LREC-COLING 2024

LREC-COLING
2024

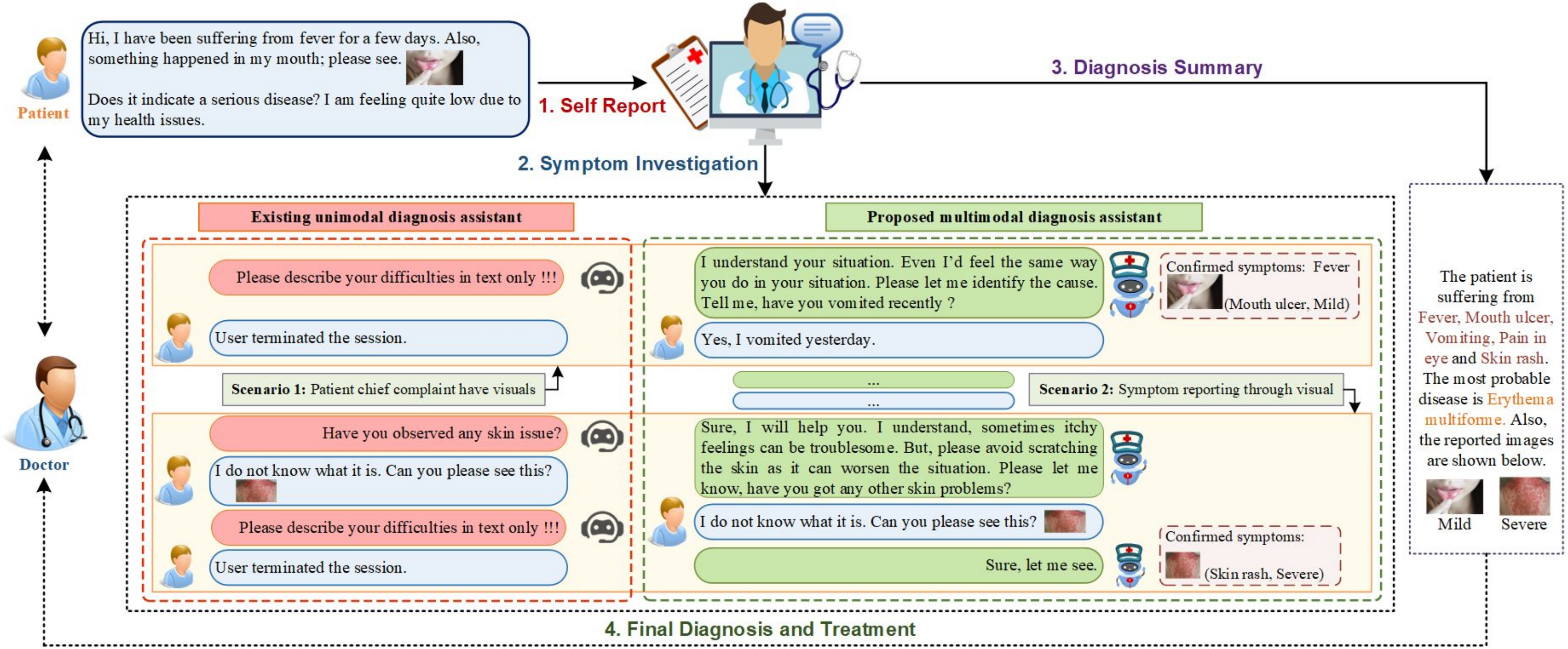
Seeing is believing !
Towards Knowledge-Infused Multi-modal Medical Dialogue Generation

LREC-COLING 2024
May 20 - May 25, 2024
Torino, Italia

Abhisek Tiwari
Research Scholar, PMRF
Dept. of Computer Science and Engineering
Indian Institute of Technology, Patna (IIT Patna)

Introduction

Problem Statement: We often describe our chief medical complaints using visuals; however, the traditional diagnosis assistants perform symptom/sign extraction through text only.



Research Questions

- **Importance of Visual expressions:** Does a diagnosis assistant diagnose patients more accurately and satisfactorily if it considers visual signs and their severity in addition to symptoms conveyed through text?

- **Role of Dialogue Context in Identifying an Image:** Can dialogue context help in identifying a sign image and its severity, which appears during the conversation?

