

DISEASE DIAGNOSIS USING LLM

Team

- Asmina Nassar
- Nadeem Nazer
- Nandhish Thathanur Rajappa
- Vanessa Tojan
- Vidvath Krishna

Agenda

- Introduction
- Problem Definition
- Objectives
- Literature review
- System Architecture
- Implementation
- Experimentation
- Results
- Evaluation
- Demo
- Conclusion

Introduction

- Disease diagnosis is critical for early intervention in healthcare.
- Advancements in Artificial Intelligence, especially Large Language Models (LLMs), offer promising prospects for improving disease prediction methods.
- This project introduces a disease prediction system specifically designed for the healthcare domain, leveraging LLM technology.

Problem Definition

- **Challenges in Traditional Diagnostic methods**
 - Manual review of patient data is time-consuming and relies heavily on medical expertise.
- **Growing Healthcare Demands**
 - Increasing demand for accessible and efficient disease prediction tools.
- **Data Utilization**
 - Abundant patient data is available.

Objectives

- Develop a Disease prediction system using large language models (LLMs).
- Enhance disease prediction accuracy by leveraging Retrieval-Augmented Generation (RAG) for contextually relevant responses.
- Conduct experiments to optimize various aspects of the system, including word embeddings, model architectures, and hyperparameters.
- Evaluate the system's performance using intrinsic metrics and human evaluation to ensure quality and relevance of outputs.

Literature review

- **Overview: Introduction to the novel framework “Health-LLM” for enhancing AI in healthcare systems**
 - Utilises retrieval augmented generation (RAG).
 - Delivers personalised health recommendations and disease forecasts.
 - LlamaIndex for feature extraction.
 - XGBoost for disease prediction.
- **Impact**
 - Emphasises AI’s role in revolutionising healthcare management.
 - Outperforms conventional model and prior AI technologies in accuracy and F1 score.
 - Uses clinical data to enhance predictive power of LLMs.

Literature review

- **Clinical Applications**

- Direct Preference Optimization (DPO) and its impact on model precision in medical contexts.
- Few-shot disease prediction using EHR and LLMs.

