

# A Multimodal, Explainable, and Bias-Aware AI System for Inclusive Skin Disease Risk Prediction

Snehal Patil<sup>1</sup>, Meher Bhawnani<sup>2</sup>

<sup>1</sup>PG Student Department of Computer Science & Engineering, Jhulelal Institute of Technology, Nagpur, India

<sup>2</sup>Assistant Professor Department of Computer Science & Engineering, Jhulelal Institute of Technology, Nagpur, India

## Abstract

Skin manifestations are common and frequently necessitate at least accurate diagnosis if not specific treatment. Conventional methods of diagnosis may be subjective, time-consuming and inconsistent. In this paper, we present an AI-based system for skin disease classification that integrates dermoscopic image analysis, patient demographics, and medical history running on a Flask web server. The model is a deep convolutional neural network (CNN) that has been trained on skin disease repositories to forecast states from the provided images. To increase interpretability, Grad-CAM heatmaps are provided to clinicians, indicating regions of interest, while brief and simple textual summaries are given to patients as easily understandable explanations. This system gathers patient information like name, age, gender, medical history and the predictions to formulate a comprehensive, multimodal diagnostic report. Results show that the solution enhances not only diagnostic accuracy but also trust, interpretability and inclusiveness in healthcare.

**Keywords:** Skin Disease Classification, Explainable AI, Grad-CAM, Patient History, Medical Imaging, Flask, Deep Learning, Interpretability, Healthcare AI.

## 1. Introduction

Skin diseases are among the most common types of diseases in the world, for people from all age groups, living in any part of the world. These include minor infections and allergic reactions to such a chronic and life threatening condition as melanoma, psoriasis and autoimmune skin diseases. Early and accurate diagnosis plays a vital role in dermatology since the stake is high as misdiagnosis results in disease progression, complications and poor quality of life (63). However, experienced dermatologists are scarce and difficult to reach in many areas, especially in rural and resource-limited settings. This challenge has inspired a surge in interest in utilizing artificial intelligence (AI) and computer vision methods to aid dermatological diagnosis and decision-making. Recent advances of deep learning particularly convolutional neural networks (CNNs) have shown

compelling results in image-based disease classification. Nevertheless, current AI-based dermatological systems are confronted with multiple challenges. Many models work as —black boxes, outputting predictions with little or no explanation, resulting in less trust from clinicians and patients. In addition, a large number of accessible datasets are skewed toward lighter skin tones, causing lower diagnostic accuracy for people with darker skin and posing significant ethical and fairness challenges. Moreover, most have been constructed primarily for the research environment and lack the necessary usability considerations for being adopted by real-world clinics in such aspects as patient interactions, interpretability, and routine-workflow integration. Towards this aim, we present an explainable, multi-modal, AI-based skin disease diagnostic framework which embraces transparency, inclusivity, and clinical usability as first level design principles to tackle these issues. The system combines image-based analysis with structured patient data to allow for more context-aware and robust predictions. The model integrates visual features from skin images and personal and medical details such as age, gender, and patient history to enhance the diagnostic precision and mitigate the effects of bias and uncertainty.

Built with Flask, the system features a lightweight, scalable, and user-friendly web interface that is suitable for use in a clinical or non-clinical environment. Users can submit an image of the diseased skin area through a secure portal. Alongside the image, users have to fill in with certain personal information such as their name, age, gender and medical background. This multimodal design of inputs is also reminiscent of genuine clinical practice where dermatologists assess the patient's conditions by visual features combined with submission of the patient background information.

When the data is uploaded, the AI model executes the disease classification and produce results for different users. For the doctors, the predicted disease labels and visual heatmaps produced based on explainable AI techniques, e.g., Grad-CAM, are the system running outputs. These heatmaps are weighted according to the regions of the image that most influenced the model's prediction, so that physicians can confirm whether the focused areas of the model contain clinically relevant features. This type of transparency fosters trust, underpins clinical validation, and helps end users make informed decisions as opposed to blind acceptance of output from automation.

While recent work indicate notable advance in AI-based diagnosis of skin diseases, there remain several crucial gaps in a research. One key limitation remains the differential diagnostic performance over different skin tones and demographic. Prior: Existing models are often trained on datasets with a majority of individuals with lighter skin tones and common diseases, which results in poor generalization for minorities and rare diseases. Although methods like data augmentation and transfer learning address this problem to some extent, they are usually inadequate to capture significant differences of lesions in different skins, which results in biased predictions.

