MedAssist: A Retrieval-Augmented Language Model for Integrating Patient Reviews and FDA Data in Medication Guidance

Koushik Rameshbabu, Medha Malavalli Hareesh

1. Introduction:

The MedAssist system is an innovative AI-powered medical assistant that assists users in identifying relevant medications for their health issues via an easy-to-use conversational interface. At its core, the system makes use of the Mistral 7B big language model, which has been optimized with Unsloth for improved performance and 4-bit quantization to reduce memory footprint while preserving responsiveness. This foundation is supplemented with PubMedBERT embeddings, which give domain-specific semantic comprehension of medical language and allow for high-precision matching between user-described symptoms and standard medical diseases. The system uses FAISS vector similarity search to quickly obtain relevant information from a huge corpus of pharmaceutical reviews and government documentation, resulting in a responsive experience even for complex medical questions.

MedAssist's design combines data from two different sources to create a bridge between clinical information and patient experience. The first source includes more than 215,000 patient pharmaceutical reviews from the UCI Machine Learning Repository, which document actual experiences such as overall satisfaction scores, side effect reports, and efficacy ratings. The second source provides reliable information on warnings, active components, suggested usage, and contraindications by using authentic FDA drug information from their public API. MedAssist helps users connect medical facts with useful outcomes that are important to people thinking about treatment alternatives by fusing these complimentary viewpoints to provide them with balanced drug information that represents both clinical guidelines and lived patient experiences.

To provide organized, understandable medicine summaries, the system goes through several computational steps in the background, including embedding construction, condition matching, medication retrieval, review analysis, and natural language summarization. Effectiveness for particular symptoms, frequent side effects, the intensity of adverse reactions, and specific usage warnings are just a few of the important elements that are highlighted in these summaries. They are also customized to the user's condition and are written in plain language, which helps to close the communication gap that is frequently present in healthcare information systems.

2. Motivation:

This was developed in response to the increased demand for accessible medical information in an era where patients are increasingly involved in healthcare decisions. Many people struggle to comprehend the enormous array of medicine possibilities, sometimes relying on fragmented online reviews or complex medical websites that contain overwhelming quantities of data. There is a significant disparity between the technical medical information accessible and the patient-friendly assistance that most people require when choosing pharmaceutical options. It bridges this gap by combining patient experiences with formal medical standards. The system is intended to provide users with comprehensive, balanced information about potential medications in a conversational format, allowing them to engage in more informed discussions with healthcare providers and gain a better understanding of their treatment options while not replacing professional medical advice.

3. Problem Statement:

The healthcare information ecosystem provides several significant issues, which these addresses. First, patients often struggle to evaluate medical information about potential treatments, particularly the real-world consequences of side effects and efficacy rates. Second, prescription information is distributed across multiple sources, including clinical recommendations, patient evaluations, and pharmaceutical paperwork, making it difficult for individuals to gain a comprehensive knowledge. Third, personal experiences with drugs differ greatly, resulting in an uneven environment of anecdotal knowledge that is difficult to comprehend. It addresses these issues by developing a unified system that recognizes medical diseases from natural language descriptions, locates relevant drugs for specific ailments, and creates detailed summaries that balance government standards with patient experiences.

3. Dataset:

It employs a hybrid dataset produced by combining two independent yet complementing data sources. The basic data comes from the UCI Machine Learning Repository dataset,

which is available on Kaggle and includes around 215,000 patient reviews of drugs, as well as their conditions, ratings, and usefulness scores. This dataset contains detailed, real-world patient experiences that include subjective elements such as side effect intensity, efficacy for specific complaints, and overall satisfaction.

The second dataset comes from the FDA's open API. It retrieves crucial medical information from this API, such as brand names, generic names, active substances, specific population usage guidelines, warnings and cautions, contraindications, adverse reactions, usage directions, and doctor consultation recommendations. These two datasets are integrated using a matching technique that aligns medications from both sources based on their names, resulting in a comprehensive resource that combines clinical guidelines and patient experiences. This integrated dataset serves as the LLM's knowledge foundation, allowing it to provide balanced and informative pharmaceutical summaries that include both official safety data and real-world effectiveness reviews.

4. System Design Diagram

The system architecture (figure 1) employs a sophisticated pipeline that routes user queries through multiple steps before producing pharmaceutical suggestions. Starting with user input, the system runs the natural language query through the PubMedBERT embedding model, which converts the text to a high-dimensional vector representation. This embedding is then used to search an FAISS index of medical condition vectors, determining the most appropriate condition based on semantic similarity. Once the ailment has been recognized, the system searches for potential pharmaceuticals by querying a second FAISS index that maps conditions to appropriate drugs, utilizing pre-computed embeddings that capture the link between conditions and medications. For each retrieved medicine, the system collects relevant data from both the patient review dataset and the FDA information dataset. This combined data is then formatted into a prompt that directs the Mistral 7B big language model, which is performance-optimized with Unsloth, to provide a detailed description of the medication's benefits, drawbacks, side effects, and patient experience.

