

Research paper

Efficient symptom inquiring and diagnosis via adaptive alignment of reinforcement learning and classification

Hongyi Yuan, Sheng Yu *

Center for Statistical Science, Tsinghua University, Haidian District, 100084, Beijing, China

Department of Industrial Engineering, Tsinghua University, Haidian District, 100084, Beijing, China

ARTICLE INFO

Dataset link: <https://github.com/Yuanhy1997/Auto-Diagnosis-by-RL-and-Classification>

Keywords:

Medical automatic diagnosis
Reinforcement learning
Adaptive alignment criterion

ABSTRACT

Medical automatic diagnosis aims to organize real-world diagnostic processes similar to those from human doctors and to achieve accurate diagnoses by interacting with patients. The task is formulated as a sequential decision-making problem with a series of information inquiry steps (asking about symptoms and ordering examinations) and the final diagnosis. Recent research has studied incorporating reinforcement learning for information inquiry and classification techniques for disease diagnosis, respectively. However, studies on efficiently and effectively combining the two procedures are still lacking. To address this issue, we devised an adaptive mechanism to align reinforcement learning and classification methods using distribution entropy as the medium. Additionally, we created a new dataset for patient simulation to address the lack of large-scale evaluation benchmarks. The dataset is extracted from the MedlinePlus knowledge base and contains significantly more diseases and more comprehensive symptom and examination information than existing datasets. Experimental evaluation shows that our method outperforms three current state-of-the-art methods on different datasets by achieving higher medical diagnostic accuracy with fewer inquiring turns.

1. Introduction

Medical automatic diagnosis (MAD) is one of the main research areas in the application of artificial intelligence in healthcare. An MAD system can organize the diagnostic processes similar to those from human doctors which includes interacting with the patients, collecting their symptom information, and making accurate diagnoses. Currently, people perform online searches for self-diagnosis. However, many results obtained from these searches may be inaccurate or irrelevant. Besides, in real clinical encounters, diagnostic accuracy can be low in many rural areas due to the limited experience of the local doctors. Therefore, a well-developed MAD system has great potential in supporting clinical diagnostic decisions or self-diagnosis and improving healthcare.

Many previous studies utilized structured or unstructured medical data to develop a disease classification system. A large variety of machine learning or deep learning methods have been used to build such systems [1,2]. However, these methods can only make diagnoses given rich medical histories. Doctors may meet patients with no clinical history and limited initial symptom information from chief complaints in a clinical setting. These methods will not perform well in such scenarios.

Therefore, in such a clinical scenario, a medical diagnosis process can be formulated as a sequential decision-making problem [3]. The MAD system collects patients' symptom information by posing symptom-related queries or assigning medical examinations to patients. When the system collects enough information, it gives a final diagnosis of the disease. While more information is always beneficial for diagnostic accuracy, asking too many questions diminishes the quality of the patient's experience, and taking too many examinations raises medical costs. The goal of MAD is to achieve high diagnostic accuracy and reduce the costs of collecting symptom information.

Collecting patients' condition information requires multi-round interactions with patients by inquiring about symptoms or assigning examinations while giving the diagnosis is a single-step decision. Considering the difference between symptom-inquiring and diagnosis, in our work, we regard symptom-inquiring actions as a sequential decision-making problem that is approached using reinforcement learning (RL), and the final diagnosis is handled with a classification model. To bridge the two parts, we monitor the entropy of disease distribution at each interaction turn and the entropy thresholds for transitioning are also assumed to be different for each disease. Finding a proper threshold for

* Corresponding author at: Center for Statistical Science, Tsinghua University, Haidian District, 100084, Beijing, China.

E-mail address: syu@tsinghua.edu.cn (S. Yu).

each disease is critical to the performance of the MAD system. We propose an adaptive approach that learns disease-independent thresholds along with training.

Training RL models requires an interactive environment for the agent to explore the diverse actions given various symptom states. The previous works trained and evaluated their methods on simulated patients based on medical knowledge bases or medical dialogue histories. However, these datasets contained very few diseases and simplified symptoms, which made the simulated patients simple and unrealistic. In this work, we propose a new dataset for more realistic patient simulation. The main contributions of this work are as follows:

- We publish a new dataset for synthetic patient simulation. Compared to other widely used datasets, the proposed dataset contains more comprehensive disease information, and we include medical examinations for the first time.
- We propose a simple yet effective method to automatically adjust entropy thresholds that incorporate an RL agent and a classifier. The thresholds change along with the training of the RL agent and the classifier. Through extensive experiments, we show the efficiency and superior performance of the proposed method compared to other MAD methods, as well as the efficiency and effectiveness of the threshold selection.

2. Related work

Automatic medical diagnosis or clinical diagnosis inference has drawn much attention in healthcare research. Many previous studies developed disease classification systems by machine learning or deep learning methods [1,2,4,5]. For example, Choi et al. [6] and Choi et al. [7] developed recurrent neural network (RNN)-based methods to perform differential diagnoses given the patient's medical histories and current conditions. Hosseini et al. [8] used graph neural networks for disease diagnosis based on medical histories and knowledge graphs. Some recent works were inspired by TREC-CDS tasks. They applied deep RL to retrieve medical disease concepts from clinical narratives with external knowledge bases (e.g., Wikipedia) [9,10]. However, these methods can only make diagnoses given comprehensive medical histories and patients' conditions. These methods will fail in real clinical scenarios where doctors may have limited awareness of patients' conditions from their chief complaints.

Some studies developed dialogue systems to interact with patients and give more personalized medical instructions [11,12]. These methods focused on natural language processing and could only provide a one-step dialogue interaction with the patients based on real dialogue histories. Our work does not aim to generate comprehensive dialogues or one-shot instructions but to approach the inner decision-making logic of real diagnostic processes to achieve a highly effective and efficient automatic diagnosis.