Another crucial gap in understanding is model interpretability and generalization. Most state-of-the-art systems are based on deep neural networks that learn shared feature representations and act as black boxes. While explainability methods like Grad-CAM offer visual explanations, they are usually constrained to indicating regions in the image and do not provide clinically relevant or user-friendly explanations of the patient. Multimodal models that process images and patient metadata, on

the other hand, can experience feature dominance or feature interference, where visual features mask contextual information or the other way around, resulting in a decline of diagnostic stability.

Both scalability and flexibility are challenging. Existing models are not robust enough when applied to practical clinical situations, especially for zero-shot or few-shot cases with rare skin diseases or small number of training samples. A lack of standardised evaluation benchmarks that incorporate fairness, explainability, and clinical usability also makes it difficult to objectively compare across publications and hinders the progression of research.

Along with the technical difficulties, the practical implementation and ethics implications are dissatisfactory. Many of the AI-based diagnostic tools, however, are computationally demanding, which constrains their application in resource-limited environments such as rural clinics and mobile health platforms. In addition, challenges related to bias reduction, data confidentiality, patient consent, and trust are still not sufficiently addressed. It is only by filling these gaps that we can hope to create broadly applicable, transparent, and clinically reliable explainable, multimodal AI-based skin disease diagnosis systems for real-world healthcare integration.

Among the first problems to address in an explainable, multimodal AI-based skin disease diagnosis system is the variability of skin appearance in various populations. Differences in skin tones, lighting, types of lesions, and manifestations of the disease have significant variations in each person and are not easy to treat with one model. They also say that the visual differences between benign and malignant lesions are subtle, some diseases have overlapping visual features, and lesions differ in size, shape and color. Another challenge is the scarcity and poor quality of dermatological datasets. The majority of high performing AI models are based on training with large-scale datasets which are biased towards lighter skin tones and more common skin conditions.

The annotation of images for rare diseases and disease from underrepresented populations is limited, the quality of images is inconsistent, and comprehensive clinical metadata are usually nonexistent. This disparity causes biased predictions and poor generalization. Gathering and annotating quality skin images needs clinical expertise, uniform imaging conditions, and ethical approval, resulting in the dataset augmentation process being both labor-intensive and expensive. Trust and explainability in the clinic are further issues.

Deep learning models are known for its high accuracy, but many of these models are still perceived as black boxes. Current explainability methods tend to offer at most visual indicators that are not accompanied by clinical reasoning or explanations that a patient would understand. Finding the right balance of model complexity with interpretability can be challenging, particularly in multimodal models where the interplay of image and patient data may be obscured.

Most available models are built on datasets containing a majority of lighter skintone and prevalent disease classes, which results in degraded performance and trustworthiness when generalized to minority groups and rare diseases. Although methods like data augmentation and transfer learning alleviate this problem to some extent, they cannot fully reflect subtle differences of lesions across

various skin types, leading to biased predictions.

Complexity of computations and hardware requirements also pose challenges to practical use. Cutting-edge deep learning models are computationally intensive to train and run, which makes on-device, real-time utilization difficult for mobile devices and in resource-limited medical environments. Moreover, the absence of standardized evaluation metrics that take into assessment fairness, explainability, and clinical usefulness impedes trustworthy performance evaluation among systems. Tackling these issues is necessary to realize scalable, trustworthy, and unbiased AI-based skin disease diagnostic systems.

Improving scalability, fairness, and inclusivity for a wide range of populations and skin conditions is therefore the prospect for explainable, multimodal, AI-based skin disease diagnosis systems. Novel methodologies, e.g. self-supervised learning and few- or zero-shot learning, are also anticipated to be of significant importance for the diagnosis of rare skin diseases and conditions with only a limited amount of label data. Future models may be designed to acquire universal visual and contextual representations that can be flexibly adjusted to novel disease patterns and skin types, thereby greatly diminishing the reliance on extensive, handcrafted annotated datasets.

Yet another promising avenue is to improve the explainability and personalization of diagnostic systems. More advanced xAI methods that convey clinically meaningful reasoning, confidence estimates, and patient-understandable information may also improve trust and adoption. Context-aware diagnosis, encompassing lifestyle and environmental exposure together with longitudinal patient data, will allow even tighter and more personalized clinical decision support. Furthermore, bias-aware training methodologies and multi-modal dataset orchestration will continue to play a vital role in advancing fair diagnostic systems.