Key elements condensed:

1. Biomedical NLP (PubMedBERT)
2. Two-phase vector retrieval (FAISS)
3. Multi-source data fusion
4. Optimized LLM processing
5. Structured clinical output
6. End-to-end efficiency focus

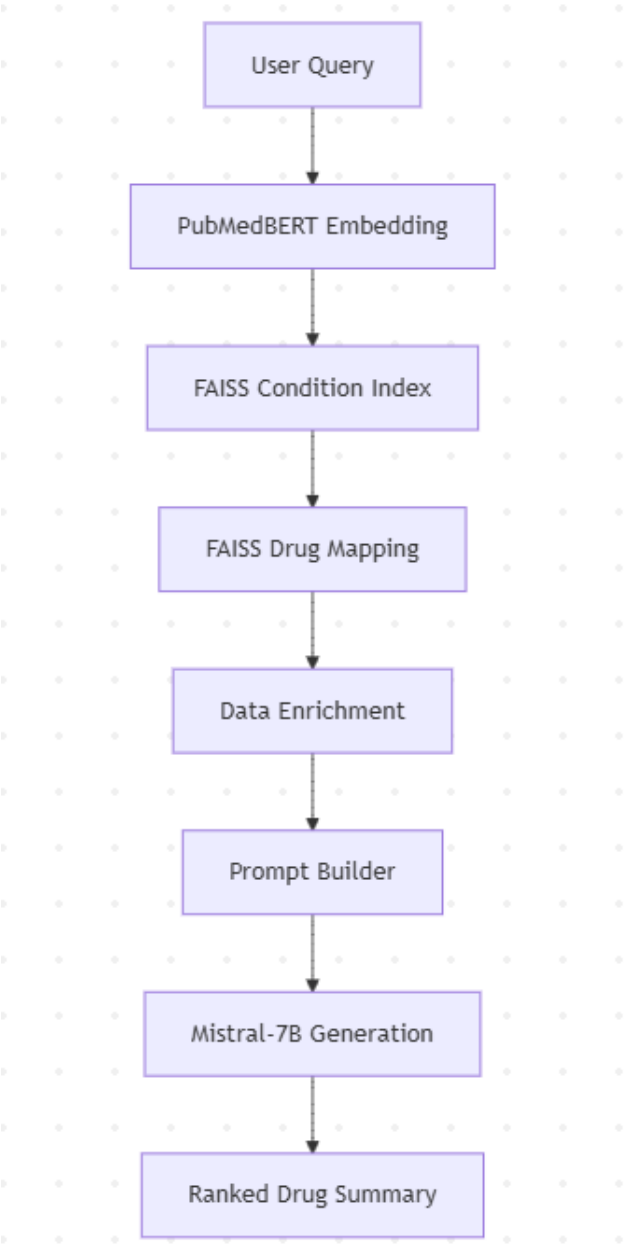

Figure 1: MedAssist Architectural Diagram

5. Approach

The technical solution uses a sophisticated multi-stage pipeline to translate user inquiries into relevant medicine information via information retrieval and natural language creation. In the first stage, the system preprocesses the datasets to provide clean, structured representations appropriate for embedding generation. The patient review dataset is

thoroughly cleaned to eliminate HTML artifacts and normalize text standards, followed by a chunking procedure that divides lengthier reviews into semantically understandable parts of around 500 characters. This preprocessing guarantees that the information retrieval engine may obtain more detailed insights rather than just document-level matches. Simultaneously, FDA data is formatted to extract key medical information like as contraindications and active components into standardized fields. The system also constructs a bidirectional mapping between brand and generic drug names to ensure comprehensive retrieval regardless of which version the user references.

The retrieval process is based on a dual-index architecture that is specifically designed for medical queries. The condition detection pipeline employs PubMedBERT to produce high-dimensional embeddings (768-dimensional vectors) for each medical condition in the dataset, resulting in a specific semantic space that encodes biomedical linkages. These embeddings are indexed using an FAISS L2 distance index, which allows for sub-millisecond similarity searches across hundreds of different medical conditions. When a user enters a query, the system develops an embedding and uses nearest-neighbor search to find the condition that is most semantically comparable to it. In addition, the system maintains a second FAISS index that map drug-condition pairings to aggregated review embeddings, allowing it to discover pharmaceuticals that are frequently used for specific diseases. This approach leverages the collective wisdom of patient reviews to surface medications with strong relevance to the detected condition, even when the connection isn't explicitly documented in formal medical literature.

The final stage of the pipeline utilizes the Mistral 7B instruction-tuned large language model to transform retrieved information into coherent, balanced summaries. For each candidate medication, the system retrieves the top reviews based on a composite scoring function that weighs both rating and usefulness metrics, along with corresponding FDA information.

$$0.6 \times rating + 0.4 \times usefulCount$$

This diverse data is then put into a structured prompt that directly tells the model to handle crucial features such as efficacy, side effects, contraindications, and administration instructions. To ensure consistent output formatting, the prompt design includes a few-shot examples with explicit section demarcations. To improve computational performance, the model employs 4-bit quantization via the Unsloth optimization package, resulting in lower memory requirements while retaining generation quality. Furthermore, the system employs batched processing for review embeddings and GPU acceleration for both embedding generation and language model inference, resulting in response times appropriate for interactive use even when processing several drug choices. This multi-stage method ensures that users obtain information that is relevant to their reported condition while also striking a balance between clinical accuracy and patient experience.

6. Method of Running code

The following provides the method to run the code:

1. Install MedAssist.ipynb from the provided zip file.
2. Download the datasets from the following drive link: Dataset_MedAssist
3. Run the entire code on a GPU enabled (Device = Cuda) system.
4. Enter the condition and receive list of drugs for the particular condition with a detailed summary.

7. Results

The following are the results obtained for the query with condition as ‘Acne’ (Results are read from left to right)

```
📝 SAMPLE PATIENT REVIEWS:
• "I was a 48 y/o female with nodular acne.My pimples were painful,
humiliatuing, embarrassing and just plain ugly.I had used every over the
counter treatment there was.Once I was diagnosed with nodular acne (severe
acne) my dermatologist prescribed Doxy 150mg.By using coupons, It is
affordable.I am 52 now and I continue to use Doxy 50 a day at this time.My
thought is A CLEAR FACE IS PRICELESS.No more pockets of puss or inflammation
on my face." (Rating: 10/10, Votes: 29)

⚠️ DISCLAIMER: This information is for educational purposes only and not
intended as medical advice.

Processing 2/5: Clindamycin
get_drug_fda_info is running now
============================================================
Clindamycin for Acne
============================================================

💊 MEDICATION OVERVIEW:
  Drug: Clindamycin
  Condition: Acne
  Average Rating: 8.9/10
  FDA Data: Available

📊 PATIENT REVIEW SUMMARY:

Based on the provided patient reviews and FDA data, Clindamycin appears to be
an effective treatment for acne based on the positive experiences reported by
patients. The patient reviews consistently report significant improvement in
acne symptoms, with some reporting complete clearance of acne within a few
weeks of use. However, it is important to note that the FDA data does not
specifically mention acne as a condition for which Clindamycin is approved,
but rather lists it under the adverse reactions section for daptomycin, which
is a different antibiotic.

The common side effects mentioned by patients include no side effects, the
appearance of tiny breakouts after initial improvement, and the need to use a
moisturizer to prevent drying.

The key benefits of Clindamycin according to patients include its
effectiveness in treating acne, the absence of side effects for some
individuals, and the improvement in overall skin health and confidence.
However, a potential drawback mentioned by some patients is the appearance of
new, smaller breakouts after initial improvement.

Overall, the patient experiences suggest that Clindamycin is an effective
treatment for acne, with a good safety profile for many individuals. However,
it is important for individuals to be aware of the potential for new
breakouts and to monitor for any signs of serious side effects, such as
anaphylaxis or myopathy.

📝 SAMPLE PATIENT REVIEWS:
• "I am a 60 year male with adult acne.I feel like I am reliving my teen
years.My scalp, my face and my back and neck started to break out.I tried
several prescription meds, when I finally used clindamycin topical gel I saw
a huge difference in the condition.Now I am mostly all clear and I am also
very happy.Also no side effects.I truly hope that everyone could have these
results especially you teens.God bless.truly yours, Porkchops 5/21/2015."
(Rating: 10/10, Votes: 66)
• "I love love love this product.It is truly the only thing that has worked
for my acne.I am lucky and havent experienced any side effects.I know that
every body is different but I recommend trying it." (Rating: 10/10, Votes:
60)
• "I had never had acne until I turned 22, an occasional pimple but a
suddenly I began getting cystic acne; it was not severe but bad.Honestly, I
had tried nearly everything for it.Thankfully I was only getting it on my
face, but the cysts kept growing and after trying months of prescription
levels and regimens of retin-a, benzoyl peroxide, doxycycline, and a few
others- I had no to very little improvement.I was about to be put on acutane-
thankfully I asked my doctor about clindamycin." (Rating: 10/10, Votes: 54)

⚠️ DISCLAIMER: This information is for educational purposes only and not
intended as medical advice.

Processing 3/5: Salicylic acid / sulfur
============================================================
Salicylic acid / sulfur for Acne
============================================================

💊 MEDICATION OVERVIEW:
  Drug: Salicylic acid / sulfur
  Condition: Acne
  Average Rating: 0.0/10
  FDA Data: Not Available

📊 PATIENT REVIEW SUMMARY:

No reviews found for Salicylic acid treating Acne.

📝 SAMPLE PATIENT REVIEWS:

⚠️ DISCLAIMER: This information is for educational purposes only and not
intended as medical advice.

Processing 4/5: Pernox
get_drug_fda_info is running now
============================================================
Pernox for Acne
============================================================

💊 MEDICATION OVERVIEW:
  Drug: Pernox
  Condition: Acne
  Average Rating: 10.0/10
  FDA Data: Not Available

📊 PATIENT REVIEW SUMMARY:

Based on the patient reviews, Pernox appears to be highly effective in
treating acne for adults. The patient reports a significant improvement in
skin appearance, with the medication leaving the skin refreshed and silky
smooth rather than overly dry. No specific symptoms or side effects were
mentioned in the reviews.

PROS:
- Effective in treating adult acne
- Leaves skin feeling refreshed and silky smooth
- Does not overly dry out the skin

CONS:
- Limited information available as no FDA data was provided
- Only one patient review was analyzed

SUMMARY:
The patient review suggests that Pernox is an effective treatment for adult
acne, leaving the skin feeling refreshed and silky smooth without over-
drying. However, further research is needed to confirm these findings as only
one patient review was analyzed and no FDA data was available.

📝 SAMPLE PATIENT REVIEWS:
• "This is a wonderful product for adult acne.It does not leave your skin
overly dry, but rather refreshed and silky smooth.I use it twice a day and
have experienced the best skin appearance since I developed acne over 18
years ago." (Rating: 10/10, Votes: 7)

⚠️ DISCLAIMER: This information is for educational purposes only and not
intended as medical advice.

Processing 5/5: Aczone
get_drug_fda_info is running now
============================================================
Aczone for Acne
============================================================

💊 MEDICATION OVERVIEW:
  Drug: Aczone
  Condition: Acne
  Average Rating: 9.1/10
  FDA Data: Available

📊 PATIENT REVIEW SUMMARY:

EFFECTIVENESS:
The patient reviews suggest that Aczone is highly effective in treating acne,
with most patients experiencing a significant reduction in acne bumps and
cysts within a few days to a few months of use. This aligns with some
clinical study findings mentioned in the FDA data, which state that Aczone
was effective in reducing inflammatory lesions and non-inflammatory lesions
in clinical trials. However, it is important to note that there is no
available data on Aczone's use in pregnant women from the FDA, and animal
reproduction studies have shown potential risks for adverse developmental
outcomes with oral doses of dapsone.

SIDE EFFECTS:
Common side effects mentioned by patients include application site dryness
and pruritus (itching).
```

