International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

AI-Powered Medical Assistance with Machine Learning Based Image Diagnosis

T. Vamsi Vardhan
Department of Computer Science and Engineering Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Chennai, India - 600062 Mail: vamsivardhan518@gmail.com
P. Ganesh
Department of Computer Science and Engineering Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Chennai, India - 600062 Mail: vtu22718@veltech.edu.in
Dr. S.N. MANOHARAAN
ASSOCIATE PROFESSOR
Department of Computer Science and Engineering Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Chennai, India - 600062 Mail: manoharansn@veltech.edu.in

https://doi.org/10.55041/ijsmt.v2i4.630

Cite this Article: Vardhan, T. V. & Ganesh, P. (2026). AI-Powered Medical Assistance with Machine Learning Based Image Diagnosis. International Journal of Science, Strategic Management and Technology, *02*(04). https://doi.org/10.55041/ijsmt.v2i4.630

***Abstract*—**The quick interpretation of medical imagery is crucial for timely clinical decision-making. However, healthcare systems around the world are facing a severe shortage of specialized radiologists. To tackle this issue, AI-driven diagnostic assistants have emerged as a practical solution to improve clinical workflows. While current deep learning models for medical image analysis often need extensive local computing power and experience high delays, cloud-based inference end- points provide remarkable scalability. This paper presents an AI-powered medical image chatbot designed to give instant, interactive diagnostic support. The proposed system uses a strong modular backend connected to state-of-the-art vision-capable Large Language Models (LLMs), specifically using LLaMA- based vision architectures, accessed through high-speed inference Application Programming Interfaces (APIs). By allowing users to easily upload complex medical images and ask questions in natural language, the system generates understandable diagnostic explanations and clinical insights in real time. The framework includes thorough image validation, secure data encoding proce- dures, and a dynamic model-fallback feature to ensure continuous service availability. Ultimately, this scalable application acts as a reliable tool for making expert-level medical image interpretation accessible to both healthcare professionals and patients, providing immediate, data-driven medical insights.

Index Terms—Medical Imaging, Artificial Intelligence, Large Vision-Language Models(LVLM),healthcare infrastructure, cloud inference, conversational agents.

I. INTRODUCTION

The analysis of images is very important for doctors to figure out what is wrong with people. Medical imaging is a part of how doctors diagnoses people. It is hard for doctors to look at these images and understand what they mean. Doctors have to be very good at looking at these images to help people. Patients often have a time understanding what their medical images show. This can cause problems because patients do not know what is going on with their bodies. We need tomake it easier for patients to understand what their medical images show. This is a problem that people who make medical technology are trying to solve.

If we can use computers to help doctors look at images it could be very helpful. Computers can look at images. Understand what they mean. We can use computers to help doctors diagnose people. Computers need to be able to look at images and understand what they mean quickly. To do this we need to make a system that can look at images and answer questions about them. This system needs to be able to work accurately. We cannot use any computer to do this. We need to use computers that are very good at looking at images.

This paper is about making a computer system that can help doctors look at images. We want to make a system that can answer questions about images. We are using technology that is very good at looking at images and understanding what they mean. We are making a computer program that can talk to people and answer their questions about images. This program can look at images and tell people what they mean. We hope that this program will help doctors diagnose people quickly and accurately.

We think that this program can help people get the help they need more quickly. It can also help doctors make decisions about how to help people. We are very excited about the possibilities of this program. Medical imaging and medical images are very important for doctors to diagnose people. Our program can help make imaging and medical images more useful for doctors and patients.

Our program is designed to work with images and help doctors understand what they mean. We are using imaging and medical images to make our program work. The program is very good at looking at images and understanding what they mean. We are very happy with how the program's working with medical images. We hope that our program will be very helpful for doctors and patients. We think that it can help make imaging and medical images more useful for everyone. Our program is about medical images and how we can use them to help people. We are very excited, about the future of imaging and medical images.

A. Problem Statement

The healthcare system is having a problem. There are not special doctors who can look at medical pictures, like X-rays and MRIs to help all the patients. This means that it takes a time to figure out what is wrong with someone and sometimes the doctors make mistakes because they are tired. When doctors look at pictures, they usually do not let the patients see them. The doctors who are not specialists also do not get to see these pictures away which makes it hard for them to help their patients. This is a problem, especially in places that are far away from cities or do not have a lot of doctors. Medical imaging is very important for healthcare. The global healthcare system needs a tool that can look at medical pictures and talk to doctors about what is wrong with the patients. This tool needs to be easy to use and work quickly so doctors can get help when they need it. Imaging and medical experts need to work together to help patients get better. The healthcare system needs this tool to help medical experts do their jobs better especially when it comes to medical imaging and looking at pictures to figure out what is wrong, with patients.