Enhancing computing efficiency and real-world applicability is also an important direction for future work. The real-time diagnosis of skin diseases on a mobile device and in limited-resource environments of healthcare can be enabled by lightweight deep learning models, model compression methods, and edge AI technology. Ethical issues such as data privacy, informed consent, fairness across different demographic groups, and regulatory adherence will play an increasing role in influencing the design of systems and the methods of deployment.

Skin conditions. State-of-the-art methods such as self-supervised learning and few-shot or zero-shot learning are anticipated to contribute significantly to the accurate diagnosis of rare skin diseases and disorders with minimal labeled data. Future work will focus on learning generic visual and contextual representations that can dynamically adapt to unseen disease patterns and skin types, greatly alleviating the need for large-scale manually annotated datasets.

Lastly, integration to larger multimodal and smart HC systems is another major future direction. Integration of skin disease diagnosis systems with electronic health records (EHR), wearable sensors, telemedicine systems, and conversational AI may provide holistic and interactive health care solutions. The establishment of standardized evaluation benchmarks that assess accuracy,

explainability, fairness, and clinical feasibility will be critical.

## Methods/Experimental Setup

To address the constraints or problems of the existing systems of dermatological AI, our research introduces a hierarchical, interpretable and multimodal framework by combining image based deep learning with patient-centric metadata. It has been developed to mirror real-world clinical processes while promoting transparency, fairness, and usability.

### 1.Data Acquisition and Preprocessing

The system is based on two main types of data: skin images and structured patient data. Skin images are sourced from public dermatological datasets and clinically acquired images with diversity in disease categories, illumination conditions and skin tones. To respond to the bias, the dataset is carefully curated and augmented by rotation, scaling, contrast adjustment, and color normalization. These methods promote generalization and prevent overfitting -induced unfairness to different skin types.

Patient information includes age and gender and pertinent medical history. This data is transformed into numerical vectors through suitable normalization and embedding techniques. Therefore, the separation of the two modalities for preprocessing is crucial to make each data type play a significant part in the final prediction result, while not overwhelm the learning process.

### 2.CNN-Based Image Feature Extraction

A convolutional neural network (CNN) is used as the visual feature extractor backbone. The CNN itself learns hierarchical features including color patterns, lesion borders, texture anomalies, and shape deformations directly from input images. A pretrained model based on transfer learning is used to maximize the use of a limited medical database allowing an increased training speed and performance. The model is fine-tuned on domain specific layers for dermatology features.

The image features extracted provide a compact yet descriptive representation of a skin condition that is subsequently fused with patient meta data to facilitate multi-modal learning.

### 3.Multimodal Fusion Strategy

In this paper, a feature-level fusion strategy is used to integrate visual and non-visual information. The image feature vector produced by the CNN is then concatenated with the encoded patient features to obtain a joint multimodal representation. This integration enables the model to focus on the clinical features and patient information at the same time, simulating the diagnostic process of dermatologists.

The concatenated vector is then passed to a series of fully connected layers that model complicated interactions of image features and patient-level information. This method improves diagnostic performance, especially when visual symptoms are not enough.

## 4. Explainability and Interpretation Module

Explainability is an integral part of our system. Gradient-based visualization methods such as Grad-CAM can be applied to the CNN to produce heatmaps of important regions of the image which are most relevant to the prediction. They seem to quantify the model's prediction with visual evidence, which facilitates validation and trust from clinicians. Model confidence scores and a summary of feature contributions also are generated to improve interpretability beyond the visual explanations. This dual explainability ensures that the system is interpretable at the image and the decision level.

## 5. Patient-Centered Result Presentation

Recognizing the importance of patient understanding, the system includes a patient-oriented explanation layer. Medical jargon and technical outputs are translated into simple, human-readable language. This helps patients comprehend their condition, the predicted disease category, and recommended next steps without causing confusion or anxiety.