Considering the real clinical scenarios, recent research treated automatic diagnosis as sequential decision-making problems. With the promising capacity of reinforcement learning (RL) methods in solving such problems in various fields, some works developed symptom checkers for online healthcare services by RL. Tang et al. [13] divided the diseases and related symptoms into 11 anatomical parts and applied deep Q-learning (DQN) [14] methods on each part to inquire about symptoms. Wei et al. [15] also used DQN methods to infer diagnosis. Kao et al. [16] and Liao et al. [17] used hierarchical RL: a master agent is assigned to decide which anatomical part agent of the lower hierarchy to inquire about symptoms and when to give the diagnosis. Kao et al. [16] also added contextual information about a patient (e.g., sex and age) to enhance the diagnostic accuracy through a Bayesian approach. Splitting diseases into several anatomical parts can reduce task complexity. With appropriate joint policy training, the above works outperform naïve DQN methods. Xu et al. [18] proposed a method called KR-DQN that embedded entity relations into

a DQN agent. A trainable correlation matrix was added to the agent outputs to help inquire about related symptoms and give diagnoses. Peng et al. [19] proposed a method based on the policy gradient [20] method, called REFUEL. They designed auxiliary rewards and used a dual neural network architecture to rebuild the sparse features of the patients to help discover the positive symptoms more quickly.

All of the above-mentioned works simply combined symptom-inquiring and disease diagnosis into the same action space. This unified framework neglects the difference between these two kinds of actions. Symptom-inquiring actions aim to obtain informative symptoms with as few steps (or other kinds of costs) as possible, while the diagnosis action is single-step and aims to obtain an accurate diagnosis at the end of each diagnostic process. The former actions may fall into the RL framework, while the latter is naturally a classification problem. Forcing diagnosis as part of the sequential decision making process increases the search space for the agent, limiting not only the diagnostic accuracy but also the symptom inquiring efficiency. In addition, they ignored the fact that negative symptoms also contain rich information and are critical in accurate clinical decision-making.

More recent works used a bipartite framework and solved the two kinds of actions separately. They used RL methods to inquire about symptoms and employed another disease classifier to give diagnoses given the current symptom information. A variety of methods were proposed to integrate RL agents and classifiers. Xia et al. [21] used a generative adversarial network [22] framework to balance symptom inquiring and disease diagnosis, and also designed entropy-based rewards to guide the agent to inquire informative and rational questions. Their method trivially gives the diagnosis when the maximum inquiring step length is reached, or the entropy of the disease distribution is smaller than a fixed threshold. The threshold was chosen intuitively, and all diseases shared the same threshold. However, a proper threshold may differ across datasets or even diseases. Lin et al. [23] proposed a method called INS-DS comprising two cooperative modules: a symptom inquiry module and an introspective module. The introspective module intervenes in the potential responses of the inquiry and decides to give the diagnosis if the diagnoses of these interventions remain unchanged. Lin et al. [24] designed a criterion that the model gives the diagnosis when the probability of the preferred disease is beyond the upper bound of the 6σ confidence interval of the other diseases' probabilities. The confidence interval is estimated by bootstrapping the classifiers, which would be computationally costly. Although they achieve good automatic diagnosis performance, the above methods are only tested on datasets containing few diseases or symptoms. These methods may fail or be inefficient when given a large decision space.

A more recent work proposed a non-RL competitive bipartite framework, called FIT, using a multi-modal variational autoencoder model and a two-step sampling strategy for disease prediction [25]. They reported state-of-the-art results on different datasets.

3. Method

In general, an MAD system is comprised of two parts: (1) the interaction part that parses and analyzes the input language from the patient and outputs the inquiry to the patient; (2) the decision-making part that decides the information to inquire about and when to give the diagnosis according to the collected information. In this paper, we only focus on the decision-making part, as it is central to MAD.

Fig. 1 is an overview of the proposed MAD method, and its pseudo-code is given in Algorithm 1. The model consists of a symptom-inquiring agent, a disease-diagnosing module, and a stopping criterion. The symptom-inquiring agent is a policy network inquiring about the patient's symptoms, given the symptoms already known by the model. The disease-diagnosing module is a neural network classifier that outputs a disease classification given the currently known symptoms. The stopping criterion sets adjustable thresholds for each disease and

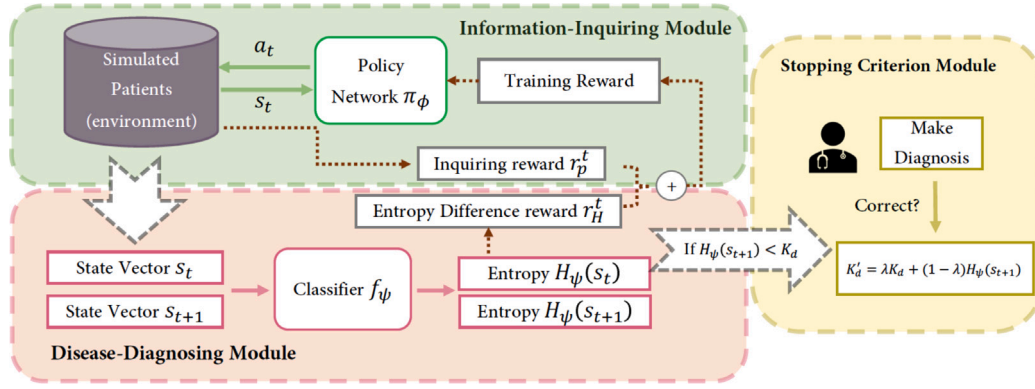


Fig. 1. Algorithm overview.