- **Impact of External Knowledge:** What impact might global knowledge, such as knowledge of symptom-disease associations, have on the diagnosis ability of diagnosis assistants? Does the mechanism of knowledge infusion influence its efficacy?

Key Contributions

- **Empathy and severity-aware Multi-Modal Medical Dialogue (ES-MMD):** We curate an Empathy and severity-aware Multi-Modal Medical Dialogue (ES-MMD) corpus in English, where each utterance is annotated with its corresponding intent, sign, symptom, and severity level.

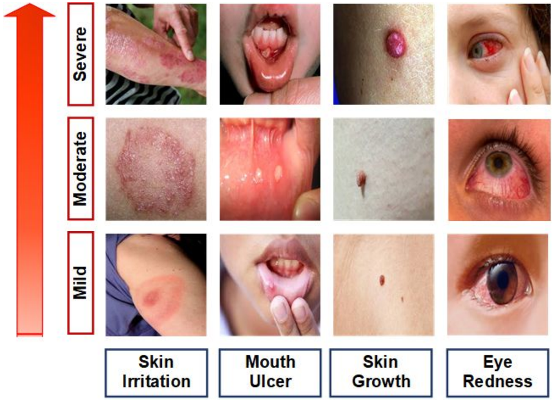
- **Knowledge-Infused, Multi-modal Medical Dialogue Generation (KI-MMDG):** We propose a transformer-based Knowledge-Infused, Multi-modal Medical Dialogue Generation (KI-MMDG) framework, which leverages a discourse-aware selective filtering strategy for knowledge distillation and a natural language understanding (NLU) module for semantic understanding of textual-visual utterances.

- **State-of-the-art Performance:** Our proposed, KI-MMDG, exhibits a substantial performance improvement over several non-knowledge infused uni-modal medical dialogue generation models across a variety of evaluation metrics, including human evaluation. Additionally, the DII model surpasses existing pre-trained image models in both symptom identification (by 7.84%) and severity recognition (by 2.63%).

