

A CASANDE details

CASANDE’s code can be found at <https://github.com/mila-iqia/Casande-RL>.

A.1 Reward shaping schedulers

We introduced time-dependent coefficients $w_{\text{Ex}}(t)$ and $w_{\text{Co}}(t)$ to control the importance of the exploration and confirmation reward components through time (see Equations 2 and 4). Figure 4 shows an example of those coefficients, based on some fixed parameters.

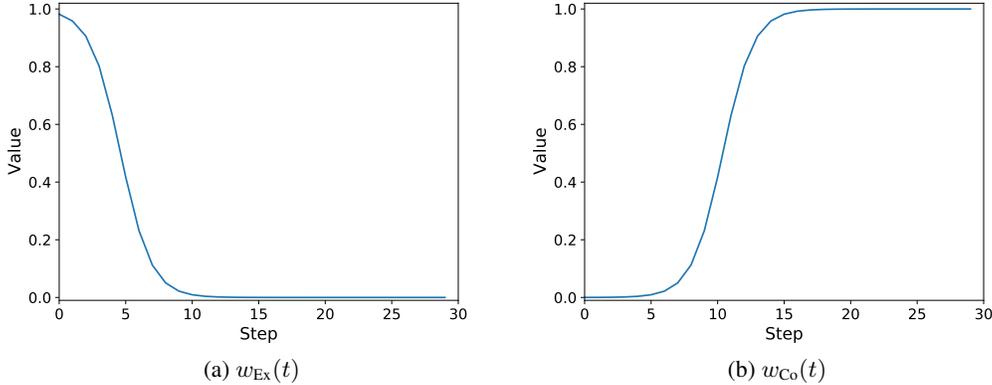


Figure 4: Examples of the time-dependent weights $w_{\text{Ex}}(t)$ and $w_{\text{Co}}(t)$ that control the importance of the exploration and confirmation reward components. Here, $\bar{x}_{\min} = -13$, $\bar{x}_{\max} = 13$, $T = 30$, $\Delta_{\text{Ex}} = 9$, and $\Delta_{\text{Co}} = 4$.

A.2 Algorithm

The two branches of our agent are updated alternatively using the loss functions introduced in Equations 8 and 9. Let θ and ϕ be the agent and target network parameters. Let π_{θ} be the agent policy characterized by θ . And let ρ be the target network soft update rate. Then, Algorithm 1 depicts a pseudo-code of the process used to train the agent.

Algorithm 1: Training Algorithm

```

Randomly initialize  $\theta$  and  $\phi$ ;
Initialize the set  $\mathbf{E}$  of all the environments;
Initialize the replay buffer  $\mathbf{Mem}$ ;
while not converged do
  for  $env \in \mathbf{E}$  do
    Simulate one step with  $\epsilon$ -greedy policy based on  $\pi_{\theta}$ ;
    Add the resulting transition in the replay buffer  $\mathbf{Mem}$ ;
  end
  Sample a batch  $B$  from the replay buffer  $\mathbf{Mem}$ ;
  Compute the policy loss  $Loss_Q(B)$  and update  $\theta$ ;
  Compute the classifier loss  $Loss_C(B)$  and update  $\theta$ ;
  Update parameters  $\phi \leftarrow (1 - \rho)\phi + \rho\theta$ ;
end

```

A.3 Input data representation

In this section, we describe how we encode input data to handle different types of evidences, namely binary, categorical, and multi-choice evidences from the DDXPlus dataset.

Let u_e be the dimension induced by the e^{th} evidence in the state space. We have u_e equals 1 for binary evidences or numerical categorical evidences (i.e., categorical evidences whose options are numbers), while u_e corresponds to the number of available options n_e for categorical evidences that are not numerical or for multi-choice evidences. Let w be the dimension induced by socio-demographic data such as the age and the sex of the patient. We assume in our state representation that the socio-demographic data are encoded before the evidences. Let $v_e = w + \sum_{t=1}^{e-1} u_t$ be the cumulative dimension induced by the set of evidences indexed before the e^{th} evidence, with $v_1 = w$. Let $f_{\text{Pres}}(e)$ be the function which returns the evidence e 's options that are experienced by the underlying patient, or, in case e is binary, whether or not it is experienced by the patient. The evidence e is then encoded as follows in the state \mathbf{s}_t based on its type:

- **Binary evidences**

$$\mathbf{s}_{t,v_e+1} = \begin{cases} 0 & \text{if } e \text{ is not yet inquired,} \\ 1 & \text{if } f_{\text{Pres}}(e), \\ -1 & \text{otherwise.} \end{cases}$$

- **Numerical categorical evidences**

$$\mathbf{s}_{t,v_e+1} = \begin{cases} 0 & \text{if } e \text{ is not yet inquired,} \\ \frac{f_{\text{Pres}}(e)+1}{M_e+1} & \text{otherwise, where } M_e \text{ is the maximum value for evidence } e. \end{cases}$$

- **Non-numerical categorical evidences**

$$\forall j \in [1, u_e], \mathbf{s}_{t,v_e+j} = \begin{cases} 0 & e \text{ is not yet inquired,} \\ 1 & \text{if } j \in f_{\text{Pres}}(e), \\ -1 & \text{otherwise.} \end{cases}$$

- **Multi-choice evidences**

$$\forall j \in [1, u_e], \mathbf{s}_{t,v_e+j} = \begin{cases} 0 & e \text{ is not yet inquired,} \\ 1 & \text{if } j \in f_{\text{Pres}}(e), \\ -1 & \text{otherwise.} \end{cases}$$

B Baseline details

B.1 BASD

BASD (short for Baseline ASD) is inspired by the work done by Luo et al. (2020). In that work, the authors propose to build an Automatic Symptom Detection (ASD) module to collect evidences from patients using supervised learning while leveraging a knowledge graph which encodes relations among symptoms and diseases. We follow the setup introduced in (Fansi Tchango et al., 2022). More specifically, we attach to the evidence acquisition module a classifier network whose goal is to predict the patient's underlying disease at the end of the acquisition process based on the collected evidences. More specifically, the BASD agent consists of an MLP network with 2 prediction branches:

- a policy branch whose role is to predict whether to stop or continue the interaction, and if the latter, which evidence to inquire about next;
- a classifier branch whose role is to predict the underlying patient disease.

The knowledge graph is not used in BASD unlike the work done by Luo et al. (2020).

To train the network, we simulate dialogue states together with their target values. In other words, let us assume a given patient has n evidences that he/she is experiencing. We simulate a dialogue state as follows:

1. Randomly select $p \in [1, n]$ representing the number of positive evidences already acquired. Sample p evidences from the ones experienced by the patient and set them in the simulated dialog state.

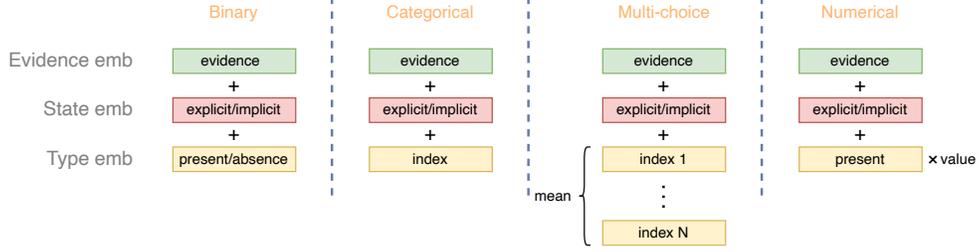


Figure 5: Illustration of Diaformer’s input representations

2. Randomly select $q \in [0, T - p)$ representing the number of negative evidences already inquired where T is the maximum number of allowed dialog turns. Sample q evidences from the ones not experienced by the patient and set them in the simulated dialog state.
3. If $p = n$, set the target of the policy branch to "stop"; otherwise set the target to be one of the experienced evidences that was not sampled at step 1).
4. Set the classifier branch target to be the ground truth pathology.

Both branches are trained using the cross-entropy loss function and the classifier branch is only updated when the target of the policy branch is set to "stop". We use the same input data representation described in Appendix A.3 for this baseline.

B.2 Diaformer

Diaformer (Chen et al., 2022) is a recent Transformer-based model that models the symptom acquisition process as a generation task. It is a supervised model, and its training objective is similar to BERT’s Masked Language Modelling (MLM) (Devlin et al., 2019). The model is trained to maximize the likelihood of synthesized trajectories, each consisting of a patient’s disease, initial complaints (called explicit evidences), and evidences that need to be collected through the interaction with the patient (called implicit evidences). At test time, the model is provided with explicit evidences only, and needs to iteratively inquire about implicit ones, until it decides to end the interaction and predict the disease.

Diaformer represents each evidence with three types of embeddings, which respectively indicate the *evidence Id*, the evidence’s *state* (i.e., whether it is explicit or implicit), and the evidence’s *type* (i.e., whether it is present or not). All embeddings are learnable, and the overall evidence’s representation is the sum of its three embeddings. To represent the data in the DDXPlus dataset, in particular categorical, multi-choice, and numerical evidences, we modify Diaformer’s input representations such that different types of evidences are represented differently. Figure 5 illustrates the modified input representations. Specifically, we define a learnable embedding for each possible option defined for non-binary evidences. Then, for categorical evidences, we use the embedding corresponding to the evidence option experienced by the underlying patient as the type embedding. Similarly, for multi-choice evidences, we take the average of the embeddings corresponding to all the options experienced by the patient for that evidence. Finally, for numerical evidences, we multiply the embedding corresponding to the “present” type with the numerical evidence value.

B.3 BED

BED (Guan and Baral, 2021) is an approach that does not need training and which exploits prior knowledge regarding evidence-disease relationships (i.e., the Quick Medical Reference belief network) to decide which evidence to inquire about next and to update its disease prediction. More specifically, the objective of BED at each turn is to inquire about the evidence e that maximizes a utility function $U(e)$ defined for binary evidences as:

$$U(e) = \sum_{v \in \{0,1\}} p(e = v | \mathbf{e}^+, \mathbf{e}^-) \times D_{\text{KL}}(p(d | \mathbf{e}^+, \mathbf{e}^-, e = v) || p(d | \mathbf{e}^+, \mathbf{e}^-)). \quad (10)$$

\mathbf{e}^+ is the set of already inquired evidences which are experienced by the patient. Similarly, \mathbf{e}^- is the set of already inquired evidences which are not experienced by the patient. $p(e = v | \mathbf{e}^+, \mathbf{e}^-)$ captures

the probability that the outcome of inquiry about e results in v given the sets \mathbf{e}^+ and \mathbf{e}^- . Finally, $p(d|\mathbf{e})$ represents the probability of the diseases given the set of inquired evidences \mathbf{e} . The inquiry process ends when the utility of all the remaining evidences is below a predefined threshold.

To be able to use BED on the DDXPlus dataset, we extend the notion of utility function on categorical and multi-choice evidences.

For categorical evidences, the extension is easy and we simply have:

$$U(e) = \sum_{v \in \mathcal{C}_e} p(e = v | \mathbf{e}^+, \mathbf{e}^-) \times D_{\text{KL}}(p(d | \mathbf{e}^+, \mathbf{e}^-, e = v) || p(d | \mathbf{e}^+, \mathbf{e}^-)), \quad (11)$$

where \mathcal{C}_e is the set of possible options for the categorical evidence e .

For multi-choice evidences, since each option can independently be experienced by the patient, we treat each such option as a binary evidence and define $U(e)$ as the maximum of the utility of the different options. Thus, we have

$$U(e) = \max_{v \in \mathcal{C}_e} U(v), \quad (12)$$

where \mathcal{C}_e is the set of possible options for the multi-choice evidence e and $U(v)$ is defined according to Equation 10. We chose to use the maximum value instead of the sum of all values to keep the magnitude of the utility values comparable among the different evidence types. The mean value wasn't considered as it might have diluted the impact of options with high utility value.

B.4 AARLC

AARLC (Yuan and Yu, 2021) is an RL-based approach consisting of two branches, just like our approach, which proposes an adaptive method to align the tasks performed by the two branches using the entropy of the distributions predicted by the classifier branch. To use this baseline with the DDXPlus dataset, we use the input data representation described in Appendix A.3.

To train AARLC with differentials, we make several changes, in addition to replacing the ground truth pathology with the ground truth differential probabilities as the classifier's training objective:

- Instead of updating the stopping threshold K when the predicted pathology matches the ground truth pathology, we update it when the set D^i of diseases in the ground truth differential is identical to the set of top- $|D^i|$ predicted diseases. We make this change because AARLC designs the threshold to be updated when the predicted disease is correct, and therefore if the differential is replacing the ground truth pathology as the target, it should also replace it as the standard of correctness.
- Second, now that the agent does not focus on predicting a single disease, it is no longer reasonable to only update the threshold associated with one disease. Therefore, we instead use one global threshold that is not associated with any particular disease, and update it every time the aforementioned condition is met.
- Similar to the condition of updating the threshold, we change the condition under which a positive reward is given to the agent, as part of r_p , for making the correct diagnosis. We give the positive reward when the set D^i is identical to the set of top- $|D^i|$ predicted diseases.

There are several differences between AARLC and CASANDE:

- CASANDE has the exploration, confirmation and severity reward functions that integrate in the training process elements that mimic the reasoning of doctors. Those functions do not exist in AARLC.
- AARLC uses separate models for the classifier and policy networks while in CASANDE, those 2 networks share the same encoder and exchange information.
- The classifier network is updated at each interaction turn in AARLC. In CASANDE, it is only updated at the end of the interaction as this is when the differential diagnosis is needed and can be accurately predicted. Forcing the classifier to predict the ground truth differential at each turn can confuse the classifier, in particular at the start of the interaction, when the number of collected evidences is small.