Algorithm 1**Input:** Disease knowledge base $\mathcal{D}$, Initial threshold K_{init} **Output:** Optimal policy π_{ϕ^*} , classifier f_{ψ^*} , threshold K^*

- 1: Initialize π_ϕ and f_ψ with random weights ϕ and ψ .
- 2: Initialize the threshold $K = K_{\text{init}}$.
- 3: Generate synthetic patients using $\mathcal{D}$.
- 4: **repeat**
- 5: Sample a patient with disease d_{True} and generate initial state vector s_0 .
- 6: Calculate the initial entropy $H_\psi(s_0)$.
- 7: **while** $t \leq T$ **do**
- 8: Inquire a symptom $a_t \sim \pi_\phi(s_{t-1})$.
- 9: Interact with the patient and obtain the next state s_t and reward r_t .
- 10: Give a disease diagnosis attempt d^t and calculate the disease distribution entropy $H_\psi(s_t)$.
- 11: Store a training sample $(s_{t-1}, a_t, r_t, d_{\text{True}})$.
- 12: **if** $H_\psi(s_t) < K_{d^t}$ **then**
- 13: End the diagnostic process.
- 14: **end if**
- 15: **end while**
- 16: The final diagnosis $d_{\text{diag}} = d^t$.
- 17: Update the policy and classifier network parameters using collected samples.
- 18: **if** $d_{\text{diag}} = d_{\text{gold}}$ **then**
- 19: Update threshold $K_{d_{\text{diag}}}$ by Equation (1).
- 20: **end if**
- 21: **until** End of epochs

decides when to stop inquiring about symptoms and to give a diagnosis. At each turn, the disease-diagnosing module receives the current patient's symptom information vector and predicts the current disease distribution. Then we use this disease distribution to check and adjust the stopping criterion.

3.1. Notations

We denote the number of all symptoms as N and the number of all diseases as M . We design our symptom-inquiring agent based on an RL framework. The symptom-inquiring part is a finite-horizon MDP with a state space $\mathcal{S}$, an action space $\mathcal{A}$, and a policy π_ϕ . Each state $s \in \mathcal{S}$ is a sparse vector of length N , and each entry of the vector indicates the status of the corresponding symptom, *positive*, *negative*, or *unknown*, encoded by 1, -1, and 0, respectively. Each action $a \in \mathcal{A}$ is a symptom to inquire about. The size of the action space is N . The disease-diagnosing part is formulated as a classification task. The classifier f_ψ receives the current patient's state s as input and outputs a disease

distribution denoted as $p_\psi(d|s)$ and $p_\psi(d|s) = f_\psi(s)$, where d denotes the diseases. Then, the entropy $H_\psi(s)$ of $p_\psi(d|s)$ is calculated to check the stopping criterion, formulate the reward for the symptom-inquiring agent and adjust the entropy thresholds of each disease.

3.2. Disease-diagnosing model

The disease-diagnosing model is a Multi-layer Perceptron (MLP) classifier. Given the patient's state s_t at time step t , the state vector is fed to the classifier to generate the disease distribution. We derive the current diagnosis attempt d^t and the entropy of the disease distribution $H_\psi(s_t)$ for the distribution as follows:

$$H_\psi(s_t) = - \sum_{i=1}^M p_\psi(d_i|s_t) \log(p_\psi(d_i|s_t)),$$

$$d^t = \operatorname{argmax}_{d \in \{d_i, 1 \leq i \leq M\}} p_\psi(d|s_t).$$

The entropy $H_\psi(s_t)$ and the diagnosis attempt d^t are used to check the inquiring stopping criterion. If the $H_\psi(s_t)$ is smaller than the threshold of d_t , the MAD system terminates inquiring about patients' symptoms and returns d^t as the final diagnosis.

The MLP classifier is trained together with the RL agent in a supervised fashion. At each inquiry turn, the current state vector is paired with the true disease label to form a training sample (s_t, d_{True}) . Then, we train the classifier using these collected samples with cross-entropy loss. Thus, the classifier is accustomed to predicting the disease with partial symptom information.

3.3. Symptom-inquiring agent

The symptom-inquiring agent π_ϕ generates actions concerning the patients' undiscovered symptoms given the patients' current states. The agent is expected to inquire about the most informative symptoms for obtaining an accurate diagnosis.

The objective of RL is to maximize the expected accumulated rewards given a current state s :

$$\pi_{\phi^*} = \operatorname{argmax}_{\pi_\phi} \mathbb{E}_{\pi_\phi(a|s)} [R(s, a)],$$

where $R(s, a) = \sum_{t=s}^T \gamma^t r_t$ is the accumulated reward from the step t_s with state s and action a from $\pi_\phi(a|s)$, γ is the discount factor and r_t is the reward at each decision step t . We design a composite reward r of two parts: r_p and r_H to incentivize informative symptom inquiring for accurate diagnosis. The details of the reward design will be discussed in Section 3.4.

We also introduce an entropy regularization term

$$H(\pi_\phi(a|s)) = - \sum_{i=1}^N \pi_\phi(a_i|s) \log(\pi_\phi(a_i|s)),$$

to the objective function to help the symptom-inquiring agent explore high-reward actions and escape from local optima at the beginning of the training stage. The final objective becomes:

$$\max_{\phi} \mathbb{E}_{\pi_{\phi}(a|s)}[R(s, a)] + \beta H(\pi_{\phi}(a|s)),$$

Following the policy gradient algorithm [20], the parameter updating rule with respect to the objective is:

$$\phi' = \phi + \alpha \left[\frac{R(s, a) \nabla_{\phi} \pi_{\phi}(a|s)}{\pi_{\phi}(a|s)} + \beta \nabla_{\phi} H(\pi_{\phi}(a|s)) \right],$$

where α is the learning rate and β is a weight parameter.

