MedGraph+: Empowering Medical Learning through Interactive Knowledge Graphs

Pengcheng Jiang[1], Megan Amber Lim[2], Adam Cross[3], Jimeng Sun[1,4]

[1]Department of Computer Science, University of Illinois Urbana-Champaign, Illinois, United States

[2]Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Illinois, United States

[3]University of Illinois College of Medicine Peoria, Illinois, United States

[4]Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, Illinois, United States

Corresponding Author:

Jimeng Sun, PhD

Department of Computer Science and Carle Illinois College of Medicine

University of Illinois, Urbana-Champaign

201 North Goodwin Avenue Urbana,

Illinois, 61801, United States

Email: jimeng@illinois.edu

Abstract

Objective: Medical education entails the daunting task of comprehending intricate medical textbooks, assimilating vast amounts of medical knowledge, and forming a cohesive mental knowledge graph encompassing all clinical concepts within a limited timeframe. To alleviate this challenge, we aim to create a medical reading assistant that harnesses the power of a large language model in conjunction with a medical ontology. The objective is to develop a tool that constructs and visually presents a clinical knowledge graph tailored to the specific context of the medical textbook being read. By doing so, learners can enhance their mental knowledge graph while actively engaging with the material. To gauge the effectiveness of our study assistant, we will conduct comprehensive user studies, carefully evaluating its impact on learning outcomes.

Materials and Methods: The developed tool, named MedKG, is an interactive software designed to generate knowledge graphs (KGs) from medical textbooks in PDF format, utilizing the advanced capabilities of large language models (LLMs). It offers a flexible and user-friendly interface that allows users to select specific pages or documents for KG generation and to customize prompts for LLMs to meet their individual needs. Additionally, MedKG includes features for users to interact with and modify the nodes and edges in the generated graphs, highlighting the importance of human-in-the-loop interaction. Our user studies involve medical students, assessing MedKG 's effectiveness in enhancing their study efficiency and comprehension of medical knowledge from various textbooks.

Results: The results indicate that MedKG successfully translates complex medical texts into clear, understandable knowledge graphs. This visualization aids medical students in grasping intricate medical concepts and theories more effectively.

Conclusions: The study concludes that MedKG, leveraging the synergy of large language models and user interaction, significantly aids medical information comprehension and retention. It presents a promising advancement in medical education tools, facilitating a deeper understanding of complex medical knowledge through innovative technology.

Keywords: Medical Education, Knowledge Graph, Large Language Models, Medical Ontology, Interactive Learning Tool, Human-in-the-Loop, Medical Textbook Comprehension.

Introduction

Medical education presents a unique challenge to students, requiring the assimilation of vast, diverse, and complex information within a constrained timeframe. This knowledge, drawn from various sources including textbooks, research articles, and lectures, is textually dense and conceptually intricate. Crucial to this educational process is the ability of students to form mental knowledge graphs. These graphs are conceptual frameworks that interlink various medical concepts, illustrating relationships vital for application in future clinical scenarios. However, the existing educational tools often fall short of aiding students to extract and organize this information into coherent mental structures effectively.

Artificial Intelligence (AI) has opened new frontiers in various fields, including medical education. Specifically, the emergence of Large Language Models (LLMs) like ChatGPT has introduced novel possibilities in processing and interpreting extensive textual data. These models can potentially deconstruct and synthesize complex medical texts, offering significant advantages in educational contexts. The ability of LLMs to navigate and interpret dense academic material can be pivotal in transforming the traditional methods of medical learning.

In light of these developments, we introduced MedKG, a tool that integrates the capabilities of ChatGPT in constructing knowledge graphs from medical educational resources. MedKG was designed to analyze various text-based medical resources, including pathology and pharmacology textbooks, USMLE board-style textbooks, research articles, and clinical vignettes. This approach aimed to encompass the broad spectrum of materials used in medical training.

Our research primarily involved first and second-year medical students at an allopathic medical school with an organ-system-based curriculum. The objective of the study was to evaluate the effectiveness of MedKG in enhancing the educational experience of these students. We aimed to obtain both qualitative and quantitative data reflecting MedKG's performance, focusing on its usefulness in facilitating the comprehension and retention of medical information. The ultimate goal of this project was to develop a significant and innovative tool for medical education, aligning with the evolving needs and challenges of modern medical training.

Background and Significance

The success of Large Language Models (LLMs) provides great opportunities for many domain specific applications, including medical education. This section underscores the pivotal role of LLMs, emphasizing their influence through key studies and developments. LLMs' foundational contributions, including dialogue agents and conversational models, are instrumental in fostering interactive learning in medical education [1]. Their advanced natural language processing abilities have proven effective in facilitating more engaging interactions between learners and complex medical texts.

In medical education, LLMs present numerous opportunities and encounter unique challenges. These models excel at distilling intricate medical information into more accessible formats, aiding students in navigating the dense and voluminous medical literature [2]. Such capabilities are crucial in educational settings, where students often grapple with the enormity and intricacy of their study materials. LLMs can also play a significant role in enhancing patient communication, especially in sensitive areas like addiction or sexually transmitted diseases [3]. LLMs can aid healthcare professionals in improving patient interactions, an aspect crucial for effective care. In digital health applications, LLMs have also shown promise in engaging and educating patients despite current technological constraints. Regarding clinical knowledge encoding, LLMs have demonstrated remarkable proficiency in comprehending and interpreting medical data [4], an essential feature for medical students who must master a broad spectrum of clinical concepts for practical application. Moreover, ethical considerations, particularly in data privacy and domain adaptation, are critical when integrating LLMs into healthcare settings [5]. This underscores the importance of careful deliberation in deploying AI in sensitive domains such as healthcare. Exploring LLMs in medical examinations, including their assessment capabilities in tests like the United States Medical Licensing Examination (USMLE), provides valuable insights into their role in medical education and knowledge evaluation [6, 7]. LLMs have also been proposed for creating simulated patient scenarios and didactic assessments, illustrating their adaptability in enriching traditional medical education methods [8].

Regarding the use of knowledge graphs (KGs) in education [9], previous studies [10, 11, 12, 13] have endeavored to create complex pipelines involving numerous processing steps such as named entity recognition (NER) [14], entity linking [15], and relation extraction (RE) [16]. These approaches, while comprehensive, presented challenges in practical educational system deployment. However, the emergence of LLMs and their superior natural language processing capabilities [17, 18, 19, 20], including advanced NER [20, 21] and RE [22, 23], has obviated the need for traditional, intricate KG construction methods. The use of prompting methods [24, 25, 26], a well-researched approach for querying LLMs to address specific problems, has recently been shown to effectively generate KGs directly from text [27] or even from the hidden parameters of LLMs themselves [28]. Thus, using LLMs for KG generation marks a significant leap forward in medical education. The ability of our developed MedKG to process, simplify, and interpret complex medical information, combined with its interactive capabilities, makes it an invaluable asset in the educational toolkit.

