

Multimodal AI-Driven Telemedicine Assistant with Dermatology Image Analysis

Varshil Bhovani¹, Vraj Patel²,

¹ Computer Engineering, Gandhinagar Institute Of Technology, Gandhinagar University Gujarat, India.

² Computer Engineering, Gandhinagar Institute Of Technology, Gandhinagar University Gujarat, India.

Abstract - We built a telemedicine assistant that combines a Large Language Model with a skin image classifier, mainly because we noticed that most existing platforms simply do not help patients describe what is wrong before the doctor joins. The system takes text, voice, and skin photographs as input, replies in multiple Indian languages, and keeps a structured summary ready for the doctor. For the image side, we trained a MobileNet-v2 model on eight dermatology conditions and reached an overall accuracy of 0.98 on our test set. The conversational part runs on Gemini and generates follow-up questions, safe explanations, and brief clinical notes. We tested both components together as a working telemedicine product, and responses came back within 1.5 to 2.5 seconds for text and 3 to 5 seconds when an image was also involved. Throughout everything, we were careful to keep the AI in a supporting role. The doctor still makes every medical decision; the system just helps them get there faster.

1. INTRODUCTION

When we started thinking about this project, one thing stood out immediately. Patients in rural areas or smaller towns often cannot easily travel to see a specialist. Telemedicine helps with that, and over the past few years a lot of platforms have come up offering video consultations. But when we actually looked at how those platforms work, most of them are basically just video calls with a scheduling layer on top.

There is no help for the patient to describe symptoms properly, no intelligent triage, and doctors end up asking the same questions at the start of every call. We thought that was a problem worth trying to solve, and that is what pushed us toward building something that goes a step further.

2. Body of Paper

We reviewed quite a few papers on telemedicine systems before starting and the pattern was consistent. Platforms improve access, they let you book an appointment and do a video call, but they stop there. Patients still have to explain everything from scratch each time, and doctors still spend the first few minutes of a consultation just gathering basic information. Researchers in [1] and [9] specifically pointed out that telemedicine tools need better pre-consultation automation, but as far as we could see, not many actual products had implemented that idea yet.

Table -1: Classification Report for the Eight Dermatology Classes

Class	Precision	Recall	F1-Score	Support
0	0.97	0.97	0.97	34
1	1.00	1.00	1.00	20
2	0.94	0.97	0.95	32
3	0.97	1.00	0.99	33
4	1.00	1.00	1.00	23
5	1.00	0.88	0.94	25
6	1.00	1.00	1.00	34
7	0.97	1.00	0.99	33
Accuracy			0.98	234
Macro Avg	0.98	0.98	0.98	234
Weighted Avg	0.98	0.98	0.98	234

The classifier did better than we honestly expected. On our test set of 234 samples, overall accuracy came out at 0.98 and both the macro and weighted F1 scores matched that. Most classes had precision and recall above 0.94, which tells us the model was not just doing well on the easy cases. Class 5 had the lowest recall at 0.88, and looking at the confusion matrix that class was being confused with two others that have similar textures. It is the kind of mistake that even a dermatologist might make from a photo alone.

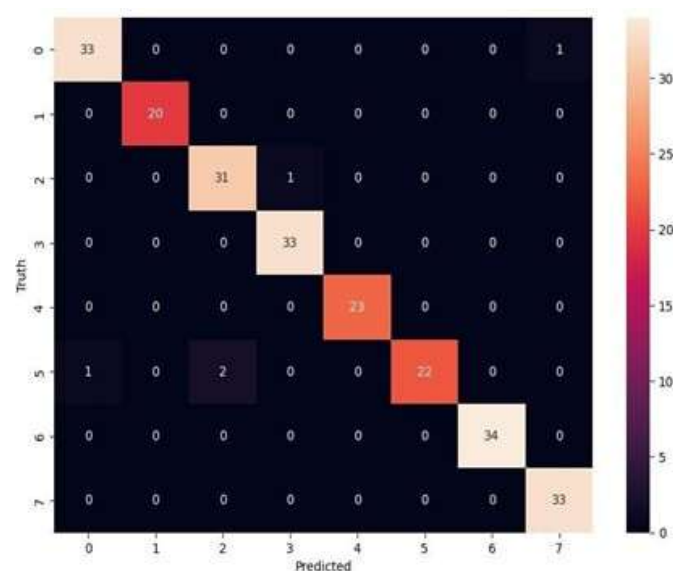
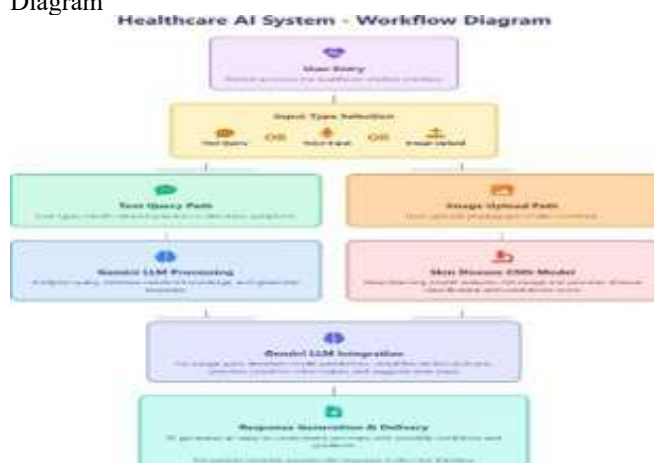


Fig -1: Figure

Diagram



3. Experiment & Result

We tested the system in two phases. First, each component, the CNN, the LLM, and the backend APIs, was evaluated on its own. Then we ran end-to-end tests that simulated realistic patient-doctor interactions. We focused on three things: how accurately the classifier performs, how quickly the conversational agent responds, and how stable the video and chat features are under repeated use. Running component tests first made it much easier to trace problems when the integrated version had issues.

4. Discussion

Looking back at how the system performed overall, we think the combination of image classification and conversational AI genuinely adds value to a telemedicine workflow. Patients who uploaded images got a preliminary explanation before the doctor joined, which gave them something concrete to refer to during the call. The multilingual support mattered more than we initially expected. During testing we noticed that responses in the patient's own language felt noticeably more natural and were better received than English responses that were grammatically correct but stilted.

5. Training

We made sure the class distribution was balanced across training, validation, and test splits. Early in our experiments we had an imbalanced split and the accuracy looked great but the model was basically ignoring the smaller classes. Once we balanced it, the numbers were more honest. We normalized pixel values to [0, 1], used Adam with categorical cross-entropy, and monitored validation loss carefully. Early stopping helped us avoid training runs where the model kept going even after it had already peaked.

- **Not a Diagnostic Tool:** The AI only assists with triage. It does not and should not diagnose anything, and we made that clear to users at every step.
- **Limited Dermatology Dataset:** The CNN covers only eight conditions and we already noticed its confidence drops for poorly lit or oddly angled photos. A broader and more diverse dataset would help.
- **Internet Dependency:** Everything depends on a stable internet connection. In areas with unreliable connectivity, the video and AI features would degrade, which is a real problem for the rural use case we originally had in mind.
- **No EHR Integration:** We have no access to hospital records or lab results. Patient profiles only contain what was entered on our platform, which is a serious limitation for any real clinical use.
- **Voice Input Accuracy:** Voice input works reasonably well but is not reliable in noisy environments or with strong regional accents. The transcription errors sometimes affected the quality of the LLM response downstream.
- **Single-Specialty Focus:** Right now the AI side only handles dermatology. Adding other specialties would require new models and significant changes to the prompt structure.