3.4. Reward design

In clinical encounters with medical professionals, asking about many negative symptoms is a waste of time, and taking many medical examinations is a waste of medical resources and brings additional cost and even physical harm to the patient. Hence, we give a negative reward for each query. If the query discovers an unknown symptom, we assign an additional positive reward for the agent. The magnitude of positive rewards depends on the feedback from the patients. The goal of collecting symptom information is to give confident diagnoses. When our method gives a correct diagnosis, we provide the agent a positive reward, while when the method gives an incorrect diagnosis or fails to diagnose within the maximum step length T , a negative reward is given. This part of the reward denoted as r_p is designed to incentivize the agent to inquire about unknown and positive symptoms, pose fewer queries, and give correct diagnoses.

An informative symptom is not necessarily positive. Negative symptom feedback can also lower the uncertainty of diagnosis. Such negative symptoms can help rule out some candidate diseases. In information theory, entropy measures the level of uncertainty of a distribution. Thus, we track the entropy given by the disease classifier $\psi(s_t)$ at each time step t . The amount of information acquired by a symptom query is quantified by the entropy difference in the two successive inquiring steps. We introduce this to the reward design to guide the model to consider the symptom queries with more potential information feedback. The entropy difference reward is formulated as:

$$r_H = \max((H_{\psi}(s_t) - H_{\psi}(s_{t+1}))/H_{\psi}(s_0), 0).$$

where s_t is the patient current state and s_{t+1} denotes the next state. $H_{\psi}(s_0)$ is the disease distribution entropy based on the patient's initial self-reports. It serves as a normalizing term to make the entropy difference reward consistent across different patients because the entropy difference can vary from very large to small given different patients' self-reports. The maximum operator is to avoid negative information gain in the early stage of training when the classifier is imperfect because the entropy should decrease monotonically as more symptom statuses are collected.

Thus, the overall reward for training the symptom-inquiring RL agent is a weighted combination of r_p and r_H :

$$r = r_p + \nu r_H,$$

where ν balances the two kinds of rewards. We run grid search on difference choices of ν , and found that the choice of the value did not bring significant difference to the final accuracy, and that they all outperformed the baselines. In the main experiments, we use $\nu = 2.5$.

3.5. Stopping criterion

To decide when to stop posing queries to give diagnoses. It is intuitive to make the transition when the entropy of the disease distribution is lower than a threshold. The threshold indicates an attainable uncertainty degree for a disease, which is difficult to determine manually. Our method contains an adaptive design to automatically find the proper thresholds.

We regard the thresholds for each disease as tunable parameters. In the training stage, we initialize the thresholds with arbitrary values. For each training episode, the disease classifier network will give a disease diagnosis when the distribution entropy is lower than the current threshold of the disease or the preset max step T is reached. If the given diagnosis is correct, we use the final disease distribution entropy to adjust the threshold of the diagnosis. Otherwise, the threshold remains unchanged. The threshold is updated according to the following equation, for a disease $d \in D$,

$$K_d = \lambda K_d + (1 - \lambda) H_{\psi}(s_{\text{fin}}), \quad (1)$$

where K_d is the adjustable threshold for the disease d , s_{fin} denotes the final patients' state vector, and λ is a Polyak parameter controlling the speed of updating the threshold. As the training progresses and the classifier becomes more accurate, the disease distribution entropy will reflect the diagnosis confidence adequately. With the above updating scheme, the automatically learned thresholds will set proper uncertainty levels for MAD to transit from symptom inquiring to diagnosing. The updating scheme also makes the system robust to initial values of the thresholds. If the initial value is too large, it will be decreased because when updating it with a correct diagnosis, the predicted disease distribution generally should have a small entropy (i.e., high confidence). If the initial value is too small, it may initially force the agent to make the maximum number of inquiries, then the classifier will tell it what entropy value the agent should expect from a correct diagnosis.

4. Experiments

4.1. Datasets

Due to the highly sensitive nature of medical data and strict laws on their use, it is difficult to access real-world datasets on patients' symptoms, examinations, and diagnoses. Therefore, we evaluate our method using two medical knowledge sources, SymCat¹ and MedlinePlus² to generate simulated patients. A simulated patient contains one disease and a set of relevant symptoms or medical examinations. We note that there are several early works (e.g., [18,21,23]) that have evaluated their methods on two public medical dialogue datasets, namely the MuZhi Medical Dialogue Dataset [15] and the Dxy Medical Dialogue Dataset [18]. However, these two datasets only contain a small number of diseases (Dxy contains 5 diseases and MuZhi contains 4 diseases) and are not appropriate benchmarks for realistic patient simulations because the real clinical encounters are more complicated. We discuss the details of the patient simulation procedures and training parameters in Appendix A.

4.2. Proposed dataset: MedlinePlus

The symptom information in the SymCat dataset is incomplete. Some diseases in SymCat do not have enough symptom information. For example, 'esophageal cancer' only has superficial and ambiguous symptom descriptions such as 'fatigue' and 'vomiting', and the conditional prevalence on the disease is very low. This will result in an abnormal situation where a synthetic patient with 'esophageal cancer' only has 1 or 2 uninformative symptoms. Moreover, the dataset does not have disease-related medical examination information. Thus the patients simulated using SymCat are not sufficiently realistic for testing clinical MAD systems.

We propose a new dataset extracted from MedlinePlus to simulate more realistic clinical encounters.

¹ <http://www.symcat.com>

² <https://medlineplus.gov>

Table 1

Final diagnosis accuracies and inquiring turns on different datasets. The ↓ symbol indicates better performance for lower values. For each result of our method and baselines, we ran three replicates to calculate means and standard errors. For the diagnosis accuracies, the improvements of our method over REFUEL and INS-DS are all significant with p-values less than 1×10^{-4} under independent two samples t-tests.

