iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79616

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Multimodal AI Healthcare Chatbot with Symptom Analysis, Image-Based Detection and Severity Classification

Hiba Anjum[1], Saikumar Chilla[2], Shivateja Kommu[3], Rahul B[4]
[1]Assistant Professor, [2 3 4]Students, Department of Computer Science and Engineering(Artificial Intelligence and Machine Learning), Methodist College of Engineering and Technology Abids, Hyderabad, Telangana, 500001, India.

Abstract: The Multimodal AI Healthcare Chatbot with Symptom Analysis, Image-Based Detection and Severity Classification aims at providing intelligent assistance in healthcare through Multimodal Large Language Model (MLLM) based reasoning of AI. In the proposed solution, the user inputs the symptoms and also uploads images which are processed by the algorithm for detection of possible health issues and categorizing them into three classes: Mild, Moderate and Severe. Based upon the level of severity, the algorithm makes suitable medical recommendations for health decisions. While LLM-based healthcare bots depend only on text-based processing of symptoms for their operations, the Multimodal Healthcare AI is characterized by symptom analysis, image-based detection and severity classification. Experimental results prove that the proposed approach achieves an accuracy of 92%, whereas the LLM based bot performs with 85% accuracy, giving an improvement of 7%.Top of Form Bottom of Form

Keywords: Multimodal HealthcareAI,LLM-based Chatbot ,Symptom Analysis,Severity Classification, Medical Recommendation System, Intelligent Medical Assistance .

I. INTRODUCTION

Healthcare provision is another area that faces considerable obstacles, especially in underprivileged and rural areas. According to the World Health Organization, billions of individuals continue to lack access to adequate health services, thus making an intelligent digital health system necessary. Recent advancements in artificial intelligence have made it possible to incorporate real-time healthcare services into automated systems.

The current healthcare chatbots concentrate on symptom analysis and ignore other aspects, such as image detection, severity, and doctor suggestion. The inability of the present healthcare chatbots to offer other functionalities besides symptom analysis creates gaps for research.

This paper introduces the concept of AI Healthcare Chatbot, which integrates symptoms, image detection, severity assessment, and doctor recommendation services.

II. LITERATURE SURVEY

The existing approach referred to in the paper [1] by Sallam, M. (2023). shows that medical responses from the AI chatbots could be equally dependable in comparison with those made by physicians; hence, showing the possibility of conversational AI's use in offering primary healthcare services. In the research [2], the applicability of AI chatbots in the field of healthcare is tested and proved better access, as well as response time, although there is an increased need for improvement in the precision of decision making. The study [3] focuses on the application of conversational AI in the field of healthcare and explains improved patient engagement, even as several challenges are pointed out regarding decision making at the point of care. The research [4] investigates the use of AI chatbots in healthcare settings and shows how they can help diagnose diseases and communicate with patients, but there exist challenges in the context of decision making. The study [5] demonstrates how large language models could be used to capture clinical knowledge and facilitate decision making.

The study in [7] analyzes the effectiveness of AI chatbot usage in medical exams and proves its high capacity to comprehend the concept of healthcare; however, implementation issues persist. In [8], AI-supported decision-making in healthcare is discussed, and better recommendation production is demonstrated, although multimodal fusion remains limited.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Paper [9] discusses the acceptance of AI healthcare chatbots by users, which is shown to have enhanced accessibility and user participation. Research [10] draws attention to the increasing significance of AI chatbots in medical practice.

While all those studies prove the effectiveness of AI healthcare chatbots, most existing models, including the current system, as shown in Sallam, M. (2023), rely only on textual symptom interaction without incorporating symptom analysis, medical image recognition, severity classification, and doctor recommendation into one comprehensive solution. In contrast, the Multimodal AI Healthcare Chatbot proposed here presents an innovative AI-powered health assistance system that includes symptom interaction based on LLMs, multimodal symptom analysis, severity classification, and medical recommendation to enhance healthcare decision-making.

III. PROPOSED METHODOLOGY

The Proposed Multimodal AI Healthcare Chatbot uses Multimodal Healthcare AI framework that can analyze symptom text data and images for providing healthcare recommendations. For performing Symptom Analysis, the system adopts the LLM-based Chatbot that uses the dataset known as Medical Chatbot from Kaggle [16], which includes structured conversations related to healthcare. Moreover, Symptom-Based Disease Prediction Dataset [17] from Kaggle assists the process of Severity Classification in finding out any health condition through symptom analysis. Also, the Skin Disease Image Dataset [18] from Kaggle is used for image input processing that facilitates visual symptom analysis. This adoption of datasets from Kaggle makes the Multimodal Healthcare AI more efficient for conducting Symptom Analysis, Severity Classification, and generating Medical Recommendations. Through the use of LLM-based reasoning, multimodal input processing, and rule-based logic, the system works as a Medical Recommendation System that gives real-time healthcare suggestions. With the inclusion of TTS Response Generation feature, the process becomes more efficient for delivering healthcare assistance..

