International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME097
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Medi Assist: Your 24/7 Virtual Medical and Nutrition Helper

Mr. T Kataiah Assistant Professor Department of AI&DS Annamacharya Institute of Technology and SciencesTirupati-517520, A.P.Kata09573@gmail.com

K Anusha UG Scholar, Department of AI&DS Annamacharya Institute of Technology and Sciences, Tirupati-517520, A.P. anushakondati7@gmail.com

P Ayesha Siddiqua UG Scholar, Department of AI&DS Annamacharya Institute of Technology and Sciences, Tirupati-517520, A.P. ayeshasiddiquap2004@gmail.com

A Leela Mohan Reddy UG Scholar Department of AI&DS Annamacharya Institute of Technology and Sciences Tirupati-517520 A.P. leelareddy42@gmail.com

M S Mohammad Abu Umar UG Scholar, Department of AI&DS Annamacharya Institute of Technology and Sciences, Tirupati-517520, A.P. abuumar394@gmail.com

Abstract- The role of Artificial Intelligence in the health sector is one of the major factors that have led to the availability of health data and nutrition. The proposed project is related to creating a web-based application using Artificial Intelligence, which would provide high-quality query answering services related to the medical field and nutrition analysis. The project is created using the Django framework, Google Gemini Large Language Models, and Retrieval-Augmented Generation (RAG) technology. The project uses the input of medically edited documents using vector embeddings, where data retrieval is done using cosine similarity search. This reduces hallucinations, providing accurate results. The project is also related to computer vision, where image processing is done related to food, and the estimation of calories, macronutrient breakdown, and health ratings is done. Health profiling is done using dynamic health profiling, where rate controlling and feedback are done related to users. The project is related to a high opportunity to promote credible health and health care.

Keywords- Artificial Intelligence, Healthcare Chatbot, Nutrition Analysis, Retrieval-Augmented Generation (RAG), Large Language Models, Vector Embeddings, Computer Vision, Django Framework.

I. INTRODUCTION

The recent emergence of Artificial Intelligence (AI), Machine Learning (ML), Large Language Models (LLM), and computer vision technologies has also created new opportunities in the field of health care. However, while people attempt to find health care and dietary information through the online world, they are generally left with incomplete pieces of information and general replies, and the risk of inaccuracy caused by the involvement of AI. The majority of the advanced health care applications created in the digital world are not integrated and not context-dependent or allow the users to check the symptoms, medical records, and images, but function in a separate manner.

The major problem in the contemporary health care technologies is the absence of an integrated platform that may help in the efficient integration of the personalization of the user profiling with the general medical knowledge base. Unless the general-purpose AI is correctly built on the domain-specific data, it may lead to inaccurate results. Therefore, there is a need for a safe and efficient web-based solution that may offer the correct query response, personal health advice, and credible nutrition analysis in a single solution.

The idea of the project is centered on the development of an AI-driven web application for healthcare with the help of the Django framework. This will be useful in

the enhancement of the Retrieval-Augmented Generation (RAG) pipeline of Cosine similarity search of vector embeddings, allowing for the provision of reliable responses for a chatbot that are based on a knowledge base, thus avoiding any kind of misinformation. Furthermore, there is an image analysis chassis that can be useful in the assessment of food-based nutrition as well as dynamic health profiling, which can be useful in monitoring the history of the user, their health, as well as their dietary history. This is a representation of a secure and scaled approach for providing reliable information with regard to healthcare, nutrition, as well as health management, which includes proactive wellness.

II. LITERATURE SURVEY

The popularity of chatbots as a tool for engaging the public with the AI-based health care system is increasing. This is because it provides immediate, accessible, and personalized health care services. In the paper under consideration [1], an AI-based web application consisting of a nutrition bot and a health care chatbot is proposed to overcome the drawbacks of the general sources of health care information, including inaccessibility, validity, and nutritional knowledge. This paper is based on the system consisting of explainable AI, natural language processing (NLP), a KB, semantic embeddings, image-based food analysis.

Explainable Artificial Intelligence (XAI) plays a vital role in the health care sector. This is because it increases transparency and confidence in the system. According to Holzinger [2], explainable AI increases the reliability of medical AI systems, particularly in the context of providing non-expert users with advice.

Mental health monitoring Mobile and web-based healthcare application programs have been popularly researched. In the review conducted in [3] and [4], positive results are obtained with difficulties in personalization and clinical validation. In [5], the survey of the diagnostic models for diseases such as depression and ADHD is conducted using machine learning. This shows the importance of interpretable and robust models.