6. Future Work

- **Integration with Hospital EHR Systems:** EHR integration is the biggest missing piece. Without access to hospital records and lab history, the doctor still has an incomplete picture.
- **Expansion Beyond Dermatology:** We would want to add models for other specialties. Dermatology was a starting point but respiratory conditions, dental issues, or X-ray analysis would all be valuable additions.
- **Stronger Data Security and Compliance:** Before any real clinical deployment we would need proper encryption, audit trails, and compliance with healthcare data regulations. We did not implement that level of security for a prototype.
- **Improved Voice Interaction:** The voice input needs to be more robust. We would look at better speech recognition models and test more carefully with diverse accents and noisy environments.
- **Mobile Application Development:** A mobile app would make a big difference for the rural use case. Many patients we had in mind do not use laptops regularly, and a well-designed app would remove that barrier.

Doctor Assistance Tools: We would also want to give doctors better prioritization tools, something like a risk indicator or a trend graph for recurring symptoms so they can triage a full appointment list more efficiently.

7. Evolution Setup

We tested the LLM with a range of symptom descriptions, multilingual queries, and multi-turn conversations. Text responses consistently came back within 1.5 to 2.5 seconds, and when an image was also being processed the combined response took 3.0 to 3.5 seconds. That felt acceptable for a telemedicine context where the patient is expecting to wait a moment. Multi-turn conversations maintained context well across several exchanges. Most importantly, across all our test cases the model stuck to the safe response format and never slipped into language that sounded like a diagnosis.

8. Work Flow

The flow we designed tries to handle whatever the patient sends and always produce a single clear response at the end. Here is how it works step by step:

- The patient enters text, speaks through the microphone, or uploads a skin image.
- Text and voice inputs are routed to the LLM for interpretation and conversation handling.
- Image inputs are processed by the CNN, which outputs a predicted skin condition along with a confidence score.
- Both outputs are passed to the fusion layer, where the LLM uses the CNN result to generate a clear and safe explanation.
- The final response is sent back to the patient in the selected language.
- A summarized version of the conversation is stored for the doctor to review before or during the consultation.

We were deliberate about keeping the AI out of the decision seat. The patient gets a safe, helpful response. The doctor gets a clean summary. Nobody gets a diagnosis from the machine.

REFERENCES

- [1] "Multimodal Artificial Intelligence in Medicine," PMC, NIH.
[Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12282626/>
- [2] "Multimodal Artificial Intelligence in Medical Diagnostics," MDPI.
[Online]. Available: <https://www.mdpi.com/2078-2489/16/7/591>
- [3] "Ensemble Deep Learning and LLM-Assisted Reporting for Automated Skin Lesion Diagnosis," arXiv.
[Online]. Available: <https://arxiv.org/html/2510.06260v1>
- [4] "AI-Driven Enhancement of Skin Cancer Diagnosis: A Two-Stage Voting Ensemble Approach Using Dermoscopic Data," PMC.
[Online]. Available:

<https://pmc.ncbi.nlm.nih.gov/articles/PMC11720667/>

[5] “AI Product Disclaimer Essentials: Crafting Strategies & Legal Considerations,” LILM.

[Online]. Available:

<https://lawinfolegalmarketing.com/ai-product-disclaimer-essentials-crafting-strategies-and-legal-considerations/>

[6] S. A. Rabbani, M. El-Tanani, S. Sharma, S. S. Rabbani, Y. El-Tanani, R. Kumar, and M. Saini, “Generative Artificial Intelligence in Healthcare: Applications, Implementation Challenges, and Future Directions,”

BioMedInformatics, vol. 5, no. 3, p. 37, 2025.

doi: 10.3390/biomedinformatics5030037

[7] D. Patel, G. Raut, S. N. Cheetirala, B. Glicksberg, M. A. Levin, G. Nadkarni, R. Freeman, E. Klang, and P. Timsina,

“AI Agents in Modern Healthcare: From Foundation to Pioneer - A Comprehensive Review and Implementation Roadmap for Impact and Integration in Clinical Settings,”