Materials and Methods

Generating knowledge graphs from text with large language models

In this study, we present MedKG, a software platform that leverages the power of large language models (LLMs) for knowledge graph (KG) generation from user-provided texts, catering to the demands of data-driven research. An illustration of it is shown in Figure 1.

MedKG comprises four integral components: (1) the document manipulation panel enables the upload and handling of multiple documents; (2) the prompt designing panel displays a textbox for the user-defined prompts that facilitate KG generation; (3) the text area panel where users input the target text for KG construction; and (4) the KG panel, which not only generates the KG but also allows for its refinement.
We provide a review of MedKG, outlining its operational logic in a manner that mirrors the user's journey: beginning with document upload and culminating in the generation of knowledge graphs. This sequential description aims to offer a clear and practical understanding of the platform's functionality.

Step 1: Uploading Documents and Defining Target Context

MedKG offers a robust feature for uploading multiple documents, thereby allowing for the creation of knowledge graphs that amalgamate varied content across different pages and documents. Users, post-upload, have the flexibility to choose specific pages to be transformed into knowledge graphs. The chosen content can be transferred to the text area panel by either directly copying text (ideal for predominantly text-based documents) or using an optical character recognition (OCR) tool for pages with images or tables. The text area panel acts as the source for generating the knowledge graph. This area is dynamic and user-editable, providing the option to modify the content by adding or subtracting information as necessary.

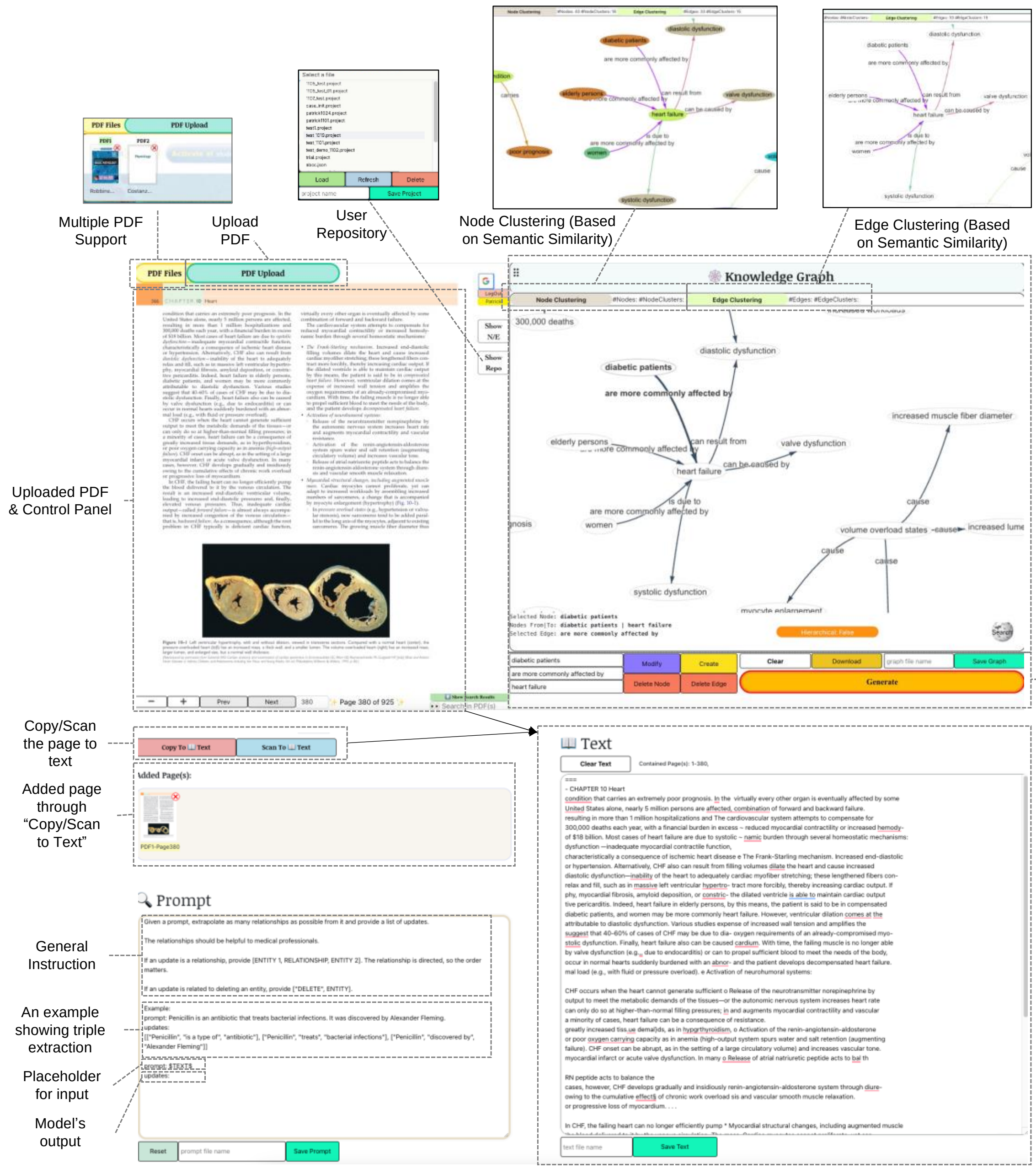


Figure 1. The overview of our developed MedKG software. We use dotted grey boxes and text explainable to illustrate each component of the framework. We use "Robbins basic pathology 9th edition – page 366, Lippincott pharmacology – diuretics" as the input in this example.

Step 2: Employing LLM for Knowledge Graph Triple Extraction

The process progresses by using a prompting technique to guide the LLM in generating a list of knowledge graph "triples." Each triple, formatted as [ENTITY1, RELATIONSHIP, ENTITY2], captures a relationship between two entities. The default template for this operation, illustrated in Figure 1, includes (1) general instructions defining the triple extraction task; (2) an example demonstrating the task execution; (3) a placeholder for the source text (added in the previous step) from which KG triples are extracted; and (4) a command prefix (“update”) signaling the LLM to commence triple extraction. We use the model “*gpt-3.5-turbo-instruct*” [29] as the LLM for this extraction process. To trigger this extraction, the user only needs to click the “Generate” button in the KG control panel.

Step 3: Post-processing and knowledge graph visualization

Upon receiving the response from the Large Language Model (LLM), we employ a regular expression technique to meticulously extract the JSON string embedded within. This string is then parsed, enabling us to concisely summarize the output. During this summarization, we focus on distilling the information specifically pertaining to nodes and edges. A representative example of this process and its resulting output is showcased in Figure 2.


```
"nodes": [
{
"id": "heart failure",
"label": "heart failure”
},
{
"id": "systolic dysfunction",
"label": "systolic dysfunction"
},
{
"id": "diastolic dysfunction",
"label": "diastolic dysfunction"
},
…]
```

```
"edges": [
{
"from": "heart failure",
"to": "systolic dysfunction",
"label": "is due to"
},
{
"from": "heart failure",
"to": "diastolic dysfunction",
"label": "can result from"
},
{
"from": "elderly persons",
"to": "heart failure",
"label": "are more commonly affected by"
},
…]
```


Figure 2. An example of a JSON file of the KG parsed from LLM’s output.

We then use “*react-graph-vis*” package [30] to visualize this data as an interactive KG, shown in the demo we present in Figure 1.

Interactive features of the MedKG

The control panel in MedKG offers users a dynamic interface to interact with the generated knowledge graphs (KGs). This interaction includes adding, modifying, or deleting nodes and edges within the graph. To add a node or edge, users are required to input relevant information into three designated textboxes. These boxes are arranged vertically and correspond to the “source node,” “edge,” and “target node,” respectively. Once the information is entered, users can click the “Create” button to add the new element to the graph. Modifying a node or edge is also streamlined. Users begin by selecting the desired node or edge on the graph. This action automatically populates the text boxes with the existing information, which users can edit as needed. To finalize these changes, clicking the “Modify” button updates the graph with the revised data. For deletion, users select the node or edge they wish to remove and then click either the “Delete Node” or “Delete Edge” button, depending on their intent. This action promptly removes the selected element from the graph. Additionally, MedKG provides the functionality to download the graph in a JSON file format, as illustrated in Figure 2. This feature is particularly useful for users who prefer or require programmatic editing of the graph.

Other features: node/edge clustering, user repository

To enhance the visualization of the generated Knowledge Graph (KG), MedKG includes an optional node/edge clustering feature. Users can activate this function by clicking on two dedicated buttons at the top of the KG panel. Upon activation, MedKG’s backend initiates a request to an embedding service [31], specifically utilizing the 'test-embedding-ada-002' model. This request aims to acquire embedding vectors for all nodes and edges within the KG. Subsequently, an *agglomerative clustering* algorithm [32] is applied, which utilizes cosine similarity between embedding vectors to group the elements. This process results in nodes and edges with similar semantic meanings clustered together, indicated by matching colors. This process of embedding retrieval and clustering is swift. For instance, in the case where the graph contains around 100 nodes (a considerably large graph for human cognition), the entire operation is completed in approximately 1.5 seconds.

Moreover, MedKG features a user repository function, leveraging Google's Firebase [33] to securely store users' data. This functionality allows users to save various elements, including PDF files, user-specific prompts, extracted texts, and generated Knowledge Graphs (KGs). Users can save these elements individually or collectively as part of a project using the "Save" buttons provided on the interface. Users wishing to revisit their previous work can easily do so by selecting the desired file and clicking the "Load" button. This feature efficiently reloads the selected data, significantly streamlining the process and saving users the effort of recreating previous progress. Additionally, MedKG has more functions, such as within-pdf and within-graph searching, wiki-searching, and hierarchical visualization.

Conducting user studies with first- and second-year medical students

To gather feedback and assess the efficacy of the tool on medical education, first and second year medical students at Carle Illinois College of Medicine were recruited. The first version of the tool was released in the first week, and based on the feedback gathered, updates were added to the tool and released as a newer version for additional feedback. The third stage of user studies is currently in progress and entails each student taking two separate quizzes on different topics within medicine. For one quiz they will have access to a PDF with associated literature and for the other a PDF as well as a corresponding KG generated with the tool.

Results

Sources of text

After exploring a variety of learning materials utilized by medical students in their pre-clinical education, we discovered that the most helpful graphs were generated from research articles and textbooks rather than United States Medical Licensing Examination (USMLE) board preparation resources. Denser sources of text, such as Robbins Basic Pathology and Costanzo Physiology, offered a richer basis of knowledge from which relations could be extracted. This is opposed to First Aid, a textbook with high-yield content that medical students use frequently in preparation for their USMLE exams, which contains knowledge already synthesized in the forms of tables and charts.

Figures 3 and 4 below are knowledge graphs generated separately from the First Aid 2020 edition and Robbins Basic Pathology 9th edition with their corresponding sections on the topic of heart failure. As seen in Figure 3, the tool generated separate graphs corresponding to each fragment of information provided on page 309 of the First Aid textbook. Many of the graphs consist of no more than two edges, which can largely be attributed to the "bullet-point" structure of text in First Aid. For example, the text "ACE inhibitors or angiotensin II receptor blockers and spironolactone decrease mortality" was correctly converted into a graph with the nodes "ACE inhibitors or angiotensin II receptor blockers" and "spironolactone" correctly pointing via separate edges of "decrease" to the common node of "mortality". Interestingly, we note that the LLM could recognize the symbol "↓" in "*spironolactone ↓ mortality*" and extracted the relationship "<*spironolactone, can reduce, mortality*>".

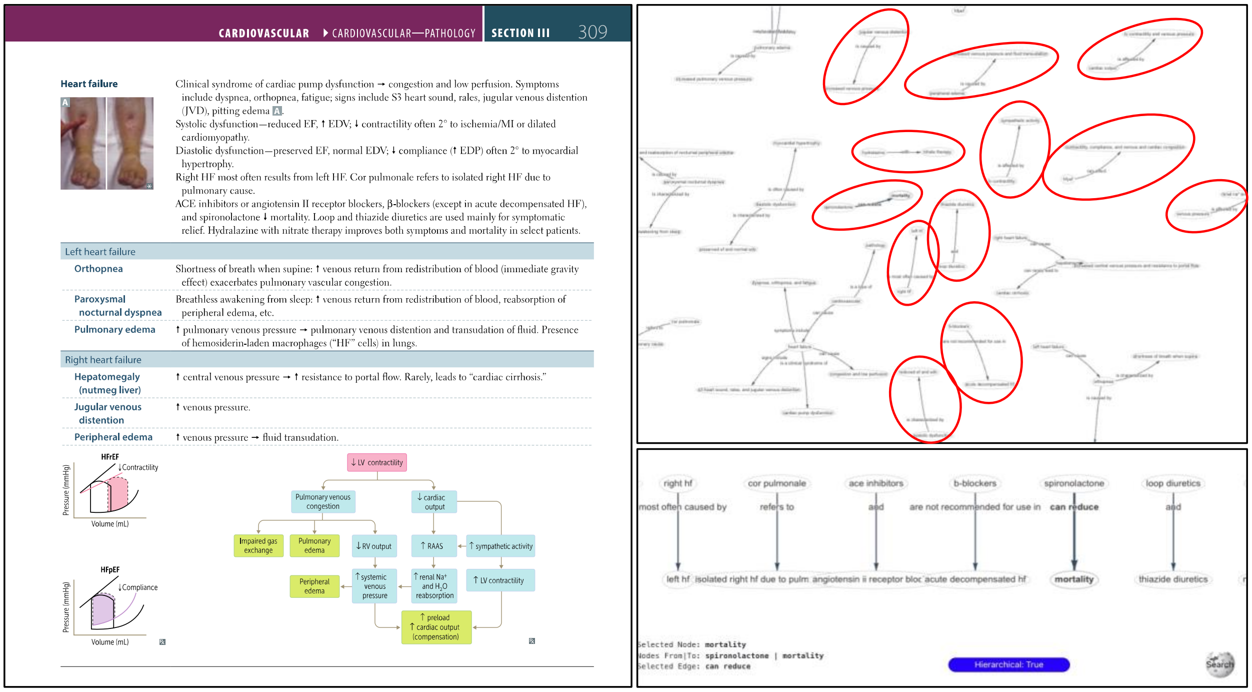


Figure 3. Knowledge graph generated from First Aid (Page 309) – Congestive Heart Failure.

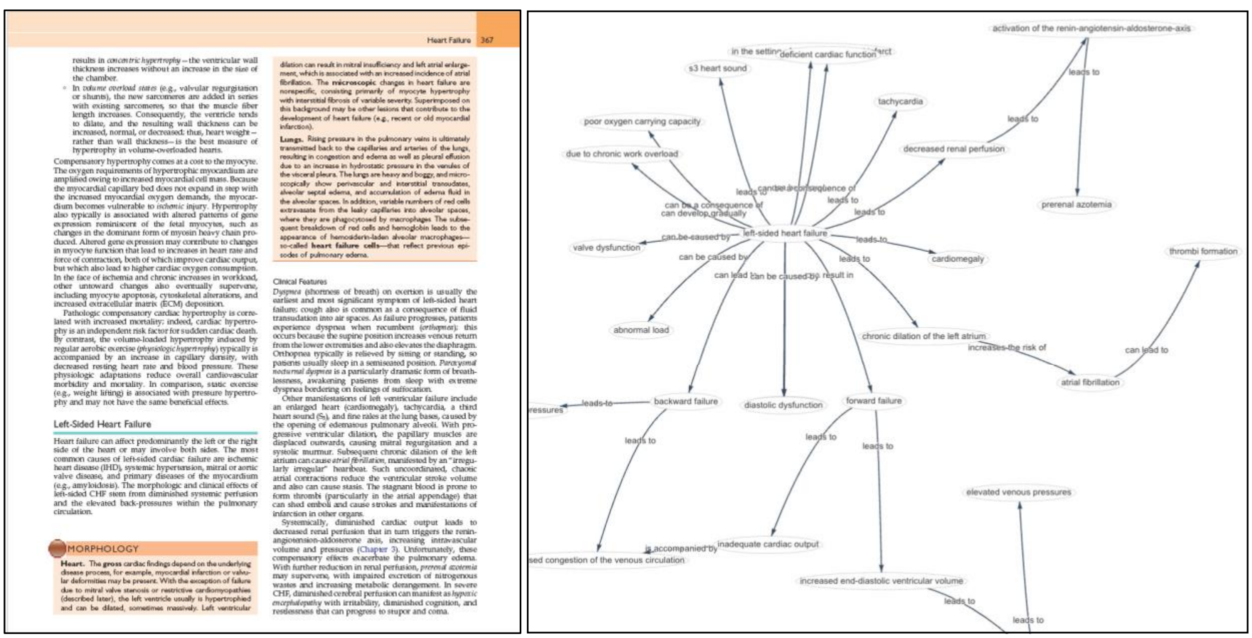


Figure 4. Knowledge graph generated from Robbins Basic Pathology (Page 366 to 368). Page 367 shown.

In comparison, the graphs based on the heart failure section in Robbins Basic Pathology (Figure 4) consist of more connections with shared clinical correlation between the central and neighboring nodes. One graph clearly illustrates the common effects of left-sided heart failure and the downstream consequences if any of them were to evolve. With "left-sided heart failure" serving as the central node, the peripheral nodes of tachycardia, cardiomegaly, and S3 heart sound indicate some of the example consequences of this clinical indication.