Model	SymCat 200		SymCat 300		SymCat 400		SymCat Common		MedlinePlus	
	Accu.(%)	Step↓	Accu.(%)	Step↓	Accu.(%)	Step↓	Accu.(%)	Step	Accu.(%)	Step↓
Ours	68.8 _{±0.4}	8.98 _{±0.01}	60.4 _{±0.2}	9.39 _{±0.01}	55.5 _{±0.1}	10.02 _{±0.03}	58.1 _{±0.2}	9.61 _{±0.01}	96.2 _{±0.1}	3.41 _{±0.02}
REFUEL	59.9 _{±0.1}	8.45 _{±0.02}	48.5 _{±0.2}	9.63 _{±0.03}	33.6 _{±0.2}	10.18 _{±0.03}	47.6 _{±0.4}	8.66 _{±0.04}	00.0 _{±0.0}	15.00 _{±0.00}
INS-DS	46.2 _{±0.1}	2.77 _{±0.03}	39.7 _{±0.4}	4.47 _{±0.03}	31.1 _{±0.6}	3.25 _{±0.03}	36.1 _{±0.1}	2.99 _{±0.02}	71.4 _{±0.1}	1.34 _{±0.02}
FIT ^a	55.7	12.0	48.2	13.1	44.6	14.4	–	–	–	–

^a We use the results reported in their paper.

MedlinePlus is a publicly accessible website with various articles on different aspects of healthcare information (e.g., Drugs & Supplements, Genetics, and Medical Tests). We construct the dataset from the disease articles on the MedlinePlus dataset. The construction procedure is as follows.

1. We require that articles must contain *Symptoms* and *Tests* sections. This results in 1267 articles on different diseases.
2. We use the vocabulary in BIOS,³ a machine-learned comprehensive biomedical knowledge graph, as a biomedical term dictionary to extract the symptom and examination entities for each disease by maximum forward matching. All the extracted entities can be mapped to a concept (CID) in BIOS.
3. We constrain the extracted symptom entities to the *Sign*, *Symptom*, or *Finding* semantic type in BIOS and the extracted medical examination-related entities to the *Laboratory Procedure* and *Diagnostic Procedure* semantic types. Finally, we manually check every extracted entity and remove those ambiguous entities such as *CN00003338 Diagnose* and *CN00080784 Laboratory Tests*.
4. To ensure the richness of information for each disease, we remove the disease with no medical examination concept and the number of symptom concepts less than 3.

The final dataset contains 893 diseases, 1252 symptoms, and 860 medical examinations. Each disease connects to 13.61 concepts on average, of which 8.31 are symptoms, and 5.30 are medical examinations. Our proposed dataset includes disease-related medical examination information for the first time and provides more comprehensive information than previous ones.

4.3. Baselines

We compare our method with three baselines namely, REFUEL [19], FIT [25] and INS-DS [23]. REFUEL is a state-of-the-art RL method on SymCat. FIT achieves state-of-the-art accuracies on synthetic patients simulated by different datasets. INS-DS is a recently proposed method that splits the symptom-inquiring and disease-diagnosing actions and uses a different stopping criterion. INS-DS showed the best performance on the Dxy and MuZhi datasets. However, these three previous works did not provide open-source code. Thus, we re-implemented REFUEL and INS-DS according to the details described in their papers. FIT did not provide enough details for re-implementation. Therefore, we use the results reported in their papers for their performance.

4.4. Main results

A good MAD system should achieve high diagnostic accuracy and pose as few symptom inquiries as possible. Thus, we use diagnostic accuracy and inquiring turns to evaluate our method. We show the results in Table 1.

On different disease sets from SymCat, our method achieves the best diagnosis accuracies, significantly outperforming other methods.

Our methods use similar or fewer inquiring steps to achieve this high accuracy. Although the randomly selected 200, 300, and 400 disease sets are not exactly the same as those used in the FIT paper, our method surpasses FIT's reported results by a large margin. INS-DS achieves the minimum inquiring turns, and its diagnostic accuracy is significantly lower than other methods. INS-DS tends to give erroneous diagnostic decisions without collecting enough symptom information.

On the MedlinePlus dataset, our method also outperforms all baselines in diagnostic accuracy. REFUEL fails to give diagnosis actions and keeps inquiring about symptoms until the maximum step. INS-DS gives diagnoses only based on patients' self-reports without collecting inquiring about symptoms carefully.

We also report the match rates used by previous works as another evaluation metric. The match rate measures the number of positive symptoms or examinations that the MAD system discovered at each inquiring turn. As shown in Table 2, our method exhibited the highest positive rate on the MedlinePlus dataset and is second to REFUEL on SymCat datasets. INS-DS was highly incapable of discovering patients' positive symptoms on all datasets, which led to its low accuracy.

On the datasets from SymCat, REFUEL achieved higher match rates than our method, which indicates more positive symptoms are discovered by REFUEL. However, as shown in Table 1, REFUEL failed to achieve higher diagnosis accuracies. These results on SymCat show that the match rate is a biased metric because it is not the purpose of designing a MAD system to discover patients' symptoms but to give accurate diagnoses. As explained previously, negative symptoms may also be helpful by ruling out other potential diseases. These results empirically verify this insight. We also illustrate this through a case study.

On MedlinePlus, our method achieved better results in terms of both diagnosis accuracies and symptom match rates. This reveals that it is difficult for a unified MAD framework to learn both information-inquiring and disease-diagnosing on a larger scale automatic diagnosis benchmark MedlinePlus. Separately approaching information-inquiring and disease-diagnosing as our method can well alleviate such drawbacks.

4.5. Case study

In this section, we present the diagnostic processes conducted by our method. The details of a positive case are in Table 3, including the patient's self-reports, posed queries, symptom status feedback, and corresponding entropy changes. We also show the marginal prevalence of each queried symptom or examination among all diseases. Through the scope of entropy, we can see that our model primarily poses queries about high existing probability symptoms to narrow down the potential diseases and then inquires about possible symptoms shared among fewer diseases to reinforce the diagnosis decisions. Our model terminates the diagnostic process when the uncertainty for diagnosing is low enough (lower than the entropy of only 0.043) and a correct diagnosis is given. Through this example, it is verified that negative feedback for a disease can also bring much information (i.e., a significant drop in entropy). After posing *CN00066522 Blood Tests* and getting negative feedback, the disease distribution entropy drops from 1.344 to 0.414.