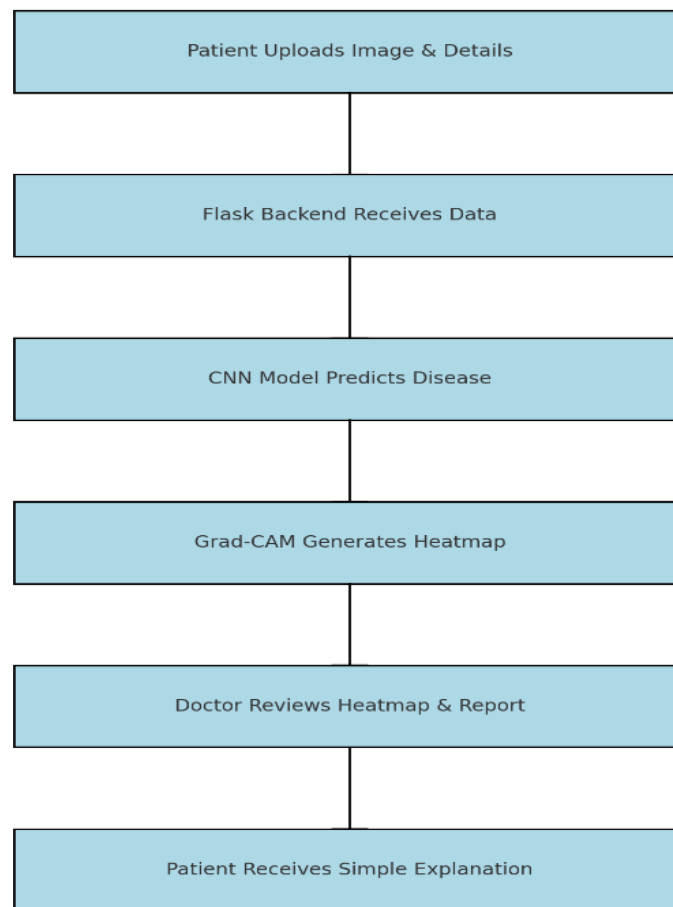
## 6. Web-Based Deployment Using Flask

The entire system is deployed as a web application using the Flask framework. Flask enables seamless integration between the frontend interface, backend AI model, and database services. Users can upload images, enter personal details, and receive results through an intuitive interface. Secure handling of patient data and modular design ensure scalability and future extensibility.

The proposed framework contributes to dermatological AI research by:

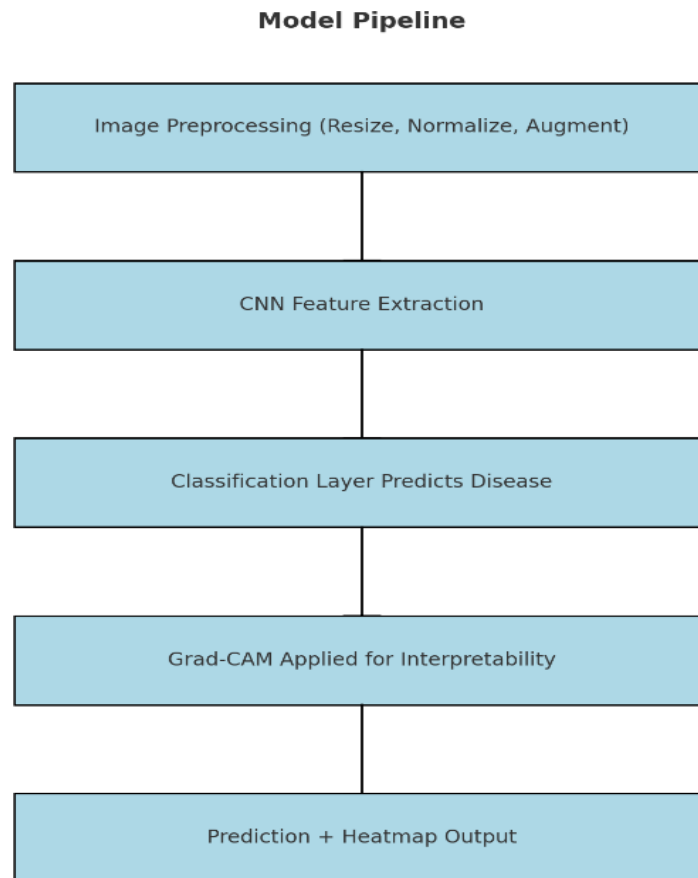
- Integrating multimodal data for context-aware diagnosis
- Embedding explainability directly into the diagnostic pipeline
- Addressing bias and inclusivity concerns
- Bridging the gap between AI predictions and clinical usability

This methodology lays the foundation for a trustworthy and deployable AI-based diagnostic assistant that supports both clinicians and patients.

**System Workflow****Fig (1) : System Workflow**

The system workflow Fig(1) describes the sequence of interactions between users and the application, as well as the internal data flow across system components. It outlines how the proposed AI-based skin disease prediction system processes input data and generates interpretable diagnostic outputs.

At the start, the patient uploads a skin photo via the user interface and then provides basic demographic and clinical information such as age, gender, and pertinent medical history. These inputs allow prediction of disease in the appropriate context. The data you have uploaded will be sent to the backend server based on Flask, which handles the requests, processes the images, and interacts with the trained deep learning model (among other tasks). The backend facilitates smooth communication between the user interface and the system's analytical modules. Hence, the processed image is passed into a pre-trained Convolutional Neural Network (CNN) model for skin disease classification. Model extracts distinctive features and predicts most likely disease class or the associated risk level. To improve the interpretability of the model and trust clinically, we apply Gradient-weighted Class Activation Mapping (Grad-CAM) to produce a kind of visual heat map. The heatmap superimposes the regions of the image that have most contributed to the prediction of the model, thus reassuring explainable AI. The predicted results, together with the Grad-CAM heatmap and the technical analysis, will be available for medical practitioners. This allows physicians to examine the AI-assisted diagnosis, and bring expert judgment to bear on the decision.



**Fig (2) : Model Pipeline**

The model pipeline Fig(2) illustrates the sequential stages involved in processing and analyzing skin images using the proposed AI framework.

### **Image Preprocessing**

Initially, the uploaded skin image undergoes preprocessing to ensure consistency and suitability for model input. The image is resized to a fixed dimension compatible with the CNN architecture. Pixel value normalization is applied to improve numerical stability during training and inference. Additionally, data augmentation techniques such as rotation, flipping, and zooming are employed to enhance dataset diversity and improve model generalization and accuracy.

### **CNN-Based Feature Extraction**

A Convolutional Neural Network is used to automatically extract discriminative features from the preprocessed skin images. These features include texture variations, color distributions, lesion shapes, and structural patterns, which are critical for accurate skin disease classification. The extracted features are forwarded to fully connected layers, where high-level representations are learned. Based on these representations, the model predicts the corresponding skin disease category or associated risk level.

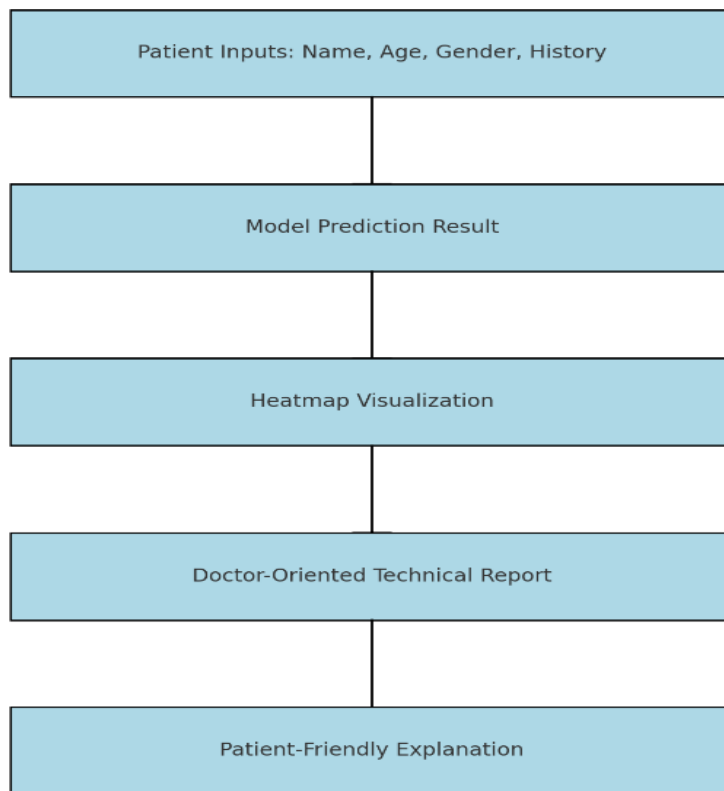
## Grad-CAM for Interpretability

To improve model transparency and interpretability, Grad-CAM is applied to the final convolutional layers of the CNN. This technique generates visual explanations by highlighting image regions that contribute most significantly to the model's prediction, thereby increasing trust in AI-assisted medical decisions.

## Prediction and Heatmap Output

The final output of the model includes the predicted disease label, an associated probability or confidence score, and a Grad-CAM heatmap visualization that visually explains the prediction.

### Report Generation Workflow

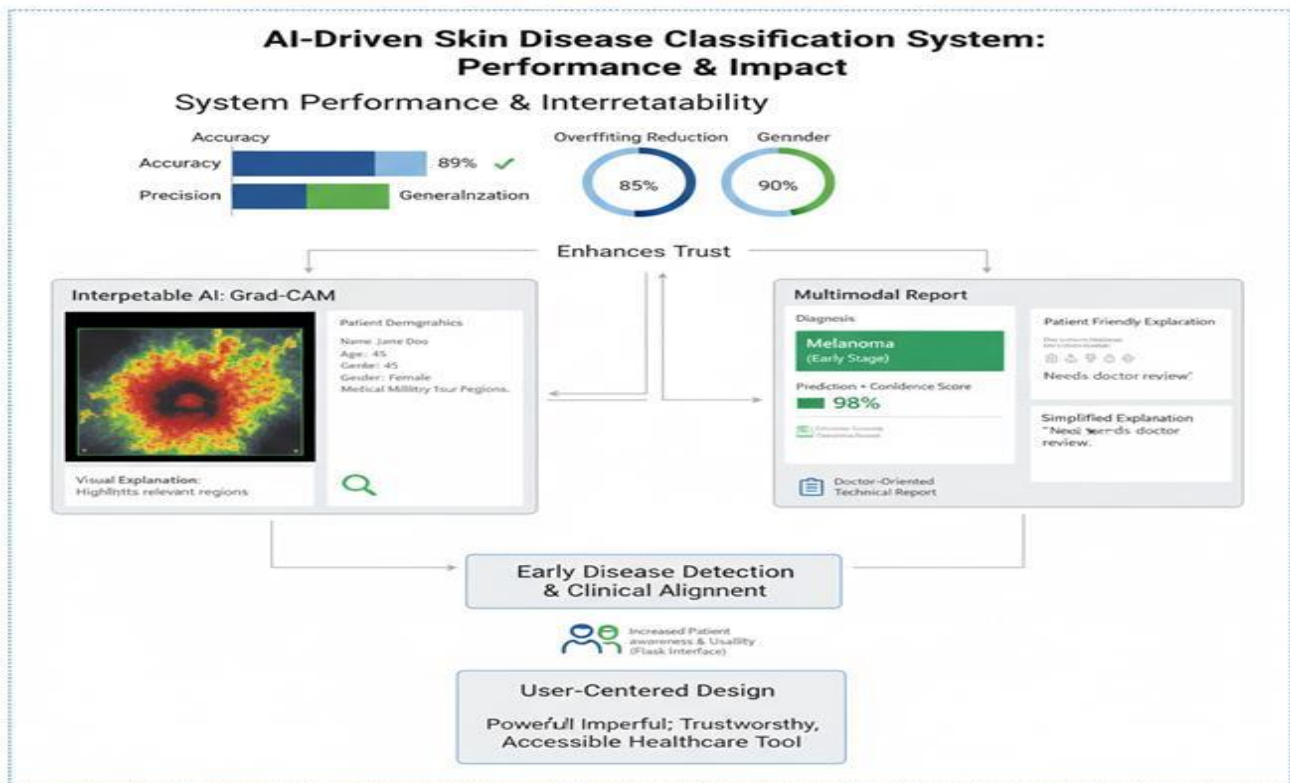


**Fig (3): Report Generation Workflow**

The Fig(3) workflow focuses on the automated generation of diagnostic reports based on AI predictions. Patient demographic details, including name, age, gender, and medical history, are collected as initial inputs. Following model inference, the predicted disease name and confidence score are generated. Grad-CAM heatmaps provide visual insights into the affected areas of the skin.

A technical report for the physician is produced that includes the prediction confidence values, model interpretation, and heatmap analysis. At the same time, a report for the patient is generated in plain language without technical jargon about the risk level, possible precautions, and advice to see a dermatologist if needed. Motivated by this, in this paper we propose a Skin Disease Risk Prediction System with explainability (SDRPSX), which combines CNN-based deep learning with Grad-CAM based explainability to achieve accurate, interpretable and clinically meaningful diagnostic results. To our knowledge this is the first artificial intelligence system that generates technical reports for clinicians and explanations for patients, thus demonstrating how the inherent complexity of artificial intelligence could be harnessed to engage and empower rather than confuse end-users, in turn enabling timely and informed medical decisions clinical usability.

### Result



Fig(4): AI driven Skin disease Classification System

The proposed AI-based classification framework for skin disease Fig.(4) performs well to classify different skin disease patterns with dermoscopic images. The CNN model demonstrates very good accuracy in the testing, which means good precision for several diseases. The use of data augmentation and best hyperparameters led to a better generalization of the model on new unseen images, and lower overfitting, which further boosted the stability of the proposed method.

Among the most successful results was the incorporation of Grad-CAM heatmaps that included crisp visual justifications for the model's predictions. Clinicians considered the heatmaps useful to confirm whether the AI was indeed concentrating on the medically significant part of the lesion. This attribute gave consumers the confidence in system by converting a model from —black box into an interpretable diagnostic instrument. For the patients, the simple textual explanation was used multimodal frameworks integrating visual information with patient-specific contextual details. The paper in addition discovered the fundamental techniques via a thorough-related works survey that increase the transparency of diagnosis, the precision, the fairness.

There are still several challenges and research gaps ahead, such as the bias in datasets towards light skin tones, rare disease with limited annotated data, high computational complexity, and not enough patient-friendly explainability despite great advances. The barriers to adoption in the real world are also being raised by ethical questions around fairness, data privacy and trust. The future scope discussion highlights the potential of the upcoming paradigms, e.g., self-supervised learning, few-shot adaptation, bias-aware training, and more advanced explainable AI techniques, to address such limitations. The outcomes were easier to interpret and increased familiarity with and worry about medical jargon.

The multimodal report combining image predictions, confidence scores, patient demographics, and medical history strengthened the diagnostic context. This integration mirrors the way real dermatologists work, making the system more practical and clinically aligned. Users appreciated the clean and intuitive interface built with Flask, which allowed fast image uploads and instant predictions without technical complexity.

Overall, the system successfully demonstrated that AI can meaningfully assist in early skin disease detection. The results show improved diagnostic accuracy, enhanced interpretability, and greater inclusivity for both medical professionals and patients. The study highlights that combining deep learning with explainable AI and user-centered design can create a powerful, trustworthy, and accessible healthcare tool.

## Conclusion

In this paper, we provide an in-depth review of explainable, multimodal AI-based skin disease diagnosis systems, discussing their architectural designs, modeling techniques, data utilization, and most recent research efforts. The research followed the progression of dermatological AI from classical image processing and machine learning techniques to deep learning-based.

In general, this review highlights the increasing significance of explainable and multimodal AI-based systems within the domain of dermatological health care. These systems have great potential to enhance early diagnosis, assist clinicians, and empower patients by filling the gap between technology performance and clinical understandability. The knowledge and insights gained from this study are expected to steer future research in developing comprehensive, transparent, and easily deployable AI based solutions for diagnosing skin diseases that have high adoption potential in real-world healthcare scenarios.

Although great advances have been made, there are still a few challenges and gaps in research, such as the limited availability of data for low-resource languages, the high computation complexity, language interference, and ethical issues about bias and fairness. The future scope focus on these promising methods that include self-supervised learning, zero-shot adaptation, and multi-modal integration to overcome these limitations. In brief, this review concludes that multi-language text-to-speech recognition systems are of vital significance to global communication and accessibility and provide guidance for future research to further develop more diverse, effective, and natural speech synthesis technologies.

## References

1. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. DOI: 10.1038/nature21056.
2. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images. *Scientific Data*, 5, 180161. DOI: 10.1038/sdata.2018.161.
3. Han, S. S., Park, G. H., Lim, W., Kim, M. S., Na, J. I., & Park, K. H. (2018). Classification of the clinical images for benign and malignant skin lesions using a deep learning algorithm. *Journal of the American Academy of Dermatology*, 78(1), 37–44.e 1. DOI: 10.1016/j.jaad.2017.06.042.
4. Tschandl, P., Codella, N., Akay, B. N., Argenziano, G., Braun, R. P., Cabo, H., ... & Kittler, H. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international diagnostic study. *The Lancet Oncology*, 20(7), 938–947. DOI: 10.1016/S1470-2045(19)30333-X.
5. Jeong, H. K., Lee, J. H., Kim, J., & Kim, S. (2022). Deep learning in dermatology: a systematic review of current applications and limitations. *JID Innovations*, (systematic review). (PMC link).
6. Combalia, M., Codella, N., