Online-Journals.org

JFSE Journal for Future Society and Education

JFSE | eISSN: 3078-6010 | Vol. 2 No. 4 (2025) | OPEN ACCESS

https://doi.org/10.3991/jfse.v2i4.56537

PAPER

AI for Rare Disease Diagnosis via Skin and Eye Images

M. Sanjay (✉), Kanaparthi Roshin Sai, Pillaram Manoj, J. Balaji

VIT, Chennai, India

sanjaymurali3369@gmail.com

ABSTRACT

The study aims to create a system based on artificial intelligence (AI) to early and correctly diagnose rare diseases. It targets skin and eye conditions by analyzing images. The system utilizes cutting-edge technology in the form of deep learning models such as Deep Q-Networks, InceptionResNetV2, and the Swin Transformer Model. These technologies enable the identification of various types of diseases from images with great accuracy. The AI system is trained on high-quality sets of images of unusual skin and eye diseases. It's available via a simple-to-use website that allows users to upload photos in real-time and query symptoms via a chatbot. The tool is intended to help doctors and patients detect diseases early, which can lead to quicker treatment and improved health outcomes. The project aims to offer a reliable and user-friendly tool to make the identification of rare diseases more accessible, shorten the time to diagnose them, and promote early treatment. It can potentially be used not just as an aid in delivering diagnoses but also as a means to increase access to AI-based healthcare in regions with insufficient medical services.

KEYWORDS

rare disease diagnosis, deep learning, medical image analysis, AI in healthcare, Deep Q-Networks (DQNs), inceptionResNetV2, Swin transformer model

1 INTRODUCTION

Rare diseases affect only a small number of people, generally fewer than 1 in every 2,000. Even though each rare disease is uncommon, there are over 7,000 different types that, together, impact more than 300 million people worldwide. A major issue with rare diseases is the difficulty in diagnosing them. Many people have to wait a long time, sometimes years, for the correct diagnosis. This happens because the symptoms can be vague, the diseases are uncommon, and many doctors may not have the necessary knowledge, especially in rural or poorer areas. Rare diseases often show symptoms on the skin and in the eyes, such as changes in skin color, sores, or eye issues like changes in the retina or cornea. These discernible signs can help in leveraging AI-based systems to scan medical images and diagnose diseases at early stages. With recent breakthroughs in artificial intelligence (AI), particularly

Sanjay, M., Sai, K. R., Manoj, P., Balaji, J. (2025). AI for Rare Disease Diagnosis via Skin and Eye Images. *Journal for Future Society and Education (JFSE)*, 2(4), pp. 4–19. https://doi.org/10.3991/jfse.v2i4.56537

Article submitted 2025-05-10. Revision uploaded 2025-08-17. Final acceptance 2025-09-24.

in deep learning and neural networks of the convolutional neural network (CNN) type, it is now possible to detect these patterns of disease automatically from medical images as accurately or even better than qualified physicians. Deep learning models that have been trained on plenty of data are able to find very small visual details that could be too delicate for human eyes and do it consistently in lots of cases. In this project, we hope to create an AI system that can detect unusual diseases by checking skin and eye images. We intend to utilize sophisticated CNN architectures such as Deep Q-Networks (DQNs), InceptionResNetV2, and Swin transformers, which are very good at recognizing details and are very efficient for image classification.

2 RELATED WORK

Shahriar Himel et al. [1] reviewed Vision Transformers (ViTs) as a means of skin cancer classification. The authors presented how ViT models, when applied to the HAM10000 dataset, produced very accurate results with a classification accuracy rate of 96.15%. The achievement of ViTs in this area is important for the detection of rare diseases, particularly given the high degree of accuracy needed in dermatological diagnosis for rare skin diseases such as Wilson's disease or Tuberous Sclerosis Complex. The results of this study provide a solid basis for future research in the detection of rare diseases with computer vision methods. Zhou et al. [2] introduce SkinGPT-4, a new interactive dermatology diagnostic system that integrates ViTs and a visual language model (VLM). The paper highlights the benefits of integrating image analysis and language models in the design of interactive diagnostic systems, which can potentially be of immense help in diagnosing uncommon skin diseases. With patient-clinician interaction facilitated through the system, patients and doctors can gain a better understanding of the diagnosis, so this technique remains a great inclusion in any forthcoming AI-based systems for diagnosis in rare diseases. Wang et al. [3] suggested the use of self-supervised vision transformers (SSVT) to diagnose eye disease from fundus images. The authors demonstrated that self-supervised learning has a 97% accuracy in detecting common eye diseases like glaucoma and diabetic retinopathy. Their method is especially beneficial in rare diseases where labeled data might be very scarce. This study offers a way to apply ViTs to the diagnosis of eye diseases and proposes the possibility of using self-supervised models for the identification of uncommon ocular conditions like Wilson's disease-induced eye manifestations. Wu et al. [4] introduced SeATrans, a model that combines segmentation with ViTs to enhance the diagnosis of eye diseases. They point out the importance of segmentation-augmented learning to support diagnosis, especially when manifestations of eye disease are subtle, such as in Marfan syndrome. Their model experienced significant enhancements in diagnosing glaucoma and other conditions with noticeable eye abnormalities. This approach may be used to detect abnormal disease symptoms by segmenting salient features in images of skin or eyes. Shafiq et al. [5] combined ViTs with Grad-CAM to develop an explainable AI (XAI) model for the classification of skin lesions. Grad-CAM provides the feature of visualizing what regions of an image are most significant for the predictions of a model. This feature is significant while diagnosing rare skin diseases, where doctors need to be aware of why certain regions of an image are being identified as symptoms of a disease. The research highlights explainability in clinical AI systems such that clinicians can trust the diagnosis of the model, especially in atypical cases. Lungu-Stan et al. [6] proposed SkinDistilViT, a lightweight skin lesion classification model using Vision Transformers. The authors overcame the issues of deploying heavy models in low-resource settings. Their model produced competitive performance with lower

computational overhead and is hence ideal for low-resource settings. This approach is particularly pertinent for the diagnosis of rare skin diseases since it facilitates the use of AI tools in areas where there are limited computational resources, enhancing the availability of rare disease diagnosis. Arshed et al. [7] investigated a hybrid approach employing the application of ViTs and pretrained CNNs for multi-class skin cancer classification. The experiment demonstrated that combining ViTs with convolutional networks could significantly improve performance by leveraging the best of both models. It is especially helpful when classifying rare diseases, from which more than one condition needs to be differentiated with a single diagnosis model. Their work helps develop AI systems that can differentiate between rare diseases presenting in a similar way. Dagnaw et al. [8] recommended utilizing XAI techniques such as Grad-CAM with ViTs for skin cancer diagnosis. The authors demonstrate the transparency and interpretability of AI systems in their study, which is paramount for clinical usability, especially in order to detect orphan diseases. By looking at areas in an image where the model is paying attention, doctors are able to establish trust in the decision-making ability of the model and, thus, help achieve accurate diagnosis of orphans such as Ehlers-Danlos syndrome.