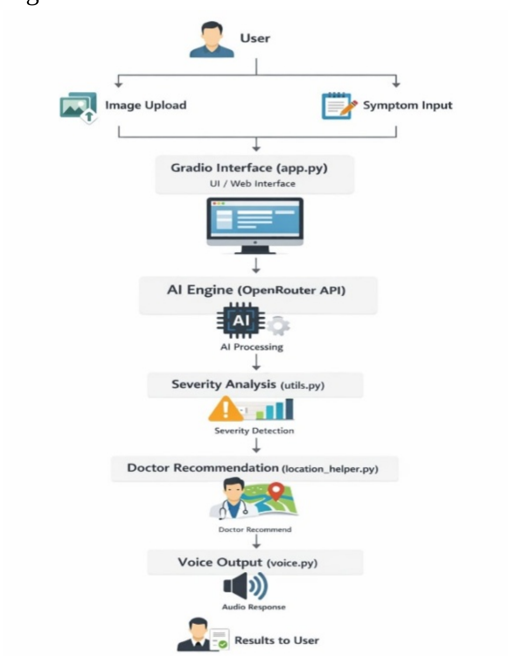


Fig 1: Proposed Architecture

The process for the Multimodal AI Healthcare Chatbot is illustrated in Fig 1. The user inputs their symptoms or uploads an image related to their medical condition. This information is analyzed by the AI engine, which performs a severity assessment. Depending on the analysis, the AI engine recommends doctors and generates voice output, which finally produces healthcare outputs.

A. Multimodal LLM (MLLM)-Based AI Reasoning

The main approach to be used in the designed system is that of MLLM-based AI reasoning via Llama 3 Vision via the OpenRouter API. Symptoms in the form of text provided by the user will be analyzed using the Llama 3 Vision, along with any medical image input from the user.

Since Llama 3 Vision utilizes a multimodal approach to analyzing medical information presented in the form of text and an image, it will be able to diagnose the health condition of the person based on these two sources of information. Using Llama 3 Vision does not require the usage of traditional machine learning models since no prior training will be needed to carry out intelligent reasoning.

B. Ai Inference using Groq

The system implements AI inference with the help of the Groq Cloud API to allow for quick and effective healthcare response generation. It uses an API Key for Groq Cloud API to be able to authenticate and access the Llama-3 Vision (Multimodal LLM) model. This API Key helps the application send input data (symptom text and medical images) for the Groq inference engine to process and return healthcare recommendations quickly. The Groq API key (gsk- style secret key) is used for authentication purposes on the server side to provide secure communication between the Gradio UI and AI model.

C. Gradio-Based User Interface

The Gradio User Interface in the proposed Multi-modal AI Healthcare Chatbot serves as the interactive user interface for symptom entry and image uploading. The process begins when the user enters his/her symptom information along with the option to upload any medical images through the Gradio interface. After this, the system forwards the entered information to the multi-modal LLM (Llama 3 vision) to analyze the information provided by the user along with performing the AI inference with Groq. The model performs the necessary analysis on the entered information for identifying health conditionsAfter the analysis, the result of the prediction of health condition, its severity level, and related medical advice gets displayed on the Gradio User Interface.s, as well as classifying their severity levels.

D. ImageInput Processing

The system uses image input processing to analyse images uploaded to the medical portal. The image analysis is performed using artificial intelligence-based multimodal reasoning, allowing the system to understand visible symptoms and abnormalities from the images.

E. Rule-BasedSeverity Detection and Recommendation Logic

Following the AI analysis, the system then applies a rule-based severity detection to identify the gravity of the identified problem. Severity can be classified into three categories: mild, moderate, and severe. Using the severity level and the output from the AI, the recommendation logic provides suggestions for the suitable doctors.

F. Text-to-Speech (TTS) Response Generation

The TTS is utilized in the suggested system by incorporating gTTS and ElevenLabs for translating the generated healthcare advice into audio format. Once the Multimodal LLM (Llama 3 Vision) evaluates the patient input and outputs the medical guidance, the generated text data is then transferred to the TTS component. The use of gTTS (Google Text-to-Speech) technology translates the text data into speech audio by creating an audio file from the response text. Moreover, ElevenLabs is incorporated to generate human-like voice synthesis. The generated audio data is then outputted on the user interface where the patient can listen to their healthcare guidance.