- AARLC provides a small positive reward when the agent asks questions about evidence that the patient doesn’t experience. CASANDE doesn’t because it isn’t clear when questions about negative evidence are useful: some questions are useful because they can help rule out pathologies from the differential; others are not informative. It might be worth exploring in future work a reward about negative evidence that is based on the impact of this information on the differential.
- AARLC has the r_H reward which encourages the model to reduce the entropy of the differential from one interaction turn to the next, as more information is accumulated. In CASANDE, we use another strategy (in the consolidation phase), which consists in ensuring that the differential at the next interaction turn is closer to the ground truth differential than at the previous turn. We therefore consider that it is not sufficient to reduce the entropy of the differential if the differential is not being pushed in the right direction.
- AARLC’s policy network doesn’t predict an *exit* action; instead, AARLC compares the entropy of the differential at each interaction turn to a learnable threshold and stops the interaction when the entropy is smaller than the threshold, irrespective of whether the differential is correct. CASANDE’s policy network can directly predict the *exit* action, and its loss (Equation 8) provides feedback to the network at each interaction turn about the value of predicting this action; it also has a reward at the end of the interaction that depends on the quality of the predicted differential diagnosis.

C Training details

The cluster used for training purposes is a mixture of NVIDIA A100, K80, M40, RTX 8000, TITAN RTX and V100 GPUs. Except for BED which does not require training, each training session is conducted with one GPU allocated by the scheduler of the cluster.

C.1 CASANDE

We rely on the rlpyt framework (Stooke and Abbeel, 2019) and use the Rainbow algorithm (Hessel et al., 2018) to train our agent. We allow interactions to have a maximum of $T = 30$ turns and set γ to 0.99. For the reward, we set r_i to 0.5, r_p to 2, and r_n to 0. This means that we do not penalize the agent when it asks about evidence not being experienced by the patient as it is sometimes necessary to inquire about such evidence; given that we do not know which negative evidence is pertinent, we neither penalize nor reward the agent when inquiring about such evidence. We use 16 environment instances to collect data during training. We use Adam (Kingma and Ba, 2015) as an optimizer with a learning rate of 6.25×10^{-5} . $[\bar{x}_{min}, \bar{x}_{max}]$ is set to $[-13, 13]$ whereas Δ_{Ex} and Δ_{Co} are respectively set to 9 and 4, unless stated otherwise. The probability threshold used to decide if a pathology is part of a differential when computing $SevOut_t$, $SevIn_t$ and Sev_y is 0.01. We perform hyper-parameter tuning and the following values were selected based on the performance on the validation dataset: $w_{si} = 1$, $\alpha_{Ex} = 12$, $\alpha_{Co} = 1$, $\alpha_{sev} = 0.75$, and $\alpha_{Cl} = 1$.

The architectural details of our agent together with the parameters of the Rainbow algorithm are described in Table 3.

C.2 BASD

The models used are MLPs with hidden layers of size 2048. We use a batch size of 2657, a patience of 20, and we tune the number of layers as well as the learning rate. For the model trained to predict the ground truth pathology, the number of hidden layers and the learning rate leading to the optimal validation performance are respectively 2 and 2.801×10^{-4} . For the model trained to predict the differential diagnosis, the optimal set of parameters corresponds to a number of hidden layers of 3 and a learning rate of 3.469×10^{-4} .

C.3 Diaformer

We reuse the same setup as in Chen et al. (2022) for the DX dataset, except for the batch size and the learning rate. We use a batch size of 256 subject to GPU memory, and we tune the learning rate. The

Table 3: The details of the network architecture and the parameters of the Rainbow algorithm used in the CASANDE experiments. The numbers in brackets represent MLP layer sizes. For all MLPs, we use the ReLU activation function. The dataset has 49 pathologies and 223 evidences.

Components	Description
Encoder	
MLP	[4096, 2048, 2048]
Classifier	
MLP	[1024, 512, 49]
Policy	
Number of atoms	51
Dueling Advantage Network (MLP)	[1024, 512, 223 x 51]
Dueling Value Network (MLP)	[1024, 512, 51]
Rainbow Algorithm	
Number of atoms	51
V_{min}	-90
V_{max}	70
N-step Q-learning	3

optimal learning rate is 1.0136×10^{-5} for Diaformer trained to predict the ground truth pathology, and 5.0257×10^{-4} for Diaformer trained to predict the differential.

C.4 BED

BED is deterministic and the only parameter to be set in addition to the maximum number of turns is the utility threshold. We follow (Guan and Baral, 2021) and set it to 10^{-2} .

C.5 AARLC

We use the same setup as in (Yuan and Yu, 2021) with a batch size of 2657. We tune the ν and λ parameters together with the learning rate. The optimal set of parameters for the model trained to predict the ground truth pathology is $\{\nu = 3.337, \lambda = 0.97, lr = 5.175 \times 10^{-4}\}$. For the model trained to predict the differential, we obtain $\{\nu = 2.826, \lambda = 0.99, lr = 3.52 \times 10^{-4}\}$.

D Evaluation metrics

This section describes the metrics used to evaluate the different agents. Let N be the number of patients, \mathcal{D} the set of diseases, and $S \subseteq \mathcal{D}$ be the set of severe pathologies. Also, let E^i be the set of evidences (i.e., symptoms and antecedents) experienced by the i^{th} patient, and \hat{E}^i be the set of evidences an agent inquired about when interacting with that patient. Additionally, let y^i be the ground truth differential, d^i be the ground truth pathology, and \hat{y}^i be the last pathology distribution (i.e., belief) generated by the agent for that patient. Besides, let $Traj^i = \{bel_1^i, bel_2^i, \dots, bel_{|\hat{E}^i|}^i = \hat{y}^i\}$ be the predicted differentials made by the agent throughout the interaction process. We further post-process both the ground truth differentials and the predicted ones to remove pathologies whose mass is less than or equal to 0.01. This threshold is selected to reduce the size of the differentials by removing highly unlikely pathologies. Let \mathcal{D}^i and $\hat{\mathcal{D}}^i$ be the resulting set of pathologies after applying the post-processing on y^i and \hat{y}^i respectively. In what follows, we use $|\cdot|$ to denote the size of a set.

D.1 End-performance metrics

Interaction length (IL) The average interaction length is defined as:

$$IL = \frac{1}{N} \sum_{i=1}^N |\hat{E}^i|. \quad (13)$$

Differential diagnosis F1 score (DDF1) This metric measures how aligned the predicted differential is with respect to the ground truth differential. We first define the differential diagnosis recall (DDR) and precision (DDP) for the i^{th} patient as

$$DDR^i = \frac{|\hat{\mathcal{D}}^i \cap \mathcal{D}^i|}{|\mathcal{D}^i|}, \quad (14)$$

and

$$DDP^i = \frac{|\hat{\mathcal{D}}^i \cap \mathcal{D}^i|}{|\hat{\mathcal{D}}^i|}. \quad (15)$$

Finally, DDF1 is defined as

$$DDF1 = \frac{1}{N} \sum_{i=1}^N DDF^i, \quad (16)$$

where

$$DDF^i = 2 \frac{DDR^i \times DDP^i}{DDR^i + DDP^i}. \quad (17)$$

Harmonic mean score of severe pathologies (DSHM) This metric corresponds to the harmonic mean of the rule-in and rule-out rates of severe pathologies. We first compute the final rule-in and rule-out rate of severe pathologies for the i^{th} patient as

$$\text{rule-in}^i = \frac{|\hat{\mathcal{D}}^i \cap \mathcal{D}^i \cap S|}{|\mathcal{D}^i \cap S|}, \quad (18)$$

and

$$\text{rule-out}^i = \frac{|S \cap (\mathcal{D} \setminus (\mathcal{D}^i \cup \hat{\mathcal{D}}^i))|}{|S \cap (\mathcal{D} \setminus \mathcal{D}^i)|}. \quad (19)$$

In other words, the final rule-in rate captures the ratio of severe pathologies that are rightfully included in the final predicted differential while the final rule-out rate measures the ratio of severe pathologies that are rightfully excluded from the final predicted differential.

DSHM is then defined as

$$DSHM = \frac{1}{N} \sum_{i=1}^N DSH^i, \quad (20)$$

where

$$DSH^i = 2 \frac{\text{rule-in}^i \times \text{rule-out}^i}{\text{rule-in}^i + \text{rule-out}^i}. \quad (21)$$

Ground truth pathology accuracy at k (GTPA@ k): This metric measures whether the top k entries of the differential diagnosis predicted by an agent contain the patient's ground truth pathology:

$$GTPA@k = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{d^i \in \hat{\mathcal{D}}_k^i}, \quad (22)$$

where $\hat{\mathcal{D}}_k^i$ is the set of top- k pathologies extracted from \hat{y}^i .

Ground truth pathology accuracy (GTPA): This metric measures whether the patient's ground truth pathology is part of the differential diagnosis predicted by an agent:

$$GTPA = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{d^i \in \hat{\mathcal{D}}^i}. \quad (23)$$

Positive evidence recall (PER) The average recall of the evidences experienced by patients is computed as:

$$PER = \frac{1}{N} \sum_{i=1}^N PER^i, \quad \text{where } PER^i = \frac{|E^i \cap \hat{E}^i|}{|\hat{E}^i|}. \quad (24)$$

D.2 Trajectory quality metrics

In this section, we introduce the *confirmation score* and the *exploration score* used to assess the quality of a trajectory. The exploration score captures how distant two consecutive agent predictions are, while the confirmation score measures how close the agent prediction is to the ground truth differential. Thus, given two consecutive agent predictions bel_{t-1} and bel_t together with the target differential diagnosis y , we have:

$$\text{confirmation_score}_t = \exp(-\text{KL}(y||bel_t)), \quad \text{and} \quad (25)$$

$$\text{exploration_score}_t = 1 - \exp(-\text{KL}(bel_{t-1}||bel_t)), \quad (26)$$

where $\text{KL}(.||.)$ is the Kullback–Leibler divergence. The more bel_t deviates from the prediction at the previous time step bel_{t-1} , the higher the exploration score. Also, the closer bel_t to y , the higher the confirmation score. Finally, both scores are within the range of 0 to 1.

E Ablation studies

In this section, we present additional ablation analyses to further demonstrate the properties of the proposed approach.

E.1 Exploration and confirmation schedulers

To analyse the impact of the exploration and confirmation schedulers on the performance of the proposed approach, we consider two additional settings with different scheduling parameters. The first one, referred to as “Shif2”, uses $\Delta_{\text{Co}} = 2$ and $\Delta_{\text{Ex}} = 7$. Basically, this shifts the initial schedule forward by two steps. The second setting, referred to as “Uni”, uses $\Delta_{\text{Co}} = 20$ and $\Delta_{\text{Ex}} = -20$, and is designed in such a way that both auxiliary rewards are active during the entire interaction process. As shown in Table 4, the “Shif2” setting results in an agent having a slightly better performance in terms of DDF1 when compared to our original setting. However, this improvement comes at the expense of an increase of the interaction length. Also, the exploration score decreases slowly for the trajectories followed by the “Shif2” agent when compared to the original setting (see Figure 6a). In the “Uni” setting, the agent is tasked with simultaneously optimizing contradictory rewards, one encouraging it to explore different differentials, and one encouraging it to confirm a differential. This leads to higher interaction length, worse evidence recall, and a smaller exploration score at the beginning of the interaction (see Figure 6b).

Table 4: Impact of the exploration and confirmation schedulers on interaction length, differential prediction, severe pathology handling, and evidence collection, as measured on the validation set. Values are in % except for IL. Values in **bold** indicate the best performance for a column. Only one experiment was performed for each setting.

	IL	DDF1	DSHM	PER	GTPA
Uni	23.02	94.17	74.14	97.72	99.76
Shif2	22.82	94.53	74.04	98.35	99.80
Casande	19.92	94.12	74.10	98.80	99.81

E.2 Impact of the severity reward weight

Table 5 shows the performance of the agent when considering different values of α_{Sev} and Figure 7 depicts how the rule-out rate of severe pathologies evolves over time throughout the interactions. As α_{Sev} increases, DDF1 and DSHM improve up at to the point where α_{Sev} is equal to 0.75, after which both scores go down. This is likely due to the fact that the severity reward focuses on ruling out severe pathologies the patient is not experiencing but doesn’t focus on ruling in the relevant severe pathologies (which are handled by the classification reward). α_{Sev} also affects the pace at which severe pathologies are ruled out from the differential predicted by the agent throughout the interaction process. Indeed, the higher α_{Sev} , the quicker the severe pathologies are ruled out.

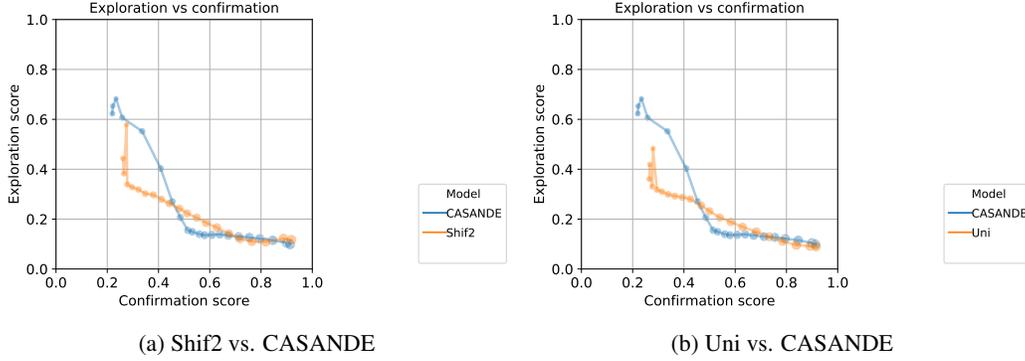


Figure 6: Plots of confirmation score vs. exploration score for different schedulers on the validation set.

Table 5: Impact of the severity reward weight α_{Sev} on the differential prediction and the handling of severe pathologies, as measured on the validation set. Values are in %. Values in **bold** indicate the best performance for a row. Only one experiment was performed for each setting.

	$\alpha_{Sev} = 0.0$	$\alpha_{Sev} = 0.5$	$\alpha_{Sev} = 0.75$	$\alpha_{Sev} = 1.0$
DDF1	93.24	93.76	94.12	93.66
DSHM	73.52	73.78	74.10	73.91

E.3 Impact of the confirmation reward weight

Table 6 and Figure 8 show the results obtained when varying the values of α_{Co} . It is noticeable that a low value of α_{Co} tends to shorten the interaction length. Conversely, a high value of α_{Co} tends to increase the PER metric. Focusing on Figure 8, we observe that, for $\alpha_{Co} = 0.5$, the exploration score tends to move upward towards the end of the interaction, a behavior that is not desirable. On the other hand, for $\alpha_{Co} = 1.0$ and $\alpha_{Co} = 2.0$, the resulting trajectories are similar and exhibit the desired properties as their exploration scores follow a downward trend at the end of the interaction.

Table 6: Impact of the confirmation reward weight α_{Co} on the performance metrics, as measured on the validation set. Values are in % except for IL. Values in **bold** indicate the best performance for a column. Only one experiment was performed for each setting.

	IL	DDF1	DSHM	PER	GTPA
$\alpha_{Co} = 0.5$	18.90	93.55	73.76	97.75	99.80
$\alpha_{Co} = 1.0$	19.92	94.12	74.10	98.80	99.81
$\alpha_{Co} = 2.0$	19.58	93.86	73.81	98.65	99.81

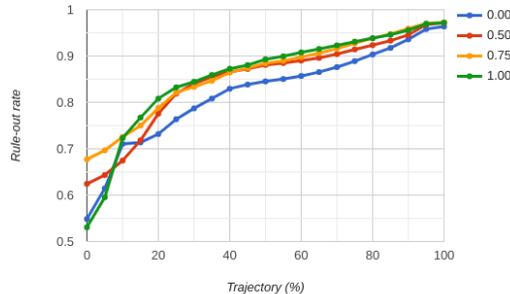


Figure 7: Evolution of the severe pathologies' rule-out rate for different values of α_{Sev} on the validation set.

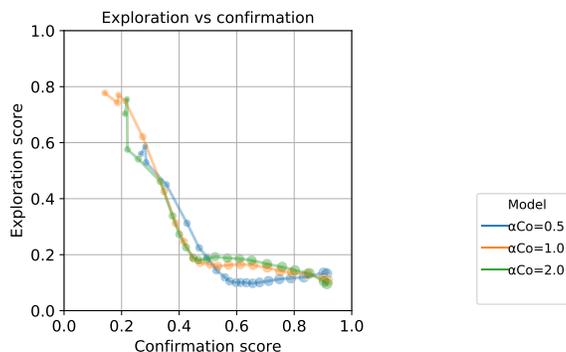


Figure 8: Plots of the confirmation score vs. exploration score for different values of α_{Co} on the validation set.

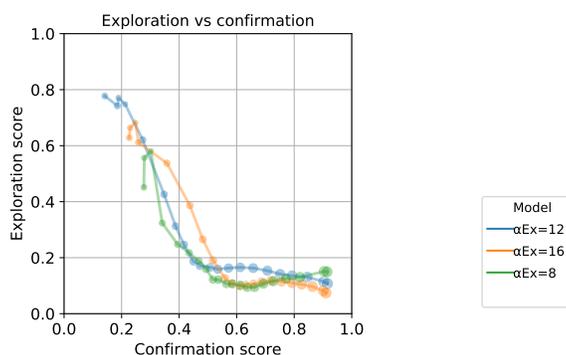


Figure 9: Plots of the confirmation score vs. exploration score for different values of α_{Ex} on the validation set.

E.4 Impact of the exploration reward weight

Table 7 and Figure 9 show the results obtained when varying the values of α_{Ex} . As expected, the interaction length tends to increase with the values of α_{Ex} . Also, the higher the value of α_{Ex} , the higher the exploration score is in the initial phase of the interaction with the patient (see Figure 9).

Table 7: Impact of the exploration reward weight α_{Ex} on the performance metrics, as measured on the validation set. Values are in % except for IL. Values in **bold** indicate the best performance for a column. Only one experiment was performed for each setting.

	IL	DDF1	DSHM	PER	GTPA
$\alpha_{Ex} = 8$	18.69	93.90	73.89	97.80	99.81
$\alpha_{Ex} = 16$	19.67	93.60	73.91	98.33	99.85
$\alpha_{Ex} = 12$	19.92	94.12	74.10	98.80	99.81

E.5 Disabling of the reward functions

We presented in Table 2 the results of an ablation study when different subsets of the exploration, confirmation, severity, and classification reward functions are disabled, and described the impact on DDF1, DSHM and PER. We now analyze the impact of those ablations on the exploration-confirmation score trajectories. Those trajectories are depicted in Figure 10. We observe several patterns:

- When using all rewards (brown curve), the trajectory starts with the highest exploration score and a small confirmation score and slowly shifts towards the lower right corner with a

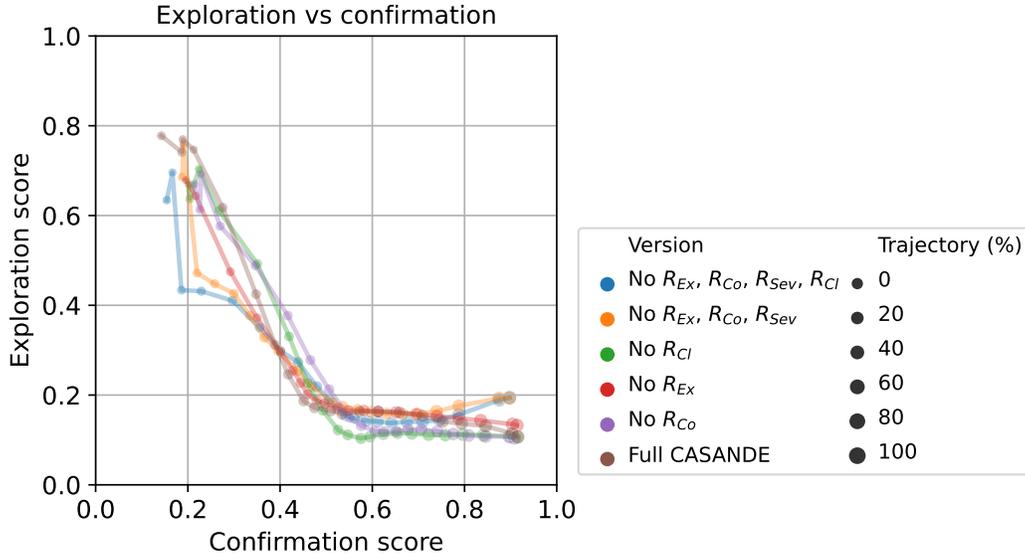


Figure 10: Exploration score vs confirmation score on the validation set for the ablation study presented in Table 2. The scores correspond to averages over 3 runs using different random seeds.

small exploration score and a high confirmation score. This trajectory corresponds to the desired behavior.

- When the exploration reward is disabled (blue, orange, and red curves), the agent quickly reduces the amount of exploration it does (with the third point on each curve getting close to 40% of exploration while the other curves are at higher exploration scores at the same stage in the interactions).
- When the confirmation reward is disabled (blue, orange, and purple curves), the agent is not as strongly constrained towards the end of the interaction to consolidating its belief and can instead also decide to increase its exploration of differentials.
- When only the classification reward is disabled (green curve), the agent manages to confirm the differential at the end of the interaction thanks to the confirmation reward.
- When all the reward components are disabled (blue curve), we observe good confirmation because the agent is trained to recover the ground truth differential by construction (as part of the classifier training). As for the exploration, the agent starts from some initial distribution and naturally moves toward the ground truth differential.

F Qualitative evaluation

We asked the doctor supporting us to qualitatively evaluate the trajectories generated by CASANDE. The doctor defined the following evaluation criteria, with a score on a 5-point Likert scale:

- Q1: The agent asks relevant questions.
- Q2: The questions asked allow me to establish a differential diagnosis.
- Q3: The agent asks enough questions to make a differential diagnosis.
- Q4: The questions asked are similar to what I would have asked.
- Q5: The information collected is useful for me to continue assessing the patient.
- Q6: The sequence of questions seems logical to me.

The Likert scale is defined as follows:

1. strongly disagree,

2. disagree,
3. neutral,
4. agree,
5. strongly agree.

Table 8 shows the scores on 20 patients that were randomly sampled from the test set. Those scores are next commented by the doctor.

Table 8: Qualitative evaluation of CASANDE on 20 randomly selected patients from the test set. The evaluation is based on 6 questions whose scores are defined on a 5-point Likert scale. The last column of the table shows the average score with the standard deviation, in brackets, for each question.

Questions	Patient IDs																				Average
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Q1	4	4	4	5	2	5	4	4	4	4	3	4	4	4	3	4	4	4	4	4	3.90 (± 0.64)
Q2	4	5	4	5	3	5	4	5	4	4	3	4	5	4	4	4	5	3	4	5	4.20 (± 0.70)
Q3	4	5	5	5	3	5	4	5	5	5	3	4	4	4	4	5	5	3	4	5	4.35 (± 0.75)
Q4	3	4	4	4	2	4	3	4	4	4	2	3	4	3	3	4	4	3	4	4	3.50 (± 0.69)
Q5	5	5	5	5	3	5	4	5	5	4	3	4	5	5	4	5	4	4	5	5	4.50 (± 0.69)
Q6	4	3	3	4	1	4	4	3	3	3	2	3	3	3	3	4	4	4	4	3	3.25 (± 0.79)

Evaluating physician’s comments

The evaluation of the interactions was done as if the agent was trained on an extensive set that would include most of the pathologies. As a doctor, it would be very difficult for me to evaluate it otherwise: my training and clinical experience in acute care setting shaped how I optimize patient evaluation. It would be impractical to deconstruct and optimize my framework on a smaller set of diseases, as I would require to train this skill. It mostly explains the lower general scores attributed to Q1 and Q4.

That being formulated, we also have to keep in mind that a tool such as CASANDE has to be considered as a mean to improve the patient experience and improve clinical outcome, by providing the care team relevant medical information from which they can build on.

The differential diagnosis approach is the one we use in practice when interacting with patients. The differential helps us explore potential pathologies and converge toward a most likely differential towards the end of the interaction. That differential will guide the next steps in the evaluation and help choose the best investigations and treatments. The clinical context influences the differential. In an acute care setting, more emphasis will be put on diseases carrying higher short term mortality and/or morbidity.

Regarding Q2, the questions asked by the agent generally allowed me to build a good differential that I would find useful in clinical practice.

For Q3, the amount of questions seemed reasonable. Of course, since the agent is not trained on the same amount of pathologies than I am, I would naturally seek out a wider array of diseases and negative evidences in the initial exploration.

Q5 is somewhat correlated to Q2. I found the information relevant and inclusive, even seeking past medical history elements and risk factors, all of that useful in a clinical context. It would be very easy to go to that patient and complete the questionnaire.

Finally, for Q6, as mentioned earlier, the optimal strategy used by CASANDE is derived from the breadth of the dataset trained on. I did not expect the agent to use a similar sequence. The order is less relevant *during* the patient evaluation in a clinical context, as the collected evidences are usually reordered by systems and sections in the patient’s chart *a posteriori* and then presented to the attending physician. Still with the questions asked, I could understand what differential was being evaluated at each step, which resembles what I would normally do.

The trajectories corresponding to those 20 patients are shown in what follows.

Patient 1

Sex: F, Age: 45
Initial evidence:

I have recently had stools that were black (like coal)
Agent inquiries:

*Turn 1:
Q: Do you have pain somewhere, related to your reason for consulting?
A: Y
*Turn 2:
Q: Does the pain radiate to another location?
A:
 * nowhere
*Turn 3:
Q: Do you have swelling in one or more areas of your body?
A: N
*Turn 4:
Q: Characterize your pain:
A:
 * sensitive
 * tugging
 * a cramp
 * exhausting
*Turn 5:
Q: Do you feel pain somewhere?
A:
 * back of head
 * forehead
*Turn 6:
Q: How precisely is the pain located?
A:
 * 4
*Turn 7:
Q: How intense is the pain?
A:
 * 1
*Turn 8:
Q: How fast did the pain appear?
A:
 * 4
*Turn 9:
Q: Are you taking any new oral anticoagulants ((NOACs)?
A: Y
*Turn 10:
Q: Have you ever had a diagnosis of anemia?
A: Y
*Turn 11:
Q: Have you noticed light red blood or blood clots in your stool?
A: Y
*Turn 12:
Q: Do you have chronic kidney failure?
A: Y
*Turn 13:
Q: Do you have a poor diet?
A: N
*Turn 14:
Q: Do you have any family members who have been diagnosed with anemia?
A: Y
*Turn 15:
Q: Is your BMI less than 18.5, or are you underweight?
A: Y
*Turn 16:
Q: Have you traveled out of the country in the last 4 weeks?
A: N
*Turn 17:
Q: Do you feel so tired that you are unable to do your usual activities or are you stuck in your bed all day long?
A: Y

*Turn 18:
 Q: Do you think you are pregnant or are you currently pregnant?
 A: Y

*Turn 19:
 Q: Do you feel slightly dizzy or lightheaded?
 A: Y

*Turn 20:
 Q: Do you feel lightheaded and dizzy or do you feel like you are about to faint?
 A: Y

*Turn 21:
 Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?
 A: N

*Turn 22:
 Q: Is your skin much paler than usual?
 A: Y

Predicted differential:

 Anemia, Atrial fibrillation, Cluster headache, HIV (initial infection), PSVT
 Ground truth differential:

 Anemia: 0.349, Atrial fibrillation: 0.243, Cluster headache: 0.153, HIV (initial infection): 0.152, PSVT: 0.103

Patient 2

Sex: F, Age: 19
 Initial evidence:
 I have a sore throat

Agent inquiries:

*Turn 1:
 Q: Do you have swelling in one or more areas of your body?
 A: N

*Turn 2:
 Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
 A: N

*Turn 3:
 Q: Do you have pain somewhere, related to your reason for consulting?
 A: Y

*Turn 4:
 Q: Do you feel pain somewhere?
 A:
 * top of the head
 * forehead
 * cheek(R)
 * occiput
 * temple(R)

*Turn 5:
 Q: Characterize your pain:
 A:
 * sensitive

*Turn 6:
 Q: How precisely is the pain located?
 A:
 * 1

*Turn 7:
 Q: How intense is the pain?
 A:
 * 6

*Turn 8:
 Q: How fast did the pain appear?
 A:

* 3

*Turn 9:

Q: Do you live with 4 or more people?

A: Y

*Turn 10:

Q: Do you have a cough that produces colored or more abundant sputum than usual?

A: N

*Turn 11:

Q: Do you have a cough?

A: Y

*Turn 12:

Q: Do you have nasal congestion or a clear runny nose?

A: Y

*Turn 13:

Q: Are you exposed to secondhand cigarette smoke on a daily basis?

A: Y

*Turn 14:

Q: Do you have a fever (either felt or measured with a thermometer)?

A: N

*Turn 15:

Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?

A: Y

*Turn 16:

Q: Do you smoke cigarettes?

A: Y

*Turn 17:

Q: Do you attend or work in a daycare?

A: Y

*Turn 18:

Q: Have you had significantly increased sweating?

A: N

*Turn 19:

Q: Do you have diffuse (widespread) muscle pain?

A: Y

*Turn 20:

Q: Do you have any family members who have asthma?

A: N

Predicted differential:

URTI, Bronchitis, Cluster headache, Chronic rhinosinusitis, Acute rhinosinusitis,
Influenza, Anemia, Chagas

Ground truth differential:

URTI: 0.230, Cluster headache: 0.174, Bronchitis: 0.167, Chronic rhinosinusitis:
0.141, Acute rhinosinusitis: 0.101, Chagas: 0.066, Anemia: 0.066, Influenza:
0.054

Patient 3

Sex: M, Age: 36

Initial evidence:

I feel pain

Agent inquiries:

*Turn 1:

Q: Do you have swelling in one or more areas of your body?

A: N

*Turn 2:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: Y

*Turn 3:

Q: What color is the rash?

A:
 * yellow

*Turn 4:
 Q: Characterize your pain:
 A:
 * sensitive
 * a cramp

*Turn 5:
 Q: Do you feel pain somewhere?
 A:
 * flank(L)

*Turn 6:
 Q: Where is the affected region located?
 A:
 * epigastric

*Turn 7:
 Q: How intense is the pain?
 A:
 * 6

*Turn 8:
 Q: How precisely is the pain located?
 A:
 * 5

*Turn 9:
 Q: How severe is the itching?
 A:
 * 3

*Turn 10:
 Q: How intense is the pain caused by the rash?
 A:
 * 1

*Turn 11:
 Q: Is the lesion (or are the lesions) larger than 1cm?
 A:
 * Y

*Turn 12:
 Q: Is the rash swollen?
 A:
 * 1

*Turn 13:
 Q: How fast did the pain appear?
 A:
 * 2

*Turn 14:
 Q: Have you had an involuntary weight loss over the last 3 months?
 A: Y

*Turn 15:
 Q: Are there members of your family who have been diagnosed with pancreatic cancer?
 A: Y

*Turn 16:
 Q: Have you had diarrhea or an increase in stool frequency?
 A: Y

*Turn 17:
 Q: Do you have diabetes?
 A: N

*Turn 18:
 Q: Do you have chronic pancreatitis?
 A: Y

*Turn 19:
 Q: Are you significantly overweight compared to people of the same height as you?
 A: N

*Turn 20:
 Q: Does the pain radiate to another location?
 A:

* thoracic spine

*Turn 21:
 Q: Do you have pale stools and dark urine?
 A: Y

*Turn 22:
 Q: Do you constantly feel fatigued or do you have non-restful sleep?
 A: Y

*Turn 23:
 Q: Do you smoke cigarettes?
 A: Y

*Turn 24:
 Q: Are you feeling nauseous or do you feel like vomiting?
 A: N

*Turn 25:
 Q: Do you have a cough?
 A: Y

Predicted differential:

 Pancreatic neoplasm, Bronchitis, Anaphylaxis, Scombroid food poisoning, Chagas,
 Pulmonary neoplasm

Ground truth differential:

 Pancreatic neoplasm: 0.265, Anaphylaxis: 0.213, Pulmonary neoplasm: 0.151, Scombroid
 food poisoning: 0.142, Chagas: 0.118, Bronchitis: 0.111

Patient 4

Sex: F, Age: 39
 Initial evidence:
 I am coughing

Agent inquiries:

*Turn 1:
 Q: Do you have any lesions, redness or problems on your skin that you believe are
 related to the condition you are consulting for?
 A: N

*Turn 2:
 Q: Do you have pain somewhere, related to your reason for consulting?
 A: Y

*Turn 3:
 Q: Does the pain radiate to another location?
 A:
 * posterior chest wall(R)
 * posterior chest wall(L)

*Turn 4:
 Q: Do you feel pain somewhere?
 A:
 * side of the chest(R)
 * breast(R)
 * breast(L)
 * posterior chest wall(R)
 * posterior chest wall(L)

*Turn 5:
 Q: Characterize your pain:
 A:
 * heartbreaking
 * haunting
 * violent
 * sharp

*Turn 6:
 Q: How precisely is the pain located?
 A:
 * 10

*Turn 7:

Q: Do you have pain that is increased when you breathe in deeply?
A: Y

*Turn 8:
Q: Do you have pain that is increased with movement?
A: Y

*Turn 9:
Q: How intense is the pain?
A:
* 10

*Turn 10:
Q: Are you being treated for osteoporosis?
A: Y

*Turn 11:
Q: How fast did the pain appear?
A:
* 9

*Turn 12:
Q: Do you drink alcohol excessively or do you have an addiction to alcohol?
A: N

*Turn 13:
Q: Do you have intense coughing fits?
A: Y

*Turn 14:
Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?
A: N

Predicted differential:

Spontaneous rib fracture, Bronchitis, Possible NSTEMI / STEMI, Unstable angina, Pulmonary embolism, Whooping cough, Stable angina, Spontaneous pneumothorax

Ground truth differential:

Spontaneous rib fracture: 0.205, Bronchitis: 0.187, Whooping cough: 0.133, Pulmonary embolism: 0.128, Possible NSTEMI / STEMI: 0.126, Unstable angina: 0.119, Stable angina: 0.061, Spontaneous pneumothorax: 0.040

Patient 5

Sex: M, Age: 25
Initial evidence:
I had an involuntary weight loss over the last 3 months

Agent inquiries:

*Turn 1:
Q: Do you have pain somewhere, related to your reason for consulting?
A: Y

*Turn 2:
Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
A: Y

*Turn 3:
Q: What color is the rash?
A:
* pale

*Turn 4:
Q: Characterize your pain:
A:
* a pulse
* exhausting

*Turn 5:
Q: Do you feel pain somewhere?
A:
* cervical spine
* forehead

* temple(R)
 * temple(L)

*Turn 6:
 Q: Where is the affected region located?
 A:
 * lower gum
 * upper gum
 * labia majora(R)
 * labia majora(L)
 * internal cheek(R)

*Turn 7:
 Q: How precisely is the pain located?
 A:
 * 2

*Turn 8:
 Q: How intense is the pain caused by the rash?
 A:
 * 7

*Turn 9:
 Q: How intense is the pain?
 A:
 * 4

*Turn 10:
 Q: Is the rash swollen?
 A:
 * 0

*Turn 11:
 Q: Is the lesion (or are the lesions) larger than 1cm?
 A:
 * Y

*Turn 12:
 Q: How fast did the pain appear?
 A:
 * 4

*Turn 13:
 Q: Do you have a fever (either felt or measured with a thermometer)?
 A: Y

*Turn 14:
 Q: Have you had sexual intercourse with an HIV-positive partner in the past 12 months?
 A: Y

*Turn 15:
 Q: Have you had unprotected sex with more than one partner in the last 6 months?
 A: Y

*Turn 16:
 Q: Do you feel so tired that you are unable to do your usual activities or are you stuck in your bed all day long?
 A: N

*Turn 17:
 Q: Have you ever had a sexually transmitted infection?
 A: Y

*Turn 18:
 Q: Are you currently using intravenous drugs?
 A: Y

*Turn 19:
 Q: Do you have swollen or painful lymph nodes?
 A: N

*Turn 20:
 Q: Have you had diarrhea or an increase in stool frequency?
 A: Y

*Turn 21:
 Q: Have you had significantly increased sweating?
 A: Y

*Turn 22:
 Q: Do you have diffuse (widespread) muscle pain?

A: N
*Turn 23:
Q: Do you have a sore throat?
A: N
*Turn 24:
Q: Are you feeling nauseous or do you feel like vomiting?
A: N
*Turn 25:
Q: Do you have painful mouth ulcers or sores?
A: N
Predicted differential:

HIV (initial infection), Chagas, Scombroid food poisoning
Ground truth differential:

HIV (initial infection): 0.544, Chagas: 0.353, Scombroid food poisoning: 0.103

Patient 6

Sex: F, Age: 56
Initial evidence:
I am experiencing shortness of breath or difficulty breathing in a significant way
Agent inquiries:

*Turn 1:
Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
A: N
*Turn 2:
Q: Do you have pain somewhere, related to your reason for consulting?
A: Y
*Turn 3:
Q: Does the pain radiate to another location?
A:
* side of the chest(R)
* side of the chest(L)
* scapula(R)
* breast(R)
* posterior chest wall(R)
*Turn 4:
Q: Do you feel pain somewhere?
A:
* side of the chest(R)
* side of the chest(L)
* scapula(R)
* scapula(L)
* posterior chest wall(R)
*Turn 5:
Q: Characterize your pain:
A:
* sensitive
* a knife stroke
*Turn 6:
Q: How precisely is the pain located?
A:
* 8
*Turn 7:
Q: How intense is the pain?
A:
* 6
*Turn 8:
Q: How fast did the pain appear?
A:

* 4

*Turn 9:

Q: Do you have swelling in one or more areas of your body?

A: Y

*Turn 10:

Q: Have you had surgery within the last month?

A: Y

*Turn 11:

Q: Do you have pain that is increased when you breathe in deeply?

A: N

*Turn 12:

Q: Have you been coughing up blood?

A: Y

*Turn 13:

Q: Did you lose consciousness?

A: Y

*Turn 14:

Q: Where is the swelling located?

A:

* posterior aspect of the ankle(R)

* posterior aspect of the ankle(L)

* calf(L)

*Turn 15:

Q: Have you ever had deep vein thrombosis (DVT)?

A: N

*Turn 16:

Q: Have you been unable to move or get up for more than 3 consecutive days within the last 4 weeks?

A: Y

*Turn 17:

Q: Have you traveled out of the country in the last 4 weeks?

A: N

*Turn 18:

Q: Do you currently take hormones?

A: Y

*Turn 19:

Q: Do you have an active cancer?

A: N

Predicted differential:

Pulmonary embolism, Acute pulmonary edema, Panic attack, Possible NSTEMI / STEMI, Unstable angina, Myocarditis, Stable angina, Pulmonary neoplasm, Guillain-Barre syndrome, Atrial fibrillation, Acute dystonic reactions, Myasthenia gravis, Anemia, Sarcoidosis, PSVT, Anaphylaxis

Ground truth differential:

Pulmonary embolism: 0.114, Acute pulmonary edema: 0.094, Panic attack: 0.090, Possible NSTEMI / STEMI: 0.084, Unstable angina: 0.080, Myocarditis: 0.070, Pulmonary neoplasm: 0.070, Stable angina: 0.065, Guillain-Barre syndrome: 0.063, Atrial fibrillation: 0.060, Acute dystonic reactions: 0.044, Myasthenia gravis: 0.044, Anemia: 0.044, Sarcoidosis: 0.036, PSVT: 0.024, Anaphylaxis: 0.016

Patient 7

Sex: F, Age: 45

Initial evidence:

I feel pain

Agent inquiries:

*Turn 1:

Q: Do you have swelling in one or more areas of your body?

A: N

*Turn 2:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
A: N

*Turn 3:
Q: Does the pain radiate to another location?
A:
 * nowhere

*Turn 4:
Q: Do you feel pain somewhere?
A:
 * side of the chest(L)
 * flank(R)
 * hypochondrium(R)
 * hypochondrium(L)
 * belly

*Turn 5:
Q: Characterize your pain:
A:
 * a cramp
 * sharp

*Turn 6:
Q: How precisely is the pain located?
A:
 * 6

*Turn 7:
Q: How intense is the pain?
A:
 * 7

*Turn 8:
Q: How fast did the pain appear?
A:
 * 7

*Turn 9:
Q: Have you ever been diagnosed with gastroesophageal reflux?
A: N

*Turn 10:
Q: Have you ever felt like you were suffocating for a very short time associated with inability to breathe or speak?
A: N

*Turn 11:
Q: Are you infected with the human immunodeficiency virus (HIV)?
A: N

*Turn 12:
Q: Do you have a known severe food allergy?
A: N

*Turn 13:
Q: Have you ever had a spontaneous pneumothorax?
A: N

*Turn 14:
Q: Do you have a deviated nasal septum?
A: N

*Turn 15:
Q: Do you suffer from chronic anxiety?
A: Y

*Turn 16:
Q: Do you have diabetes?
A: N

*Turn 17:
Q: Have you had diarrhea or an increase in stool frequency?
A: N

*Turn 18:
Q: Have you ever had a head trauma?
A: Y

*Turn 19:
Q: Have you ever been diagnosed with depression?

A: Y
*Turn 20:
Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?
A: Y
*Turn 21:
Q: Do you feel your heart is beating fast (racing), irregularly (missing a beat) or do you feel palpitations?
A: Y
*Turn 22:
Q: Have you recently had numbness, loss of sensation or tingling, in both arms and legs and around your mouth?
A: Y
*Turn 23:
Q: Do you suffer from fibromyalgia?
A: N
*Turn 24:
Q: Do you feel like you are dying or were you afraid that you were about to die?
A: Y
*Turn 25:
Q: Do you have asthma or have you ever had to use a bronchodilator in the past?
A: N
*Turn 26:
Q: Do you feel like you are detached from your own body or your surroundings?
A: Y
*Turn 27:
Q: Are you feeling nauseous or do you feel like vomiting?
A: N
*Turn 28:
Q: Do you feel anxious?
A: Y
*Turn 29:
Q: Have you ever had a migraine or is a member of your family known to have migraines?
A: Y
*Turn 30:
Q: Do you feel like you are (or were) choking or suffocating?
A: N

Predicted differential:

Panic attack, PSVT, Guillain-Barre syndrome, Spontaneous pneumothorax, Possible NSTEMI / STEMI, Atrial fibrillation, Unstable angina, Acute pulmonary edema, Anaphylaxis, Pulmonary embolism, Anemia, GERD, Stable angina, Scombroid food poisoning, Acute dystonic reactions, Myasthenia gravis, Sarcoidosis, Pancreatic neoplasm, SLE

Ground truth differential:

Panic attack: 0.101, PSVT: 0.085, Guillain-Barre syndrome: 0.077, Spontaneous pneumothorax: 0.067, Possible NSTEMI / STEMI: 0.063, Atrial fibrillation: 0.063, Anemia: 0.061, Anaphylaxis: 0.060, Pulmonary embolism: 0.055, Acute pulmonary edema: 0.054, Unstable angina: 0.051, Stable angina: 0.049, GERD: 0.042, Scombroid food poisoning: 0.039, Acute dystonic reactions: 0.033, Myasthenia gravis: 0.033, Sarcoidosis: 0.027, Pancreatic neoplasm: 0.023, SLE: 0.015

Patient 8

Sex: M, Age: 16

Initial evidence:

I have nasal congestion

Agent inquiries:

*Turn 1:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: N

*Turn 2:
 Q: Do you have pain somewhere, related to your reason for consulting?
 A: Y

*Turn 3:
 Q: Characterize your pain:
 A:
 * tedious
 * sensitive

*Turn 4:
 Q: Do you feel pain somewhere?
 A:
 * forehead
 * cheek(R)
 * temple(L)

*Turn 5:
 Q: How precisely is the pain located?
 A:
 * 1

*Turn 6:
 Q: How intense is the pain?
 A:
 * 4

*Turn 7:
 Q: How fast did the pain appear?
 A:
 * 2

*Turn 8:
 Q: Do you have a cough?
 A: Y

*Turn 9:
 Q: Are you exposed to secondhand cigarette smoke on a daily basis?
 A: Y

*Turn 10:
 Q: Do you have a cough that produces colored or more abundant sputum than usual?
 A: N

*Turn 11:
 Q: Do you have a sore throat?
 A: N

*Turn 12:
 Q: Do you live with 4 or more people?
 A: Y

*Turn 13:
 Q: Do you have a fever (either felt or measured with a thermometer)?
 A: N

*Turn 14:
 Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?
 A: Y

*Turn 15:
 Q: Do you have diffuse (widespread) muscle pain?
 A: N

*Turn 16:
 Q: Have you had significantly increased sweating?
 A: Y

*Turn 17:
 Q: Do you attend or work in a daycare?
 A: Y

*Turn 18:
 Q: Do you smoke cigarettes?
 A: N

Predicted differential:

 Bronchitis, URTI, Cluster headache, Chronic rhinosinusitis, Acute rhinosinusitis, Chagas, Anemia

Ground truth differential:

URTI: 0.230, Bronchitis: 0.223, Cluster headache: 0.186, Acute rhinosinusitis:
0.130, Chagas: 0.086, Anemia: 0.086, Chronic rhinosinusitis: 0.059

Patient 9

Sex: M, Age: 81

Initial evidence:

I feel pain

Agent inquiries:

*Turn 1:

Q: Do you have swelling in one or more areas of your body?

A: N

*Turn 2:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: N

*Turn 3:

Q: Do you feel pain somewhere?

A:

- * tonsil(R)
- * tonsil(L)
- * palace
- * under the jaw

*Turn 4:

Q: Characterize your pain:

A:

- * sensitive
- * burning

*Turn 5:

Q: How precisely is the pain located?

A:

- * 10

*Turn 6:

Q: How intense is the pain?

A:

- * 5

*Turn 7:

Q: How fast did the pain appear?

A:

- * 1

*Turn 8:

Q: Do you have nasal congestion or a clear runny nose?

A: Y

*Turn 9:

Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?

A: Y

*Turn 10:

Q: Do you have a fever (either felt or measured with a thermometer)?

A: Y

*Turn 11:

Q: Do you attend or work in a daycare?

A: Y

*Turn 12:

Q: Do you have a cough?

A: N

*Turn 13:

Q: Are you immunosuppressed?

A: N

*Turn 14:

Q: Do you live with 4 or more people?

A: N
*Turn 15:
Q: Have you been coughing up blood?
A: N
*Turn 16:
Q: Do you smoke cigarettes?
A: Y
Predicted differential:

Viral pharyngitis, Possible NSTEMI / STEMI, Epiglottitis, Unstable angina, Acute laryngitis, URTI, Chagas, Stable angina
Ground truth differential:

Viral pharyngitis: 0.254, Possible NSTEMI / STEMI: 0.158, Unstable angina: 0.148, Epiglottitis: 0.144, URTI: 0.090, Acute laryngitis: 0.082, Chagas: 0.070, Stable angina: 0.054

Patient 10

Sex: F, Age: 76
Initial evidence:
I am coughing
Agent inquiries:

*Turn 1:
Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
A: N
*Turn 2:
Q: Do you have pain somewhere, related to your reason for consulting?
A: Y
*Turn 3:
Q: Do you feel pain somewhere?
A:
* posterior chest wall(L)
*Turn 4:
Q: Characterize your pain:
A:
* a knife stroke
* exhausting
*Turn 5:
Q: How precisely is the pain located?
A:
* 0
*Turn 6:
Q: How intense is the pain?
A:
* 1
*Turn 7:
Q: How fast did the pain appear?
A:
* 0
*Turn 8:
Q: Have you been coughing up blood?
A: Y
*Turn 9:
Q: Have you had an involuntary weight loss over the last 3 months?
A: Y
*Turn 10:
Q: Are you a former smoker?
A: Y
*Turn 11:
Q: Do you smoke cigarettes?
A: Y

*Turn 12:
 Q: Are you exposed to secondhand cigarette smoke on a daily basis?
 A: N

*Turn 13:
 Q: Do you have family members who have had lung cancer?
 A: Y

*Turn 14:
 Q: Do you constantly feel fatigued or do you have non-restful sleep?
 A: Y

*Turn 15:
 Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?
 A: Y

*Turn 16:
 Q: Do you have an active cancer?
 A: N

*Turn 17:
 Q: Have you recently had a loss of appetite or do you get full more quickly than usually?
 A: N

*Turn 18:
 Q: In the last month, have you been in contact with anyone infected with the Ebola virus?
 A: N

*Turn 19:
 Q: Do you have swelling in one or more areas of your body?
 A: N

*Turn 20:
 Q: Do you feel slightly dizzy or lightheaded?
 A: N

Predicted differential:

 Pulmonary neoplasm, Tuberculosis, Acute pulmonary edema, Possible NSTEMI / STEMI, Bronchitis, Pneumonia, Pulmonary embolism, Unstable angina, Myocarditis, Anemia, Bronchiectasis, Bronchospasm / acute asthma exacerbation, Stable angina, Atrial fibrillation, Guillain-Barre syndrome, Spontaneous rib fracture, Pancreatic neoplasm, Myasthenia gravis, Acute dystonic reactions

Ground truth differential:

 Pulmonary neoplasm: 0.091, Tuberculosis: 0.074, Bronchitis: 0.069, Acute pulmonary edema: 0.069, Pneumonia: 0.066, Pulmonary embolism: 0.064, Possible NSTEMI / STEMI: 0.059, Unstable angina: 0.057, Bronchiectasis: 0.050, Myocarditis: 0.050, Anemia: 0.048, Stable angina: 0.046, Guillain-Barre syndrome: 0.044, Bronchospasm / acute asthma exacerbation: 0.043, Atrial fibrillation: 0.042, Spontaneous rib fracture: 0.038, Acute dystonic reactions: 0.031, Myasthenia gravis: 0.031, Pancreatic neoplasm: 0.027

Patient 11

Sex: M, Age: 49

Initial evidence:

I feel pain

Agent inquiries:

 *Turn 1:

Q: Do you have swelling in one or more areas of your body?

A: N

*Turn 2:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: Y

*Turn 3:

Q: What color is the rash?

A:

* yellow

*Turn 4:
Q: Characterize your pain:
A:
* a cramp

*Turn 5:
Q: Where is the affected region located?
A:
* epigastric

*Turn 6:
Q: Do you feel pain somewhere?
A:
* flank(L)
* hypochondrium(L)
* epigastric

*Turn 7:
Q: How intense is the pain?
A:
* 4

*Turn 8:
Q: How precisely is the pain located?
A:
* 5

*Turn 9:
Q: How severe is the itching?
A:
* 3

*Turn 10:
Q: How intense is the pain caused by the rash?
A:
* 0

*Turn 11:
Q: Is the rash swollen?
A:
* 2

*Turn 12:
Q: Is the lesion (or are the lesions) larger than 1cm?
A:
* Y

*Turn 13:
Q: How fast did the pain appear?
A:
* 1

*Turn 14:
Q: Are there members of your family who have been diagnosed with pancreatic cancer?
A: Y

*Turn 15:
Q: Does the pain radiate to another location?
A:
* thoracic spine

*Turn 16:
Q: Do you have chronic pancreatitis?
A: Y

*Turn 17:
Q: Are you significantly overweight compared to people of the same height as you?
A: Y

*Turn 18:
Q: Do you have diabetes?
A: Y

*Turn 19:
Q: Do you constantly feel fatigued or do you have non-restful sleep?
A: Y

*Turn 20:
Q: Do you smoke cigarettes?

A: N
*Turn 21:
Q: Do you have pale stools and dark urine?
A: Y
*Turn 22:
Q: Do you have a cough?
A: N
*Turn 23:
Q: Have you had an involuntary weight loss over the last 3 months?
A: N
*Turn 24:
Q: Are you feeling nauseous or do you feel like vomiting?
A: Y
*Turn 25:
Q: Have you had diarrhea or an increase in stool frequency?
A: Y
Predicted differential:

Boerhaave, Pancreatic neoplasm, Anaphylaxis, Scombroid food poisoning, Possible
NSTEMI / STEMI, Unstable angina, GERD, Chagas
Ground truth differential:

Pancreatic neoplasm: 0.190, Unstable angina: 0.150, Anaphylaxis: 0.132, Boerhaave:
0.131, Scombroid food poisoning: 0.126, Possible NSTEMI / STEMI: 0.096, Chagas:
0.090, GERD: 0.084

Patient 12

Sex: M, Age: 62
Initial evidence:
I feel pain
Agent inquiries:

*Turn 1:
Q: Do you have swelling in one or more areas of your body?
A: N
*Turn 2:
Q: Do you have any lesions, redness or problems on your skin that you believe are
related to the condition you are consulting for?
A: N
*Turn 3:
Q: Does the pain radiate to another location?
A:
* nowhere
*Turn 4:
Q: Do you feel pain somewhere?
A:
* tonsil(L)
* thyroid cartilage
* palace
* under the jaw
*Turn 5:
Q: Characterize your pain:
A:
* sensitive
* burning
*Turn 6:
Q: How precisely is the pain located?
A:
* 10
*Turn 7:
Q: How intense is the pain?
A:
* 6

*Turn 8:
 Q: How fast did the pain appear?
 A:
 * 2

*Turn 9:
 Q: Do you attend or work in a daycare?
 A: Y

*Turn 10:
 Q: Do you live with 4 or more people?
 A: Y

*Turn 11:
 Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?
 A: Y

*Turn 12:
 Q: Do you have a fever (either felt or measured with a thermometer)?
 A: Y

*Turn 13:
 Q: Do you have a cough?
 A: Y

*Turn 14:
 Q: Are you immunosuppressed?
 A: N

*Turn 15:
 Q: Do you have nasal congestion or a clear runny nose?
 A: N

*Turn 16:
 Q: Have you been coughing up blood?
 A: Y

*Turn 17:
 Q: Do you smoke cigarettes?
 A: N

Predicted differential:

 Bronchitis, Tuberculosis, Viral pharyngitis, Acute laryngitis, Possible NSTEMI / STEMI, Unstable angina, Pneumonia, Influenza, Stable angina
 Ground truth differential:

 Bronchitis: 0.197, Tuberculosis: 0.183, Ebola: 0.152, Acute laryngitis: 0.131, Viral pharyngitis: 0.130, Possible NSTEMI / STEMI: 0.107, Unstable angina: 0.100

Patient 13

Sex: F, Age: 60
 Initial evidence:
 I have a sore throat
 Agent inquiries:

*Turn 1:
 Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
 A: N

*Turn 2:
 Q: Do you have pain somewhere, related to your reason for consulting?
 A: Y

*Turn 3:
 Q: Does the pain radiate to another location?
 A:
 * nowhere

*Turn 4:
 Q: Do you feel pain somewhere?
 A:
 * top of the head
 * forehead

* cheek(R)
 * occiput
 * temple(R)

*Turn 5:
 Q: Characterize your pain:
 A:
 * sensitive
 * heavy

*Turn 6:
 Q: How precisely is the pain located?
 A:
 * 3

*Turn 7:
 Q: How intense is the pain?
 A:
 * 8

*Turn 8:
 Q: How fast did the pain appear?
 A:
 * 3

*Turn 9:
 Q: Do you have nasal congestion or a clear runny nose?
 A: N

*Turn 10:
 Q: Do you live with 4 or more people?
 A: Y

*Turn 11:
 Q: Do you attend or work in a daycare?
 A: Y

*Turn 12:
 Q: Do you have a cough?
 A: Y

*Turn 13:
 Q: Do you have a cough that produces colored or more abundant sputum than usual?
 A: N

*Turn 14:
 Q: Do you have a fever (either felt or measured with a thermometer)?
 A: N

*Turn 15:
 Q: Do you have diffuse (widespread) muscle pain?
 A: Y

*Turn 16:
 Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?
 A: N

*Turn 17:
 Q: Are you exposed to secondhand cigarette smoke on a daily basis?
 A: Y

*Turn 18:
 Q: Have you had significantly increased sweating?
 A: Y

*Turn 19:
 Q: Do you smoke cigarettes?
 A: Y

Predicted differential:

 Bronchitis, URTI, Influenza, Cluster headache, Anemia, Chagas
 Ground truth differential:

 Bronchitis: 0.260, URTI: 0.255, Cluster headache: 0.153, Influenza: 0.144, Chagas:
 0.094, Anemia: 0.094

Patient 14

Sex: F, Age: 63
Initial evidence:
I have nasal congestion
Agent inquiries:

*Turn 1:

Q: Do you have pain somewhere, related to your reason for consulting?

A: Y

*Turn 2:

Q: Do you have swelling in one or more areas of your body?

A: N

*Turn 3:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: N

*Turn 4:

Q: Do you feel pain somewhere?

A:

- * top of the head
- * forehead
- * cheek(L)
- * occiput
- * temple(L)

*Turn 5:

Q: Characterize your pain:

A:

- * tedious

*Turn 6:

Q: How precisely is the pain located?

A:

- * 6

*Turn 7:

Q: How intense is the pain?

A:

- * 4

*Turn 8:

Q: How fast did the pain appear?

A:

- * 3

*Turn 9:

Q: Do you live with 4 or more people?

A: Y

*Turn 10:

Q: Do you have a sore throat?

A: Y

*Turn 11:

Q: Do you have a cough?

A: Y

*Turn 12:

Q: Are you exposed to secondhand cigarette smoke on a daily basis?

A: Y

*Turn 13:

Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?

A: Y

*Turn 14:

Q: Do you smoke cigarettes?

A: Y

*Turn 15:

Q: Do you attend or work in a daycare?

A: Y

*Turn 16:

Q: Do you have a cough that produces colored or more abundant sputum than usual?

A: N

*Turn 17:

Q: Do you have a fever (either felt or measured with a thermometer)?
A: Y
*Turn 18:
Q: Have you had significantly increased sweating?
A: Y
*Turn 19:
Q: Do you have diffuse (widespread) muscle pain?
A: N
Predicted differential:

URTI, Influenza, HIV (initial infection), Bronchitis, Tuberculosis, Chronic
rhinosinusitis, Chagas, Acute rhinosinusitis
Ground truth differential:

URTI: 0.218, Influenza: 0.174, HIV (initial infection): 0.164, Bronchitis: 0.155,
Tuberculosis: 0.119, Chronic rhinosinusitis: 0.071, Chagas: 0.052, Acute
rhinosinusitis: 0.046

Patient 15

Sex: F, Age: 14
Initial evidence:
I have diffuse muscle pain
Agent inquiries:

*Turn 1:
Q: Do you have any lesions, redness or problems on your skin that you believe are
related to the condition you are consulting for?
A: N
*Turn 2:
Q: Do you have pain somewhere, related to your reason for consulting?
A: Y
*Turn 3:
Q: Does the pain radiate to another location?
A:
* nowhere
*Turn 4:
Q: Characterize your pain:
A:
* tedious
* sensitive
* heavy
*Turn 5:
Q: How precisely is the pain located?
A:
* 5
*Turn 6:
Q: Do you feel pain somewhere?
A:
* top of the head
* forehead
* cheek(R)
* cheek(L)
* occiput
*Turn 7:
Q: How intense is the pain?
A:
* 7
*Turn 8:
Q: How fast did the pain appear?
A:
* 5
*Turn 9:
Q: Do you have a cough that produces colored or more abundant sputum than usual?

A: N
*Turn 10:
Q: Do you have a cough?
A: Y
*Turn 11:
Q: Do you have a fever (either felt or measured with a thermometer)?
A: N
*Turn 12:
Q: Do you live with 4 or more people?
A: N
*Turn 13:
Q: Do you have a sore throat?
A: Y
*Turn 14:
Q: Do you smoke cigarettes?
A: Y
*Turn 15:
Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?
A: N
*Turn 16:
Q: Are you exposed to secondhand cigarette smoke on a daily basis?
A: N
*Turn 17:
Q: Do you have nasal congestion or a clear runny nose?
A: Y
*Turn 18:
Q: Do you attend or work in a daycare?
A: Y
*Turn 19:
Q: Have you had significantly increased sweating?
A: N

Predicted differential:

URTI, Bronchitis, Influenza, Cluster headache, Chronic rhinosinusitis, Chagas, Anemia, Acute rhinosinusitis, Pneumonia

Ground truth differential:

URTI: 0.230, Cluster headache: 0.174, Bronchitis: 0.167, Chronic rhinosinusitis: 0.141, Acute rhinosinusitis: 0.101, Chagas: 0.066, Anemia: 0.066, Influenza: 0.054

Patient 16

Sex: M, Age: 49

Initial evidence:

I have nasal congestion

Agent inquiries:

*Turn 1:
Q: Do you have swelling in one or more areas of your body?
A: N
*Turn 2:
Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
A: N
*Turn 3:
Q: Do you have pain somewhere, related to your reason for consulting?
A: Y
*Turn 4:
Q: Does the pain radiate to another location?
A:
* nowhere
*Turn 5:

Q: Characterize your pain:
A:
 * sensitive

*Turn 6:
Q: Do you feel pain somewhere?
A:
 * top of the head
 * forehead
 * cheek(R)
 * occiput
 * temple(R)

*Turn 7:
Q: How precisely is the pain located?
A:
 * 7

*Turn 8:
Q: How intense is the pain?
A:
 * 2

*Turn 9:
Q: How fast did the pain appear?
A:
 * 5

*Turn 10:
Q: Do you have a sore throat?
A: N

*Turn 11:
Q: Do you live with 4 or more people?
A: N

*Turn 12:
Q: Do you attend or work in a daycare?
A: Y

*Turn 13:
Q: Do you have a cough that produces colored or more abundant sputum than usual?
A: N

*Turn 14:
Q: Do you have a cough?
A: Y

*Turn 15:
Q: Have you been in contact with a person with similar symptoms in the past 2 weeks?
A: Y

*Turn 16:
Q: Are you exposed to secondhand cigarette smoke on a daily basis?
A: N

*Turn 17:
Q: Do you have a fever (either felt or measured with a thermometer)?
A: Y

*Turn 18:
Q: Do you have diffuse (widespread) muscle pain?
A: N

*Turn 19:
Q: Have you had significantly increased sweating?
A: Y

*Turn 20:
Q: Do you smoke cigarettes?
A: N

Predicted differential:

URTI, Bronchitis, Influenza, Tuberculosis, Chronic rhinosinusitis, Chagas, Acute rhinosinusitis

Ground truth differential:

URTI: 0.236, Influenza: 0.202, Bronchitis: 0.201, Tuberculosis: 0.149, Chronic rhinosinusitis: 0.089, Chagas: 0.065, Acute rhinosinusitis: 0.058

Patient 17

Sex: F, Age: 34

Initial evidence:

I feel pain

Agent inquiries:

*Turn 1:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: N

*Turn 2:

Q: Does the pain radiate to another location?

A:

* nowhere

*Turn 3:

Q: Do you feel pain somewhere?

A:

* lower chest

* side of the chest(L)

* upper chest

* breast(L)

* posterior chest wall(R)

*Turn 4:

Q: Characterize your pain:

A:

* sensitive

*Turn 5:

Q: How precisely is the pain located?

A:

* 7

*Turn 6:

Q: How intense is the pain?

A:

* 5

*Turn 7:

Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?

A: N

*Turn 8:

Q: How fast did the pain appear?

A:

* 0

*Turn 9:

Q: Have you ever had pneumonia?

A: N

*Turn 10:

Q: Do you smoke cigarettes?

A: Y

*Turn 11:

Q: Do you have a cough?

A: Y

*Turn 12:

Q: Do you have a chronic obstructive pulmonary disease (COPD)?

A: Y

*Turn 13:

Q: Do you have nasal congestion or a clear runny nose?

A: Y

*Turn 14:

Q: Do you attend or work in a daycare?

A: N

*Turn 15:

Q: Do you have a cough that produces colored or more abundant sputum than usual?

A: Y

*Turn 16:

Q: Do you have a fever (either felt or measured with a thermometer)?
A: N
*Turn 17:
Q: Are your symptoms more prominent at night?
A: N
*Turn 18:
Q: Do you have a sore throat?
A: N
*Turn 19:
Q: Have you noticed a wheezing sound when you exhale?
A: Y
*Turn 20:
Q: Are your vaccinations up to date?
A: N
Predicted differential:

Bronchitis, Acute COPD exacerbation / infection, GERD, Pneumonia, Bronchiectasis,
Bronchospasm / acute asthma exacerbation, Spontaneous rib fracture, Possible
NSTEMI / STEMI, Unstable angina, Boerhaave, URTI, Pericarditis, Stable angina,
Allergic sinusitis, Scombroid food poisoning
Ground truth differential:

Acute COPD exacerbation / infection: 0.121, Bronchitis: 0.118, GERD: 0.092,
Pneumonia: 0.090, Bronchospasm / acute asthma exacerbation: 0.084,
Bronchiectasis: 0.079, Spontaneous rib fracture: 0.074, Possible NSTEMI / STEMI:
0.070, Unstable angina: 0.065, Boerhaave: 0.062, URTI: 0.045, Pericarditis:
0.043, Stable angina: 0.036, Scombroid food poisoning: 0.020

Patient 18

Sex: F, Age: 18
Initial evidence:
I feel pain
Agent inquiries:

*Turn 1:
Q: Do you have any lesions, redness or problems on your skin that you believe are
related to the condition you are consulting for?
A: N
*Turn 2:
Q: Does the pain radiate to another location?
A:
* nowhere
*Turn 3:
Q: Do you feel pain somewhere?
A:
* lower chest
* side of the chest(L)
* breast(L)
*Turn 4:
Q: Characterize your pain:
A:
* heartbreaking
* a knife stroke
*Turn 5:
Q: How precisely is the pain located?
A:
* 6
*Turn 6:
Q: How intense is the pain?
A:
* 8
*Turn 7:
Q: Do you have pain that is increased when you breathe in deeply?

A: N
*Turn 8:
Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?
A: N
*Turn 9:
Q: Have you ever had a spontaneous pneumothorax?
A: Y
*Turn 10:
Q: How fast did the pain appear?
A:
* 7
*Turn 11:
Q: Do you smoke cigarettes?
A: Y
*Turn 12:
Q: Do you have nasal congestion or a clear runny nose?
A: N
*Turn 13:
Q: Do you have symptoms that are increased with physical exertion but alleviated with rest?
A: Y
*Turn 14:
Q: Do you have chest pain even at rest?
A: N
*Turn 15:
Q: Do you feel your heart is beating fast (racing), irregularly (missing a beat) or do you feel palpitations?
A: N
Predicted differential:

Unstable angina, Stable angina, Possible NSTEMI / STEMI, Spontaneous pneumothorax, Atrial fibrillation, Pericarditis
Ground truth differential:

Unstable angina: 0.272, Stable angina: 0.243, Possible NSTEMI / STEMI: 0.199, Pericarditis: 0.124, Atrial fibrillation: 0.097, Spontaneous pneumothorax: 0.064

Patient 19

Sex: F, Age: 35
Initial evidence:
I feel pain
Agent inquiries:

*Turn 1:
Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?
A: N
*Turn 2:
Q: Does the pain radiate to another location?
A:
* lower chest
*Turn 3:
Q: Characterize your pain:
A:
* haunting
* sensitive
* a knife stroke
* tugging
* burning
* sickening
*Turn 4:

Q: Do you feel pain somewhere?
A:
* lower chest
* upper chest
* hypochondrium(R)
* hypochondrium(L)
* epigastric

*Turn 5:
Q: How precisely is the pain located?
A:
* 6

*Turn 6:
Q: How intense is the pain?
A:
* 3

*Turn 7:
Q: Are your symptoms worse when lying down and alleviated while sitting up?
A: Y

*Turn 8:
Q: Do you have a cough?
A: N

*Turn 9:
Q: Do you have a hiatal hernia?
A: N

*Turn 10:
Q: Do you have symptoms that get worse after eating?
A: N

*Turn 11:
Q: Do you have a burning sensation that starts in your stomach then goes up into your throat, and can be associated with a bitter taste in your mouth?
A: Y

*Turn 12:
Q: How fast did the pain appear?
A:
* 0

*Turn 13:
Q: Do you have asthma or have you ever had to use a bronchodilator in the past?
A: N

*Turn 14:
Q: Do you smoke cigarettes?
A: Y

*Turn 15:
Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?
A: N

*Turn 16:
Q: Do you think you are pregnant or are you currently pregnant?
A: Y

*Turn 17:
Q: Have you recently thrown up blood or something resembling coffee beans?
A: N

*Turn 18:
Q: Do you drink alcohol excessively or do you have an addiction to alcohol?
A: Y

*Turn 19:
Q: Are you significantly overweight compared to people of the same height as you?
A: N

*Turn 20:
Q: Have you recently had stools that were black (like coal)?
A: N

Predicted differential:

GERD, Possible NSTEMI / STEMI, Pericarditis, Unstable angina, Stable angina,
Pancreatic neoplasm, Anaphylaxis

Ground truth differential:

GERD: 0.297, Possible NSTEMI / STEMI: 0.221, Pericarditis: 0.151, Unstable angina:
0.145, Stable angina: 0.116, Anaphylaxis: 0.070

Patient 20

Sex: F, Age: 43

Initial evidence:

I have had significantly increased sweating

Agent inquiries:

*Turn 1:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: N

*Turn 2:

Q: Do you have pain somewhere, related to your reason for consulting?

A: Y

*Turn 3:

Q: Does the pain radiate to another location?

A:

- * biceps(L)
- * thyroid cartilage
- * thoracic spine
- * shoulder(R)
- * shoulder(L)

*Turn 4:

Q: Do you feel pain somewhere?

A:

- * biceps(R)
- * thoracic spine
- * side of the chest(L)
- * upper chest
- * epigastric

*Turn 5:

Q: Characterize your pain:

A:

- * tedious
- * heavy

*Turn 6:

Q: How precisely is the pain located?

A:

- * 4

*Turn 7:

Q: How intense is the pain?

A:

- * 5

*Turn 8:

Q: How fast did the pain appear?

A:

- * 5

*Turn 9:

Q: Have you ever had a heart attack or do you have angina (chest pain)?

A: Y

*Turn 10:

Q: Do you have diabetes?

A: Y

*Turn 11:

Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?

A: Y

*Turn 12:

Q: Do you have close family members who had a cardiovascular disease problem before the age of 50?

A: Y

*Turn 13:
 Q: Do you have high blood pressure or do you take medications to treat high blood pressure?
 A: Y

*Turn 14:
 Q: Are you significantly overweight compared to people of the same height as you?
 A: Y

*Turn 15:
 Q: Do you have high cholesterol or do you take medications to treat high cholesterol?
 A: Y

*Turn 16:
 Q: Do you have swelling in one or more areas of your body?
 A: N

*Turn 17:
 Q: Do you smoke cigarettes?
 A: Y

*Turn 18:
 Q: Do you exercise regularly, 4 times per week or more?
 A: Y

*Turn 19:
 Q: Do you drink alcohol excessively or do you have an addiction to alcohol?
 A: N

*Turn 20:
 Q: Do you have symptoms that are increased with physical exertion but alleviated with rest?
 A: N

*Turn 21:
 Q: Are you feeling nauseous or do you feel like vomiting?
 A: N

*Turn 22:
 Q: Do you constantly feel fatigued or do you have non-restful sleep?
 A: N

*Turn 23:
 Q: Do you have chest pain even at rest?
 A: N

*Turn 24:
 Q: Do you find that your symptoms have worsened over the last 2 weeks and that progressively less effort is required to cause the symptoms?
 A: N

*Turn 25:
 Q: Are you infected with the human immunodeficiency virus (HIV)?
 A: N

Predicted differential:

Possible NSTEMI / STEMI, SLE, GERD, Guillain-Barre syndrome, Atrial fibrillation, Myocarditis, Anemia, Acute dystonic reactions, Pericarditis, Panic attack, Acute pulmonary edema, Myasthenia gravis, Anaphylaxis, Pancreatic neoplasm, Scombroid food poisoning, Pulmonary neoplasm, PSVT, Chagas

Ground truth differential:

Possible NSTEMI / STEMI: 0.127, SLE: 0.087, GERD: 0.086, Guillain-Barre syndrome: 0.081, Atrial fibrillation: 0.077, Panic attack: 0.065, Myocarditis: 0.063, Pericarditis: 0.058, Acute dystonic reactions: 0.057, Myasthenia gravis: 0.057, Anemia: 0.057, Acute pulmonary edema: 0.053, Pancreatic neoplasm: 0.043, Scombroid food poisoning: 0.036, PSVT: 0.030, Pulmonary neoplasm: 0.013, Unstable angina: 0.009

G Example interactions

We showcase three example interactions of CASANDE with synthetic patients from the DDXPlus test set. For each example, we show the sex, age and initial symptom of the patient, followed by

the questions asked by CASANDE, the answers given by the patient, and the differential diagnosis at each turn. Finally, we show the ground truth differential diagnosis. Diseases that are severe are marked with (severe) after their names.

In all three examples, we observe a common pattern: the predicted differential diagnosis tends to be larger and changes more frequently at the beginning of the interaction than at the end. We also see several questions at the end with the same differential. This is aligned with the desired exploration-confirmation behavior. Notably, there are often severe pathologies present in the predicted differential at some turns which are later excluded. This could indicate the agent is actively considering those diseases and gathering evidence to rule them out.

Example 1

Sex: F, age: 31

Initial evidence: I feel slightly dizzy or lightheaded.

Differential diagnosis: ['Anemia', 'Atrial fibrillation', 'Cluster headache', 'HIV (initial infection)']

Turn 1:

Q: Have you traveled out of the country in the last 4 weeks?

A: North America

Differential diagnosis: ['Anemia', 'Atrial fibrillation', 'Cluster headache', 'Pulmonary embolism (severe)', 'Anaphylaxis (severe)', 'Guillain-Barre syndrome (severe)', 'Acute dystonic reactions (severe)', 'Acute pulmonary edema (severe)', 'HIV (initial infection)', 'Myocarditis (severe)', 'Myasthenia gravis', 'PSVT (severe)', 'Chagas', 'Scombroid food poisoning (severe)', 'Pulmonary neoplasm', 'SLE']

Turn 2:

Q: Do you have pain somewhere, related to your reason for consulting?

A: No

Differential diagnosis: ['Atrial fibrillation', 'Panic attack', 'Cluster headache', 'Myocarditis (severe)', 'Anemia', 'Pulmonary embolism (severe)', 'PSVT (severe)', 'Guillain-Barre syndrome (severe)', 'Acute pulmonary edema (severe)', 'Anaphylaxis (severe)', 'Acute dystonic reactions (severe)', 'Myasthenia gravis', 'Scombroid food poisoning (severe)', 'SLE', 'HIV (initial infection)', 'Bronchospasm / acute asthma exacerbation', 'Pulmonary neoplasm', 'Acute COPD exacerbation / infection']

Turn 3:

Q: Do you feel your heart is beating very irregularly or in a disorganized pattern?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Pulmonary embolism (severe)', 'Anemia', 'Anaphylaxis (severe)', 'Acute dystonic reactions (severe)', 'Acute pulmonary edema (severe)', 'Scombroid food poisoning (severe)', 'Guillain-Barre syndrome (severe)', 'Myasthenia gravis', 'Myocarditis (severe)', 'Bronchospasm / acute asthma exacerbation', 'Acute COPD exacerbation / infection', 'Panic attack']

Turn 4:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: No

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 5:

Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?

A: No

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 6:

Q: Do you have severe Chronic Obstructive Pulmonary Disease (COPD)?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 7:

Q: Do you have high blood pressure or do you take medications to treat high blood pressure?

A: No

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 8:

Q: Do you have a known issue with one of your heart valves?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 9:

Q: Have you ever had a heart attack or do you have angina (chest pain)?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 10:

Q: Have you been diagnosed with hyperthyroidism?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 11:

Q: Have you ever had a pericarditis?

A: No

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 12:

Q: Are you significantly overweight compared to people of the same height as you?

A: No

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 13:

Q: Do you feel your heart is beating fast (racing), irregularly (missing a beat) or do you feel palpitations?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 14:

Q: Do you have a known heart defect?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 15:

Q: Do you have diabetes?

A: Yes

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Turn 16:

Q: Do you have symptoms that are increased with physical exertion but alleviated with rest?

A: No

Differential diagnosis: ['Atrial fibrillation', 'PSVT (severe)', 'Anemia', 'Scombroid food poisoning (severe)']

Ground truth differential diagnosis: ['Atrial fibrillation: 0.416', 'PSVT (severe): 0.331', 'Anemia: 0.165', 'Scombroid food poisoning (severe): 0.088']

Example 2

Sex: M, age: 20

Initial evidence: I have pain that is increased when I breathe in deeply.

Differential diagnosis: ['Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)', 'Panic attack', 'Pericarditis', 'Guillain-Barre syndrome (severe)', 'Boerhaave (severe)', 'Myocarditis (severe)', 'Possible NSTEMI / STEMI (severe)', 'Anaphylaxis (severe)', 'GERD', 'Atrial fibrillation', 'Myasthenia gravis', 'Acute pulmonary edema (severe)', 'Anemia', 'Acute dystonic reactions (severe)', 'Scombroid food poisoning (severe)', 'Sarcoidosis', 'HIV (initial infection)', 'Chagas', 'Pancreatic neoplasm', 'PSVT (severe)', 'SLE', 'Unstable angina (severe)']

Turn 1:

Q: Do you have any lesions, redness or problems on your skin that you believe are related to the condition you are consulting for?

A: No

Differential diagnosis: ['Localized edema', 'Possible NSTEMI / STEMI (severe)', 'GERD', 'Pulmonary embolism (severe)', 'Unstable angina (severe)', 'Panic attack', 'Pericarditis', 'Spontaneous pneumothorax (severe)', 'Boerhaave (severe)', 'Cluster headache', 'Guillain-Barre syndrome (severe)', 'Myocarditis (severe)', 'Acute pulmonary edema (severe)', 'Anaphylaxis (severe)', 'Atrial fibrillation', 'Stable angina (severe)', 'Anemia', 'Scombroid food poisoning (severe)', 'Acute dystonic reactions (severe)', 'Bronchospasm / acute asthma exacerbation', 'Sarcoidosis', 'PSVT (severe)', 'HIV (initial infection)', 'Bronchitis', 'Myasthenia gravis', 'Chagas', 'Pneumonia', 'Pancreatic neoplasm', 'Bronchiectasis', 'Larygospasm (severe)', 'SLE', 'Spontaneous rib fracture']

Turn 2:

Q: Do you have pain somewhere, related to your reason for consulting?

A: Yes

Differential diagnosis: ['Possible NSTEMI / STEMI (severe)', 'Localized edema', 'Pulmonary embolism (severe)', 'Unstable angina (severe)', 'Spontaneous pneumothorax (severe)', 'Anaphylaxis (severe)', 'Stable angina (severe)', 'Pericarditis', 'Scombroid food poisoning (severe)', 'GERD', 'Acute COPD exacerbation / infection', 'Chagas', 'Boerhaave (severe)', 'Sarcoidosis']

Turn 3:

Q: Does the pain radiate to another location?

A: nowhere

Differential diagnosis: ['Pancreatic neoplasm', 'Scombroid food poisoning (severe)', 'Chagas', 'Anaphylaxis (severe)', 'Anemia']

Turn 4:

Q: Do you have swelling in one or more areas of your body?

A: No

Differential diagnosis: ['Cluster headache', 'Anemia', 'Possible NSTEMI / STEMI (severe)', 'Pancreatic neoplasm', 'GERD', 'Pulmonary embolism (severe)', 'Chagas', 'Unstable angina (severe)', 'Stable angina (severe)', 'Scombroid food poisoning (severe)']

Turn 5:

Q: Do you feel pain somewhere?

A: {'lower chest', 'upper chest', 'breast(R)'}
'

Differential diagnosis: ['Bronchitis', 'Possible NSTEMI / STEMI (severe)', 'Panic attack', 'Pulmonary embolism (severe)', 'Unstable angina (severe)', 'Pericarditis', 'Pneumonia', 'Spontaneous pneumothorax (severe)', 'Acute

pulmonary edema (severe)', 'Spontaneous rib fracture', 'GERD', 'Bronchiectasis', 'Boerhaave (severe)', 'Guillain-Barre syndrome (severe)', 'Atrial fibrillation', 'Myocarditis (severe)', 'Cluster headache', 'Whooping cough', 'Stable angina (severe)', 'Anaphylaxis (severe)', 'Acute COPD exacerbation / infection', 'Bronchospasm / acute asthma exacerbation', 'Acute dystonic reactions (severe)', 'Anemia', 'Scombroid food poisoning (severe)', 'PSVT (severe)', 'Sarcoidosis', 'Myasthenia gravis', 'Acute otitis media', 'HIV (initial infection)', 'Pulmonary neoplasm', 'URTI', 'Chagas']

Turn 6:

Q: Characterize your pain:

A: {'a knife stroke', 'heartbreaking', 'violent'}

Differential diagnosis: ['Unstable angina (severe)', 'Possible NSTEMI / STEMI (severe)', 'Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)', 'Stable angina (severe)', 'Guillain-Barre syndrome (severe)', 'Pericarditis', 'GERD', 'Myocarditis (severe)', 'Acute pulmonary edema (severe)', 'Panic attack', 'Atrial fibrillation', 'Acute dystonic reactions (severe)', 'Sarcoidosis', 'Anemia', 'Myasthenia gravis', 'Boerhaave (severe)', 'Scombroid food poisoning (severe)']

Turn 7:

Q: How precisely is the pain located?

A: 5

Differential diagnosis: ['Unstable angina (severe)', 'Possible NSTEMI / STEMI (severe)', 'Spontaneous pneumothorax (severe)', 'Pulmonary embolism (severe)', 'Stable angina (severe)', 'Guillain-Barre syndrome (severe)', 'Pericarditis', 'GERD', 'Myocarditis (severe)', 'Acute pulmonary edema (severe)', 'Panic attack', 'Atrial fibrillation', 'Acute dystonic reactions (severe)', 'Sarcoidosis', 'Anemia', 'Myasthenia gravis', 'Scombroid food poisoning (severe)', 'Boerhaave (severe)', 'PSVT (severe)']

Turn 8:

Q: How intense is the pain?

A: 5

Differential diagnosis: ['Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)', 'Possible NSTEMI / STEMI (severe)', 'Unstable angina (severe)', 'Pericarditis', 'Stable angina (severe)', 'Guillain-Barre syndrome (severe)', 'Panic attack', 'Acute pulmonary edema (severe)', 'GERD', 'Myocarditis (severe)', 'Boerhaave (severe)', 'Atrial fibrillation', 'Acute dystonic reactions (severe)', 'PSVT (severe)', 'Sarcoidosis', 'Anemia', 'Myasthenia gravis', 'Scombroid food poisoning (severe)', 'Pneumonia']

Turn 9:

Q: How fast did the pain appear?

A: 5

Differential diagnosis: ['Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)', 'Possible NSTEMI / STEMI (severe)', 'Unstable angina (severe)', 'Stable angina (severe)', 'Pericarditis', 'Guillain-Barre syndrome (severe)', 'Panic attack', 'Acute pulmonary edema (severe)', 'GERD', 'Myocarditis (severe)', 'Atrial fibrillation', 'Boerhaave (severe)', 'Acute dystonic reactions (severe)', 'PSVT (severe)', 'Sarcoidosis', 'Anemia', 'Myasthenia gravis', 'Scombroid food poisoning (severe)']

Turn 10:

Q: Have you ever had a spontaneous pneumothorax?

A: Yes

Differential diagnosis: ['Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)', 'Possible NSTEMI / STEMI (severe)', 'Unstable angina (severe)', 'Pericarditis', 'Stable angina (severe)', 'Guillain-Barre syndrome (severe)', 'Panic attack', 'Acute pulmonary edema (severe)', 'GERD', 'Myocarditis (severe)', 'Atrial fibrillation', 'Boerhaave (severe)', 'Acute dystonic reactions (severe)', 'PSVT (severe)', 'Sarcoidosis', 'Anemia', 'Myasthenia gravis', 'Scombroid food poisoning (severe)']

Turn 11:

Q: Have any of your family members ever had a pneumothorax?

A: Yes

Differential diagnosis: ['Spontaneous pneumothorax (severe)', 'Pulmonary embolism (severe)', 'Possible NSTEMI / STEMI (severe)', 'Unstable angina (severe)', 'Pericarditis', 'Stable angina (severe)', 'Guillain-Barre syndrome (severe)', 'Panic attack', 'GERD', 'Myocarditis (severe)', 'Atrial fibrillation', 'Boerhaave (severe)', 'Acute pulmonary edema (severe)', 'Acute dystonic reactions (severe)', 'PSVT (severe)', 'Sarcoidosis', 'Anemia', 'Myasthenia gravis', 'Scombroid food poisoning (severe)']

Turn 12:

Q: Do you have a chronic obstructive pulmonary disease (COPD)?

A: Yes

Differential diagnosis: ['Spontaneous pneumothorax (severe)', 'Pulmonary embolism (severe)', 'Possible NSTEMI / STEMI (severe)', 'Unstable angina (severe)', 'Pericarditis', 'Stable angina (severe)', 'Guillain-Barre syndrome (severe)', 'Panic attack', 'Acute pulmonary edema (severe)', 'GERD', 'Myocarditis (severe)', 'Atrial fibrillation', 'Boerhaave (severe)', 'Acute dystonic reactions (severe)', 'PSVT (severe)', 'Sarcoidosis', 'Anemia', 'Myasthenia gravis', 'Scombroid food poisoning (severe)']

Turn 13:

Q: Do you have symptoms that are increased with physical exertion but alleviated with rest?

A: No

Differential diagnosis: ['Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)', 'Possible NSTEMI / STEMI (severe)', 'Guillain-Barre syndrome (severe)', 'GERD', 'Boerhaave (severe)', 'Pericarditis', 'Atrial fibrillation', 'Myocarditis (severe)', 'Panic attack', 'Sarcoidosis', 'Acute dystonic reactions (severe)', 'Unstable angina (severe)', 'Anemia', 'Myasthenia gravis', 'Scombroid food poisoning (severe)', 'Stable angina (severe)', 'PSVT (severe)']

Turn 14:

Q: Do you smoke cigarettes?

A: No

Differential diagnosis: ['Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)', 'Possible NSTEMI / STEMI (severe)', 'Guillain-Barre syndrome (severe)', 'GERD', 'Boerhaave (severe)', 'Myocarditis (severe)', 'Atrial fibrillation', 'Unstable angina (severe)', 'Pericarditis', 'Sarcoidosis', 'Anemia', 'Acute dystonic reactions (severe)', 'Panic attack', 'Myasthenia gravis', 'Scombroid food poisoning (severe)', 'Stable angina (severe)', 'PSVT (severe)']

Turn 15:

Q: Do you have chest pain even at rest?

A: No

Differential diagnosis: ['Possible NSTEMI / STEMI (severe)', 'Spontaneous pneumothorax (severe)', 'Pulmonary embolism (severe)', 'GERD', 'Guillain-Barre syndrome (severe)', 'Myocarditis (severe)', 'Pulmonary neoplasm', 'Boerhaave (severe)', 'Atrial fibrillation', 'Sarcoidosis', 'Pericarditis', 'Anemia', 'Scombroid food poisoning (severe)', 'Acute dystonic reactions (severe)', 'Myasthenia gravis', 'Panic attack', 'Unstable angina (severe)', 'PSVT (severe)', 'Stable angina (severe)']

Turn 16:

Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?

A: No

Differential diagnosis: ['Possible NSTEMI / STEMI (severe)', 'Pulmonary embolism (severe)', 'Spontaneous pneumothorax (severe)']

Ground truth differential diagnosis: ['Possible NSTEMI / STEMI (severe): 0.395', 'Pulmonary embolism (severe): 0.381', 'Spontaneous pneumothorax (severe): 0.222']

Example 3

Sex: M, age: 49

Initial evidence: I have felt like I was suffocating for a very short time associated with inability to breathe or speak.

Differential diagnosis: ['Bronchitis', 'Possible NSTEMI / STEMI (severe)', 'Panic attack', 'Pneumonia', 'Inguinal hernia', 'GERD', 'Pericarditis', 'Pulmonary embolism (severe)', 'Acute pulmonary edema (severe)', 'Unstable angina (severe)', 'Myocarditis (severe)', 'Spontaneous pneumothorax (severe)', 'Atrial fibrillation', 'Boerhaave (severe)', 'Guillain-Barre syndrome (severe)', 'Bronchiectasis', 'Spontaneous rib fracture', 'Whooping cough', 'Anaphylaxis (severe)', 'URTI', 'Anemia', 'Acute COPD exacerbation / infection', 'Bronchospasm / acute asthma exacerbation', 'Localized edema', 'Acute dystonic reactions (severe)', 'Stable angina (severe)', 'Pancreatic neoplasm', 'Myasthenia gravis', 'PSVT (severe)', 'HIV (initial infection)', 'Scombroid food poisoning (severe)', 'Acute otitis media', 'Sarcoidosis', 'Chagas', 'Cluster headache', 'SLE', 'Pulmonary neoplasm', 'Larygospasm (severe)', 'Tuberculosis']

Turn 1:

Q: Have you traveled out of the country in the last 4 weeks?

A: N

Differential diagnosis: ['Acute otitis media', 'URTI', 'Localized edema', 'Influenza', 'Anemia', 'Cluster headache', 'HIV (initial infection)', 'GERD', 'Bronchitis', 'Pancreatic neoplasm', 'Chagas', 'Boerhaave (severe)', 'Panic attack', 'Pneumonia', 'Bronchiectasis', 'SLE', 'Tuberculosis', 'Myocarditis (severe)', 'Guillain-Barre syndrome (severe)', 'Possible NSTEMI / STEMI (severe)', 'Atrial fibrillation', 'Pulmonary embolism (severe)', 'Acute pulmonary edema (severe)', 'Unstable angina (severe)']

Turn 2:

Q: Do you have pain somewhere, related to your reason for consulting?

A: No

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Pulmonary embolism (severe)', 'Atrial fibrillation', 'Pulmonary neoplasm', 'Anemia', 'Anaphylaxis (severe)', 'Scombroid food poisoning (severe)', 'Bronchitis', 'PSVT (severe)', 'Myasthenia gravis', 'Possible NSTEMI / STEMI (severe)', 'Acute COPD exacerbation / infection', 'Chagas', 'Bronchospasm / acute asthma exacerbation', 'Tuberculosis', 'Guillain-Barre syndrome (severe)', 'Pneumonia', 'Acute pulmonary edema (severe)', 'Myocarditis (severe)', 'Whooping cough']

Turn 3:

Q: Do you have a hard time opening/raising one or both eyelids?

A: Yes

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Pulmonary embolism (severe)', 'Myocarditis (severe)', 'Myasthenia gravis', 'Acute pulmonary edema (severe)', 'Atrial fibrillation', 'Anaphylaxis (severe)', 'Anemia', 'Pulmonary neoplasm', 'Scombroid food poisoning (severe)', 'SLE', 'Possible NSTEMI / STEMI (severe)', 'PSVT (severe)', 'Guillain-Barre syndrome (severe)', 'Chagas']

Turn 4:

Q: Do you feel that muscle spasms or soreness in your neck are keeping you from turning your head to one side?

A: No

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Pulmonary neoplasm', 'Atrial fibrillation', 'Anemia', 'Scombroid food poisoning (severe)', 'Myasthenia gravis', 'Myocarditis (severe)', 'Possible NSTEMI / STEMI (severe)', 'Anaphylaxis (severe)']

Turn 5:

Q: Do you regularly take stimulant drugs?

A: No

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Atrial fibrillation', 'Pulmonary neoplasm', 'Myasthenia gravis', 'Anemia', 'Scombroid food poisoning (severe)', 'Anaphylaxis (severe)', 'Possible NSTEMI / STEMI (severe)', '']

Myocarditis (severe)', 'PSVT (severe)', 'Guillain-Barre syndrome (severe)', 'Chagas', 'Acute pulmonary edema (severe)']

Turn 6:

Q: Do you have difficulty articulating words/speaking?

A: No

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Myasthenia gravis', 'Anemia']

Turn 7:

Q: Have you been treated in hospital recently for nausea, agitation, intoxication or aggressive behavior and received medication via an intravenous or intramuscular route?

A: Yes

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Myasthenia gravis', 'Atrial fibrillation', 'Anemia', 'Pulmonary neoplasm', 'Guillain-Barre syndrome (severe)', 'Scombroid food poisoning (severe)', 'Myocarditis (severe)', 'Possible NSTEMI / STEMI (severe)']

Turn 8:

Q: Have you started or taken any antipsychotic medication within the last 7 days?

A: Yes

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Atrial fibrillation', 'Anemia', 'Pulmonary neoplasm', 'Myasthenia gravis', 'Myocarditis (severe)', 'Scombroid food poisoning (severe)', 'Guillain-Barre syndrome (severe)', 'Possible NSTEMI / STEMI (severe)', 'PSVT (severe)']

Turn 9:

Q: Do you have trouble keeping your tongue in your mouth?

A: Yes

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Myasthenia gravis', 'Guillain-Barre syndrome (severe)', 'Myocarditis (severe)', 'Atrial fibrillation', 'Anemia', 'Scombroid food poisoning (severe)', 'Pulmonary neoplasm', 'Pulmonary embolism (severe)', 'SLE', 'PSVT (severe)', 'Chagas', 'Possible NSTEMI / STEMI (severe)']

Turn 10:

Q: Do you have annoying muscle spasms in your face, neck or any other part of your body?

A: Yes

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Acute pulmonary edema (severe)', 'Pulmonary embolism (severe)', 'Atrial fibrillation', 'Anaphylaxis (severe)', 'Myocarditis (severe)', 'Bronchospasm / acute asthma exacerbation', 'Guillain-Barre syndrome (severe)', 'Bronchitis', 'Myasthenia gravis', 'Pulmonary neoplasm', 'Anemia', 'Scombroid food poisoning (severe)', 'Acute COPD exacerbation / infection', 'SLE', 'Pneumonia', 'Possible NSTEMI / STEMI (severe)', 'PSVT (severe)', 'Bronchiectasis', 'Tuberculosis', 'Panic attack', 'Influenza']

Turn 11:

Q: Are you experiencing shortness of breath or difficulty breathing in a significant way?

A: No

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Myasthenia gravis']

Turn 12:

Q: Do you suddenly have difficulty or an inability to open your mouth or have jaw pain when opening it?

A: No

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Myasthenia gravis']

Turn 13:

Q: Are you unable to control the direction of your eyes?

A: No

Differential diagnosis: ['Acute dystonic reactions (severe)', 'Myasthenia gravis']

Ground truth differential diagnosis: ['Acute dystonic reactions (severe): 0.627', '
Myasthenia gravis: 0.373']