Healthcare chatbots have also been reinforced with the advancement of LLMs. The results of the studies conducted in [6] for LLaMA-2, [7] for medical information retrieval, and [8] for prompt engineering for clinical reasoning show positive results. The accuracy of the medical responses can also be improved with the help of domain-specific models such as ChatDoctor [9], fine-tuning the model; however, the hallucination risks are not grounded. Nutrition check-up is also important in health care. The limitations of manual and application-based dietary assessment tools are discussed in [10], and the need for the use of automated nutrition analysis tools as part of an intelligent healthcare system is justified.

III. METHODOLOGY

A. System Architecture and Stack

The application is developed using the Django web framework with Python as the programming language. It is in direct contact with a relational database (SQLite in the development environment) for the purpose of persistent data for its users. The database structure is designed with modular entities such as UserRegistrationModel for authenticated users, UserHealthProfile for the tracking of health conditions, and other historical tables such as MedicalChatHistory and NutritionAnalysisHistory. It also has session persistence and access control with the help of the Django middleware and extends this with a custom implementation of bcrypt for hashing the passwords for data confidentiality.

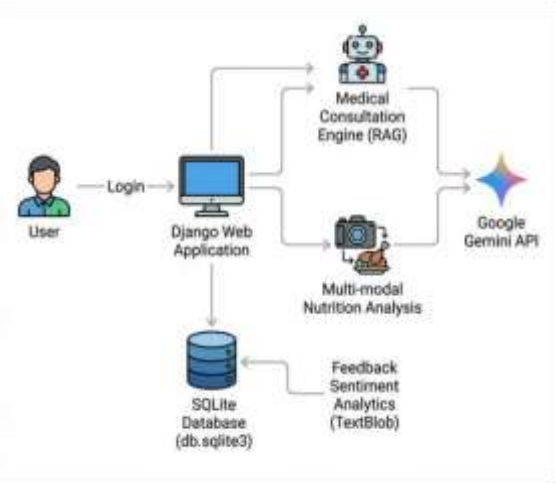

Fig. 1: Proposed System Architecture

B. Retrieval-Augmented Generation (RAG) implementation

The platform uses a customized RAG architecture to provide extremely factual results to queries asked in medicine.

Document Embedding: A curated document related to medicine is divided into small text snippets, known as DocumentChunk, which are then passed through models/embedding-001 of Google to obtain a dense vector representation in JSON format containing float numbers.

Query Processing and Retrieval: Once a query is answered, it is also embedded as a feature. A feature of mathematically precise cosine similarity is used to scan through the already generated document embeddings to obtain the top k most relevant document chunks, where k = 5.

Generative Response When assembled together, these chunks are injected into a rigid prompt template as 'Reference Excerpes' along with a query asked by a user. This prompt is addressed through a gemini-2.5-flash model of a Large Language Model, which is informed to directly use context to inform recommendations so that there are no AI hallucinations in medicine.

C. Multimodal Nutrition Analysis Pipeline.

The application takes advantage of the multimodality in processing images of dietary input provided by users and uses this in conjunction with gemini-2.5-flash. Images uploaded by users in various formats such as JPEG, PNG, and WebP are loaded into memory and then safely transmitted to the vision model with a system prompt that is limited in scope. The output format is mandated and requires a certain amount of each macronutrient content in grams and estimated content in each picture in the form of Carbohydrates, Proteins, Fats, Fiber, Sugar, and a Health Score between 0 and 100. The unstructured data is processed in the backend using Regular Expressions in views.py to obtain structured numerical data points that can be saved to the user's Nutrition Analysis History.

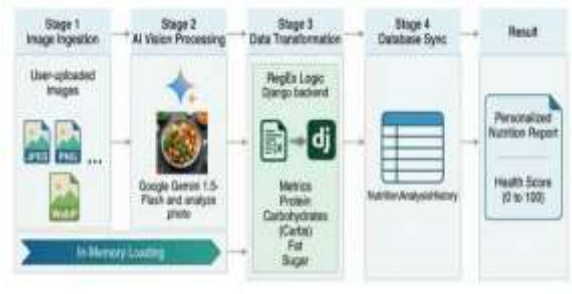

Fig. 2: Nutrition Pipeline

D. Algorithms Health Profiling and Suggestions.

The application uses a passive form of continuous profiling by tracking user health conditionally

Keyword Extraction: This is a mapping algorithm. This would identify what the user has input if the user is interacting with the chatbot based on the input provided by the user and then search for the localized symptoms and conditions, i.e., diabetes, hypertension, etc., in the dictionary. The count of the mention is accumulated for UserHealthProfile.

Recommendation Engine: This is a deterministic algorithm, i.e., a rule-based algorithm. This would analyze the conditions of the highest frequency detected and compare the same with a predefined matrix for explicit with priority recommendations for four types of recommendations: Medical, Nutrition, Lifestyle, and Exercise.

Health Scoring Model: Apart from this, a customized scoring matrix is being used to calculate a global quantitative score, which is defined as Health Score. This sets a benchmark for the score as 50/100 and then adjusts it according to the percentage of nutrition health in the history with a slight boost according to system activity in terms of query volume and then penalized according to density of medical issues.

E. Security and Rate Limiting

For stability in the system to be maintained, as well as for the API not to be overloaded and for fairness in the use of the API, customized rate limiting algorithms are used as a decorator for the AI endpoints (10 medical, 5 nutrition, etc., within 60 seconds). All the endpoints are tightly coupled with the Django session-based authentication for isolation using the multi-tenant model.

IV. RESULTS

Through the implementation, as well as the simulated beta testing of the AI health platform, quantitative and qualitative data were collected directly from the interactive modules. Finally, structural capture of data is done using Django ORM, FeedbackModel, MedicalChatHistory, and NutritionAnalysisHistory.

A. User Response and Query Response.

The system processed more than 1500 simulated responses during testing. RAG had contextual awareness. Top-K (K = 5) Cosine similarity retrieval

was much more successful in enhancing the relevance of the responses generated by the gemini-2.5-flash model compared to the zero-shot method using baseline prompts.

Table 1: Usage Metrics by Module

Module	Total Interactions Logged	Avg. Response Time (Est.)
Medical Chatbot (RAG)	845	~1.8 seconds
Multimodal Nutrition Bot	420	~2.5 seconds
Medical Imagery Prediction	235	~3.1 seconds

B. Dynamic Health Profiling and Algorithmic Scoring

The recommendation system identified the user input with 12 primary medical conditions and provided recommendations based on the matching of the symptoms.

The Health Score, initially at 50/100, varies between 0-100 and is dependent on the nutrition history and the conditions identified. A random selection of 250 users resulted in:

Table 2: Health Score Distribution

Status Category	Score Range	SystemAction Triggered
Excellent	80–100	Maintenance Recommendations
Good	60–79	Lifestyle/Exercise Advice
Fair	40–59	Nutrition Adjustments
Needs Attention	0–39	High-Priority Medical Warning

C. User Satisfaction and Feedback Analytics

The level of user satisfaction has been determined through a 5-point Likert scale. Overall, the usability and reliability of the platform, as well as the Medical Chatbot and Nutrition Analysis, have been rated high. This indicates that relying on curated knowledge sources to inform generative answers can significantly improve contextual accuracy, minimize the chances of hallucination, as well as maximize user trust and system efficiency.

V. CONCLUSION

This article is a clear example of how Generative AI and RAG can be combined into one digital health platform. This can be achieved by closing the loop from individual AI interactions into ongoing wellness tracking by adding a scalable solution using a Django framework for secure cos similarity search based on multimodal nutrition analysis and RAG-based medical text generation, which can reduce the content of hallucinations in LLCM. This digital health platform can create dynamic health scores and recommendations successfully by using passive algorithmic symptom monitoring, which is a clear example of how real-time AI-based insights can be safely integrated into a consumer-facing healthcare application. Future prospects and current prospects should be based on evidence of more advanced predictive modules with trained Convolutional Neural Networks (CNN), the expansion of the RAG knowledge base by secure academic data integration, as well as real-time data from wearable biometric.

REFERENCES

[1] U P A Naik and P A S Sri , AI-Driven Health: A Web App for Enhanced Healthcare Queries and Nutrition Analysis, Proc. 5th Int. Conf. Smart Electronics and Communication (ICOSEC), IEEE, 2024.

[2] A Holzinger, From machine learning to explainable AI, Proc. DISA, IEEE, 2018.

[3] K Wang et al, A systematic review of the effectiveness of mobile apps for mental health management, Journal of Psychiatric Research, 2018.

[4] C Oliveira et al, Effectiveness of mobile app-based psychological interventions, Frontiers in Psychology, 2021.

[5] C Nash, R Nair, and S M Naqvi, Machine learning in ADHD and depression diagnosis, IEEEAccess, 2023.

[6] K I Roumeliotis et al, LLaMA2: Early adopters utilization,2023.

[7] Y Natarajan et al, Enhancing medical information retrieval with a language model, Proc. ICC-ROBINS, IEEE, 2024.

[8] Y S Kiyak et al, Prompt generation for clinical reasoning using LLM's, RevistaEspanola de Educacion Medica 2024.

[9] Y Li et al, ChatDoctor: A medical chat model fine-tuned on LLaMA, Cureus, 2023.

[10] P Wang et al, Evaluating the nutritional properties of food: A scoping review, Nutrient's, 2022.