IV. RESULTS AND EVALUATION

The findings reveal that the AI Healthcare Chatbot using the Multi-modal Large Language Model (MLLM) performed better in classification of the conditions in health care than the Large Language Model (LLM). The Multi-modal Large Language Model (MLLM) accurately recognized more cases under the Mild, Moderate, and Severe categories. In particular, the Multi-modal Large Language Model (MLLM) classified 45 Mild, 42 Moderate, and 43 Severe cases, while the Large Language Model (LLM) classified 42 Mild, 38 Moderate, and 39 Severe cases.

A. Performance Evaluation

Metric	Existing Model	Proposed Model
Accuracy	85%	92%
Precision	83%	90%

Recall	84%	91%
Sensitivity	86%	93%
Specificity	82%	89%

Table 1:Comparition between LLM and MLLM

As shown in table 1, performance comparisons between the LLM-based model and the Multimodal LLM (MLLM)-based AI healthcare chatbot reveal improvements for all evaluation criteria. It is clear from the results obtained from the performance comparisons that Multimodal Healthcare AI, Symptom Analysis, Severity Classification, and Medical Recommendation System lead to better system performance.

1) Accuracy

The suggested Multimodal LLM (MLLM) model has an accuracy rate of 92% as opposed to the existing LLM model which had an accuracy rate of 85%, hence showing an improvement rate of 7%.

2) Precision

The MLLM model has a precision rate of 90%, whereas the existing LLM model has a precision rate of 83%, indicating an improvement of 7%.

3) Recall

The MLLM model has a recall rate of 91% compared to the existing LLM model which had a recall rate of 84%, showing an improvement rate of 7%.

4) Sensitivity

The suggested MLLM model has a sensitivity rate of 93%, whereas the existing LLM model has a sensitivity rate of 86%, showing an improvement rate of 7%.

5) Specificity

The MLLM model has a specificity rate of 89%, whereas the existing LLM model has a specificity rate of 82%, indicating an improvement rate of 7%.

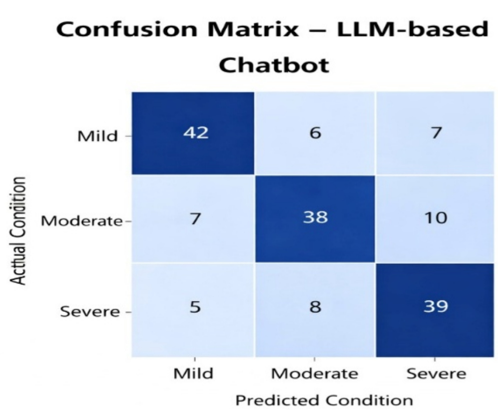


Fig2 :Confusion Matrix for LLM

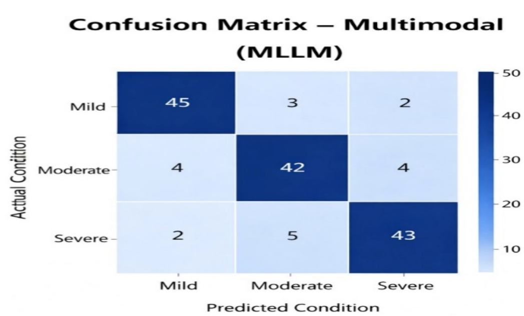


Fig 3:Confusion Matrix for Multimodal

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

The classification capability of both LLM-based Chatbot and Multimodal LLM is shown in the above Fig 2, Fig 3 for Mild, Moderate, and Severe illnesses. While the LLM model was able to successfully classify 42 cases of Mild condition, 38 cases of Moderate, and 39 cases of Severe condition, the results show that the LLM model had greater misclassification rates within each of these severity levels. On the other hand, the MLLM model was able to successfully classify 45 cases of Mild illness, 42 cases of Moderate, and 43 cases of Severe conditions while having relatively fewer misclassifications.