8.Quality Evaluation

The figure, titled "Quality Evaluation: Acne Medication Summaries," presents a three-panel assessment of five common acne treatments—Doxy 100, Clindamycin, Salicylic acid, Pernox, and Aczone—across five key dimensions: Completeness (coverage of all essential aspects), Accuracy (faithfulness to patient experiences), Usefulness (practical help for decision-making), Information Use (effective utilization of available data), and Readability (clarity and organization). A brief caption beneath the title reminds the viewer that higher scores denote better quality summaries.

In the top panel, a heatmap lays out each medication's score from 0 to 10 in a grid: Doxy 100 and Aczone earn nearly all 10s, Clindamycin hovers at 8–10, and Pernox occupies the mid-high range (5–10). By contrast, Salicylic acid scores a perfect 10 for Accuracy and Readability but plummets to 1 in Usefulness and 2 in Information Use, with only a 5 for Completeness. The center panel converts those five scores into a single average per medication: Doxy 100 and Aczone tie at 9.6, followed by Clindamycin at 9.2, Pernox at 8.2, and Salicylic acid trailing at 5.6. Finally, the bottom radar chart overlays each drug's profile around a pentagon of axes. Doxy, Aczone, Clindamycin, and Pernox form almost full, rounded shapes—evidence of uniformly strong performance—whereas Salicylic acid's shape collapses inward on the Usefulness and Information Use spokes but remains fully extended on Accuracy and Readability. Together, these panels vividly illustrate that while most acne medications deliver comprehensive, accurate, and usable summaries, Salicylic acid's descriptions, despite being clear and accurate, lack substantive guidance and data integration.