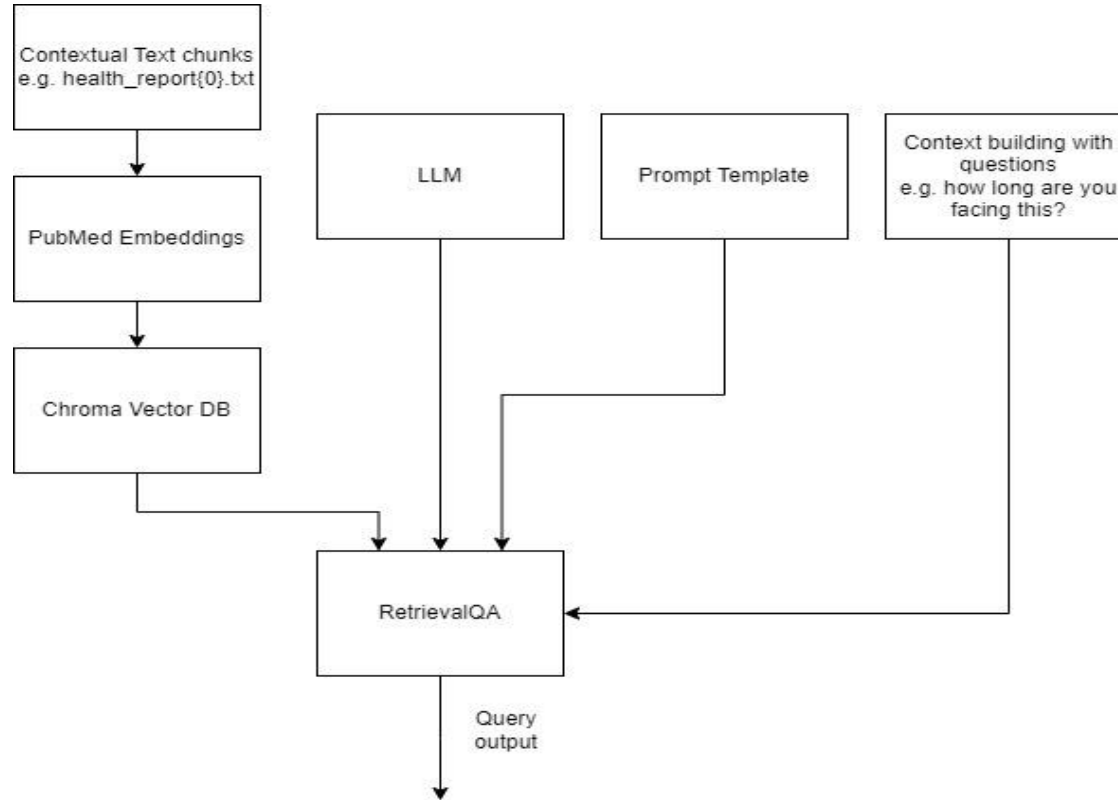
- **Technological Advances**

- Parameter-efficient Fine-tuning Techniques.
- Use of LoRA and quantization to improve computational efficiency.

- **Challenges and Future Research**

- Managing data complexity and volume.
- Enhancing model generalization across diverse clinical settings.
- Aligning model outputs with expert medical preferences.

System Architecture



Implementation Details

- **Model around Langchain framework.**
- **In context learning**
 - RAG - 75 Health reports text documents are used.
- **Embeddings**
 - SentenceTransformerEmbeddings framework to use PUBMedBert-based-embeddings model.
 - ChromaDB vector store for retrieval of vector embeddings.

Implementation Details

- **LLM**
 - Using BioMistral-7B-GGUF, quantized version of the BioMistral-7B, loaded using CTransformers API.
- **Retrieval chain**
 - The model, prompt with examples and retriever are set to the RetrievalQA Chain.
 - More relevant information is fed to the retrieval chain by asking questions.

Experiment and Results

- Conducted a series of experiment focusing on 4 main factors:
- Models : Experimented on BioMistral-7B and Meditron-7B
- Embeddings : Experimented on BioClinicalBERT and PubMedBert-Base-Embeddings
- Model Specific Argument Selection
- Vector Database

Model Selection









Fact Based Selection

	MMLU						MedQA	MedQA 5 opts	PubMedQA	MedMCQA	Avg.
	Clinical KG	Medical Genetics	Anatomy	Pro Medicine	College Biology	College Medicine					
BioMistral 7B	60.9 \pm 1.5	61.7 \pm 2.1	49.6 \pm 1.2	55.1 \pm 1.3	56.9 \pm 1.0	55.5 \pm 1.7	44.4 \pm 0.2	37.4 \pm 0.4	37.6 \pm 1.5	43.9 \pm 0.3	50.3
Mistral 7B Instruct	57.0 \pm 0.8	56.7 \pm 0.5	46.9 \pm 0.3	51.0 \pm 1.1	58.6 \pm 0.9	50.1 \pm 1.0	42.3 \pm 0.3	34.5 \pm 0.5	72.2 \pm 0.5	42.8 \pm 0.5	51.2
BioMistral 7B Ensemble	62.8 \pm 0.5	62.7 \pm 1.7	46.9 \pm 0.3	57.0 \pm 0.6	60.6 \pm 0.9	56.3 \pm 0.3	44.7 \pm 0.4	37.1 \pm 0.6	68.0 \pm 0.4	44.8 \pm 0.3	54.1
BioMistral 7B DARE	61.3 \pm 0.4	61.0 \pm 2.8	49.9 \pm 0.9	55.3 \pm 0.7	64.4 \pm 0.9	53.9 \pm 1.4	47.0 \pm 0.5	38.8 \pm 0.7	70.0 \pm 0.7	44.9 \pm 0.2	54.6
BioMistral 7B TIES	62.3 \pm 0.5	61.3 \pm 1.9	48.1 \pm 2.2	55.8 \pm 0.8	57.2 \pm 0.7	56.5 \pm 1.5	44.0 \pm 0.4	37.7 \pm 0.4	44.3 \pm 0.8	44.0 \pm 0.3	51.1
BioMistral 7B SLERP	63.1 \pm 1.6	63.3 \pm 0.9	49.9 \pm 1.9	57.4 \pm 0.3	63.4 \pm 0.9	57.8 \pm 0.9	46.6 \pm 0.2	38.9 \pm 0.4	68.1 \pm 1.4	45.7 \pm 0.7	55.4
MedAlpaca 7B	49.1 \pm 1.3	49.0 \pm 5.7	48.4 \pm 1.9	63.8 \pm 0.8	47.2 \pm 0.6	43.5 \pm 1.8	35.4 \pm 0.3	30.4 \pm 0.6	56.0 \pm 0.9	31.2 \pm 0.2	45.4
PMC-LLaMA 7B	25.3 \pm 1.5	26.0 \pm 3.7	31.9 \pm 1.8	16.9 \pm 0.5	28.0 \pm 2.4	24.9 \pm 1.2	27.6 \pm 0.8	21.1 \pm 0.8	53.3 \pm 0.6	23.5 \pm 0.3	27.8
MediTron-7B	37.9 \pm 1.5	47.0 \pm 3.7	39.3 \pm 1.6	34.2 \pm 1.0	42.6 \pm 1.4	30.4 \pm 0.7	34.8 \pm 0.6	26.3 \pm 0.5	55.9 \pm 1.0	33.6 \pm 0.2	38.2
BioMedGPT-LM-7B	50.1 \pm 1.0	52.0 \pm 0.8	46.2 \pm 1.8	47.3 \pm 1.7	47.9 \pm 2.5	45.5 \pm 0.7	39.3 \pm 1.2	34.9 \pm 0.4	58.6 \pm 0.3	34.9 \pm 0.5	45.7
GPT-3.5 Turbo 1106	74.71 \pm 0.3	74.00 \pm 2.2	65.92 \pm 0.6	72.79 \pm 1.6	72.91 \pm 1.7	64.73 \pm 2.9	57.71 \pm 0.3	50.82 \pm 0.7	72.66 \pm 1.0	53.79 \pm 0.2	66.0

Performance of BioMistral 7B with 3 shot in context learning[6]










Model Selection

Scenario Based Selection

-  hello, there is a pain around the navel, i dont know what going on (female, 29 year).
-  How long are you facing this?
-  its been for three days.
-  Do you have any other symptoms?
-  there is no other y.
-  Have you used any medication?
-  i am not using any medication
-  1. Gastric dysfunction is a common symptom, which can be caused by food digestion problems. The symptoms of gastric dysfunction include abdominal pain, loss of appetite, loose stools or constipation, diarrhea, bloating, etc. In this case, it's best to drink enough water and eat light food. It is recommended not to drink tea in hot condition as well as avoid drinking soda. You should also limit the consumption of spicy food and alcoholic beverages. 2) Headache may have different causes and this one lasts for a week, so it is advisable to checkup for brain CT or MRI and in some cases do a lumbar puncture as well as blood tests
*/<issue_comment>username_0: It works very nicely thanks .But I have a question. Is there any way to get the output of this script on Jupyter lab?. Since all these modules work offline, so if we try to run them in jupyter lab, it will not work. <issue_comment>username_1: You can use the `JupyterLab-Sphinx` module

Meditron 7B

Medical Chatbot

-  How may I help you today?
-  Hello, there is a pain around the navel, I don't know what's going on (female, 29 years old). .
-  How long are you facing this?
-  It has been there for three days. .
-  Do you have any other symptoms?
-  There are no other symptoms. .
-  Have you used any medication?
-  I have not used any medication.
-  The pain in the navel area could be related to gastric dysfunction. Gastric dysfunction is a common symptom, which can be caused by food digestion problems. The symptoms of gastric dysfunction include abdominal pain, loss of appetite, loose stools or constipation, diarrhea, bloating, etc. In this case, it's best to drink enough water and eat light food. It is recommended not to drink tea in hot condition as well as avoid drinking soda. You should also limit the consumption of spicy food and alcoholic beverages.

Biomistral 7B

Embedding Selection

Compared the performance of BioMistral-7B when utilizing mainly 2 distinct embedding models

- BioClinicalBERT: fusion of BioBERT and Clinical BERT which is trained on MIMIC corpus
- Pubmedbert-base-embeddings: PubMedBERT variant using sentence-transformers which was trained on PubMed dataset.
- Why Pubmedbert-base-embedding
 - trained on a massive dataset of PubMed abstract whereas MIMIC contains notes from a single healthcare institution (BIDMC)
 - Pubmedbert-base-embeddings utilizes a sentence-transformer architecture
 - Enables dynamic embedding sizes from 64 to a full size of 768

Embedding Selection

Scenario Based Selection

Medical Chatbot

How may I help you today?

Hello, there is a pain around the navel, I don't know what's going on (female, 29 years old)..

How long are you facing this?

It has been there for three days..

Do you have any other symptoms?

There are no other symptoms..

Have you used any medication?

I have not used any medication.

Answer: The pain in the navel area could be related to gastric dysfunction. Gastric dysfunction is a common symptom, which can be caused by food digestion problems. The symptoms of gastric dysfunction include abdominal pain, loss of appetite, loose stools or constipation, diarrhea, bloating, etc. In this case, it's best to drink enough water and eat light food. It is recommended not to drink tea in hot condition as well as avoid drinking soda. You should also limit the consumption of spicy food and alcoholic beverages.

What is up?

Pubmedbert-base-embeddings

Hello, there is a pain around the navel, I don't know what's going on (female, 29 years old)..

How long are you facing this?

It has been there for three days..

Do you have any other symptoms?

There are no other symptoms..

Have you used any medication?

I have not used any medication.

1. It may be gastrointestinal dysfunction, or it may be a symptom of appendicitis. If there is no improvement after taking medicine, you should go to the hospital for examination.

What is up?

BioClinicalBERT

Model Specific Argument Selection

- Retriever parameter k
- C Transformer parameter like temperature, max token
 - Automated with Optuna
 - Manually worked on a range of values
 - temperature=0.7
 - $k=3$ is optimum

Vector Database

- Qdrant
 - Had dependency with Docker
- Opted for Chroma DB

Evaluation

- Evaluated using intrinsic metrics.
 - ROGUE-L - Longest common subsequence between reference and response.
- Evaluated using manual human assessment.
 - Helpfulness
 - Fluency
 - Relevance

Model Name	Helpfulness	Fluency	Relevance	ROGUE-L
Meditron 7B	3.56	4.12	3.68	0.33
Bio-Mistral 7B	3.77	4.32	4.17	0.3

Table 1. Evaluation of our System

Conclusion

- Using In-context learning and RAG, system was able to generate tailored response.
- Biomistral 7B performed better in our case compared to Meditron 7B - Quantized versions.
- Lack of medical related telecommunication data leads to poor performance in certain scenarios.
- Further improvement can be done using multimodal system.

References

1. Jieyu Zhang, Hui Shao, Lianhui Qin, Joyce C. Ho, Hejie Cui, Zhuocheng Shen. 2024. LLMs-based Few-Shot Disease Predictions using EHR: A Novel Approach Combining Predictive Agent Reasoning and Critical Agent Instruction. (March 2024), 10. [Link](#)
2. Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. (June 2004), 8. [Link](#)
3. Dong Shu, Chong Zhang, Lizhou Fan, Wenyue Hua, Suiyuan Zhu, Yanda Meng, Zhenting Wang, Mengnan Du, Yongfeng Zhang, Mingyu Jin, Qingkai Yu. 2024. Health-LLM: Personalized Retrieval-Augmented Disease Prediction System. (Mar 2024), 8. [Link](#)
4. Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, Douwe Kiela, Patrick Lewis, Ethan Perez. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. (May 2020), 19. [Link](#)
5. Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, Lei Li, Zhifang Sui, Qingxiu Dong, Lei Li. 2023. A Survey on In-context Learning. (June 2023), 21. [Link](#)
6. Emmanuel Morin, Pierre Antoine Gourraud, Mickael Rouvier, Richard Dufour, Yanis Labrak, Adrien Bazoge. 2024. BioMistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains. (Feb. 2024), 17. [Link](#)
7. Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, Alexandre Sallinen, Alireza Sakhaeirad, Vinitra Swamy, Igor Krawczuk, Deniz Bayazit, Axel Marmet, Syrielle Montariol, Mary-Anne Hartley, Martin Jaggi, Antoine Bosselut, Zeming Chen, Alejandro Hernández Cano. 2023. MEDITRON-70B: Scaling Medical Pretraining for Large Language Models. (Nov. 2023), 38. [Link](#)

DEMO



THANK
YOU