ES-MMD Corpus

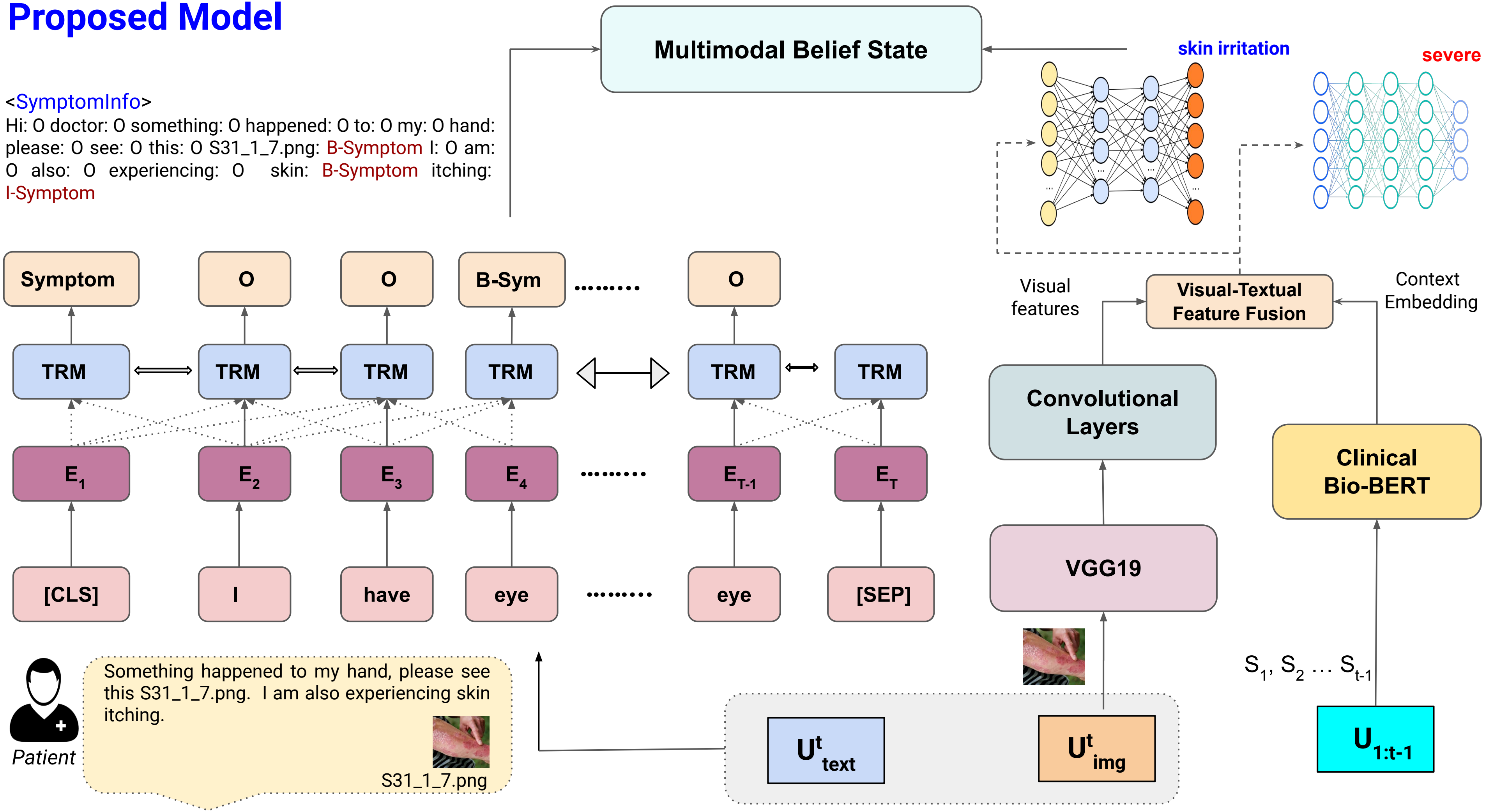
Dataset	Language	Conversation	Intent	Symptom	Multimodality	Severity
RD (Wei et al., 2018)	Chinese	×	×	×	×	×
DX (Xu et al., 2019)	Chinese	✓	×	✓	✓	×
M^2 - MedDialogue (Yan et al., 2021)	Chinese	✓	×	✓	×	×
MedDialog-EN (Zeng et al., 2020)	English	×	×	×	×	×
MedDG (Liu et al., 2020)	Chinese	✓	×	✓	×	×
SD (Zhong et al., 2022)	English	×	×	×	×	×
ES-MMD (ours)	English	✓	✓	✓	✓	✓