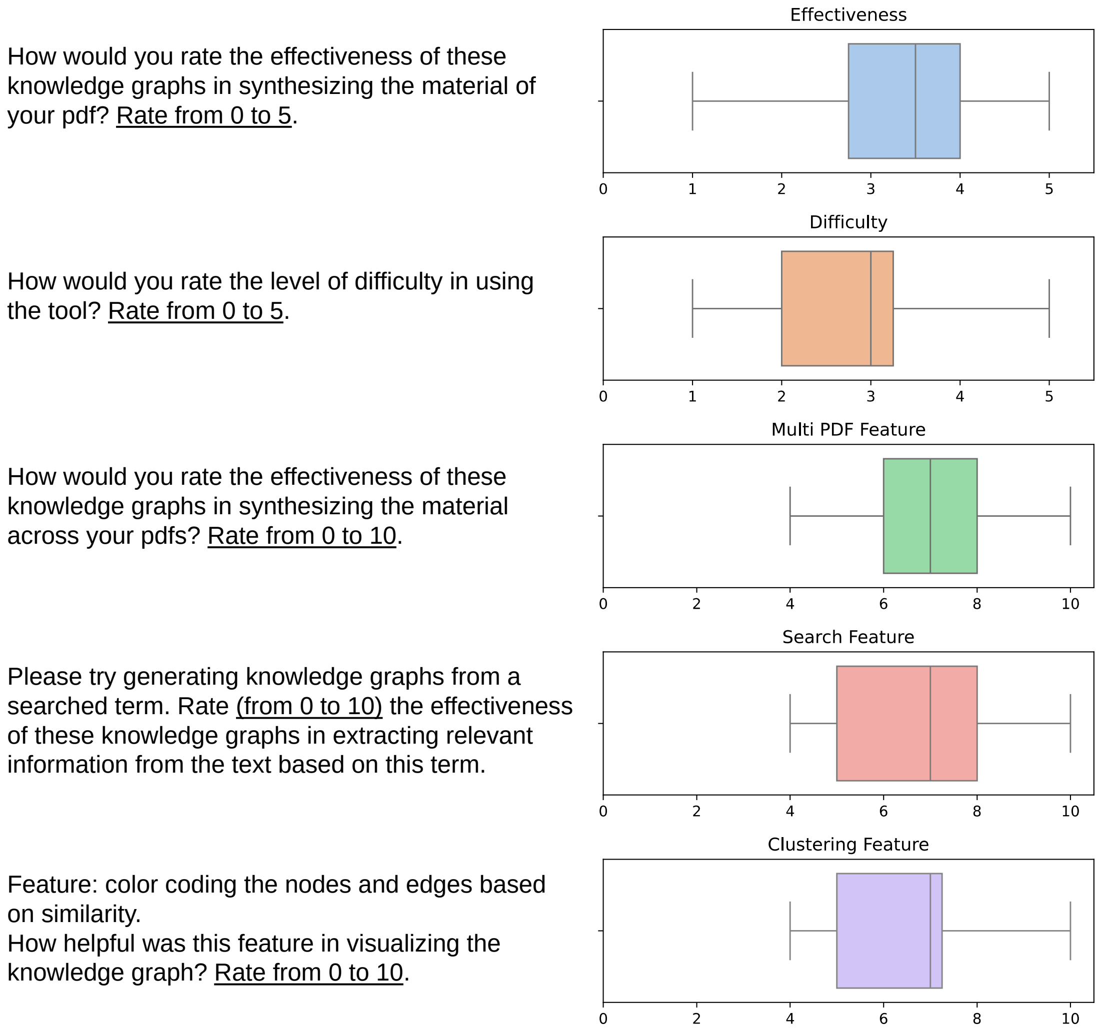


Figure 5. User studies of MedKG.

Specifically, the node of "decreased renal perfusion" leads to a series of nodes and edges that indicate activation of the renin-angiotensin-aldosterone-axis and consequence of prerenal azotemia. This correctly highlights the downstream effects of lack of blood flow to the kidneys and how, as a result, they can influence the current state the patient is in by exacerbating the fluid buildup in the lungs. The path branching from the node, "chronic dilation of the left atrium" provides another example that guides the learner through the clinical reasoning of how volume-overloaded states increase the risk of atrial fibrillation and the potential downstream consequence of deadly thrombi formation. The thought process depicted in these knowledge graphs is especially empowering for students to deploy when solving USMLE questions that probe for answers one step removed from the immediate factoid provided in First Aid.

User studies

After launching the tool to first and second year medical students at Carle Illinois College of Medicine, there are currently 63 active users. The most used resources that were uploaded to the site were research articles, First Aid, Robbins and Cotran, Pathoma, Sketchy, and Costanzo and Katsung. The feedback from the first week of release highlighted the application's ability to connect concepts together and pull relevant information from the text. Recommendations for improvement, which were added in the subsequent week's updated version of the tool, included a search feature, color coding of common nodes and edges, and the ability to generate graphs from more than one source simultaneously. The user feedback results, presented in Figure 5, reflect the practical application and effectiveness of the knowledge graph tool among first and second-year medical students at the Carle Illinois College of Medicine. In addition to the questions shown up in Figure 5, we further asked some questions with binary (*yes/no*) answers, which are discussed as follows. When students were asked, "*Did these knowledge graphs help you gain a better understanding of the topic?",* a significant majority of 71.4% responded with "yes". This suggests that the tool played a beneficial role in enhancing the comprehension of medical topics. When posed with the question, "*Did the tool help you recognize connections between concepts that you hadn't noticed prior?*", more than half of the students, 52.3%, answered "yes". This response indicates that the knowledge graphs effectively revealed new insights and relationships within the material. However, 23.4% of the students were "unsure", indicating a potential area for further refinement of the tool to assist in illuminating connections more clearly for all users.

These findings underscore the tool's potential as an innovative educational aid, capable of supporting the synthesis and visualization of complex information, thereby enriching the learning process for medical students.

Discussion

In this study, we introduce a novel tool that can supplement the medical education process...

A feature that was indicated by the students to be helpful was the search function of specific concepts because it allowed them generate graphs that extracted and formulated connections from various parts of the text.

Support multiple sources/types of pdfs – various text sources

More time saved generating graphs for dense text such as robbins, cotran versus the usmle board sources sucha as first aid

We envision that this application can be integrated within the learning process as a method for students to cross validate their understanding of the material with the generated knowledge graphs. Many medical schools are transitioning to Problem-Based Learning (PBL) [34] which is centered on discussing clinical cases corresponding to the learning objectives in the curriculum. This tool can be incorporated into the PBL environment by equipping students with knowledge graphs generated based on the clinical case of discussion for that week. By practicing explaining the connections that are illustrated with other students, the students can engage in active learning and gain exposure to a different perspective of the material.

Limitations and future work:

There is the possibility that the tool is overlooking some of the information in the text. Our goal is for the knowledge graphs to supplement the learning process of the medical student, so while they may not be completely comprehensive, they highlight core concepts most important for strengthening clinical reasoning.

MedKG, like other LLM-based tools, is susceptible to the common risk of generating hallucinations or non-factual information that may not accurately reflect the source text [35, 36, 37]. To address this issue and reduce the potential negative impact of such inaccuracies, we have implemented a specific mitigation strategy. This involves adding a clear instruction in the user prompt: "*If an update is related to deleting an entity, provide ["DELETE", ENTITY]*." This directive is designed to facilitate the removal of any incorrect or unreal relationships from the KG, thereby enhancing the accuracy and reliability of the generated content.

Since our work is a new educational tool, there is also the initial learning curve that comes when first navigating the site. The initial week of launch to the medical students involved creating detailed tutorials as well as trouble shooting. Because medical students are often faced with a myriad of different educational resources when studying, adoption of a new tool into their study regimen may be a challenge. However, as the user base expands, a library of knowledge graphs shared between medical schools and students across the nation can be constructed. Since all students preparing for the USMLE board exams are expected to understand the same learning objectives, it would be helpful for a student to see and access the various knowledge graphs generated for a particular concept by others.

The tool also has transferability to the instructor domain for assessment of written assignments submitted by students. From the graphs generated from student work, the facilitator can evaluate whether their understanding meets the learning objectives in the curriculum.

Conclusion

In this study, we present MedKG, an innovative software platform that leverages the advanced text comprehension and relationship extraction capabilities of large language models to convert user-uploaded documents into comprehensive knowledge graphs. Our objective was to assess the impact of knowledge graphs, derived from medical textbooks, on the study efficiency of medical students. The user studies conducted demonstrate a consensus among students that MedKG effectively synthesizes useful knowledge graphs. Moreover, the students acknowledged that the knowledge graphs facilitated a deeper understanding of their study material and enabled them to discover previously unobserved concepts. The features unique to MedKG were particularly well-received, suggesting that the tool not only meets the educational requirements of the students but also enhances their learning experience. Looking forward, we plan to refine MedKG, focusing on enhancing its user interface and ensuring the stability of knowledge graph generation. Our goal is to provide a more intuitive and reliable resource that further supports the educational endeavors of medical students.

Code Availability Statement

The codes of the software are publicly available at https://github.com/pat-jj/TextbookKG
The developed software is publicly accessible at https://pat-jj.github.io/TextbookKG/

Competing Interest Statement

The authors declare that there are no competing interests.

Authors Statement

PJ developed the software. ML conducted user studies on medical students. ML and AC provided clinical guidance. PJ, ML, AC and JS participated in report writing. All authors declare that they have no conflicts of interest. All correspondence can be sent to jimeng.sun@gmail.com.

Acknowledgement

Inclusion & Ethics Statement

The user studies on medical students were approved by the University of Illinois Institute Review Board with the project title "MedGraph+: Empowering Medical Learning through Interactive Knowledge Graphs" and IRBNet protocol number 24488.

References

[1] Glaese, Amelia, et al. "Improving alignment of dialogue agents via targeted human judgements." arXiv preprint arXiv:2209.14375 (2022).

[2] Venerito, Vincenzo, et al. "AI am a rheumatologist: a practical primer to large language models for rheumatologists." Rheumatology (2023): kead291.

[3] Clusmann, Jan, et al. "The future landscape of large language models in medicine." Communications Medicine 3.1 (2023): 141.

[4] Singhal, Karan, et al. "Large language models encode clinical knowledge." arXiv preprint arXiv:2212.13138 (2022).

[5] Karabacak, Mert, and Konstantinos Margetis. "Embracing Large Language Models for Medical Applications: Opportunities and Challenges." Cureus 15.5 (2023).PLOS Digital Health.

[6] Kung, Tiffany H., et al. "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models." PLoS digital health 2.2 (2023): e0000198.

[7] Kasneci, Enkelejda, et al. "ChatGPT for good? On opportunities and challenges of large language models for education." Learning and individual differences 103 (2023): 102274.

[8] Safranek, Conrad W., et al. "The role of large language models in medical education: applications and implications." JMIR Medical Education 9 (2023): e50945.

[9] Rizun, Mariia. "Knowledge graph application in education: A literature review." Acta Universitatis Lodziensis. Folia Oeconomica 3.342 (2019): 7-19.

[10] Chen, Penghe, et al. "Knowedu: A system to construct knowledge graph for education." Ieee Access 6 (2018): 31553-31563.

[11] Sun, Kai, et al. "Visualization for knowledge graph based on education data." International Journal of Software and Informatics 10.3 (2016): 1-13.

[12] Chen, Penghe, et al. "An automatic knowledge graph construction system for K-12 education." Proceedings of the fifth annual ACM conference on learning at scale. 2018.

[13] Li, Nan, et al. "MEduKG: a deep-learning-based approach for multi-modal educational knowledge graph construction." Information 13.2 (2022): 91.

[14] Li, Jing, et al. "A survey on deep learning for named entity recognition." IEEE Transactions on Knowledge and Data Engineering 34.1 (2020): 50-70.

[15] Shen, Wei, Jianyong Wang, and Jiawei Han. "Entity linking with a knowledge base: Issues, techniques, and solutions." IEEE Transactions on Knowledge and Data Engineering 27.2 (2014): 443-460.

[16] Zhou, GuoDong, et al. "Exploring various knowledge in relation extraction." Proceedings of the 43rd annual meeting of the association for computational linguistics (acl'05). 2005.

[17] Wei, Jason, et al. "Emergent Abilities of Large Language Models." Transactions on Machine Learning Research (2022).

[18] Wang, Xiaoxuan, et al. "Scibench: Evaluating college-level scientific problem-solving abilities of large language models." arXiv preprint arXiv:2307.10635 (2023).

[19] Wang, Shuhe, et al. "Gpt-ner: Named entity recognition via large language models." arXiv preprint arXiv:2304.10428 (2023).

[20] Min, Bonan, et al. "Recent advances in natural language processing via large pre-trained language models: A survey." ACM Computing Surveys 56.2 (2023): 1-40.

[21] Vīksna, Rinalds, and Inguna Skadiņa. "Large language models for Latvian named entity recognition." Human Language Technologies–The Baltic Perspective. IOS Press, 2020. 62-69.

[22] Li, Guozheng, Peng Wang, and Wenjun Ke. "Revisiting Large Language Models as Zero-shot Relation Extractors." arXiv preprint arXiv:2310.05028 (2023).

[23] Wadhwa, Somin, Silvio Amir, and Byron C. Wallace. "Revisiting relation extraction in the era of large language models." arXiv preprint arXiv:2305.05003 (2023).

[24] Liu, Pengfei, et al. "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing." ACM Computing Surveys 55.9 (2023): 1-35.

[25] Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." Advances in Neural Information Processing Systems 35 (2022): 24824-24837.

[26] Zhang, Zhuosheng, et al. "Automatic chain of thought prompting in large language models." arXiv preprint arXiv:2210.03493 (2022).

[27] Christino, Leonardo, and Fernando V. Paulovich. "ChatKG: Visualizing Temporal Patterns as Knowledge Graph." (2023).

[28] Jiang, Pengcheng, et al. "GraphCare: Enhancing Healthcare Predictions with Open-World Personalized Knowledge Graphs." arXiv preprint arXiv:2305.12788 (2023).