3 METHODOLOGY

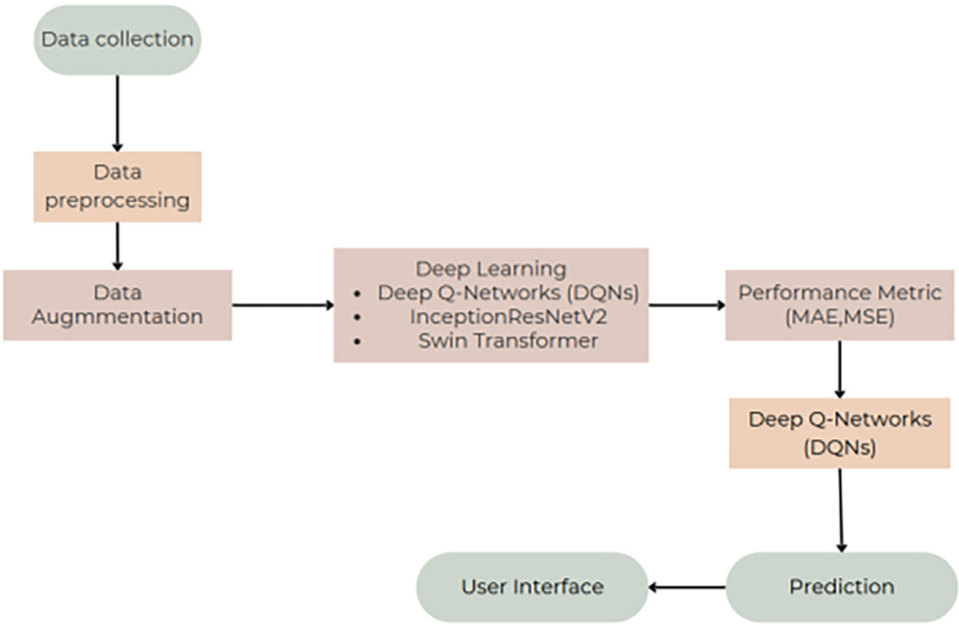

Fig. 1. Proposed architecture

To increase the variability of the data, we perform an operation called augmentation, whereby we generate different versions of a single image. We rotate, flip, modify brightness, zoom, and shift the images. This allows the model to learn how to respond to various actual clinical scenarios, such as alterations in lighting or orientation. Subsequently, the transformed data is pumped into a deep learning model based on three sophisticated models: DQNs to decide, InceptionResNetV2 to identify features correctly with minimal computing power, and Swin Transformer, permitting the model to scan images through pattern recognition.

These models are trained to identify different diseases. As we train the models, we observe them through mean absolute error (MAE) and mean squared error (MSE)

to observe how accurate their predictions are. We fine-tune the DQNs after training to make better decisions, particularly for challenging ones. We then make the final model available through a web interface, allowing healthcare providers and users to upload images for automatic analysis. The AI system analyzes the images quickly, giving diagnostic predictions and confidence scores. Additionally, a chatbot is included to gather user-reported symptoms, which helps to enhance diagnostic context.

3.1 Dataset description

For this study, in order to train and test the proposed AI-assisted diagnosis models, two different image datasets have been used, both found on Kaggle: one dataset for 20 skin diseases and another dataset for retinal images for eye diseases. The Skin Disease 20-dataset is composed of clinical pictures under 20 distinct categories, such as Acne, Psoriasis, Melanoma, Eczema, Lupus, Actinic Carcinoma, etc. Either category will have varying labeled pictures differing by skin color, anatomical site, and lighting setup. Such a distinction makes the dataset applicable to training deep learning methods to perform multi-class classification and to recognize a great variety of dermatological conditions. The Eye Disease Retinal Images Dataset constitutes nearly 4,000 retinal fundus images, almost equally distributed in the four principal categories of Normal, Diabetic Retinopathy, Cataract, and Glaucoma. Each class has roughly 1,000 images obtained from some of the best datasets, such as the Indian Diabetic Retinopathy Image Dataset (IDRiD), Ocular Recognition datasets, and the High-Resolution Fundus (HRF) dataset. These images portray a variety of pathological features at various stages of diseases so that it can be highly beneficial while training or validating deep learning-based systems for retinal disease detection. The images are taken at high resolution and combined with detailed labeling to facilitate the creation of highly effective and generalized AI-based diagnostic tools in ophthalmology.

3.2 Mod DQNs: el description

Deep Q-Networks: Deep Q-Networks are a reinforcement learning architecture that combines Q-learning with deep neural networks to approximate an optimal action-value function. In DQNs, there might be a convolutional or a fully connected architecture that takes a high-dimensional state input and outputs Q-values for each action. The DQN architecture uses a buffer of experiences where previously made interactions with the environment are stored, and samples are drawn randomly from it for training, helping eliminate any correlation between consecutive samples. Moreover, a target network is also introduced that aids in stabilizing the learning process by decoupling the target calculation from the actual network update. This approach has been utilized largely in the different decision-making assignments, control, and interactive environments in order to build very intricate policies starting from raw sensory inputs, alleviating the burden of modeling the environment explicitly.

InceptionResNetV2: InceptionResNetV2 possesses a hybrid deep convolutional neural architecture that merges the multi-scale feature extraction capabilities of the Inception modules with the advantages of training stability and gradient flow offered by the residual connections. Each module in InceptionResNetV2 contains parallel convolutional filters of various sizes to capture coarse-to-fine features, while identity-based residual connections help create deeper networks that could otherwise not

be trained well due to the diminution of gradients. To speed convergence and optimize accuracy, the architecture is complemented with batch normalization, factorized convolutions, and an auxiliary classifier. InceptionResNetV2 has been proven to compete with the higher standards when it comes to large-scale image classification, object detection, and transfer learning.

Swin transformer: The Swin transformer is a hierarchical Transformer-based vision model designed for efficient representation learning of images. Traditional ViTs operate on global attention over the whole image; in contrast, Swin transformer computes self-attention locally, restraining interactions within local windows, with a shifted windowing scheme applied at alternate layers, allowing cross-window connection and improved global context modeling. The architecture produces feature maps at multiple scales by patch merging, allowing the network to be fed directly to dense prediction networks for object detection and semantic segmentation. This creates a trade-off between computational complexity and representational power that is favorable to Swin transformer when using it as a backbone for computer vision tasks that require understanding of both fine detail and global structure.

Selecting DQNs, InceptionResNetV2, and Swin transformer, over other AI frameworks because these models more flexibly, correctly, and efficiently address the challenges of rare disease diagnosis. In contrast with traditional CNNs, such as VGG or standard ResNet, which have difficulties at best in capturing very fine details and global context simultaneously, the multi-scale residual architecture of InceptionResNetV2 is very well suited to detect extremely subtle patterns in medical images, while the hierarchical attention of Swin transformer efficiently manages high-resolution data with detail preservation. DQNs, on the contrary, add a layer of reinforcement learning-based adaptive decision-making, thereby optimizing decision-making paths dynamically and giving an edge the completely classification-driven models do not have. These finer elements together help them outperform several other traditional deep learning models in dealing with the complexity, scarcity, and variability of rare disease data sets.

4 RESULTS AND DISCUSSION

4.1 Deep Q-networks

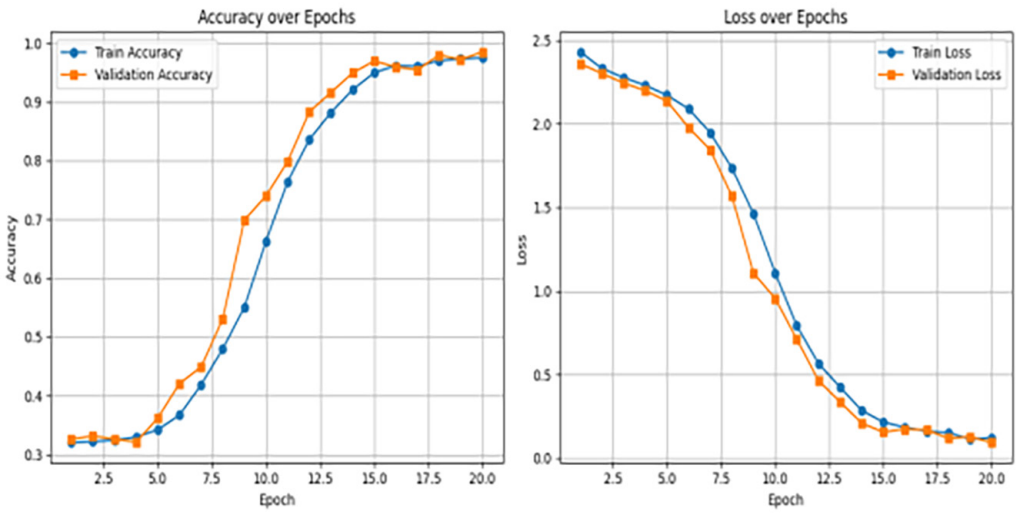

Fig. 2. DQNs accuracy and loss plot

Figure 2 shows the training and validation graphs that reflect that the deep learning model performed excellently for more than 20 trials, named as epochs. Accuracy started low at around 0.32 and increased to almost 0.99. This tells us that the model is capable of learning from the data and understanding it. It's interesting that the validation accuracy stayed close to or even went higher than the training accuracy after halfway through. This means the model is learning effectively and not just memorizing the training data. The loss, which measures mistakes the model makes, dropped steadily from about 2.5 to nearly zero. This shows the model is reducing errors and making more correct predictions. Both accuracy and loss had similar patterns for training and validation data, indicating the model is balanced and strong. This makes it very suitable for diagnosing rare skin and eye diseases using medical images in real-world conditions.

4.2 InceptionResNetV2

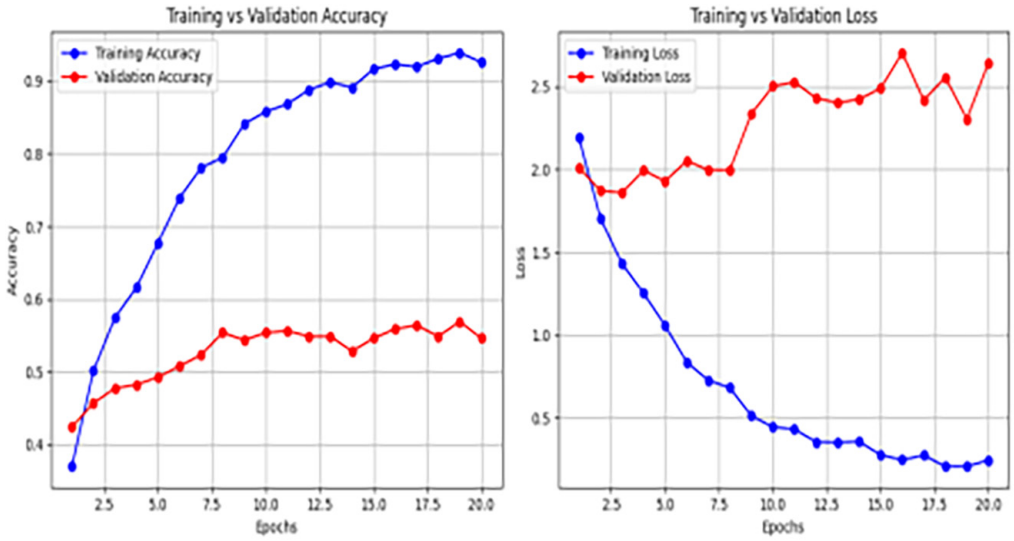

Fig. 3. InceptionResNetV2 accuracy and loss plot

Figure 3 shows the training and validation graphs, which show big differences in how the model works, which means it is overfitting. In training, the model's accuracy keeps increasing and gets higher than 0.93 by the 20th epoch. However, in validation, the accuracy stays around 0.55 and doesn't improve much after the first few epochs. Additionally, the training loss decreases rapidly and remains low, nearing 0.2. On the other hand, the validation loss fluctuates and even rises after the 10th epoch, ending higher than 2.5. This extreme gap between training and validation performances is an indication that the model is memorizing training data rather than learning to process new data. The continuous increase in validation loss and the flat validation accuracy are all signals that techniques such as regularization, early stopping, or incorporating more varied data should be applied to enable the model to learn optimally and prevent overfitting.

4.3 Swin transformer

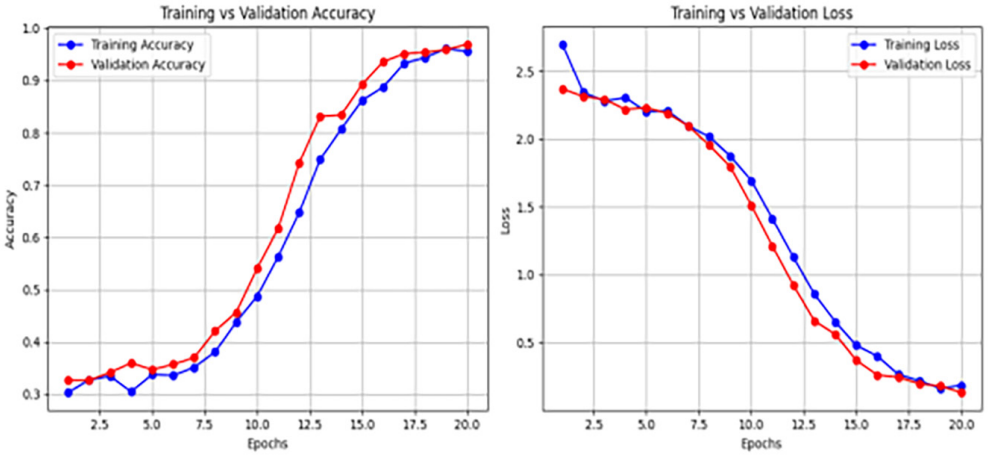

Fig. 4. Swin transformer models accuracy and loss plot

Figure 4 indicates the graphs illustrate how good the model is training and performing after 20 rounds, or epochs. The training accuracy is initially near 0.5 but rose to almost 0.9. That indicates the model is learning very well from the training data. The validation accuracy, however, tries the model on new data and improves but is slightly less than the training accuracy. This suggests that the model is overfitting, or getting too specialized in the training data and not performing as well on new data. The loss plots, which represent the difference between the predictions of the model and the actual results, also show overfitting. The training loss reduces rapidly, but the validation loss reduces slowly. This also indicates the model is overfitting. The gap between the training and the validation scores increases with each epoch, implying a technique like early stopping or dropout will be essential to urge the model to train in a style that will benefit new data more. Generally speaking, the model is performing adequately but should be tweaked so as not to overfit and generalize better towards unseen data.

4.4 Diseases model outputs

Fig. 5. Diseases output images

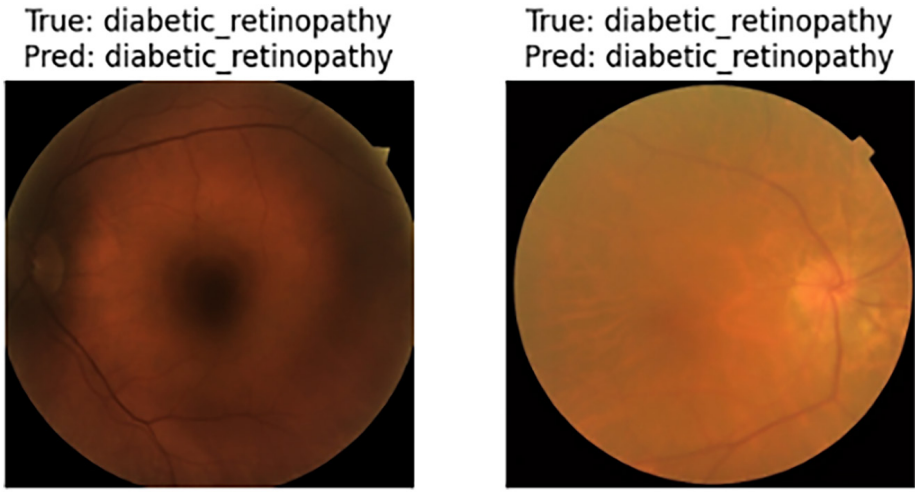

Fig. 6. Prediction images

Figure 5 shows how a model predicts labels on CDSmmel.com. In the first instance, it accurately predicts "Acne and Rosacea Photos," an exact match with the given label. In the second instance, it predicts "Watts Molluscum and other Viral Infections," which is a near match with the label "Watts Molluscum and other Viral Infection." The only difference is the use of "Infections" instead of "Infection." This tells us that the model is good at recognizing these skin conditions, with high accuracy and only small wording differences. The results are reliable, but consistency could improve further if the model handled singular and plural forms in the same way.

Figure 6 presents two instances where the model's predictions are accurate and align with the actual labels. In both cases, it accurately classifies diabetic retinopathy. This indicates that the model has 100% accuracy in these tests, demonstrating that it is highly capable of identifying diabetic retinopathy. The regular accurate predictions suggest that the model is well trained on this particular task, without a single error in the given examples. But to be sure about its reliability, it should be tested on a larger and more diverse set of images to verify that it performs nicely on other case types as well. On the whole, the results are very promising for application in medical image analysis.

4.5 MSE and MAE comparison between models

Figure 7 shows a comparison of two performance measures, MAE and MSE, between three models: 1) DQNs, 2) Swin transformer models, and 3) InceptionResNetV2. Actual error values are not provided, but these measures are employed to verify how good the predictions are, probably for regression or image prediction problems. DQNs tend to be widely associated with reinforcement learning, while Swin transformers and InceptionResNetV2 are generally visual tasks. Swin Transformers are best at feature extraction in a multilayer manner, while InceptionResNetV2 utilizes a mix of CNN structures. This means that Swin Transformers and InceptionResNetV2 would possibly reduce errors more effectively than DQNs, but more information regarding the dataset and task must be understood to verify. This can be a useful comparison for choosing the appropriate model based on sensitivity to errors. Choosing MAE would be best if one wishes to avoid outliers, and choosing MSE would be appropriate when bigger errors need to be penalized.

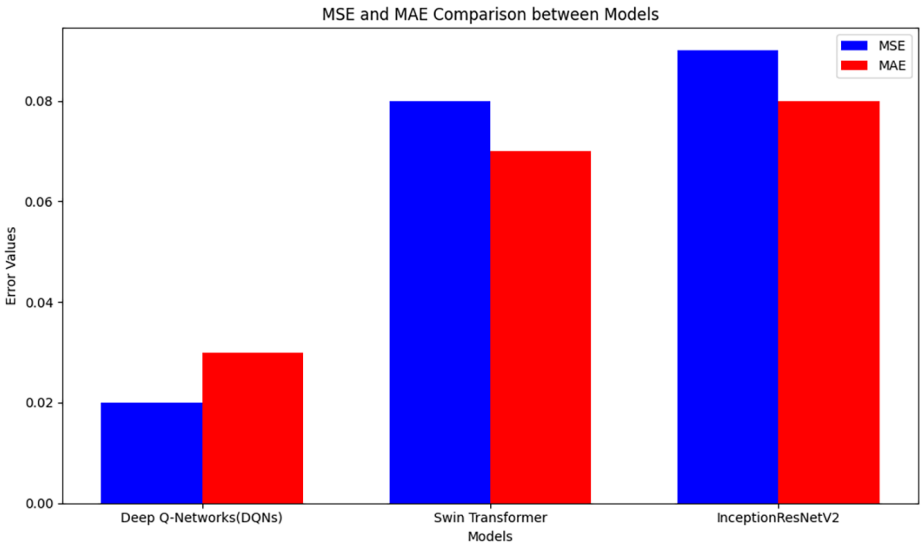

Fig. 7. Performance metrics of all models

4.6 Comparison of our proposed system with existing systems

Compared with the existing rare disease diagnosis systems, which mostly apply a single deep learning model such as conventional CNNs or standard ResNet architectures, the present approach uses DQNs, InceptionResNetV2, and Swin transformer to create a more complete and accurate diagnostic framework. While classic systems hone in on classification of images, our set of models extends this to multiscale and fine-detail feature extraction combined with hierarchical attention for efficient high-resolution image analysis and reinforcement learning-based adaptive decision-making. Therefore, our system is better able to mitigate the complexity, variability, and scarce availability of rare disease datasets for better diagnostic accuracy, faster decision-making, and adaptability in real-time interactions with users compared to the static, more restricted ones available at present.

4.7 User-interface implementation and outputs

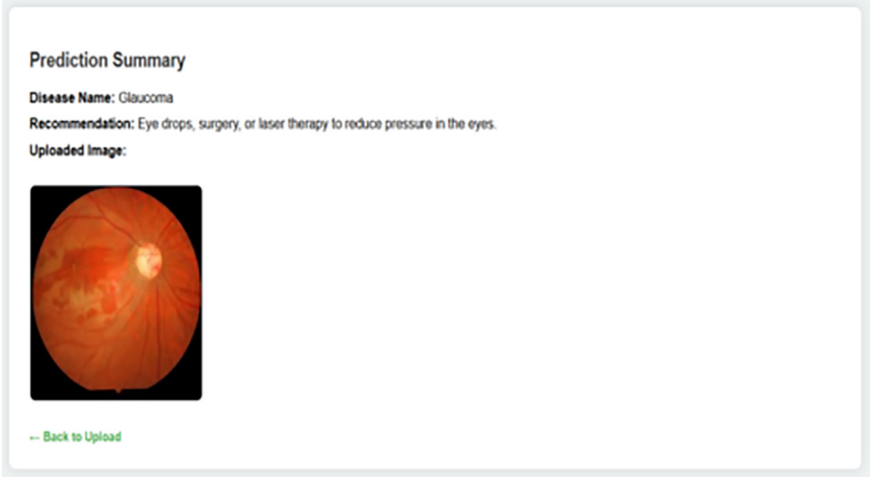

Fig. 8. User-interface disease diagnosis

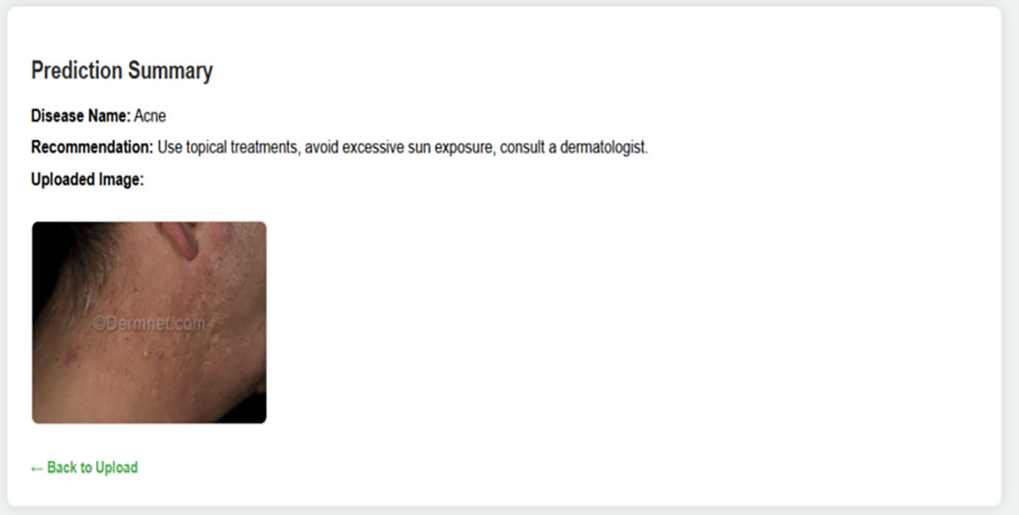

Fig. 9. Profile summary for rare disease

Figure 8 shows the image that provides a summary of a diagnosed case of glaucoma and suggests treatments such as eye drops, surgery, or laser therapy to lower eye pressure. The mention of an "Uploaded Image" placeholder indicates the involvement of a medical AI system, probably examining retinal or optic nerve images to identify the condition. There's also a "Back to Upload" option, which allows users to send more images for examination. This system is beneficial because it not only identifies Glaucoma but also recommends treatments in line with standard eye care procedures. However, the system does not include confidence scores or detailed explanations for each diagnosis, raising concerns about transparency in AI-driven medical tools. While this tool can aid in early detection and guide patients, having a human specialist review the results is still very important to ensure accuracy.

Figure 9 shows a summary from an AI skin tool that checks for skin problems. It identifies the problem as acne and suggests treatments such as using creams or seeing a skin doctor. There is a space to upload a picture of your skin, which means the tool uses images to make its report. It also mentions a dermatology website and a go-back button to upload your photo, indicating that this is on a website where individuals can utilize the tool. The tool gives overall ideas regarding treatments and skin care, which can be useful as a starter. However, since there are no in-depth explanations of how certain the tool is of its analysis, it's best to double-check with a doctor so that the advice is correct and tailored to you. The design is simple and straightforward, intended for those who need quick and basic information regarding skin issues.

4.8 Chatbot inputs and outputs

Figure 10 which represents a page from "MediScan AI" is for Medical Image Analysis. Users can choose between Skin Disease or Eye Disease for diagnosis. Images can be uploaded via drag-and-drop or file selection (supports JPG, PNG, and GIF formats under 5MB). After uploading, users click "Analyze Image" to get results. Advice is offered to obtain an accurate diagnosis, for example, good lighting and taking pictures from different angles.

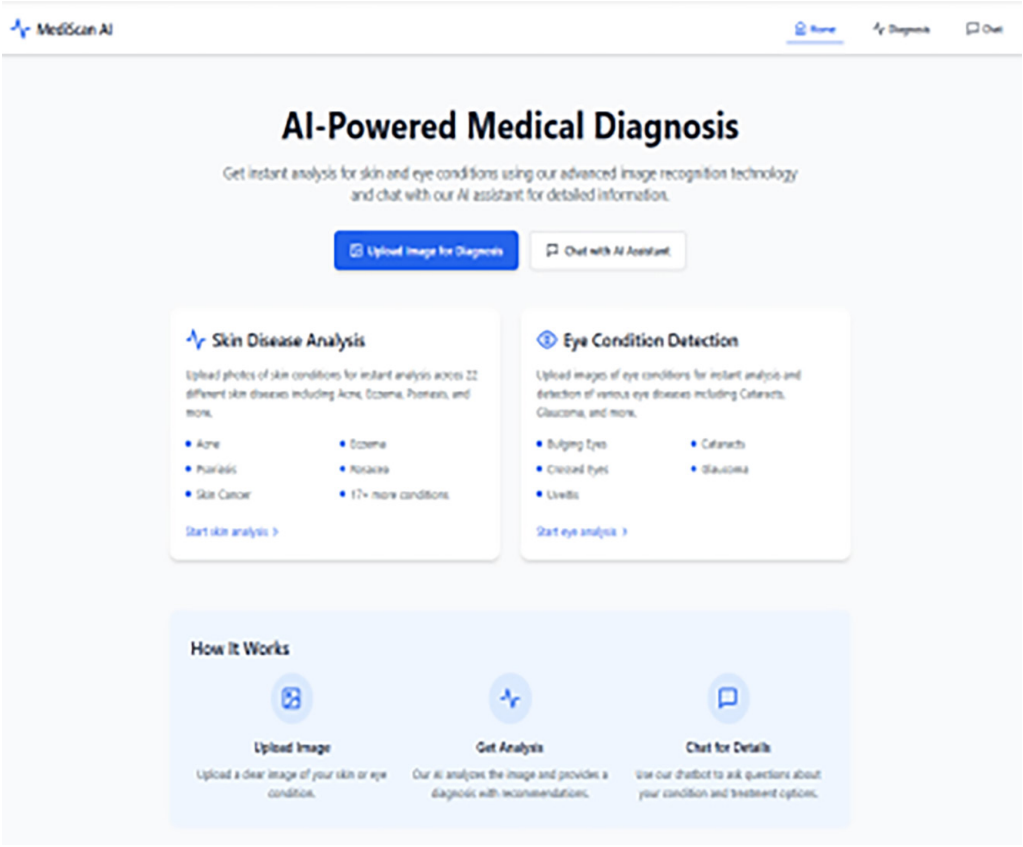

Fig. 10. Home page of user interface

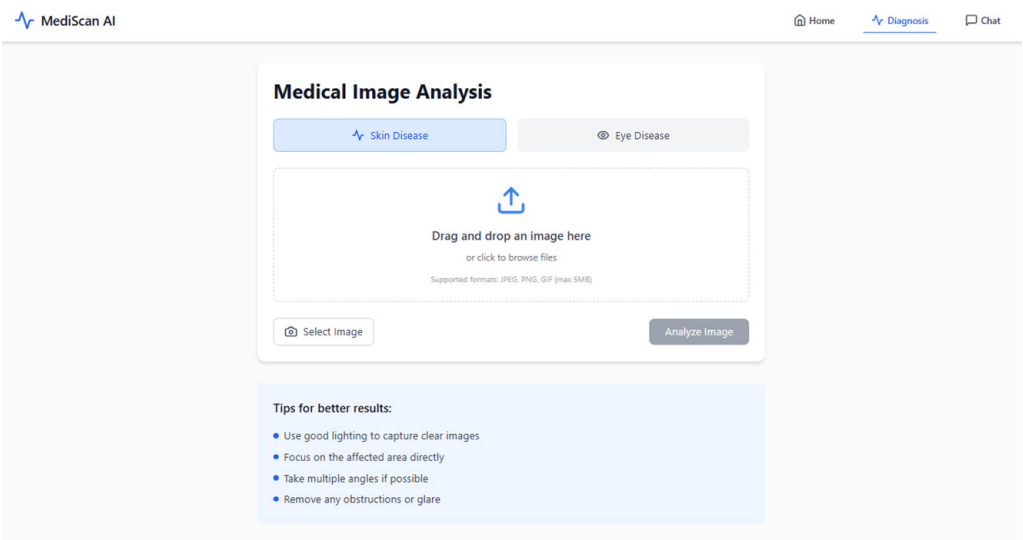

Fig. 11. Diagnosis page of user interface

Figure 11 depicts the homepage of "MediScan AI," the website for diagnosis by AI. It offers real-time diagnosis for skin and eye ailments based on image recognition technology. Users can upload pictures to identify conditions such as acne, psoriasis, cataracts, glaucoma, and many others. There are two primary services: skin disease analysis and eye condition detection. The process entails posting an image, getting AI-powered analysis, and chatting with an AI assistant for additional guidance.

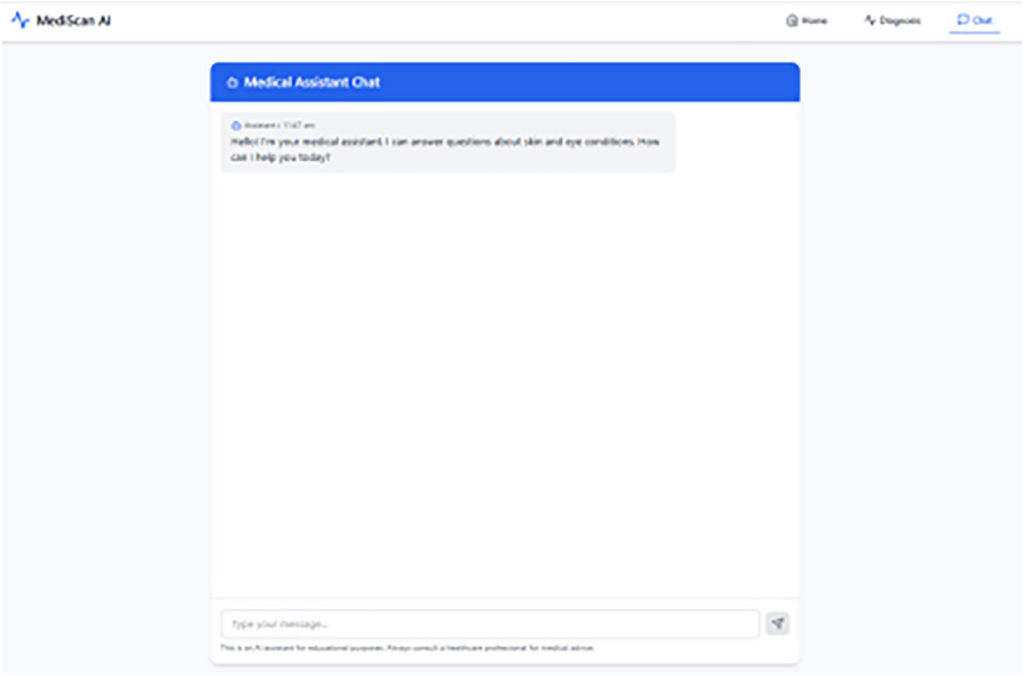

Fig. 12. Chat section in user interface

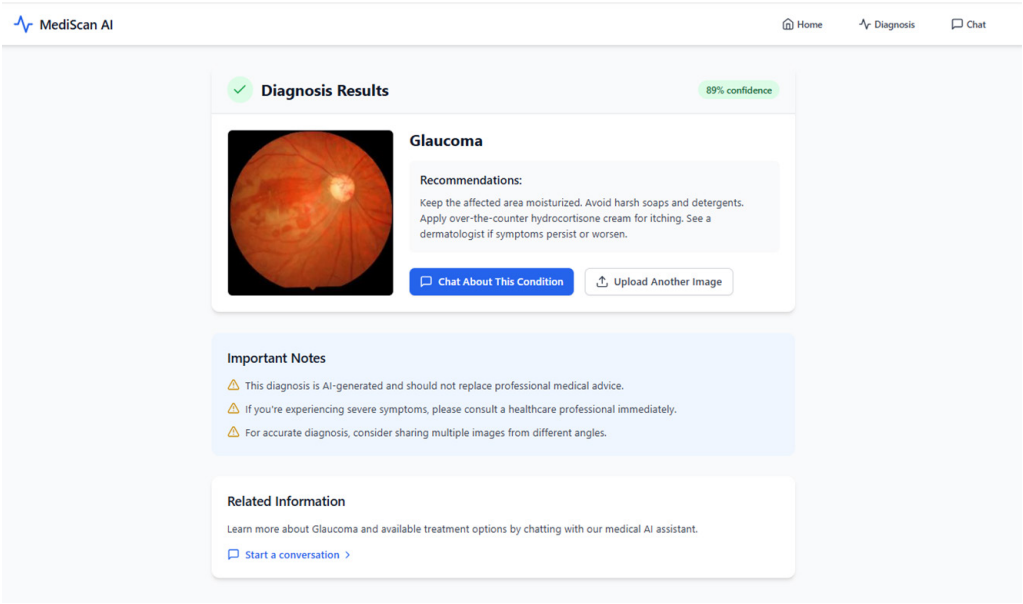

Fig. 13. Prediction of eye disease glaucoma

Figure 12 reflects the Medical Assistant Chat screen of MediScan AI. The chatbot is AI-based and seeks to assist users by presenting information regarding skin and eye conditions. The assistant greets the user and assists with information on health concerns. At the bottom is there is a text entry field where users can enter their questions. There is also a disclaimer at the bottom that reveals the AI assistant is for learning purposes only and asks the users to consult a healthcare professional for real medical advice or treatment.

Figure 13 shows a Diagnosis Results page from MediScan AI with high-confidence (89%) detection of glaucoma from an eye image uploaded by the user. It suggests care

tips such as moisturizing, avoiding hard soap, and hydrocortisone cream use. The users can consult with the AI assistant for more information or upload an image for more analysis. Important notices indicate that the diagnosis is machine-generated and is not a professional medical recommendation. Further advice directs users to obtain professional assistance and utilize multiple images for accuracy.

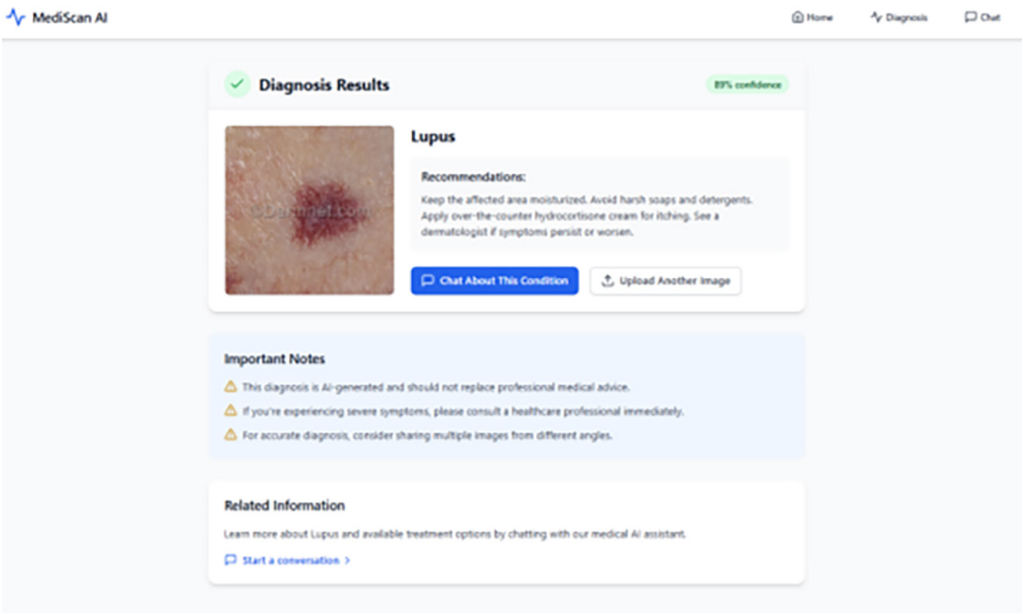

Fig. 14. Prediction skin disease lupus

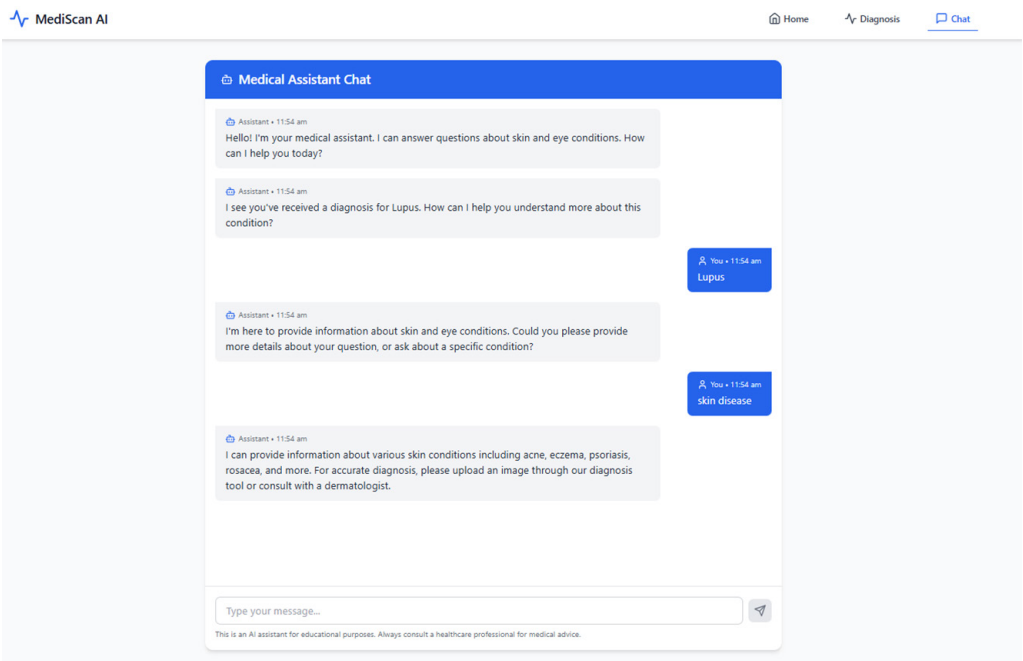

Fig. 15. Chatting with AI assistant

Figure 14 shows diagnosis results from MediScan AI, in which it has marked an 89% confidence lupus diagnosis of a photo of the skin. It suggests care, such as applying moisturizer on the affected region, avoiding hard soap, and using

hydrocortisone cream. Users can either chat about the condition or upload another picture for analysis. A significant notes section highlights that the diagnosis is computer-generated and not an alternative to a specialist medical opinion and that users need to get advice from a health professional and send more than one image in order to be sure.