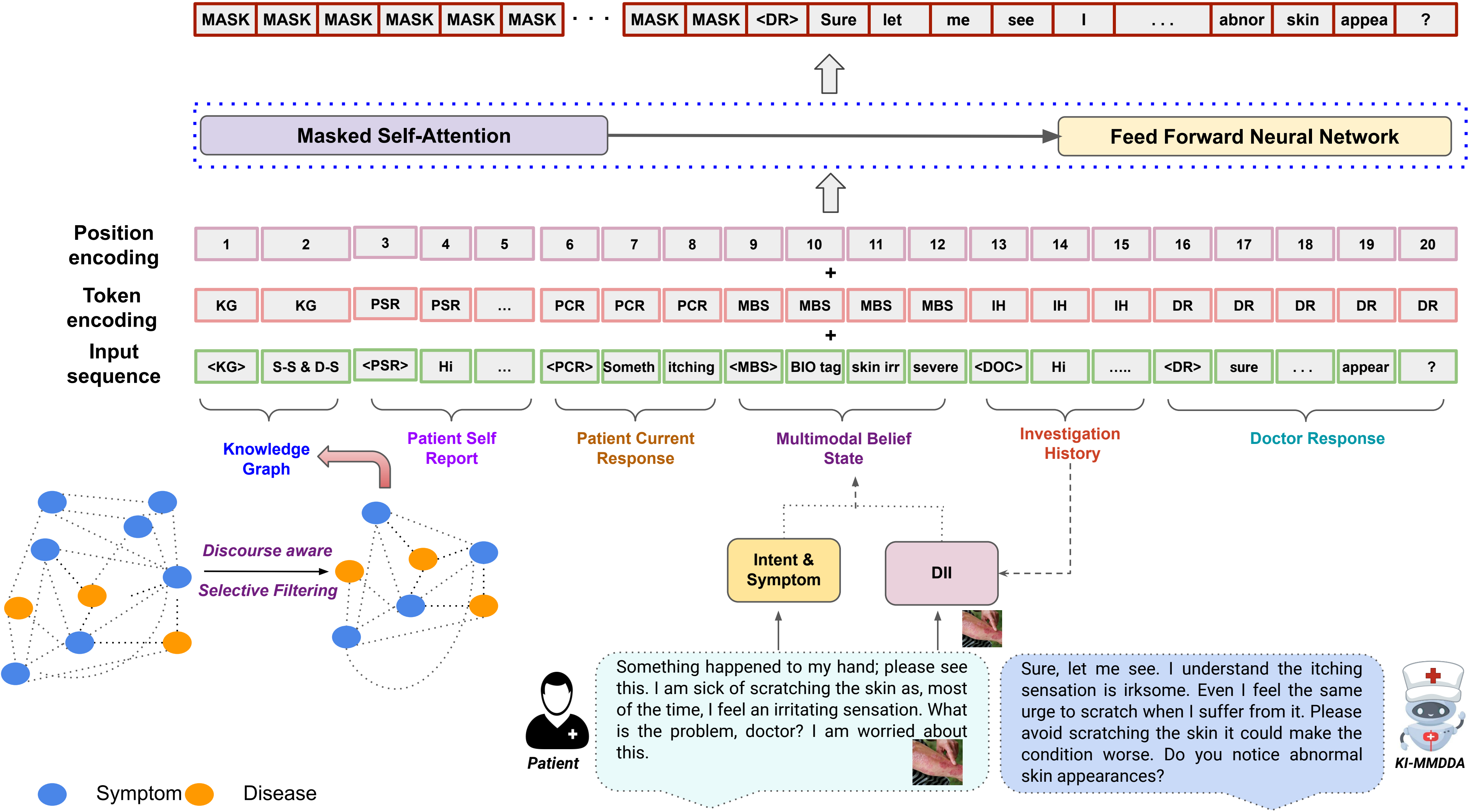
Attribute	Value
# of Dialogues	1742
# of Utterances	12466
Avg. dialogue length	7.16
# of intents	3
# of diseases	90
# of symptoms	266
# of signs	17
# of symptom images	1805



A conversation from the curated ES-MMD corpus

Proposed Model





MASK MASK MASK MASK MASK MASK · · · MASK MASK <DR> Sure let me see I . . . abnor skin appea ?
Masked Self-Attention
Feed Forward Neural Network
Position encoding
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
+
Token encoding
KG KG PSR PSR ... PCR PCR PCR MBS MBS MBS MBS IH IH IH DR DR DR DR DR
+
Input sequence
<KG> S-S & D-S <PSR> Hi ... <PCR> Someth itching <MBS> BIO tag skin irr severe <DOC> Hi ..... <DR> sure . . . appear ?
Knowledge Graph
Patient Self Report
Patient Current Response
Multimodal Belief State
Investigation History
Doctor Response
Discourse aware
Selective Filtering
Intent & Symptom
DII
Patient
Something happened to my hand; please see this. I am sick of scratching the skin as, most of the time, I feel an irritating sensation. What is the problem, doctor? I am worried about this.
Sure, let me see. I understand the itching sensation is irksome. Even I feel the same urge to scratch when I suffer from it. Please avoid scratching the skin it could make the condition worse. Do you notice abnormal skin appearances?
KI-MMDDA
Symptom
Disease

Experimental Setup

Algorithm 1 Discourse-aware Selective Filtering (DSF)

Initialization: KG = $\{(s_i, s_j, a_{ij})\}$ where s_i, s_j are nodes, a_{ij} is the edge weight.
Input: Current Knowledge Graph (KG_t), PSR: Patient Self Report, and Current Discourse (C_t)
Output: Filtered Knowledge Graph (KG_{t+1})

$C_t = \{(s_0), (s_1), (DII(i_2), s_2)...(s_t)\}$
$KG_{t+1} = KG_t$
Potential_diseases (PD) = $\Pi_{i=1}^{i=3}$ i^{th} most_associated_disease (PSR)
for d in PD **do**
  triplet = (PSR, d, a_{PSR-d}) $\Rightarrow$ a_{PSR-d}: edge (PSR, d) weight (Equation 3)
  KG_{t+1} = append(KG_{t+1}, triplet)
end for
for s in C_t[-1] **do**
  ps = $\Pi_{j=1}^{j=3}$ j^{th}-most_associated_symptom(s)
  for k in ps **do**
    triplet = (s, k, a_{s-k})
    KG_{t+1} = append(KG_{t+1}, triplet)
  end for
end for
return KG_{t+1}

- Training: (i) NLU and (ii) Dialogue Generation Module
- Train and Test split: 80% and 20%
- Total no. of epoch: 10, RTX 2080 Ti GPU, time: 2-3 hrs
- Loss func : Cross entropy
- Batch size: 4, optimizer : Adam, learning rate (6.25e-05)

Dataset and Code

Results

Model	Accuracy (%)	F1-Score
CNN (Li et al., 2014)	40.99	0.4247
Inception v3 (Xia et al., 2017)	66.14	0.6475
Inception v3 + Conv Layers	72.29	0.7163
DenseNet121 (Huang et al., 2017)	68.17	0.6712
DenseNet121 + Conv Layers	75.58	0.7412
DenseNet169 (Serte et al., 2022)	72.27	0.7157
DenseNet169 + Conv Layers	78.51	0.7734
VGG19 + Conv Layers (Gupta et al., 2022)	81.11	0.7924
DII with CW=1	82.82	0.8271
DII with CW=2	85.27	0.8505
DII with CW=3	**88.95** (**7.84** ↑)	**0.8703** (**0.0779** ↑)
DII with CW=4	81.59	0.8171

Performance of different models for visual signs identification

Model	Accuracy (%)	F1-Score
VGG19+ Conv Layers (Gupta et al., 2022)	50.65	0.5086
DII with CW=1	49.34	0.4907
DII with CW=2	52.63	0.5233
DII with CW=3	**53.28** (**2.63** ↑)	**0.5117** (**0.0091** ↑)
DII with CW=4	50.01	0.4771

DII with varying dialogue context window for severity recognition

Results

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4	BLEU	ROUGE 1	ROUGE L	METEOR
DLGNet (Oluwatobi and Mueller, 2020)	21.84	9.66	4.21	2.10	9.45	26.86	25.05	21.98
DLGNet with VSI-KG	**25.48 (3.64↑)**	**12.26 (2.60↑)**	**6.86 (2.65↑)**	**3.51 (1.41↑)**	**12.02 (2.57↑)**	**29.45(2.59↑)**	**28.82(2.57↑)**	**25.79(3.77↑)**
BART (Lewis et al., 2020)	23.19	12.34	7.32	4.37	11.80	27.77	27.37	29.66
BART with VSI-KG	**25.69(2.50↑)**	**15.07(2.73↑)**	**9.41(2.09↑)**	**5.62(1.25↑)**	**13.95(2.15↑)**	**29.93(2.23↑)**	**29.58(2.21↑)**	**32.41(2.75↑)**
DialoGPT (Zhang et al., 2020)	26.59	16.16	10.52	6.92	15.05	30.63	30.22	34.07
KI-MMDG	**28.53**(1.94↑)	**18.41**(2.25↑)	**12.28**(1.76↑)	**8.34**(1.42↑)	**16.89**(1.84↑)	**32.69**(2.06↑)	**32.25**(2.03↑)	**36.52**(2.45↑)

Performance of different baselines and proposed models incorporated with the visual sign and knowledge (VS-KG) guided disease diagnosis component

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4	BLEU	ROUGE 1	ROUGE L	METEOR
DLGNet with only KG	23.54	8.65	4.85	0.99	9.51	29.18	26.01	22.78
DLGNet with only VS	23.59	10.65	5.81	2.44	10.62	28.31	25.45	26.69
DLGNet with only VSI	24.25	10.77	5.91	3.40	11.08	29.37	27.72	25.42
BART with only KG	25.03	14.20	9.30	5.91	13.61	29.04	28.60	31.11
BART with only VS	25.37	14.85	9.39	5.63	13.81	29.58	29.12	32.35
BART with only VSI	25.69	14.98	9.65	5.93	14.06	29.92	29.43	31.88
DialoGPT with complete KG(w/o DSF)	1.58	0.73	0.48	0.28	0.77	1.99	1.95	2.20
DialoGPT with only KG	27.72	17.27	11.39	7.70	16.02	31.74	31.21	35.42
DialoGPT with only VS	27.53	17.08	11.47	7.82	15.98	31.54	31.13	35.29
DialoGPT with only VSI	27.11	16.94	11.45	7.74	15.81	31.23	30.91	35.07

Ablation study– performances of the proposed KI-MMDG model with different components. Here, KG, V S, V SI, and DSF refer to the knowledge graph, visual symptom, visual & severity information, and discourse-aware filtering

Key Findings

- **Dialogue context** significantly matters in sign image identification surfaced during dialogue

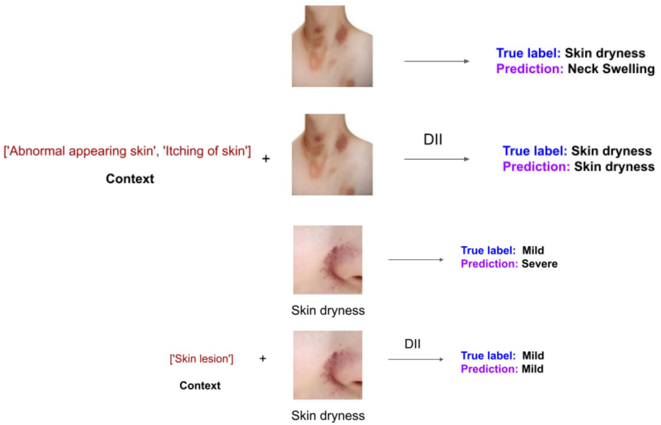
- We observed that the infusion of the symptoms and diseases relation significantly **narrows the exploration space** for the agent, resulting in context-based symptom inspection and disease diagnosis.

- While knowledge is crucial, **focusing on relevant knowledge** is more significant while solving a task

Analysis

Context	Gold	DialoGPT	DialoGPT+VSI	DialoGPT+KG	*KI-MMDG*
Hello doctor, I am suffering from Lower body pain. And because of it, I cannot sleep properly	I understand; it must be hard. Do you also have back pain?	Please be patient; it can be treated. Do you feel headache?	Please let me investigate. Do you feel headache ?	I understand that the pain would be hard. Do you also feel back pain?	I can understand how hard it is for you to sleep with this pain. Let me check, do you have back pain?
I am having weight gain. I feel like I look ugly because of it. ... Do you feel fatigue? yes	You don't have to think like that. You will be fine. You seem to have Cushing syndrome	Weight gain is a common problem these days. You will be fine if you take care of it. Do you have aches all over?	Weight gain is a common problem these days and can be completely curable as well. Are you suffering from Acne or pimples?	I can understand your concern. Weight gain can be due to various reasons. Let me ask you a few questions to get a better idea of its cause. Do you have skin lesion?	You are not alone. Weight gain is a common problem these days and can be completely curable as well. Based on your symptoms, you have Cushing syndrome.
Hello doctor, can you please check my eye? I am not sure what is wrong with it [moderate Swollen eye].	Is there any neck swelling?	And do you have any skin swelling? Please show.	Do you have any skin rash?	Do you have any skin rash? Please show me	Do you have neck swelling? Please show your neck.

Some generated responses by different models for some common test cases



- Knowledge infusion has not only improved symptom inspection relevancy but also significantly influenced the model's ability to generate appropriate and empathetic responses because of the in-depth understanding of user concerns and potential diseases.
- Dialogue context significantly matters in sign image identification surfaced during dialogue
- The proposed model suffers from sequence agnostic symptom investigation

Conclusion

- In this paper, we proposed a knowledge-infused, multi-modal medical dialogue generation (KI-MMDG) framework

- We also introduced a discourse-aware image identification (DII) model that exploits dialogue context to identify an image and its severity effectively

- The obtained improvements and detailed ablation study firmly establish the efficacy of (a) visual signs, (b) discourse-aware selective filtering (DSF) for knowledge infusion, and (c) discourse information for identifying an image surface during the conversation

सत्यमेव जयते
Ministry of Education
Government of India

LREC-COLING 2024

LREC-COLING
2024

Thank You !

Code

@abhisektiwari