³ <https://bios.idea.edu.cn/>

Table 2

Match rates on different datasets. For each result of our method and baselines, we ran three replicates to calculate means and standard errors.

	SymCat 200	SymCat 300	SymCat 400	SymCat Common	MedlinePlus
Ours	0.103 $\pm$ 0.001	0.108 $\pm$ 0.001	0.092 $\pm$ 0.002	0.100 $\pm$ 0.002	0.234 $\pm$ 0.001
REFUEL	0.166 $\pm$ 0.002	0.154 $\pm$ 0.003	0.146 $\pm$ 0.002	0.155 $\pm$ 0.002	0.138 $\pm$ 0.003
INS-DS	0.011 $\pm$ 0.001	0.003 $\pm$ 0.001	0.009 $\pm$ 0.002	0.006 $\pm$ 0.001	0.002 $\pm$ 0.001

Table 3

The diagnostic processes of disease *Sjögren Syndrome*.

Self-report		CN00022194 Cough CN00090165 Joint Pain		
Inquiring turn	Symptom	Proportion	Feedback	Entropy
1	CN00000560 Physical Exam	358/893	Positive	2.075→1.344
2	CN00066522 Blood Tests	256/893	Negative	1.344→0.414
3	CN00455502 Complete Blood Count	179/893	Positive	0.414→0.133
4	CN00015431 CT Scan	181/893	Negative	0.133→0.099
5	CN00002615 Chest X-ray	138/893	Positive	0.099→0.077
6	CN00353823 Electrocardiogram	97/893	Negative	0.077→0.043
Diagnosis		Sjögren Syndrome		

Table 4

The negative diagnostic processes of disease *Alcoholic Ketoacidosis*.

Self-report		CN00022015 Nausea		
Inquiring turn	Symptom	Proportion	Feedback	Entropy
1	CN00000560 Physical Exam	358/893	Negative	5.22→4.75
2	CN00066522 Blood Tests	256/893	Negative	4.75→4.28
3	CN00002615 Chest X-ray	138/893	Negative	4.28→4.13
4	CN00000072 Fever	247/893	Negative	4.13→3.98
5	CN00353823 ECG	97/893	Negative	3.98→3.92
6	CN00455502 Complete Blood Count	179/893	Negative	3.92→3.67
7	CN00015440 X-rays	91/893	Negative	3.67→3.62
8	CN00015431 CT Scan	181/893	Negative	3.62→3.33
9	CN00078206 Shortness of Breath	209/893	Negative	3.33→3.30
10	CN00015772 Urinalysis	101/893	Negative	3.30→3.27
11	CN00026936 Itching	54/893	Negative	3.27→3.28
12	CN00011830 CT Scan of the Head	53/893	Negative	3.28→3.19
13	CN00002937 Skin Rash	88/893	Negative	3.19→3.17
14	CN00004932 Loss of Appetite	101/893	Positive	3.17→1.27
15	CN00020172 Muscle Aches	64/893	Negative	1.27→1.26
Diagnosis		Giardia Infection		

This case shows that our method tends to assign patients to various medical examinations, which is consistent with actual diagnostic processes as examinations are more critical than symptoms for determining the diagnosis.

The details of a negative case are shown in Table 4. In this case, the diagnostic process reaches the maximum inquiring steps. The ground true disease is *Alcoholic Ketoacidosis*, while our method gives a false diagnosis of *Giardia Infection*. The final entropy is still of a high value of 1.26 which indicates a low diagnostic confidence. The possible reasons for such an error could be due to few informative symptoms are discovered where only the negative Physical Exam and Blood Tests and the positive Loss of Appetite cause relatively large entropy decreases, and the symptom set from MedlinePlus is not perfectly realistic as there are ambiguous terms CT Scan and X-rays. Through this negative case, we can still find that high existing probability symptoms are relatively inquired early.

4.6. Discussion

Final threshold values

Boxplots of the distributions of the final threshold values on different datasets are shown in Fig. 2. Among different SymCat datasets with randomly selected diseases, we see that the threshold increases with the total number of diseases, which is reasonable because having more possible diseases should increase the difficulty, resulting in higher entropy, i.e. uncertainty of diagnosis. SymCat Common has 210 diseases and

has higher final thresholds compared to SymCat 200. This is because the common diseases are hard to distinguish from each other than the randomly chosen 200 diseases, and this results in a harder diagnostic task. However, the proposed MedlinePlus dataset has 893 diseases but resulted in much lower thresholds. This is because the dataset has much more informative symptom and examination information that allows the model to reduce uncertainty and distinguish between different diseases, which is also confirmed by the accuracies in Table 1. This illustrates the advantage of the MedlinePlus dataset.

The figure also shows that the threshold values have large variances relative to their means, which indicates significant threshold differences between different diseases. Some diseases are potentially more distinguishable than others, allowing them to use lower thresholds and high thresholds can be used to identify diseases difficult to diagnose. The results show that different diseases require different thresholds for efficient diagnosis.

Robustness of Adaptive Threshold

To further illustrate the robustness of our adaptive entropy threshold updating scheme, we conduct more experiments on the MedlinePlus dataset. A robust entropy threshold updating scheme is expected to maintain similar final threshold values and diagnosis accuracies with different threshold initial values. In this experiment, we select different starting values of the entropy thresholds and repeat the experiment on the MedlinePlus dataset, keeping other training settings the same. We select 4 different starting values 0.1, 1, 2, and 4 which are unified across different diseases, and we also randomly assign different values

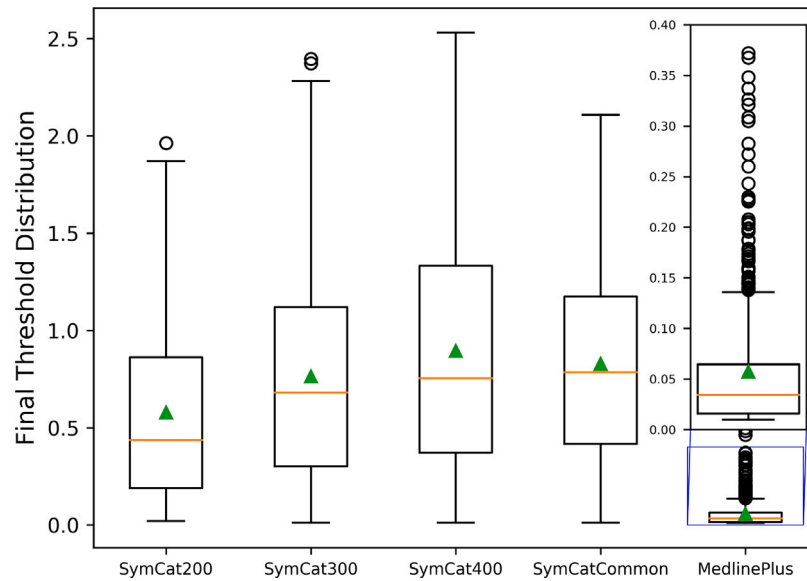


Fig. 2. The distribution of the final thresholds on different datasets. The green triangles indicate the mean of the thresholds and the orange line indicates the median. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
The final adaptive thresholds of different initial values on MedlinePlus.

Init. value	Mean	Std.	Accu.(%)	Step↓
0.1	0.059	0.103	96.1	3.37
1	0.057	0.101	96.2	3.41
2	0.057	0.105	96.1	3.31
4	0.059	0.104	96.0	3.37
Rand.1	0.057	0.097	96.0	3.36
Rand.2	0.059	0.105	96.2	3.41

between 0 and 1 to each disease as initial values. In Table 5, we list the means and standard deviations of final thresholds and the corresponding accuracies. As shown in the table, with different initial values, the entropy thresholds and the accuracies reach similar final values. This shows that our threshold updating scheme is not sensitive to the selection of starting values and robustly finds the proper entropy thresholds.

Limitations

MAD algorithms provide a mechanism to interact with data to train a diagnosis system. However, whether the system is clinically meaningful is determined by the realism of the data. Although the proposed MedlinePlus dataset is significantly more detailed and comprehensive than previous datasets such as SymCat, the information it provides is still far from sufficient. For example, it lacks information on the prior distribution of the diseases and conditional distributions of the symptoms and examination results given the disease. In addition, symptoms and examination results simulated from conditional independent distributions can be unrealistic. For example, the case study shown in Table 3 says that the patient has a negative (normal) blood test result but a positive (abnormal) complete blood count, which is strange because a complete blood count is a kind of blood test. Similarly, having a normal CT scan but an abnormal chest X-ray is also unusual. Solving this problem requires simulating joint distributions of hundreds of symptoms and examination results, which is not only difficult in terms of statistical inference but also requires real clinical data, which are hard to obtain. Another limitation is that currently the model can only treat symptom inquiries and examination inquiries in the same way. Ideally, the model should assign higher costs to examination inquiries than asking about symptoms because the time and economic cost of taking examinations are often by orders of magnitude higher

than checking symptoms. However, for RL models, using rewards of different scales often causes the model to fail to converge. Therefore, reward designs that can reflect the time and the economic cost difference between inquiring about symptoms and examination results is a direction of future research.

Future Work

Considering the limitations discussed above, it is of paramount importance to build the system on real data. One possible solution is to have the system primarily trained on simulated patient data and then further calibrate the system through RL by online feedback from human patients and doctors. The online feedback can be used in a new reward design to guide the MAD system to have higher diagnostic accuracy and more efficient symptom inquiry decisions.

5. Conclusion

In this work, we propose a new MAD model that uses an adaptive threshold learning approach to effectively align an RL agent to inquire about symptom and medical examination information and a classifier for disease diagnosis. Our method sets new state-of-the-art results across different datasets by achieving more efficient information inquiry and more accurate diagnosis. Our proposed MedlinePlus dataset contains more comprehensive information on more diseases than existing datasets and allows for more realistic patient simulation. Experimental evaluations on different datasets confirm the superiority and validity of our MAD system, and extensive experiment shows the necessity and robustness of the adaptive threshold learning approach.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and code

The data and code used in this paper are available at: <https://github.com/Yuanhy1997/Auto-Diagnosis-by-RL-and-Classification>.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 12171270) and the Natural Science Foundation of Beijing Municipality (Grant No. Z190024).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.artmed.2023.102748>.