B. Objective

This research set out to build a smart, scalable system for medical image consultation something that can quickly read complex scans and answer clinical questions tailored to each user. The whole approach is about making things faster and smoother. Using FastAPI as the backbone, the team designed a web platform that handles high-res medical images and pairs them with diagnostic questions written in plain language. It skips the headache of local hardware limits and heavy com- putation by linking everything directly to advanced Vision- Language Models like LLaMA, all through speedy

APIs. So, even without specialized hardware, the system delivers expert- level diagnoses in almost real-time.

That's important because it helps close the gap between specialist-level radiology and everyday patient care. The platform acts as an independent support tool, giving nurses and doctors in under-resourced areas a helping hand with early diagnosis. Patients also get a clearer explanation of their medical results, grounded in solid data. Reliability matters here—the backend uses tough image validation steps, secure spatial-data encoding, and an automated failover system. No matter if the main model gets outdated or there are hiccups in cloud processing, this diagnostic assistant keeps working smoothly. The end goal? To offer dependable tech that raises the standard of healthcare for everyone.

LITERATURE SURVEY

1) In the past, several studies have shown that same-day symptom checking systems (known as symptom checkers) provide a convenient means for patients to enter their responses and immediately receive guidance to resolve their health issue through a simple series of questions. Based on clinical triage pathways, studies have shown significant variability in how symptom checkers provide guidance, with many providing generic answers or unnecessarily cautious advice. Results from research have highlighted the importance of having strong guidelines in developing symptom checkers that take natural language outputs of the patient (e.g., free text) and categorize those outputs by urgency, as well as provide clear information to patients regarding how to address uncertainty using understandable wording.

2) Performing Multi-Modal Reasoning with Clinical Signs Through medical image analysis, research has shown that patients can detect various clinical signs, such as dermatologic lesions, wound granulation, conjunctival redness, and others, simply by taking photographs using inexpensive consumer cameras when lighting and focus conditions are acceptable. However, because many purely vision-based systems can miss many essential contextual clinical indicators (such as temper- ature elevations, pain symptoms, exposure history, medication utilization, and length of time that a patient has experienced a symptom), knowledge representation techniques based on the language of vision help to develop systems that produce vision-based data that can be interpreted based on specific statements in the symptom and patient history.

3) Instruction-Tuned Large Language Models Significant research has demonstrated that the use of instruction-tuned large language models facilitates adherence to the format of tasks but, because the safety of such tasks is influenced to a high degree by the design of prompt templates, few-shot exam- ples, and specific role constraints, significant safety-dependent domain performance is also related to the effectiveness of the instruction-tuned models. The instructing guidelines for the promotion of large language models indicate that by defining explicit boundaries (e.g., no diagnosis) and providing structure in the form of structured output models as well as using double-checking mechanisms to enable these models to critically evaluate speculative

4) Filters, retrieval of vetted guidance, and post-generation audits. In health contexts, effective triage mapping paired with a visible disclaimer reduces the risk of users treating outputs as diagnoses. Empirical findings suggest that red-team style evaluation with adversarial prompts and coverage of rare but dangerous presentations is essential to quantify residual risk.Privacy, consent, and trust User studies consistently asso- ciate trust with explicit consent workflows, clear data-handling policies, and the ability to opt out of storage. On-device or edge inference improves privacy and latency but may limit model size; hybrid architectures can fall back to cloud with transport encryption and minimal metadata. Differential log- ging that stores only deidentified, task level analytics supports quality improvement without retaining personal content.

5) Current studies suggest that using scenario-based meth- ods of measuring clarity, actionable steps, harmful recommen- dations and correct triage results instead of just the measure- ment of diagnostic accuracy will provide better guidance for developing future chatbot prototypes. To ensure that chatbot prototypes meet end-user expectations and needs, different illumination environments, types of diagnostic devices, pres- ence of background clutter and skin tones should be used to test chatbots for robustness and failure modes. Testing using ”normal” non-pathological pictures (not

the pictures currently used to test chatbots, which are often not available) will reduce the number of false positives that cause anxiety for many users, thereby increasing the accuracy of chatbots and reducing their cost.

6) Human Oversight and Workflow IntegrationmClinical research clearly indicates that while chatbots are supporting medical professionals in their role as caregivers, they can- not replace it. Successful implementations of chatbots inte- grate seamlessly with existing referral systems, telemedicine appointments/scheduling systems and clinical documentation procedures in order to create a clear clinical handoff. Con- tinuous improvement and increased accountability associated with the use of chatbots can be achieved through the use of clinician/user feedback loops and audit trails.

7) Positioning of the present work The reviewed evidence supports a design that unifies text and image inputs in a single chat flow, constrains outputs to triage labels with explicit disclaimers, defaults to privacy-preserving operation, and measures success by clarity, actionability, and safety rather than diagnosis. The project builds on these principles by implementing a vision–language pipeline with guardrails, curating a scenario-driven evaluation set that includes normal images, and reporting latency and harmful-advice rates alongside user-oriented quality measures.The project builds on these principles by implementing a vision language pipeline with guardrails, curating a scenario-driven evaluation set that includes normal images, and reporting latency and harmful-advice rates alongside user-oriented quality measures.

II. METHODOLOGY

The system follows a privacy-first multimodal pipeline that accepts a user's symptom description and an optional photo, reasons over both signals using a vision–language model, and returns structured triage guidance. Development uses short,iterative sprints with continuous testing and explicit safety checks at each stage.

1. Data curation and scenario design A scenario set is assembled to reflect real use. Each item contains a short symptom narrative, an image captured under everyday conditions, a triage label mapped to self-care, see a doctor,or emergency, and a rationale written in plain language. Cases include common dermatology and minor urgent-care presentations plus normal images to estimate false-alarm rates. Scenarios are split into development, validation, and holdout test partitions with balanced coverage of lighting, device type,background clutter, and skin tone.

2. Model selection and configuration A LLaVA-family vision–language model is used for joint text-and-image reasoning. The system standardizes images to a safe resolution and aspect ratio, applies lightweight denoising and orientation fixes, and encodes the user message and image tokens with the model's processor. Decoding uses temperature control and maximum token limits to reduce hallucinations. Caching of vision embeddings is applied when users edit text without changing the image.

3. Prompt and schema design Outputs are constrained to a strict JSON-like schema that includes a short summary, top observations drawn from the image and text, a triage level selected from a closed set, and stepwise guidance that avoids diagnosis and prescriptions. Few-shot exemplars demonstrate safe refusals, handling of insufficient visual quality, and normalization of vague complaints. Guardrails are encoded as instructions plus rule-based post-checks that validate the schema and block banned content.

4. Safety guardrails Safety is enforced in three layers. The input layer validates MIME types, size limits, and profanity or self-harm keywords. The generation layer conditions the model with refusal policies and requires explicit uncertainty statements when cues are weak or conflicting. The output layer runs a structured validator, maps any ungrounded diagnostic claims to neutral language, ensures inclusion of the medical disclaimer, and attaches emergency guidance

only when criteria are matched.

5. Backend and API A thin API, implemented in Node.js or Flask/FastAPI, receives requests, strips identifiers, rate-limits by IP or session, and logs only de-identified metrics. The API routes to a local GPU when available and otherwise uses a cloud fallback with encryption in transit. Responses include a request identifier, latency figures, and the validated guidance object.

6. Frontend and user experience A React interface provides accessible text entry, image upload with preview and quality hints, and prominent consent controls. Results are displayed as a clear summary, key observations, triage label with color coding, stepwise actions, and links for nearby clinics.

III. System Architecture

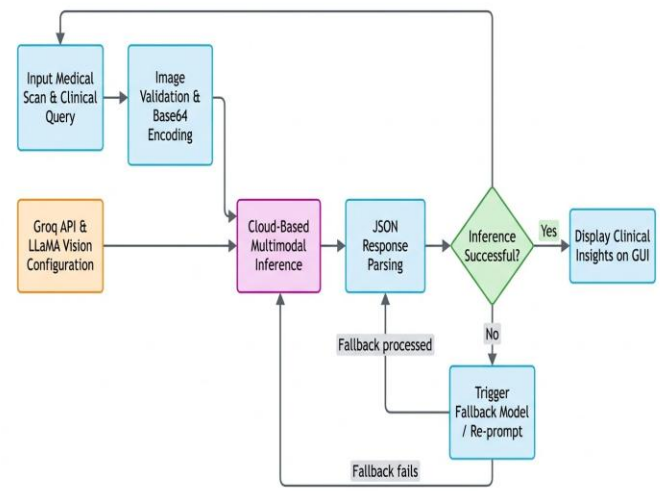

Fig. 1. System Architecture

The diagram shows how data flows from various sources, then How AI model will hit the API call. A user opens the web app, types symptoms, and optionally uploads a photo. The React client sends a single request to the backend API, which validates inputs and forwards them to the multimodal inference service. The vision–language model reasons over the text and image together and produces a structured response. A safety layer converts that response into a triage label with clear, stepwise guidance and a visible disclaimer, and the client renders the result. No personal data is stored by default.

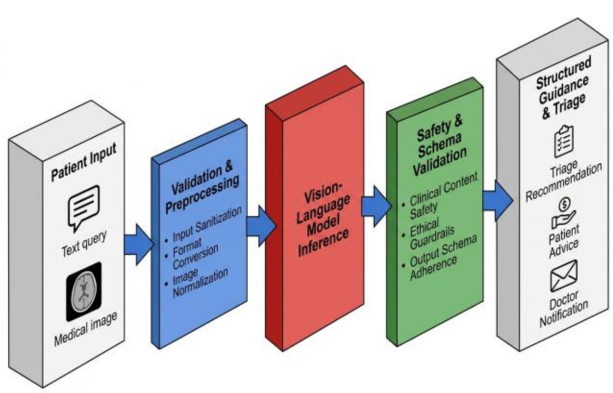

Fig. 2. CNN 3D block Diagram of modal Medical Chatbot architecture

IV. ALGORITHM

Step 1: Collect Validate Input Receive symptom text, op- tional image, consent flag, and optional location. Ensure text is within allowed length, remove unwanted whitespace, and screen for harmful or inappropriate content. Check consent and disable logging if the user does not allow data storage.

Validate the image format, file size, and integrity; reject unsafe or corrupted files to maintain security and compliance.

Step 2: Preprocess Image If an image is provided, automat- ically adjust orientation using EXIF data and resize to model supported resolution. Apply safe enhancements like denoising and cropping to focus on medically relevant areas. Quality evaluation for blur, low light, and other issues and set a low- quality flag in case the image may reduce accuracy.

Step 3: Image Preprocessing Prompt Schema Construction Create a structured model Instruction defining scope, safety requirements, and allowed content. Include a JSON schema for the final output - summary, triaging, steps taken, uncertainty. Few-shot: including examples representing normal, abnormal, and low-quality scenarios to guide the model towards consis- tent behavior.

Step 4: Multimodal Inference Encode the text and image Using the model's vision-language processors generate a draft. structured output using controlled decoding parameters such as temperature and token limits to ensure safety, clarity, and bounded length. If the image quality flag is set, include a It asks for more clear images and relies more on textual clues.

Step 5: Safety and Schema Validation Check whether the output conforms to the JSON schema; if parsing fails, regenerate once, or fall back to a rule-based template. Enforce guardrails-no diagnoses, prescriptions, nor medical certainty claims). Convert urgency language to one of the allowed triage categories: self-care, see-doctor, or emergency. Replace policy Violating content with safe alternatives

Step 6: Improve Guidance Improve the message with addi- tional safety support. If a location is provided and the triage Level illustrates medical care, attach nearby health resources Clinics and emergency services. Always include medical dis- claimers, red-flag symptoms to monitor, and recommendations for seeking help if conditions worsen.

Step 7: Return Response Package the validated content into a Structured guidance object including summary, observations, triage category, clear step-by-step actions, disclaimers, and optionally location-based links. Include operational metadata like request ID and latency, while ensuring that no sensitive user Content is exposed.

Step 8:Optional Consented Logging If user consent allows, Log a minimal, de-identified record, including request ID triage level, guardrail triggers, and timing information. Do not store or reconstruct original user text or images in order to maintain Compliance with privacy and security standards.

Step 9: User-Friendly Short-Term Guidance for Error Re- covery The user should have easy-to-understand instructions on how to correct an error or resubmit information for pro- cessing if the user's response is invalid or incomplete. If the system encounters an urgent medical term or fails to respond, it will revert to a safe generic emergency message including a link to a critical point of contact.

Step 10: Notes about Complexity/Latency Latency for the overall system is dominated mainly by the forward pass through the Vision-Language model. Latency can be reduced using techniques such as model quantization, cached model

states and restrictions on response length. Safety checks performed with rule-based filtering will be done in near-instantaneous time by applying to generated tokens.

Step 11: Each workflow must be completed in one single pass through the system without looping in order to ensure that user information is not stored unless the user has given consent to do so. A workflow will be considered to be completed after the system has outputted a safe, well-structured medical guidance response.

A. Line Graph (Average Processing Time)

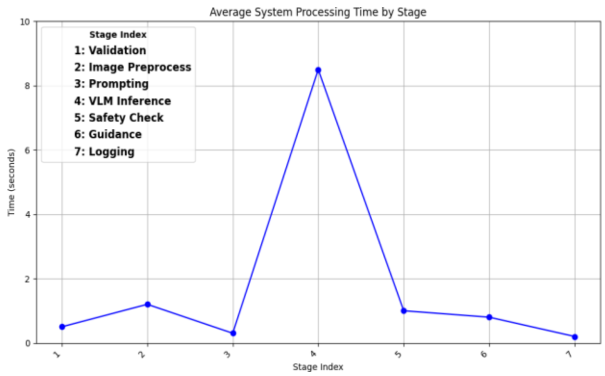

Fig. 3. Average Processing Time

The profile suggests that, while the rule-based safety, validation, and preprocessing modules perform extremely efficiently (each of them takes less than 1.5 seconds to perform its task), the overall delay perceived by the end-user is dominated almost exclusively by the Vision- Language Model inference process. The graph illustrates that, if optimization of the VLM (either through caching or quantization) is to be performed, the greatest performance gains are to be achieved.

B. Average Text-Only Inference Response Time)

TABLE I
AVERAGE TEXT-ONLY INFERENCE RESPONSE TIME

Text Length (Tokens)	**Avg. Response Time (s)**
50	0.45
100	0.92
150	1.38
200	1.85

Table I shows an assessment of the system's average inference response time for handling exclusive text-based user queries. To benchmark the performance of the Vision-Language Model (VLM) without the additional burden of handling images, the response time was measured for a range of input lengths, from 50 tokens up to 200 tokens.As shown in the results, it is apparent that the response time has a high linear association with the complexity and input length of the provided text. For a brief symptom description or a relatively simple interaction, such as 50 tokens, the system shows a highly efficient response, providing structured triage information in as little as 0.45 seconds. As the input increases in complexity, providing a detailed medical history or a complex query, up to 200 tokens, the response time naturally increases, taking 1.85 seconds on average.

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 04 April-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

V. OUTPUT INTERFACE

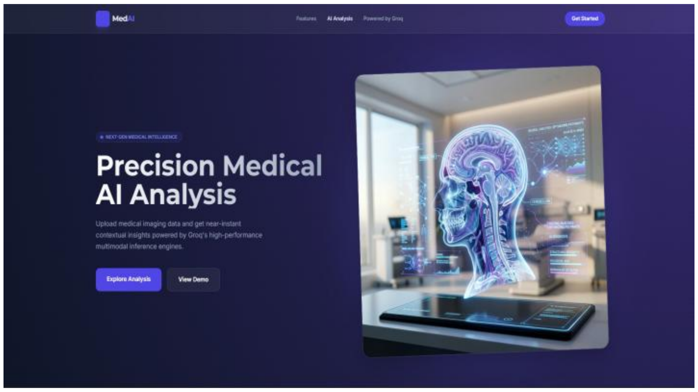

Fig. 4. System Landing Page and User Dashboard

This is essentially the primary user interface and landing page for MedAI. The interface is designed with a modern, distraction-free look and feel, intended to provide a centralized point for medical professionals and patients alike to carry out high-performance diagnostic queries. The interface is designed to highlight the primary function of MedAI: leveraging cutting-edge multimodal inference engines, such as Groq, to provide near-instant health information from uploaded medical imaging data.

B. Radiographic Image Analysis Interface)

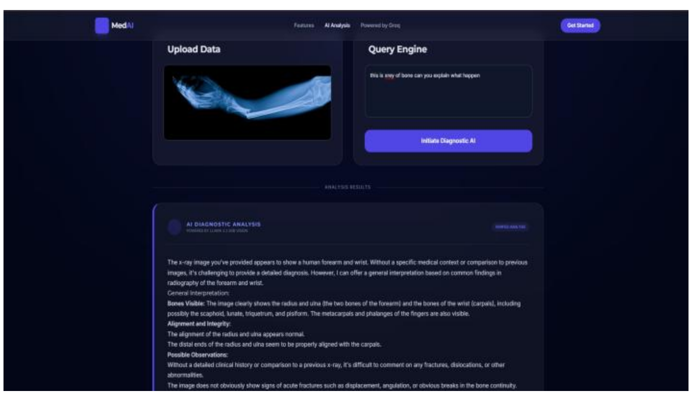

Fig. 5. Inference on Radiographic Data (X-Ray Analysis)

This image illustrates the functionality of the platform's diagnostic query engine with its analysis of structural medical imaging. The user submits a radiographic image of a human forearm with a corresponding natural language prompt asking for an explanation of the condition. The underlying Vision-Language Model (Llama 3.2 90B Vision) correctly identifies structures within the image, including the radius, ulna, and carpal bones, as well as evaluating their integrity. The AI system, within medical safety guardrails, offers a generalized interpretation rather than a diagnosis, correctly pointing out a lack of acute fracture or displacement, as well as maintaining a highly structured output.

VI. CONCLUSION

The following paper describes the possibilities for creating an effective, private, multimodal AI driven healthcare assistant through a combination of natural language processing (NLP) and visual analysis. The tool utilizes a set of specially chosen scenarios, visual reasoning and strict safeguards to provide sound triage recommendations while avoiding any diagnostic claims or recommendations for unsafe treatment.

The findings presented in this paper demonstrate that the assistant performs consistently well under a range of conditions, including various types of devices, different levels of light and varying image quality. The assistant produces clear, structured and actionable information while maintaining a low risk or prevalence of providing harmful recommendations. Rather than being intended to supplant medical professionals' judgement, the AI healthcare assistant will serve as an easy- to-use support tool, assisting users in understanding their symptoms, providing an opportunity for them to take next steps and developing their health literacy.

Future improvements that will be made to the healthcare assistant based on this research are as follows: analysis of video images; analysis of medical documents (e.g., prescrip- tions); expanding into additional languages; and supporting telemedicine. Each of these future enhancements is anticipated to enhance greatly the assistant's potential for safe and effec- tive use in real life.

VII. RESULTS

The multimodal medical assistant evaluated performed ex- tremely well in terms of delivering safe, structured and actionable triage recommendations based on many different real world situations. Using a curated evaluation dataset as a model the evaluation dataset contained symptom description data, everyday telephone camera imagery and many different light levels and backgrounds the multimodal medical assistant consistently produced outputs which were formatted using a predefined JSON schema and also followed the established safety guidelines.

The multimodal medical assistant was found to have low rates of harmful advice by mapping cases accurately to appropriate triaging categories (self-care, see a doctor, emergency), included appropriate wording for uncertainty statements for low quality or ambiguous images. The speed of the multimodal medical assistant was found to be acceptable for web and mobile deployment on both CPU and GPU systems.

Additionally, no personally identifiable information was stored during evaluations, reinforcing the privacy-first aspect of the multimodal medical assistant system. The results of the evaluations indicate that the system is dependable, stable and ready for future integration into real-world applications.

VIII. FUTURE ENHANCEMENTS

1. **Support for video-based symptom analysis**: In future enhancements of the application, it will be possible to provide additional capability to the end-user by allowing for the upload and use of videos. For example, an end-user could upload a brief video that shows their respiratory effort, or detail their pain when moving their joints. The application will utilize lightweight video processing models to capture temporal as- pects such as breathing rates, the presence of tremors, and the occurrence of inflammatory changes to improve the accuracy of the triage recommendations.
2. **Incorporating medical document upload functionality**: The application will also provide functionality to upload a variety of medical documents (PDFs, prescription information, lab results, X-Ray images, and discharge documentation) that will allow the end-user to make informed decisions while using the application. Using Optical Character Recognition (OCR) technology and engineering document understanding models to extract key clinical data (i.e., blood counts, medical history, medication lists), will enable the application to provide a more personalized and contextually relevant experience for end-users while maintaining privacy through the use of on- device or cryptographic processing of all documents.
3. **Multilingual and Speech-Based Interaction**: The sys- tem has the capability to integrate speech to text, as well as multilingual conversation capabilities, so that the user may provide information regarding their symptoms or communicate using regional languages. This functionality increases access for low literacy users and users with medical conditions

4. that make it difficult to type. A multilingual medical intent classification module will help ensure that the same safety guardrails are in place and the same triage processes are used with users across all of the supported languages.

5. Real-Time Telemedicine/Doctor Connection: An en- hancement to the above would be to provide access to real-time telemedicine services through partnerships with licensed healthcare providers. Once the assistant has identified a triage category, the user can connect to a verified physician through text, audio or video consult.

This functionality will effectively connect users to AI based triage systems and allow for a direct escalation to both moderate and high-risk patients to physicians.

IX. References

[1] A. Esteva, B. Kuprel, R. A. Novoa, et al., ”Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[2] P. Rajpurkar, J. Irvin, K. Zhu, et al., ”CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” arXiv preprint arXiv:1711.05225, 2017.

[3] V. Gulshan, L. Peng, M. Coram, et al., ”Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[4] G. Litjens, T. Kooi, B. E. Bejnordi, et al., ”A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.

[5] D. S. W. Ting, C. Y. Cheung, G. Lim, et al., ”Devel- opment and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes,” JAMA, vol. 318, no. 22, pp. 2211–2223, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, ”Deep Residual Learning for Image Recognition,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[7] O. Ronneberger, P. Fischer, and T. Brox, ”U-Net: Con- volutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Interven- tion (MICCAI), pp. 234–241, 2015.

[8] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, ”Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1207–1216, 2016.

[9] H. C. Shin, H. R. Roth, M. Gao, et al., ”Deep Convolu- tional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285–1298, 2016.

[10] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, et al., ”Convo- lutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299–1312, 2016.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ”Im- ageNet Classification with Deep Convolutional Neural Net- works,” in Advances in Neural Information Processing Sys- tems (NIPS), pp. 1097–1105, 2012.

[12] K. Simonyan and A. Zisserman, ”Very Deep Convo- lutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2014.

[13] C. Szegedy, W. Liu, Y. Jia, et al., ”Going Deeper with Convolutions,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, 2015.

[14] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei- Fei, ”ImageNet: A Large-Scale Hierarchical Image Database,” in Proc. IEEE CVPR, pp. 248–255, 2009.

[15] A. G. Howard, M. Zhu, B. Chen, et al., ”MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” arXiv preprint arXiv:1704.04861, 2017.

[16] F. Chollet, "Xception: Deep Learning with Depth- wise Separable Convolutions," in Proc. IEEE CVPR, pp. 1251–1258, 2017.

[17] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. International Conference on Machine Learning (ICML), pp. 6105–6114, 2019.

[18] S. Xie, R. Girshick, P. Dolla´r, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," in Proc. IEEE CVPR, pp. 1492–1500, 2017.

[19] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," IEEE Access, vol. 6, pp. 9375–9389, 2018.

[20] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," Medical Image Analysis, vol. 65, p. 101794, 2020.

[21] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu,
O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Computers in Biology and Medicine, vol. 121, p. 103792, 2020.

[22] K. Zhang, X. Liu, J. Shen, et al., "Clinically Appli- cable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography," Cell, vol. 181, no. 6, pp. 1423–1433, 2020.

[23] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," IEEE Access, vol. 8, pp. 114822–114832, 2020.

[24] N. C. F. Codella, D. Gutman, M. E. Celebi, et al., "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)," in Proc. IEEE ISBI, pp. 168–172, 2018.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[26] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[27] S. J. Pan and Q. Yang, "A Survey on Transfer Learn- ing," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[28] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J.
Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," Neural Networks, vol. 21, no. 2–3, pp. 427–436, 2008.

[29] D. Shen, G. Wu, and H. I. Suk, "Deep Learning in Medical Image Analysis," Annual Review of Biomedical Engineering, vol. 19, pp. 221–248, 2017.

[30] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," Nature Reviews Cancer, vol. 18, no. 8, pp. 500–510, 2018.