[29] "Models." OpenAI, OpenAI, Inc., https://platform.openai.com/docs/models. Accessed Dec. 15th, 2023.

[30] Rubier, Clément. "react-graph-vis." GitHub, https://github.com/crubier/react-graph-vis. Accessed Dec. 15th, 2023.

[31] OpenAI. "New and Improved Embedding Model." OpenAI Blog, OpenAI, [insert the date of the blog post here], https://openai.com/blog/new-and-improved-embedding-model. Accessed Dec. 15th, 2023.

[32] Murtagh, Fionn, and Pierre Legendre. "Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?." Journal of classification 31 (2014): 274-295.

[33] Firebase." Google, https://firebase.google.com/. Accessed Dec. 15th, 2023.

[34] Forbes, Heather M., Munir S. Syed, and Octavia L. Flanagan. "The Role of Problem-Based Learning in Preparing Medical Students to Work As Community Service-Oriented Primary Care Physicians: A Systematic Literature Review." Cureus 15.9 (2023).

[35] Zhang, Yue, et al. "Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models." arXiv preprint arXiv:2309.01219 (2023).

[36] Ji, Ziwei, et al. "Survey of hallucination in natural language generation." ACM Computing Surveys 55.12 (2023): 1-38.

[37] Azamfirei, Razvan, Sapna R. Kudchadkar, and James Fackler. "Large language models and the perils of their hallucinations." Critical Care 27.1 (2023): 1-2.

Supplemental Materials

Organization of Supplemental materials:

IRB materials

====

Notes from user study:

- Todo: average the feedback statistics
- Third week: knowledge-based assessment
    - Each student serves as their self-control and takes 2 quizzes
    - First quiz - just the article
    - Second quiz – with knowledge graph and article

Solutions to problems students ran into their first week:

- trouble getting the site to work – was not logged into the email account that they signed up for the study with/granted permissions
- Site appears blank page when window is not fully expanded to the screen
- Do not use on carle/hospital wifi
- if It's taking too long to generate a graph – try clicking generate again. If it's taking longer than 3 minutes message me. You can click generate more than one times to add more nodes/relations to the graph. Try playing around with it
- Graphs are nondeterministic – that is the intention. the vision is that one day there will be library of graphs already made by other students, so you won't generate the graphs but rather find the ones you like. But since you guys are the first ppl to use it, that library does not exist yet

Week 2

- Couldn't generate: too far zoomed in or out – way the page automatically
- Help button
- Didn't there was stuff to the site at the bottom of the screen
- Toggle words on/off
- Specify range of pages, rather than each one
- fear the tool isn't picking up 100% of the relations in the text, missing some important information

- Chatgpt is non-domain specific, medicine is rigorous domain is highly specific training/clinical reasoning
- concern about accuracy of chatgpt knowledge base,
- upvote downvote system, create a library/suite of knowledge graphs where the best voted ones rise to the surface

For the clinical vignettes in which the next appropriate intervention was inquired

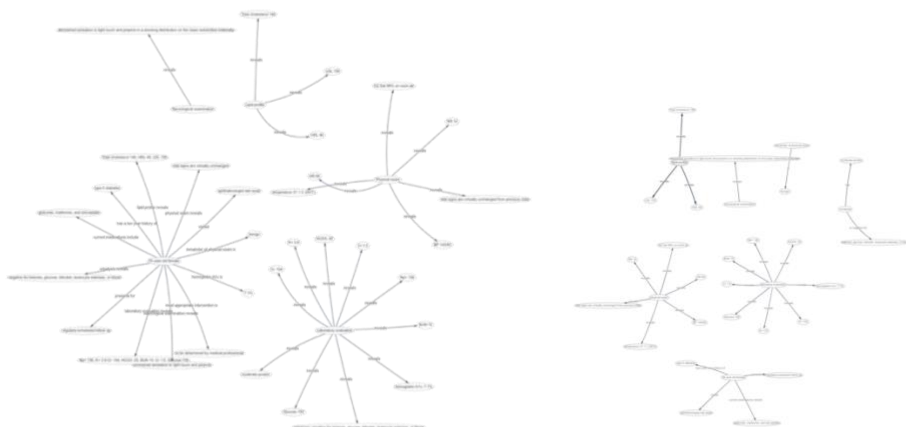
Next appropriate intervention – struggles to get correct answer, narrows down to options, but helps when feeds in additional text of guidelines – leverage multi pdf upload function njem example

q	graph
1	Add hydrochlorothiazide or lisinopril – narrowed down to two correct answers

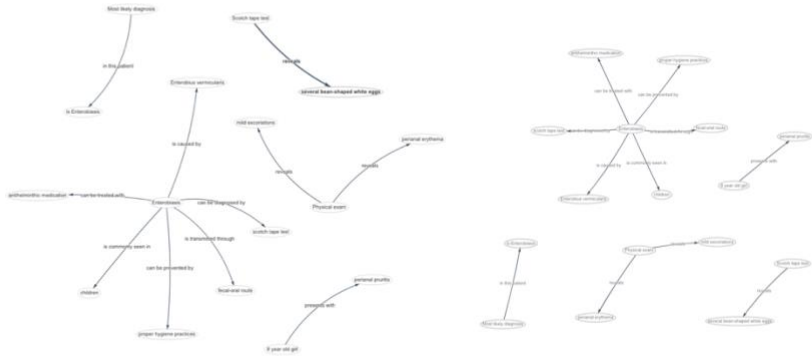
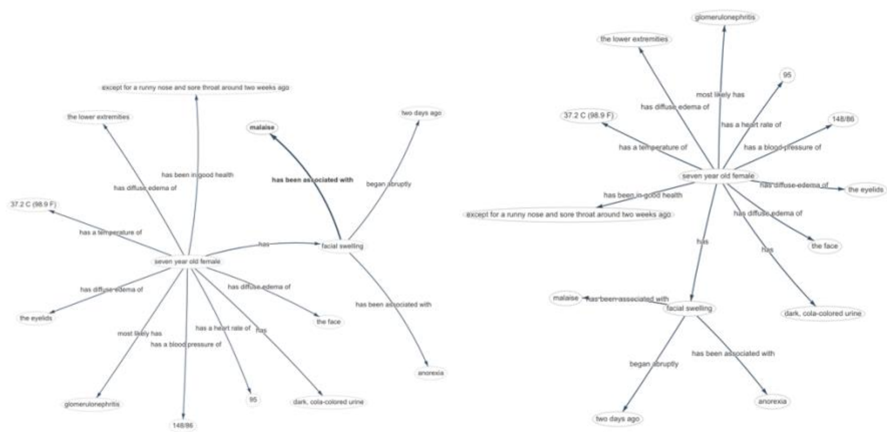
Clinical vignettes
We also investigated the tool's ability to synthesize content of standardized clinical vignettes from a Family Medicine Clerkship exam provided by the University of Virginia School of Medicine and provide clinical diagnoses illustrated in a knowledge graph form. Patient narratives consisting of chief complaint, past medical history, imaging and lab results along with the question "what is the most likely clinical diagnosis" without the multiple-choice answers were fed into the tool. It was important to indicate that the "most likely "diagnosis was requested because, without the specification, the generated knowledge graph would consist of all possible differential diagnoses including those not tailored to the patient case discussed.

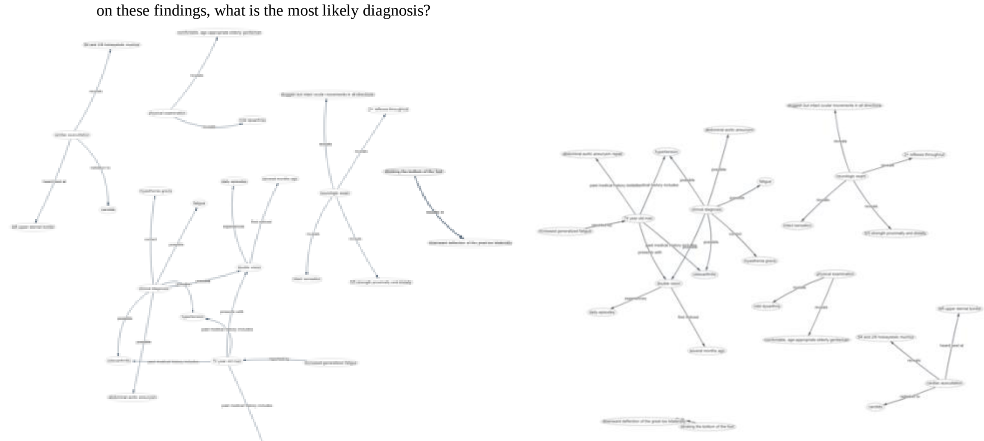
The tool was able to extract the relevant information from the question stem that was pertinent to making the correct clinical diagnosis. The restructuring of the paragraph text into the form of a graph organized the information in a visual way that highlighted aspects of the patient case not immediately evident upon reading the vignette. As seen in the figures below, components of the patient's history were distributed into separate graphs that correspond to different sections of the medical documentation notes physicians create in electronic health records. With central nodes being "laboratory evaluation", "physical exam", "lipid profile" and their corresponding results placed peripherally, it is visually easier to differentiate between normal and abnormal values.

Figure 5 illustrates graphs corresponding to a clinical vignette that makes the correct most likely diagnosis of enterobiasis, matching one of the choices provided in the original question, as well as additional information about the disease such as treatment, commonly affected demographics, and diagnostic tests. The graph of another clinical vignette correctly illustrated the most likely diagnosis of glomerulonephritis and provided other potential associations related to her facial edema. For the case of a man with ptosis and mild dysarthria, a graph depicted related clinical diagnoses as well as the indication matching the correct answer choice of myasthenia gravis.



- Fig D
  - Uva fam medicince lerkship; https://med.virginia.edu/family-medicine/education/student-programs/third-year-clerkships/
  - “A 50 year old female with a ten year history of type II diabetes presents for regularly-scheduled follow up. She has no complaints, and just visited her ophthalmologist last week. Current medications include glyburide, metformin, and simvastatin. On physical exam, vital signs are virtually unchanged from previous visits, with temperature 37.1 C (99 F), HR 80, BP 140/83, RR 15, and O2 Sat 98% on room air. Neurological examination reveals diminished sensation to light touch and pinprick in a stocking distribution on the lower extremities bilaterally. Remainder of physical exam is benign. Laboratory evaluation reveals: Na+ 136, K+ 3.9 Cl- 104, HCO3- 25, BUN 15, Cr 1.0, Glucose 150; hemoglobin A1c: 7.1%; Urinalysis: negative for ketones, glucose, bilirubin, leukocyte esterase, or blood; moderate protein; Lipid profile: Total cholesterol 146, HDL 46, LDL 100. At this time, which of the following would be the most appropriate intervention?”

-  Fig E
    - Q5: An otherwise healthy 8 year old girl presents with two weeks of perianal pruritis. She has two younger brothers, one of whom has had similar complaints for the past few days. Physical exam reveals perianal erythema with mild excoriations. The “scotch tape test” reveals several bean-shaped white eggs. What is the most likely diagnosis in this patient?
- 

    - Q9: g
    - A seven year old female is brought to the physician by her mother because of facial swelling and dark, cola-colored urine. These symptoms began abruptly two days ago and have been associated with anorexia and malaise. There have been no known sick contacts. Her mother states that the child is up to date with her immunizations and has been in good health except for a runny nose and sore throat around two weeks ago, which resolved after a few days without treatment. Vital signs are temperature 37.2 C (98.9 F), heart rate 95, and blood pressure of 148/86. There is diffuse edema of the lower extremities, face, and eyelids. Lungs and heart are clear to auscultation. Urinalysis shows moderate hematuria and proteinuria, and dysmorphic RBCs and occasional RBC casts are noted on microscopic examination. Based on these findings, what is the most likely diagnosis?
- 

Q14: A 74 year old man presents with double vision. He first noticed this several months ago, and although his symptoms wax and wane, he now experiences daily episodes of “seeing double,” most frequently in the evenings. He also reports increased generalized fatigue and notes that he sometimes gets so tired at dinner that he “can hardly chew” his food. Past medical history includes osteoarthritis, hypertension, and abdominal aortic aneurysm repair. Physical examination reveals a comfortable, age-appropriate elderly gentleman with mild dysarthria. Cardiac auscultation reveals both an S4 and a 2/6 holosystolic murmur heard best at the left upper sternal border with radiation to the carotids. On neurologic exam, the patient has 5/5 strength proximally and distally. Sensation is intact and reflexes are 2+ throughout. Ocular movements are sluggish but intact in all directions. The patient has mild bilateral ptosis, which is noted to increase with sustained upward gaze. Stroking the bottom of the foot results in downward deflection of the great toe bilaterally. What are the possible clinical diagnoses and which is the correct one?