References

- [1] Kononenko I. Machine learning for medical diagnosis: History, state of the art and perspective. *Artif Intell Med* 2001;23(1):89–109. [http://dx.doi.org/10.1016/S0933-3657\(01\)00077-X](http://dx.doi.org/10.1016/S0933-3657(01)00077-X), URL: <https://linkinghub.elsevier.com/retrieve/pii/S093336570100077X>.
- [2] Mullenbach J, Wiegrefe S, Duke J, Sun J, Eisenstein J. Explainable prediction of medical codes from clinical text. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long papers). New Orleans, Louisiana: Association for Computational Linguistics; 2018, p. 1101–11. <http://dx.doi.org/10.18653/v1/N18-1100>, URL: <https://aclanthology.org/N18-1100>.
- [3] Yu C, Liu J, Nemati S. Reinforcement learning in healthcare: A survey. 2020, [arXiv:1908.08796](https://arxiv.org/abs/1908.08796).
- [4] Kononenko I. Inductive and Bayesian learning in medical diagnosis. *Applied Artificial Intelligence an International Journal* 1993;7(4):317–37, Publisher: Taylor & Francis.
- [5] Lin R-H. An intelligent model for liver disease diagnosis. *Artif Intell Med* 2009;47(1):53–62. <http://dx.doi.org/10.1016/j.artmed.2009.05.005>, URL: <https://linkinghub.elsevier.com/retrieve/pii/S0933365709000803>.
- [6] Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting clinical events via recurrent neural networks. In: Machine learning for healthcare conference. PMLR; 2016, p. 301–18.
- [7] Choi E, Bahadori MT, Kulas JA, Schuetz A, Stewart WF, Sun J. RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. 2017, [arXiv:1608.05745](https://arxiv.org/abs/1608.05745) [cs], [arXiv:1608.05745](https://arxiv.org/abs/1608.05745).
- [8] Hosseini A, Chen T, Wu W, Sun Y, Sarrafzadeh M. HeteroMed: Heterogeneous information network for medical diagnosis. In: Proceedings of the 27th ACM international conference on information and knowledge management. Torino Italy: ACM; 2018, p. 763–72. <http://dx.doi.org/10.1145/3269206.3271805>, URL: <https://dl.acm.org/doi/10.1145/3269206.3271805>.
- [9] Ling Y, Hasan SA, Datla V, Qadir A, Lee K, Liu J, et al. Diagnostic inferencing via improving clinical concept extraction with deep reinforcement learning: A preliminary study. In: Doshi-Velez F, Fackler J, Kale D, Ranganath R, Wallace B, Wiens J, editors. Proceedings of the 2nd machine learning for healthcare conference. Proceedings of machine learning research, vol. 68, Boston, Massachusetts: PMLR; 2017, p. 271–85, URL: <http://proceedings.mlr.press/v68/ling17a.html>.
- [10] Ling Y, Hasan SA, Datla V, Qadir A, Lee K, Liu J, et al. Learning to diagnose: Assimilating clinical narratives using deep reinforcement learning. In: Proceedings of the eighth international joint conference on natural language processing (volume 1: long papers). Taipei, Taiwan: Asian Federation of Natural Language Processing; 2017, p. 895–905, URL: <https://aclanthology.org/I17-1090>.
- [11] Li D, Ren Z, Ren P, Chen Z, Fan M, Ma J, et al. Semi-supervised variational reasoning for medical dialogue generation. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. 2021, p. 544–54. <http://dx.doi.org/10.1145/3404835.3462921>, URL: <http://arxiv.org/abs/2105.06071>, [arXiv:2105.06071](https://arxiv.org/abs/2105.06071).
- [12] Lin S, Zhou P, Liang X, Tang J, Zhao R, Chen Z, et al. Graph-evolving meta-learning for low-resource medical dialogue generation. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35. 2021, p. 13362–70, Issue: 15.
- [13] Tang K-F, Kao H-C, Chou C-N, Chang EY. Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning. In: NIPS workshop on deep reinforcement learning. 2016.
- [14] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. 2013, CoRR abs/1312.5602, [arXiv:1312.5602](https://arxiv.org/abs/1312.5602), URL: <http://arxiv.org/abs/1312.5602>.
- [15] Wei Z, Liu Q, Peng B, Tou H, Chen T, Huang X, et al. Task-oriented dialogue system for automatic diagnosis. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Melbourne, Australia: Association for Computational Linguistics; 2018, p. 201–7. <http://dx.doi.org/10.18653/v1/P18-2033>, URL: <https://aclanthology.org/P18-2033>.
- [16] Kao H-C, Tang K-F, Chang EY. Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. In: AAAI. 2018.
- [17] Liao K, Liu Q, Wei Z, Peng B, Chen Q, and others WS. Task-oriented dialogue system for automatic disease diagnosis via hierarchical reinforcement learning. 2020, [arXiv:2004.14254](https://arxiv.org/abs/2004.14254).
- [18] Xu L, Zhou Q, Gong K, Liang X, Tang J, Lin L. End-to-end knowledge-routed relational dialogue system for automatic diagnosis. In: Proceedings of the AAAI conference on artificial intelligence, vol. 33. 2019, p. 7346–53, Issue: 01.
- [19] Peng Y-S, Tang K-F, Lin H-T, Chang E. REFUEL: Exploring sparse features in deep reinforcement learning for fast disease diagnosis. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. Advances in neural information processing systems, vol. 31, Curran Associates, Inc.; 2018, URL: <https://proceedings.neurips.cc/paper/2018/file/b5a1d925221b37e2e399f7b319038ba0-Paper.pdf>.
- [20] Sutton RS, McAllester D, Singh S, Mansour Y. Policy gradient methods for reinforcement learning with function approximation. In: Proceedings of the 12th international conference on neural information processing systems. Cambridge, MA, USA: MIT Press; 1999, p. 1057–63.
- [21] Xia Y, Zhou J, Shi Z, Lu C, Huang H. Generative adversarial regularized mutual information policy gradient framework for automatic diagnosis. In: Proceedings of the AAAI conference on artificial intelligence, vol. 34. 2020, p. 1062–9, Issue: 01.
- [22] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Advances in neural information processing systems, vol. 27, 2014.
- [23] Lin J, Xu L, Chen Z, Lin L. Towards a reliable and robust dialogue system for medical automatic diagnosis. 2021, URL: <https://openreview.net/forum?id=TCAmP8zKZ6k>.
- [24] Lin J, Chen Z, Liang X, Wang K, Lin L. Learning reinforced agents with counterfactual simulation for medical automatic diagnosis. 2020, [arXiv:2003.06534](https://arxiv.org/abs/2003.06534) [cs], [arXiv:2003.06534](https://arxiv.org/abs/2003.06534).
- [25] He W, Mao X, Ma C, Huang Y, Hernández-Lobato JM, Chen T. FIT: A fast and accurate framework for solving medical inquiring and diagnosing tasks. 2021, [arXiv:2012.01065](https://arxiv.org/abs/2012.01065).