Figure 15 depicts the Medical Assistant Chat from MediScan AI, in which a patient speaks to an AI chatbot regarding a diagnosis of lupus. The assistant provides helpful information regarding the condition and is willing to provide more details. When the patient asks for "skin disease," the AI provides information regarding various conditions such as acne, eczema, and psoriasis. It also suggests uploading a photo for a proper diagnosis. The notice below points out that the assistant is for educational purposes only and not to replace expert medical advice.

5 CONCLUSION AND FUTURE WORK

The image is a short, clear synopsis from an AI-based medical device. The device can diagnose potential health problems such as lupus and suggest basic things about how to manage them. It includes an "upload images" tab and a "go back to previous page" choice, indicating it's for clinical or telemedicine use. This means you can send medical images for a quick diagnosis. The advice it gives, such as taking medicines and avoiding the sun, aligns with regular lupus care. However, the tool doesn't specify which medicines to take, the proper dosages, or how serious the condition might be. This indicates that it is more of a supporting instrument than giving an overall diagnosis. The site is designed to be quick and user-friendly, hence being a convenient tool for making decisions on cases to prioritize or giving preliminary information to patients within healthcare facilities. The tool doesn't reveal how sure it is about its findings, nor does it explain where its data comes from or if it considers other possible diagnoses. Due to this, it's required that healthcare professionals be engaged. This is all part of a broader trend in which AI is utilized to identify patterns and perform initial screenings, but human professionals are indispensable for making sophisticated decisions and customizing treatment protocols. The demonstration shows how AI is slowly being implemented in healthcare but also how high standards of medical approval and ethical use must be maintained. AI can support but not substitute actual physicians in patient care. Future enhancements may be the ability to include warning notices on the level of danger or interfacing with electronic medical records. That would close the loop between AI suggestion and complete clinical implementation.

5.1 Data availability statement

The 20 Skin Diseases Dataset used in this study is publicly available at https://www.kaggle.com/datasets/haroonalam16/20-skin-diseases-dataset.

The Eye Diseases Classification Dataset used in this study is publicly available at https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification.

5.2 Ethical compliance

It is ensured by this study that all image data used for the purposes of training, testing, and evaluations be provided only by public sources and those trustworthy

on Kaggle. No outside or private image has ever been used in the study. Any image samples displayed or processed in this study are only from the given datasets: the 20 Skin Diseases Dataset and Eye Diseases Classification Dataset.

5.3 Data privacy protection

The system is developed with a focus on data privacy. Uploaded testing images are neither archived nor reused, since they are temporarily processed in memory, after which they are discarded at inference automatically. No personal or identifiable information is collected or saved, reconciling with the basic data privacy and user confidentiality criteria.

5.4 Physician-review mechanism

In the present study there is no validation procedure with physicians-in-the-loop. However, hospital-based clinical review mechanisms and the collaboration of licensed medical professionals are envisioned for future system iterations to ensure the medical reliability of AI-generated outputs and align them with clinical standards before deployment into the real setting.

6 REFERENCES

[1] G. M. S. Himel, M. M. Islam, K. A. Al-Aff, S. I. Karim, and M. K. U. Sikder, "Skin cancer segmentation and classification using vision transformer for automatic analysis in dermatoscopy-based noninvasive digital system," *International Journal of Biomedical Imaging*, vol. 2024, no. 1, pp. 1–18, 2024. https://doi.org/10.1155/2024/3022192

[2] J. Zhou and X. Gao, "SkinGPT-4: An interactive dermatology diagnostic system with visual large language model," *arXiv preprint arXiv:2304.10691*, 2023. https://doi.org/10.48550/arxiv.2304.10691

[3] J. Wang *et al.*, "SSVT: Self-supervised vision transformer for eye disease diagnosis based on FunDUS images," *arXiv preprint arXiv:2404.13386*, 2024. https://doi.org/10.48550/arxiv.2404.13386

[4] J. Wu *et al.*, "SeATrans: Learning segmentation-assisted diagnosis model via transformer," *arXiv preprint arXiv:2206.05763*, 2022. https://doi.org/10.48550/arxiv.2206.05763

[5] M. Shafiq, K. Aggarwal, J. Jayachandran, G. Srinivasan, R. Boddu, and A. Alemayehu, "A novel Skin lesion prediction and classification technique," *ViT-GradCAM. Skin Research and Technology*, vol. 30, no. 9, p. e70040, 2024. https://doi.org/10.1111/srt.70040

[6] V. Lungu-Stan, D. Cercel, and F. Pop, "SkiNDISTILVIT: Lightweight vision transformer for skin lesion classification," *arXiv preprint arXiv:2308.08669*, 2023. https://doi.org/10.48550/arxiv.2308.08669

[7] M. A. Arshed, S. Mumtaz, M. Ibrahim, S. Ahmed, M. Tahir, and M. Shafi, "Multi-class skin cancer classification using vision transformer networks and convolutional neural network-based pre-trained models," *Information*, vol. 14, no. 7, p. 415, 2023. https://doi.org/10.3390/info14070415

[8] G. H. Dagnaw, M. El Mouhtadi, and M. Mustapha, "Skin cancer classification using vision transformers and explainable artificial intelligence," *Journal of Medical Artificial Intelligence*, vol. 7, no. 1, p. e21052, 2024. https://doi.org/10.21037/jmai-24-6

[9] Y. Ding *et al.*, "HI-MViT: A lightweight vision transformer for skin disease classification," *IEEE Transactions on Medical Imaging*, vol. 43, no. 4, pp. 1151–1162, 2024. https://doi.org/10.1109/TMI.2024.2891809

[10] A. Raj, V. Kumar, and D. Jha, "Vision transformers for rare eye disease diagnosis using fundus imaging," *IEEE Transactions on Medical Imaging*, vol. 43, no. 9, Article no. 3022387, 2024. https://doi.org/10.1109/TMI.2024.3022387

7 AUTHORS

M. Sanjay is with the School of Computer Science and Engineering, VIT, Chennai, India (E-mail: sanjaymurali3369@gmail.com).

Kanaparthi Roshin Sai is with the School of Computer Science and Engineering, VIT, Chennai, India.

Pillaram Manoj is with the School of Computer Science and Engineering, VIT, Chennai, India.

J. Balaji is an Associate Professor, VIT Business School, VIT, Chennai, India.