal US: Multiple gallstones present in the gallbladder with mild dilation of the biliary tree with a thickened gallbladder wall
	Differential diagnosis	Cholecystitis, choledocolithiasis, peptic ulcer, hepatitis, liver abscess, hiatal hernia, MI	Biliary colic, pancretitis, choledocolithiasis, cholangitis, peptic ulcer, hepatitis, liver abscess, hiatal hernia, MI
	Final diagnosis	Pancreatitis	Cholecystitis
Reference context variables		/stomachache/,/nausea/, /vomit/,/loss-appetite/, /radiate-pain/,/back/, /fever/,/appendectomy/ /hypertension/,etc.	/stomachache/,/nausea/, /vomit/,/loss-appetite/, /radiate-pain/,/fever/, /vasectomy/, etc.

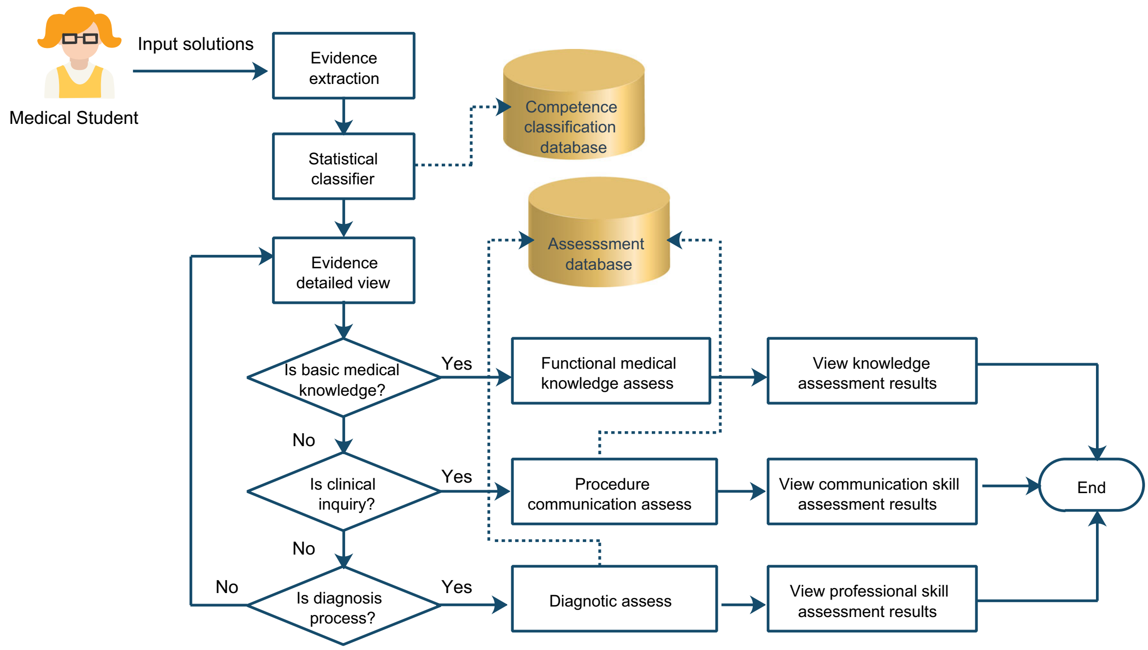
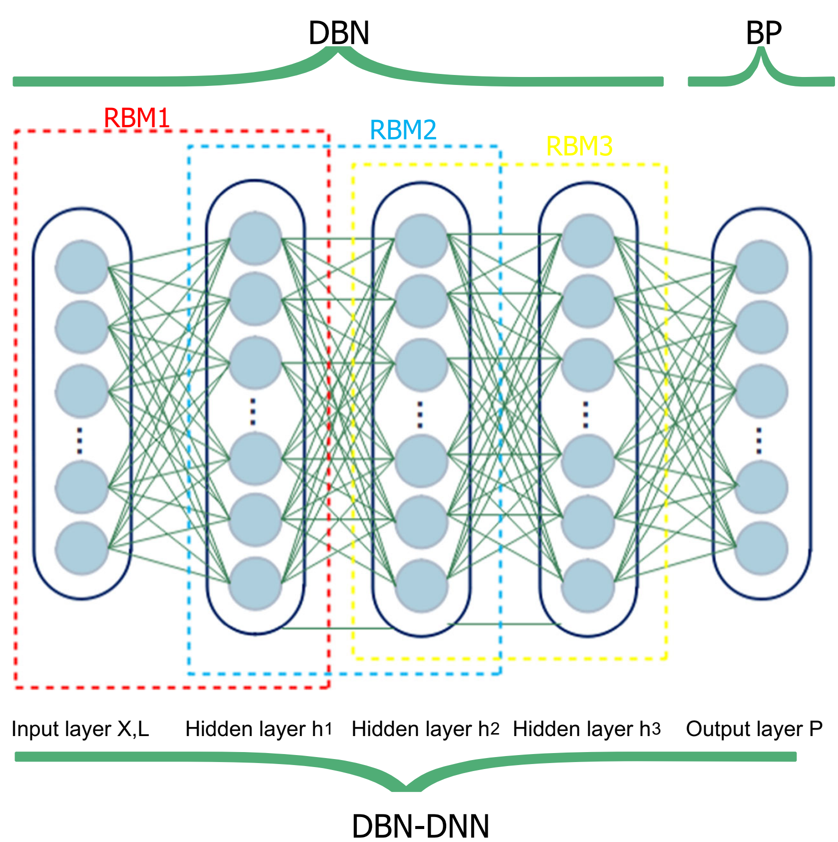


