Need to Increase Facetime Between Patients and Clinician

Harish Kumar K S[1],Y Amarnath Chowdhary[2], Charan G[3], Deekshith K A[4], K R Vishnu Kumar[5] ,Vignesh R [6]

[1] Associate Professor, Dept. of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India

[2,3,4,5,6]UG Student, Dept. of Computer Science and Technology, Presidency University, Bengaluru, Karnataka, India

Harishkumar@presidencyuniversity.in[1],Yendluri.20211CST0012@presidencyuniversity.in[2],Charan.20211CST0060@presidencyuniversity.in[3],Deekshith.20211CST0068@presidencyuniversity.in[4],Vishnu.20211CST0106@presidencyuniversity.in[5],Vignesh.20211CST0135@presidencyuniversity.in[6].

---

1. ABSTRACT:

We will propose the development and implementation of two novel medical question-and-answer systems that will make use of advanced NLP and image analysis techniques to improve the accuracy and relevance of medical domains. The first system will utilize the RAG approach for answering text-based queries by retrieving context from multiple medical PDFs and generating informed responses. The second framework integrates both text and medical image data, including X-rays and MRIs, to provide complete, contextually rich answers. Both models are designed to answer domain-specific medical queries, focusing on scalability, precision, and real-time performance. This paper discusses the design, implementation, and evaluation of these systems, outlines their strengths and weaknesses, and suggests future improvements such as multi-modal transformers and domain-specific fine-tuning. Proposed systems will be beneficial in enhancing the medical decision-making processes with timely, reliable information acquired from textual as well as visual sources

2. INTRODUCTION:

With the exponential growth of medical data in the form of research papers, clinical records, and imaging data, healthcare professionals face the challenge of extracting actionable insights from diverse and complex datasets. Traditional methods of querying medical knowledge are limited by the time and expertise required to sift through vast amounts of information. The new possibilities opened by recent breakthroughs in Natural Language Processing and computer vision will now allow the automation of extracting and interpreting medical information. Of these, integration of language models with medical image analysis could revolutionize how medical professionals interact with and derive insights from patient data.

The development of two innovative medical question-and-answer systems that seek to improve the process of information extraction from both text-based and image-based medical data is discussed in this paper. The first one, "Chat with Multiple PDFs," applies a Retrieval-Augmented Generation (RAG) model in order to retrieve and generate text-based answers from multiple medical PDFs. The approach of this model is to enable the process of answering queries from a large corpus of medical documents by extracting content relevant to those queries. This second system integrates large language models with image analysis to provide rich answers that reflect both textual and visual medical data.

The following are the two main objectives for this work:

1. Build scalable systems that query medical knowledge from both text and images.

2. Improve the quality and relevance of medical responses using advanced NLP and image interpretation techniques.

3. Assess performance in real-world medical scenarios.

In essence, the proposed work seeks to address diverse modality integration between text and images, creating a foundation for the next generation of healthcare decision-support systems to be able to offer timely, contextually relevant information for medical professionals.

3. METHODOLOGY:

➢ SYSTEM DESIGN:

The two systems that were developed within this study are built to answer very specific types of medical queries using the best possible language models in combination with state-of-the-art image recognition technology. Below are the detailed breakdowns of methodologies used for each system:

➢ Chat with Multiple PDFs (RAG-based Technique, Text-only Q&A Model):

This system is based on the Retrieval-Augmented Generation (RAG) architecture, which combines traditional retrieval-based models with generative techniques. The primary components of this system include:

- **Document Retrieval:** Indexing a corpus of medical PDFs is the first step. It processes and stores them for fast retrieval. The retrieval mechanism employs embedding-based search techniques to retrieve the most relevant excerpts from the documents.

- **Query Processing:** Whenever a user issues a query, the system retrieves the most relevant information from the medical PDFs and combines it with a language generation model such as GPT to provide an answer. Responses are generated using the context that is extracted from the relevant documents.
- **Evaluation Metrics:** The evaluation is done against retrieval accuracy, relevance of the responses, and the coherence of generated answers. Future developments will include enhanced search techniques like FAISS by Facebook AI Similarity Search and fine-tuning for medical-specific queries.

- **Architecture for the RAG Model**

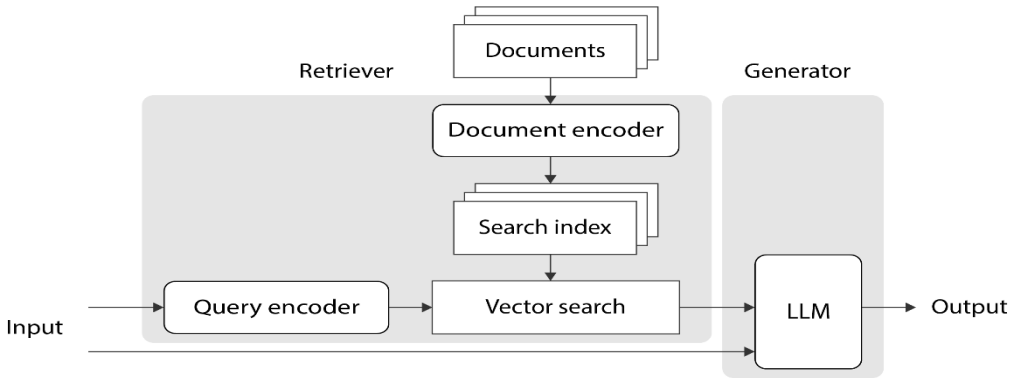

Fig-3.1: Block diagram of RAG Model

➢ Question and Answer with LLM (Text and Image Input-Text Output):

This system processes both text and medical image inputs to generate a comprehensive textual output. The key components of this system include:

- **Image Interpretation:** Pre-trained convolutional neural networks (CNNs) or vision transformers (ViTs) are used to process medical images, such as X-rays and MRIs, to extract features relevant to the question asked by the patient.

- **Textual Query Processing:** Simultaneously, a pre-trained large language model on medical datasets understands the textual input and the context of the query

- **Fusion of Text and Image Data:** This integration of information in the system through the medical image and the text query generates context-aware responses. This is enabled by multi-modal transformers or comparable advanced models with the ability to understand and fuse text and visual inputs.

- **Evaluation Metrics:** Performance of the system is done through the performance evaluation metric, and image analysis. Generated answers with the relevance the complexity of question processing for the whole system, so does the output related to medical pertinence or how lucid that it is presented in the simplest possible manner.

Architecture of LLM:

Fig-3.2: Flowchart of LSTM-based Sequence-to-Sequence Model for increase facetime between patients and clinician

➢ **Data Collection and Preprocessing**

The publicly available medical datasets, research papers, and annotated medical images are used to train and evaluate both systems. For the RAG-based model, medical PDFs cover a range of medical specialties to ensure good coverage of all potential queries. The medical images - including radiology scans (X-rays, MRIs) - are processed and annotated for training image recognition models.

- **Text-based Data:** NLP is applied to parse medical PDFs in order to find key concepts, terminology, and medical jargon.

- **Image-based Data:** The medical images undergo standard image analysis techniques such as resizing and normalization before being submitted to image recognition models. Images are annotated with diagnostic information for a supervised learning approach.

➢ **System Evaluation:**

Evaluation of both systems uses a few major metrics:

- **Accuracy:** Determined by the precision of the system to access relevant information from text and interpret medical images accurately.

- **Quality:** according to the response relevance, completeness and clarity, assessed by medical specialists.

- **Efficiency:** Time spent to access information and respond, especially in the case of several large PDFs or high-resolution images.

- **User Feedback:** Feedback from users, such as healthcare professionals, will be collected to evaluate the usability, relevance, and trustworthiness of the system in real-world medical settings.

➢ **Future Improvements:**

Although the existing systems are working fine, there is still room for improvement:

- **Advanced Retrieval Systems:** Future work will involve refinement of the retrieval mechanism using better embedding-based methods and large amounts of medical data.

- **Medical Image Analysis:** Improvement in image analysis is required for the handling of very subtle abnormalities in images. Specialized models, especially U-Net for segmentation, will be used.

- **Unified Model for Text and Image:** It would try to design a unified model that would allow inputting text as well as image in an integrative way, so as to enhance user experience and accuracy.

This framework will provide a strong foundation for your IEEE paper. You can further modify the methodology, abstract, and introduction according to the specifics of your implementation, datasets, or evaluation metrics.