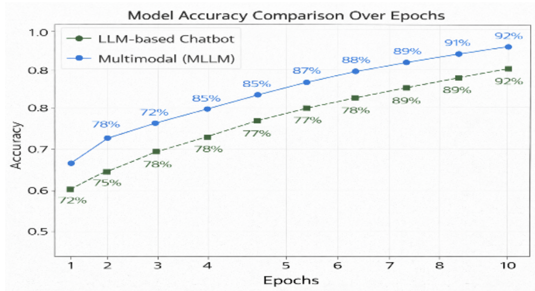


Fig 4: Model Accuracy Comparison OverEpochs

The figure below (fig. 4) represents the comparison of accuracy for both models with respect to the number of epochs. Both the models improve in accuracy as the number of epochs increases, but the Multimodal (MLLM) model outperforms the LLM-based model throughout. The MLLM model has an accuracy of 92%, whereas the LLM model has an accuracy of 85%. The improved performance is attributed to the multimodal inputs in the model, including symptoms text and medical images.

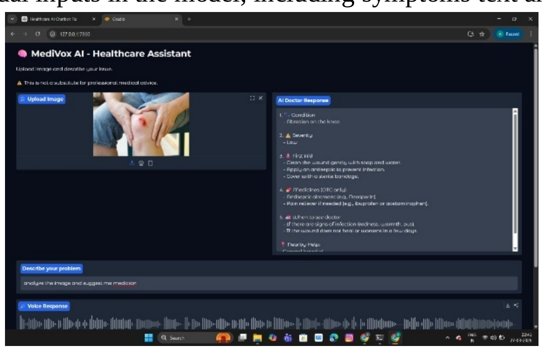


Fig5 :Application interface

Figure 5 represents the Gradio-based Multimodal AI Healthcare Chatbot interface wherein patients can upload their medical images and symptoms. The proposed solution processes the data provided by the users through the use of multimodal LLM (Llama-3 Vision via Groq), gives health recommendations powered by artificial intelligence, and speaks to users via the TTS method.

V. CONCLUSION

The Multimodal AI Healthcare Chatbot is designed to offer intelligent real-time healthcare support using Multimodal LLM-based reasoning, symptom detection, image detection, severity classification, doctor recommendation, and Text-to-Speech (TTS). It uses the inputs of the symptoms and medical images to diagnose the health issues and recommend the right medical advice, whereas the severity classification helps make informed decisions and prioritize severe cases.

According to the evaluation results, the Multimodal AI Healthcare Chatbot significantly outperforms the current LLM-based model. The system achieves an accuracy rate of 92%, a precision of 90%, a recall of 91%, a sensitivity of 93%, and a specificity of 89%. Hence, there is an average 7% improvement across all measures.

REFERENCES

[1] Ayers, J. W., Poliak, A., Dredze, M., et al. (2023). Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Internal Medicine

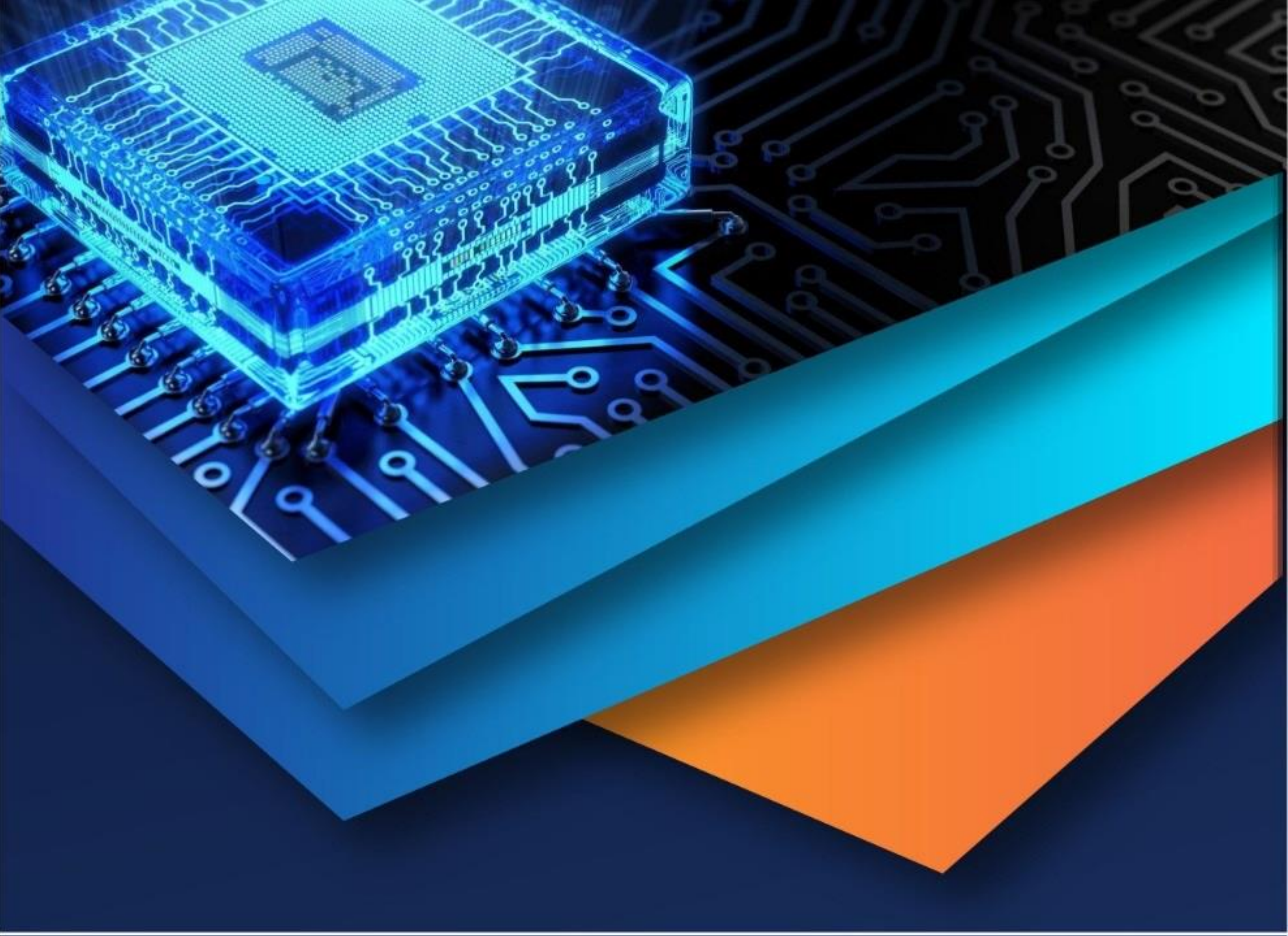
[2] Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review. Healthcare (MDPI), 11(6), 887