Fig. 8 Flow diagram of Competence-based Assessment module

training samples, etc. used in the designed cases which medical students need to practice in this report's virtual environment step-by-step. The system is able to record the student's solutions and show competency assessment results based on the generated modules. Furthermore, feedback from student is also collected through the platform. The detailed process of *Cholecystitis* dialogue training and assessment is described by the following steps.

Fig. 9 The network structure of DBN-DNN



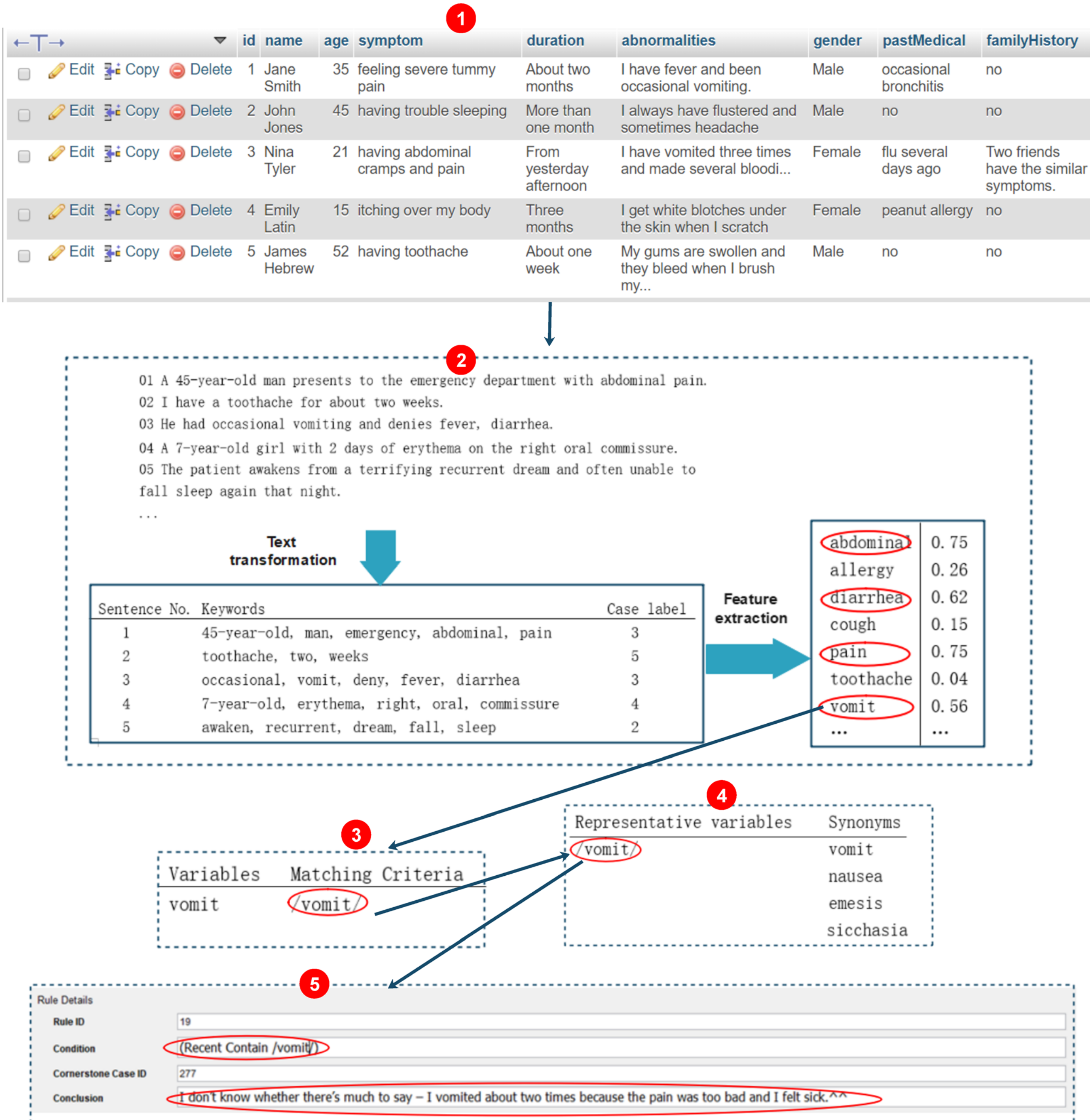


Fig. 10 Clinical dialogue creation: contain context variable */vomit/*

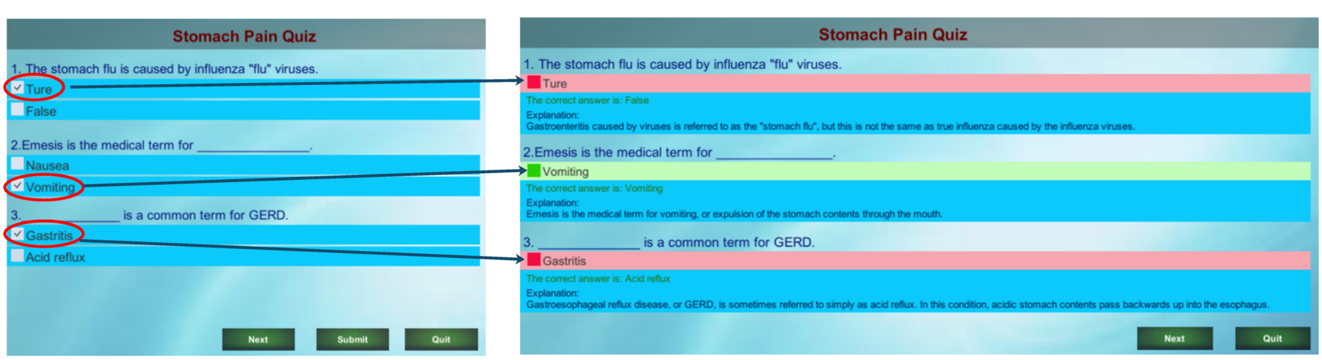


Fig. 11 Basic medical knowledge assessment referred to specific clinical case

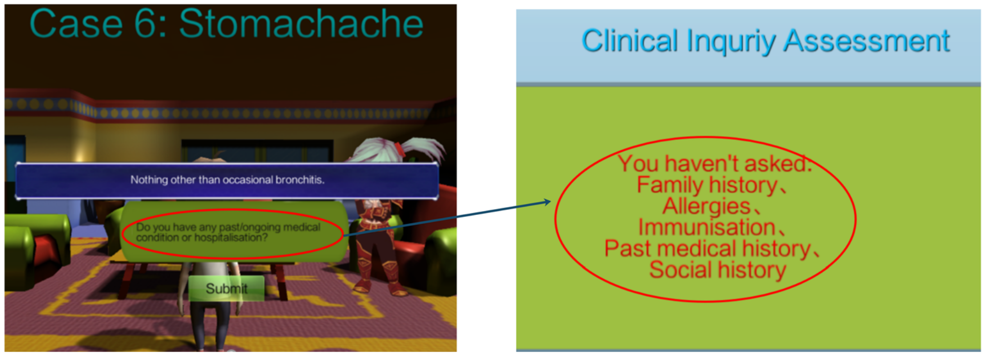


Fig. 12 Procedural communication assessment

Clinical dialogue training

This clinical training scenario is created by this report's medical tutor system semi-automatically. As described in section 3.2, each student's input inquiry was processed and presented for inference to the MCRDR knowledge-base to generate a patient's response. The five-step detailed process below are referenced in Fig. 10:

1. Design patient database according to real-world clinical trials and patient information, including patient name,

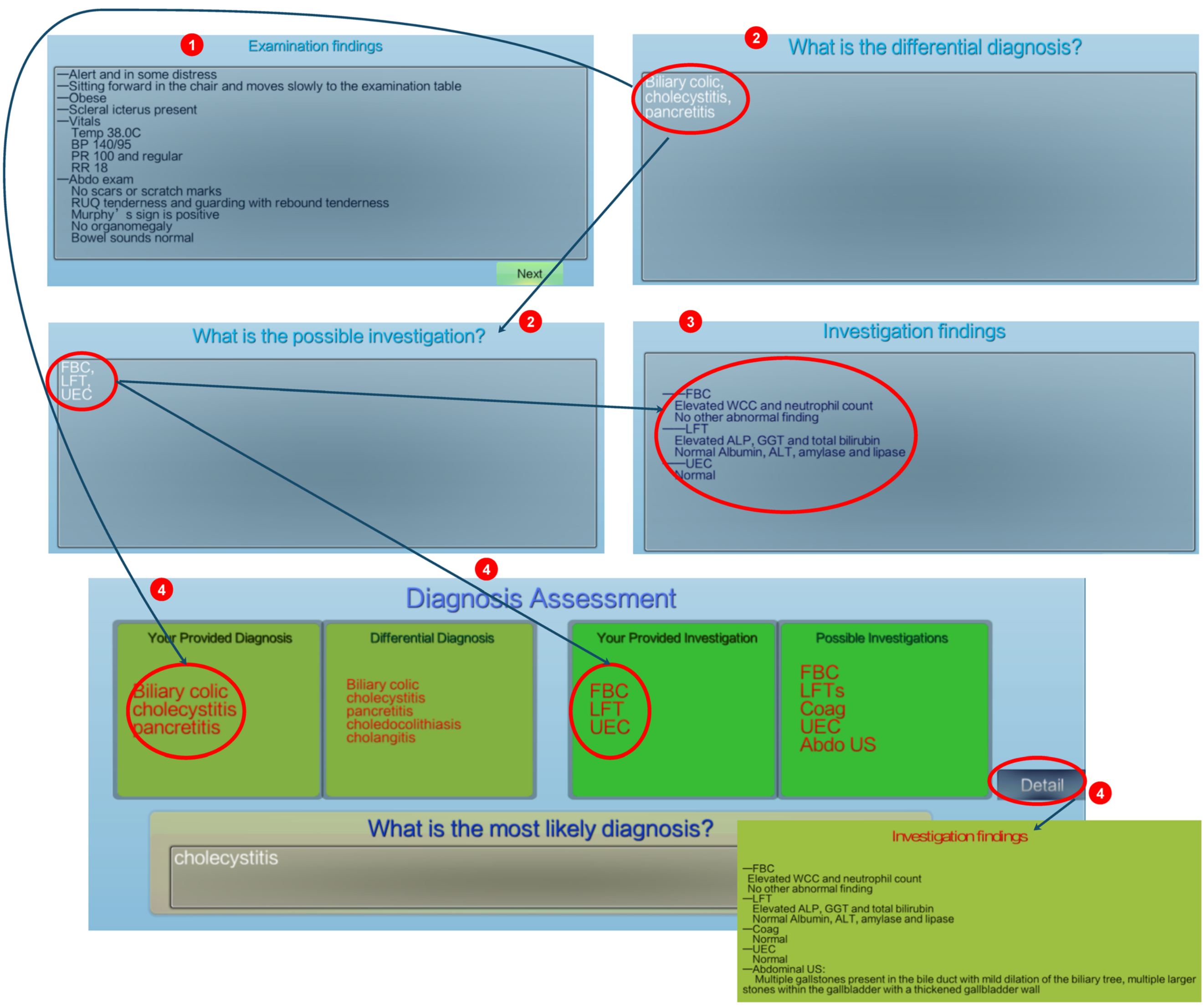


Fig. 13 Diagnosis and investigation assessment

age, gender, symptoms, duration, past medical history and family/social history, etc.

2. Select sets of keywords or phrases to serve as test samples to label specific cases for classification based on the generated supervised model.
3. Define context variables according to test results referred to in the previous step.
4. Define dictionary terms according to variable definition.
5. Define MCRDR rules that refer to student's inquiry containing context variables.

The created system simulated a doctor-patient dialogue session as below (the highlight terms are replaced as they match defined synonyms in dictionary terms):

- *Student: What is wrong with you today?*
- *Patient: I came in to ED because of this awful* ***abdominal pain.***
- *Student: Can you tell me more about this?*
- *Patient: I don't know whether there's much to say – I* ***vomited*** *about two times because the pain was too bad and I felt sick.*

Clinical skills assessment

In order to achieve efficient competency assessment, the created system includes multiple competency units including basic medical knowledge, clinical inquiry and diagnosis processing. After the input solutions with different classified competence were obtained, which includes quiz answers, clinical inquiry as well as investigation and diagnosis results, the three main parts were integrated to be the final assessment results.

1. Design an interactive quiz for referred medical knowledge in each clinical case, and complete knowledge assessment based on the student's quiz answer (Fig. 11).
2. Extract keywords or phrases from student procedural inquiry, including current medical conditions, the illness duration, medication status, the character of illness (such as how severe the pain is, pain location, etc.), family history, social history, personal habits, and so on. Analysis determines any omitted questions from the student and a procedural communication assessment results was generated automatically (Fig. 12).
3. On the basis of the case-based classification model, the professionalism skills assessment for diagnosis and investigation is intended to: 1) provide examination findings related to different cases, 2) allow student to write differential diagnosis and possible investigations, 3) based on the student's input, list the relevant investigation findings, 4) identify the keywords or phrases, and evaluate the diagnosis and investigation to help students better understand the case (Fig. 13).
4. Construct the above assessment to create competency assessment results for the specific case. In this step, the system integrates all the information as described in Step 1 to Step 3 to generate the final assessment results shown in Fig. 14.

In summary, medical students can get the assessment results with auto-generated visualization and analysis on the basis of the designed instructions, which will be stored in a domain database, and then the medical tutor can also review these records to collect and analyze the data captured from the medical students.

Evaluation

Objective

This section describes how to verify the effectiveness of the training and assessment modules. The feedback collected

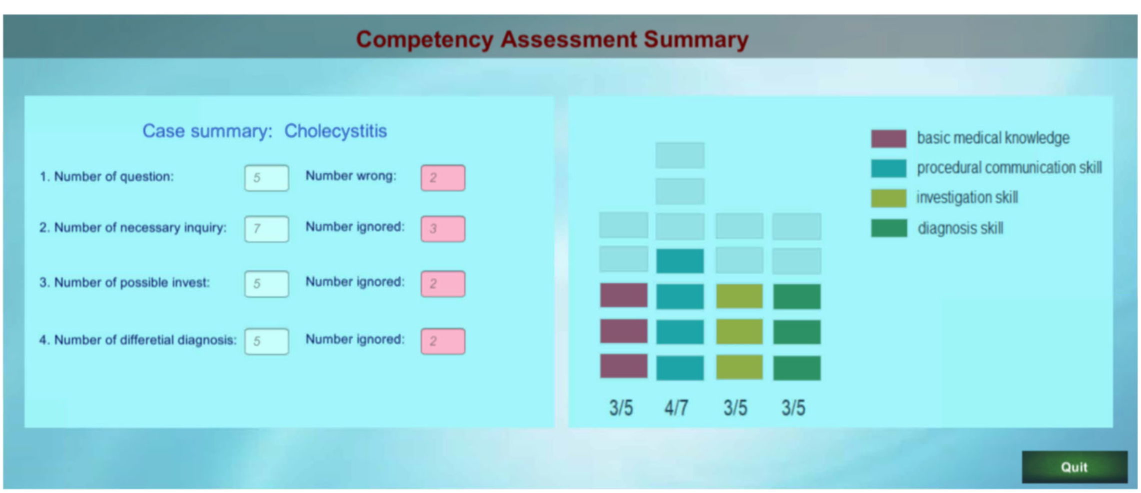


Fig. 14 Final competency assessment result

Table 6 Interactive evaluation criteria

Evaluation Object	Evaluation criteria	Evaluation criteria (for medical tutors)
Medical students	Training operation	• Appropriate for solo training. • Saves time and makes training more efficient
	Training effectiveness	• Useful for improving clinical skills • Appropriate for improving the motivation of learning • Appropriate for making students more confident in actual cases.
	Interface interaction	• Appropriate for making training more attractive and interesting. • Flexible and easy to use
Medical tutors	System capacity	• Reasonable design for clinical training • Useful for performing straightforward clinical cases • Appropriate for making assessment more effective.
	System flexibility	• Flexible data entry design • Flexible and easy to operate. • Flexible for designing different levels of clinical practices

from our system from 20 medical students, which included 5 tutors from the School of Medicine, was compiled. The feedback provided by the medical students as well as medical tutors was designed to discuss the holistic evaluation of our system, described in Table 6.

Experiment settings

The experiments were run on computers with the Microsoft Windows operating system. Table 7 lists the minimum hardware and software requirements to run the proposed platform on a Windows desktop PC.

Test of interactive efficiency

We evaluated the interactive efficiency between standard virtual patient system and the proposed system. We divide 20 students into baseline group and explore group with similar capacity. For both baseline and explore groups, students are assigned into 5 groups with 2 students of each group.

In this step, two different medical cases were selected and used as the test set. All the students were asked to read necessary instructions in advance to make preparation. For the baseline group, the student used the standard virtual patient system [6], while for the explore group they used our proposed system. During the test, we designed the same or similar questions with synonyms to ask and evaluated the responses to check the overlapping interactive accuracy by the medical experts. The detailed evaluation results of the two cases are shown in Table 8.

From the above test results, we can find the proposed MCRDR training system has higher average interactive accuracy compared with the standard virtual patient system. Its main advantage is mainly reflected in two aspects:

1) Support the multi conditions multi conclusions. For example, when we input “Do you have /vomiting/ or /diarrheal/?”, which includes two context variables, the standard virtual patient system can only get single classification conclusion” My stool seems normal”, while our proposed system can support multiple classification conclusions “I have vomited three times and my stool seems normal”.
2) Higher context variable matching. Since our proposed system applied pattern evaluation based on deep neural

Table 7 The hardware and software requirements to run our proposed 3D clinical training platform

Components	Requirements
Operating System	Windows 10 Fall Creators Update (RS3) - Home, Pro, Business, Education.
CPU Processor	Intel Core i5 7200U (7th generation mobile), dual-core with Intel® Hyper-Threading Technology enabled.
Disk Space	At least 10 GB.
RAM	8GB DDR3 dual channel (or better).
Graphics Driver	Windows Display Driver Model (WDDM) 2.2
Graphics Card	Integrated Intel® HD Graphics 620 (or greater)
Graphics Display Port	HDMI 1.4 or DisplayPort 1.2

Table 8 The comparison results of interactive efficiency

Case 1: Ear Pain	Baseline group					Explore group				
	1	2	3	4	5	1	2	3	4	5
Interaction examples	S: Can I confirm your name and age first? P: 15 S: what's wrong with you? P: My ear really hurts. S: How strong is the pain? Let's say in a 1 to 10 scale, how would you describe the intensity of the pain? P: My ear really hurts but can still sleep. S: Did you have ear pain before? P: Yes, my ear really hurts.					S: Can I confirm your name and age first? P: My name is Anna and I'm 15. S: what's wrong with you? P: My ear really hurts. S: How strong is the pain? Let's say in a 1 to 10 scale, how would you describe the intensity of the pain? P: 4–5. S: Did you have ear pain before? P: Yes, I had three times in the past one year.				
Total length	20	18	16	20	18	20	18	16	20	18
Overlapping interactive accuracy	70%	78%	68%	65%	72%	82%	83%	87%	85%	80%
Average accuracy	70.6%					83.4%				
Case 2: Stomachache										
	Baseline group					Explore group				
	1	2	3	4	5	1	2	3	4	5
Interaction examples	S: How is your feeling today P: I am Ok except this problem. S: Do you have vomiting or diarrheal? P: My stool seems normal. S: Did you eat outside yesterday? P: My diet is pretty good.					S: How is your feeling today P: I have a strong stomachache. S: Do you have vomiting or diarrheal? P: I have vomited three times and my stool seems normal. S: Did you eat outside yesterday? P: Yes, I eat some seafood yesterday.				
Total length	25	22	24	25	22	25	22	24	25	22
Overlapping interactive accuracy	76%	81%	79%	76%	77%	88%	90%	91%	88%	86%
	77.8%					88.6%				

network to select the conditions for generating MCRDR inferences, the interactive efficiency improved a lot. For example, when we input "How strong is the pain? Let's say in a 1 to 10 scale, how would you describe the intensity of the pain?", the standard virtual patient system outputted the responses "My ear really hurts, but can still sleep.", however, the detailed levels from 1 to 10 are expected response. Thus, the context variable matching of traditional method is not efficient than our proposed system.

Attitude survey

At the end of the evaluation, we also asked medical students as well as tutors to submit sentiment orientation as feedback about our proposed system as represented in Table 6, to training operations, training effectiveness, system interface design, system capacity and flexibility, etc., and the feedback was collated and stored in a feedback database for later analysis. A five-point Likert scale was used by each user to rate each

Table 9 Interactive evaluation results of medical student participants

Categories	Statements	Five-point Likert scale (proportion of responses)				
		1	2	3	4	5
Training operation	It is appropriate to make clinical training more effective.	0.00	0.05	0.05	0.45	0.45
	It is useful for performing straightforward clinical cases.	0.00	0.15	0.15	0.45	0.25
Training effectiveness	My skills are improved a lot after employing virtual training.	0.00	0.05	0.20	0.45	0.30
	It is really interesting and could improve the motivation for learning.	0.00	0.10	0.40	0.30	0.20
	It makes me more confident in actual medical cases after employing virtual system.	0.00	0.00	0.20	0.40	0.40
System interface	It is better to make medical training more attractive.	0.00	0.10	0.10	0.40	0.40
	It is flexible and easy to use.	0.00	0.05	0.05	0.40	0.50

1, Strongly disagree; 2, Disagree; 3, Undecided; 4, Agree; 5, Strongly agree

Table 10 The interactive evaluation results of medical tutors

Categories	Statements	Five-point Likert scale (overall percentage)				
		1	2	3	4	5
System capacity	It is a reasonable design for clinical training.	0.00	0.05	0.40	0.30	0.25
	It saves a lot of time to employ this virtual training system.	0.00	0.00	0.20	0.40	0.40
	It is appropriate to make assessment more effective.	0.00	0.05	0.20	0.55	0.20
System flexibility	It is flexible and easy to operate.	0.00	0.10	0.25	0.35	0.30
	It is flexible for data entry design.	0.00	0.15	0.25	0.55	0.05
	It is flexible for designing different levels of experimental practices.	0.00	0.15	0.25	0.50	0.10

1, Strongly disagree; 2, Disagree; 3, Undecided; 4, Agree; 5, Strongly agree

statement. Tables 9 and 10 show the statements and data collected from medical students and tutors respectively in the feedback.

Based on the results, it was found that the majority of both medical students and tutors gave positive feedback to all corresponding statements. Statements with the highest proportional responses from the student survey are related to making students more confident, improving skills and flexibility, while the highest three scores from the tutor survey are about efficient assessment, saving time and flexibility to design. This confirms that the proposed platform of this research not only improves the medical students' clinical skills, but also makes medical training and assessment more effective.

Furthermore, Fig. 15 represents the detailed confidence of each evaluation criteria, including training operation, training effectiveness, system interface, system capacity and flexibility, and it shows that more than 70% of all users are satisfied with the proposed system. In this way, it can be seen that appropriateness of the training operation and system interface was greater than 75%, while there is a need for system flexibility to be strengthened in future research.

Discussion

This research describes how to design a dialogue-driven simulation platform to achieve clinical training and assessment. The main aim of medical students is to communicate with virtual patients through various real-world cases based on an MCRDR knowledge-base, and then provide differential diagnoses and investigations. Furthermore, this platform supports the automatic competency assessment that can be useful for generating feedback between medical students and tutors. The evaluation experiments indicate that the platform proposed by this research will be effective in clinical training and assessment.

During the dialogue-driven simulation, it was envisioned that the MCRDR-based NLU module would be the basis for the dialogue representation that solves the classification issues faced in current dialogue-based systems. However, there are several challenges that need to be addressed prior to the MCRDR NLU model creation. Here, the key challenges are noted:

Guide to asking effective questions: the new MCRDR rules are lacking an efficient scheme to guide users asking valid questions. Due to this, the invalid questions always tend to get useless responses, such as "Sorry, I don't understand" or "Do you mean…?" Therefore, appropriate V&V needs to be

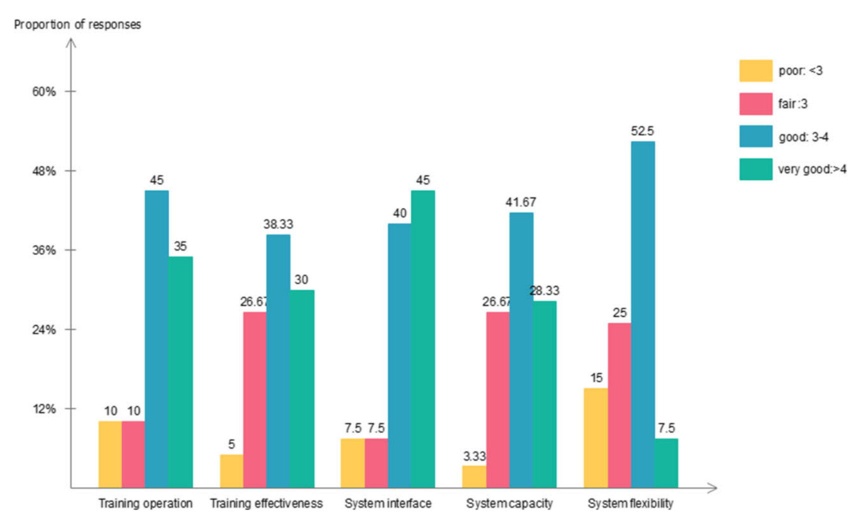


Fig. 15 Evaluation results of each criteria

considered in this context and future work will investigate any improvements for asking better questions with a dialog-based approach containing contextual meta-data.

Interoperability of multiple knowledge-bases: each expert generates an MCRDR knowledge base on the basis of their domain expertise, which may result in inter-expert disagreements. To maintain a global state of knowledge, weighted, comparison inference results between distributed knowledge bases would be helpful in order to obtain responses that are more accurate.

In addition, the system flexibility, especially for medical tutors to manage enrolments and generated dialogues also needs to be improved in future research.

Conclusions

The goal of this paper was the design and development of an interactive training and assessment platform using dialogue simulation to realize effective clinical training and collaborative assessment. To achieve this goal, we designed a natural language understanding based on pattern-matching with an MCRDR knowledge-base to create the semantic representations of requests with contextual responses from inference, then we used deep neural networks to evaluate the generated MCRDR inference results. Furthermore, a competency assessment module integrated with statistical classification model was presented to classify various solutions into multiple competencies including basic medical knowledge, as well as procedural communication and professionalism skills. Finally, the proposed platform was evaluated using a case study, which was qualitatively measured from the perspectives of both medical students and medical tutors in terms of training operation, training effectiveness, system interface, system capacity and flexibility. Based on the collected feedback, the preliminary data from the experiment indicatively supports our assumption that our proposed virtual dialogue simulation could potentially be more effective in terms of making students more confident, improving their skills, and allowing for a flexible approach to interactive training and assessment while also saving time.

Compliance with ethical standards

Declaration of conflict of interest Wenli Yang, David Herbert, Sunwoo Kim and Byeong Kang declare no conflict of interest directly related to the submitted work.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Owen, H., Follows, V., Reynolds, K., Burgess, G., and Plummer, J., Learning to apply effective cricoid pressure using a part task trainer. Anaesthesia 57(11):1098–1101, 2002.
2. Girzadas, Jr., D. V. et al., Hybrid simulation combining a high fidelity scenario with a pelvic ultrasound task trainer enhances the training and evaluation of endovaginal ultrasound skills. Acad. Emerg. Med. 16(5):429–435, 2009.
3. Jansen, D. A., Johnson, N., Larson, G., Berry, C., and Brenner, G. H., Nursing faculty perceptions of obstacles to utilizing manikin-based simulations and proposed solutions. Clinical Simulation in Nursing 5(1):e9–e16, 2009.
4. Vincent, D. S., Berg, B. W., and Ikegami, K., Mass-casualty triage training for international healthcare workers in the Asia-Pacific region using manikin-based simulations. Prehospital and Disaster Medicine 24(3):206–213, 2009.
5. Hubal, R. C., Kizakevich, P. N., Guinn, C. I., Merino, K. D., and West, S. L., The virtual standardized patient. In *Medicine Meets Virtual Reality*, pp. 133–138, 2000.
6. Kenny, P., Virtual Patient, 2012.
7. Bruce, B., Case systems for natural language. Artif. Intell. 6(4): 327–360, 1975.
8. Lehnert, W. G., Knowledge-based natural language understanding. In *Exploring artificial intelligence*: Elsevier, pp. 83–131, 1988.
9. Aha, D. W., Maney, T., and Breslow, L. A., Supporting dialogue inferencing in conversational case-based reasoning. In *European Workshop on Advances in Case-Based Reasoning*, pp. 262–273: Springer, 1998.
10. Inui, N., Ebe, T., Indurkhya, B., and Kotani, Y., A case-based natural language dialogue system using dialogue act. In: *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, vol. 1, pp. 193–198: IEEE, 2001.
11. Aleven, V., Popescu, O., and Koedinger, K. R., A tutorial dialogue system with knowledge-based understanding and classification of student explanations. In *Working Notes of 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*: Citeseer, 2001.
12. Sun, J., Karray, F., and Basir, O. A., Knowledge-based flexible natural speech dialogue system. ed: Google Patents, 2008.
13. Wanner, L. et al., Kristina: A knowledge-based virtual conversation agent. In: *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pp. 284–295: Springer, 2017.
14. Abella, A., Brown, M. K., and Buntschuh, B. M., Methods and apparatus object-oriented rule-based dialogue management. ed: Google Patents, 2000.
15. Sun, K., Chen, L., Zhu, S., and Yu, K., A generalized rule based tracker for dialogue state tracking," in *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pp. 330–335: IEEE, 2014.
16. Lhioui, C., Zouaghi, A., and Zrigui, M., A Rule-based Semantic Frame Annotation of Arabic Speech Turns for Automatic Dialogue Analysis. Procedia Computer Science 117:46–54, 2017.
17. Kang, B., Compton, P., and Preston, P., Multiple classification ripple down rules: evaluation and possibilities. In *Proceedings 9th Banff knowledge acquisition for knowledge based systems workshop*, vol. 1, pp. 17.1–17.20, 1995.
18. Gaines, B. R., and Compton, P., Induction of ripple-down rules applied to modeling large databases. J. Intell. Inf. Syst. 5(3):211–228, 1995.
19. Kim, Y. S., Park, S. S., Deards, E., and Kang, B. H., Adaptive web document classification with MCRDR. In *null*, p. 476: IEEE, 2004.
20. Han, S. C., Yoon, H.-G., Kang, B. H., and Park, S.-B., Using MCRDR based Agile approach for expert system development. Computing 96(9):897–908, 2014.

21. Herbert, D., and Kang, B. H., Intelligent Conversation System using Multiple Classification Ripple Down Rules and Conversational Context. Expert Syst. Appl., 2018.
22. BROOKES, D., Objective structured clinical examination assessment. Nurs. Times 103(43):30–31, 2007.
23. Sloan, D. A., Donnelly, M. B., Schwartz, R. W., Felts, J. L., Blue, A. V., and Strodel, W. E., The use of the objective structured clinical examination (OSCE) for evaluation and instruction in graduate medical education. J. Surg. Res. 63(1):225–230, 1996.
24. Wallace, J., Rao, R., and Haslam, R., Simulated patients and objective structured clinical examinations: review of their use in medical education. Adv. Psychiatr. Treat. 8(5):342–348, 2002.
25. Baxter, G. P., Elder, A. D., and Glaser, R., Knowledge-based cognition and performance assessment in the science classroom. Educ. Psychol. 31(2):133–140, 1996.
26. Kim, Y., Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882,* 2014.
27. Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H., Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, pp. 153–160, 2007.