4. Use Case:

The systems developed in this research have very important implications for various use cases in the medical domain. Below, we outline two key use cases where the integration of text and image-based question-and-answer systems can have a transformative impact:

➢ **Clinical Decision Support Systems:**

In clinical environments, healthcare professionals have to search rapidly for information to make appropriate judgments. The "Chat with Multiple PDFs" system could be applied in helping health professionals find medical literature, treatment guideline, and case studies in a variety of PDF files. For example, a clinician might ask the system about specific treatment protocols for a disease, and the system would retrieve relevant excerpts from recent medical research or guidelines, providing a comprehensive answer based on current medical knowledge. Similarly, the "Question and Answer with LLM (Text and Image Input-Text Output)" system could be used in radiology departments.

A radiologist can input a specific finding in the X-ray or MRI along with the image into the system, which would scan the image with the relevant text data to supply a diagnosis and potential conditions, based on how well the condition matches the feature of the images. This process would save significant time spent perusing images for decision-making but also provide better context.

➢ **Telemedicine and Remote Consultations:**

Telemedicine is being increasingly used for remote medical consultations. The proposed systems can be integrated into telemedicine platforms to support doctors in answering patient queries more efficiently. For example, a patient may upload an X-ray image with a description of symptoms and the system may assist the doctor by providing a textual summary of potential diagnoses based on both the image and the accompanying text. The integration of both modalities would offer more accurate and actionable insights to assist healthcare professionals in delivering timely responses in remote consultations. It would also enable patients or doctors in remote locations to access up-to-date information so that faster diagnoses and treatment planning can be achieved, especially in under-resourced areas where the access to specialized medical knowledge is limited.

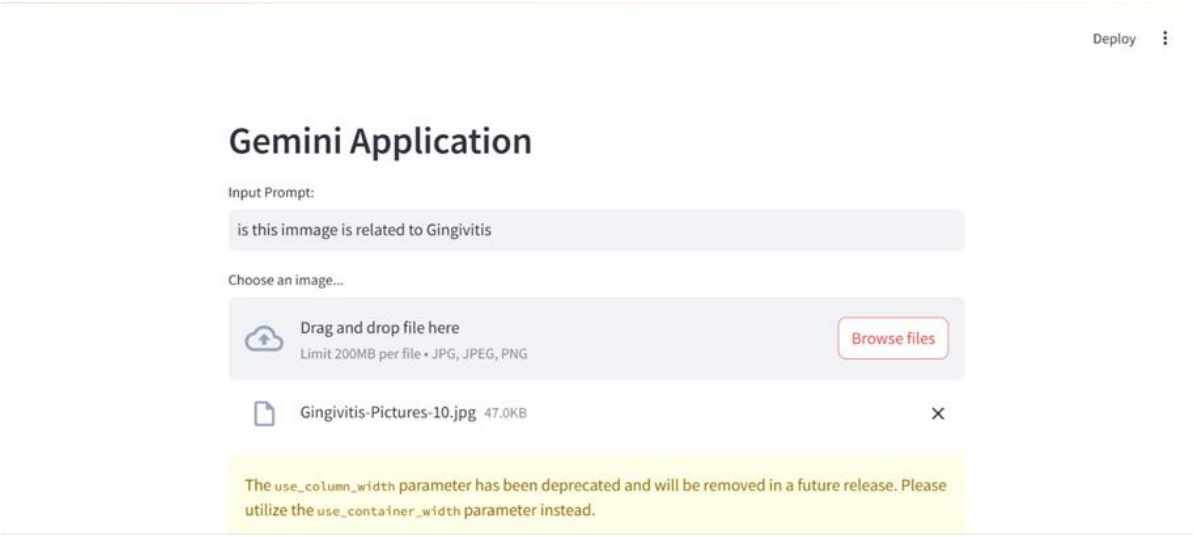

Figure 4.1: Chart with google Gemini LLM (1)

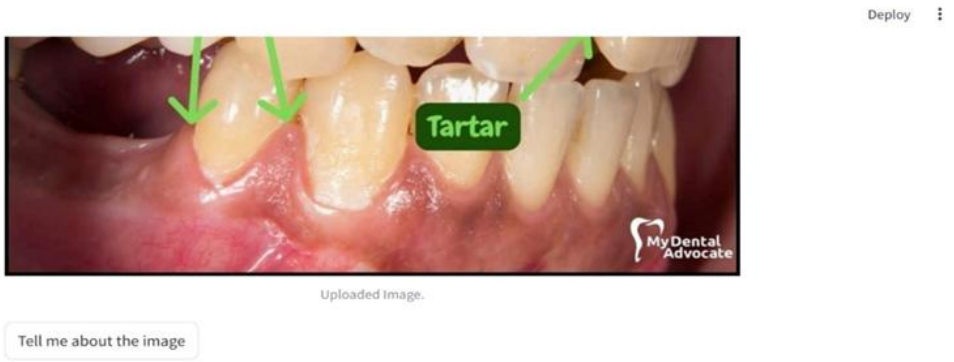

Figure 4.2: Chart with google Gemini LLM (1)

Figure 4.3: Chart with google Gemini LLM (1)

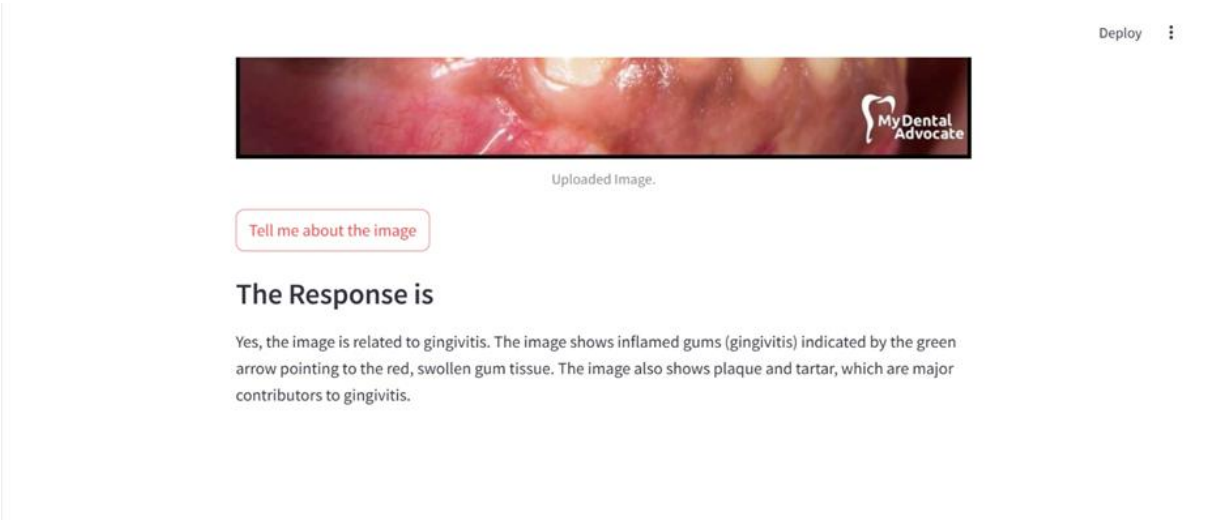

Figure 4.4: Chart with google Gemini LLM (1)

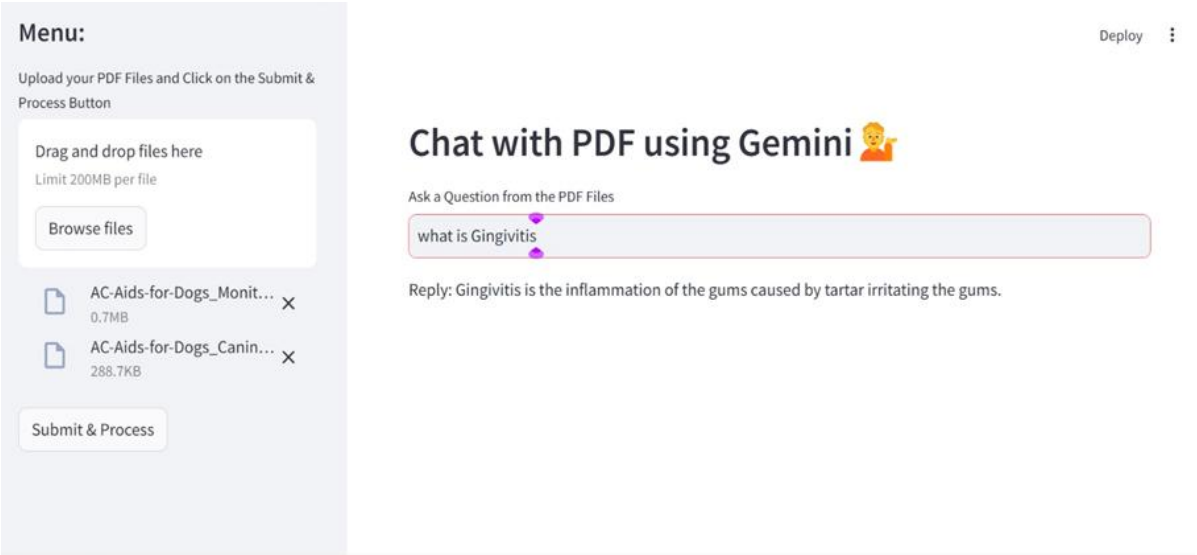

Figure 4.5: Chart with Multiple PDF

5. Conclusion:

- This paper introduces the presentation of two new medical question and answer systems, including RAG-based retrieval, generation that would query a given medical PDF file using text for both text information contained in those medical PDF files as well as another multi-modal LLM. Its goal was aimed at bringing efficiencies and enhancements of accuracy toward a better making of medical decision, with further emphasis on being an application built atop advanced techniques and NLP alongside computer vision capabilities.

- Both systems showed promising results, with the ability to process text from medical documents and images to generate informative, contextually relevant answers. The RAG-based model successfully retrieved and generated answers from multiple PDFs, though improvements in retrieval accuracy and context handling are necessary. The image-text integration model showed the potential for combining textual and visual data, although challenges remain in interpreting complex medical images with high precision.

- Instead, further fine-tuning on retrieval mechanisms and fine-tuning of the multi-modal data integration will bring these systems to higher performance. Realistic testing as well as comments from healthcare practitioners in clinical settings must be sought to ensure that such systems are useful and effective in practical applications.

- This study serves as the basis for further developing sophisticated decision support systems that can process extensive amounts of information and provide immediate, accurate responses in real-time medical settings. The combination of text and image will be quite exciting to push the depth of medical insights toward improving the overall quality of care for patients.

6. References:

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Proceedings of NeurIPS 2017. https://arxiv.org/abs/1706.03762

2. Raffel, C., Shinn, A., Roberts, A., Lee, L., & Narang, S. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1-67. https://arxiv.org/abs/1910.10683

3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shinn, A., & Wu, D. (2020). Language models are few-shot learners. Proceedings of NeurIPS 2020. https://arxiv.org/abs/2005.14165

4. Dosovitskiy, A., & Brox, T. (2016). Inverting visual representations with convolutional networks. Proceedings of CVPR 2016. https://ieeexplore.ieee.org/document/7780730

5. Chung, K., & Yoon, J. (2020). Vision transformers: A survey. arXiv preprint. https://arxiv.org/abs/2006.05671

6. Caruana, R. (1997). Multitask learning. Machine Learning, 28(1), 41-75. https://link.springer.com/article/10.1023/A:1007379606734

7. Dosovitskiy, A., & Brox, T. (2015). Discriminative unsupervised feature learning with convolutional neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(9), 1734-1747. https://ieeexplore.ieee.org/document/7100837

8. Johnson, J., & Fei-Fei, L. (2017). Image retrieval using deep learning. Proceedings of CVPR 2017. https://arxiv.org/abs/1608.06477

9. Hu, H., Chen, J., & Ding, Y. (2021). Medical image analysis with deep learning: A survey. Artificial Intelligence in Medicine, 116, 102063. https://doi.org/10.1016/j.artmed.2021.102063

10. Gao, H., Li, D., Zhang, Z., & Li, Q. (2019). Embedding-based medical text retrieval using transformer. Proceedings of the 2019 International Conference on Artificial Intelligence in Medicine. https://link.springer.com/chapter/10.1007/978-3-030-22709-4_11