[3] Cascella, M., Montomoli, J., Bellini, V., et al. (2023). Evaluating the Feasibility of ChatGPT in Healthcare: Systematic Review. Healthcare (MDPI), 11(4), 547

[4] Kung, T. H., Cheatham, M., Medenilla, A., et al. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted Medical Education. PLOS Digital Health

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

[5] Thirunavukarasu, A. J., Hassan, R,Mahmood, S., et al. (2023). Trialling ChatGPT in Healthcare: Review of Applications. BMJ Health & Care Informatics
[6] Lee, P., Bubeck, S., Petro, J. (2023). Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. New England Journal of Medicine
[7] Nadarzynski, T., Miles, O., Cowie, A., et al. (2022). Acceptability of Artificial Intelligence-Led Chatbots in Healthcare. Digital Health Journal
[8] Kocaballi, A. B., Quiroz, J. C., Rezazadegan,D.,etal. (2022). Conversational Agents for Healthcare: Systematic Review. npj Digital Medicine WHO (2023). Ethics and Governance
[9] Gilson, A., Safranek, C. W., Huang, T., et al. (2023)How Does ChatGPT Perform on the United States Medical Licensing Examination?PLOS Digital Health
[10] Singhal, K., Azizi, S., Tu, T., et al. (2023). Large Language Models Encode Clinical Knowledge. Nature.
[11] Patel, S. B., & Lam, K. (2023). ChatGPT: The Future of Medical Education and Clinical Practice? The Lancet Digital Health.
[12] Rao, A., Kim, J., Kamineni, M., et al. (2023). Evaluating ChatGPT as an Adjunct for Radiologic Decision-Making. Radiology.
[13] Meskó, B., & Topol, E. (2023). The Imperative for Artificial Intelligence in Healthcare. npj Digital Medicine.
[14] Anjum, Hiba, B. Sasidhar, and T. Rajesh. "Lung Nodule Segmentation and Classification using Image Processing and Deep Learning Techniques." Grenze International Journal of Engineering & Technology (GIJET) 11 (2025).
[15] Medical Chatbot Dataset, Kagglehttps://www.kaggle.com/datasets/saifulislamsarfaraz/medical-chatbot-dataset
[16] Symptom-Based Disease Prediction Dataset, Kaggle https://www.kaggle.com/datasets/miltonmacgyver/symptom-based-disease-prediction-dataset
[17] Skin Disease Image Dataset, Kaggle https://www.kaggle.com/datasets/ismailpromus/skin-diseases-image-dataset



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY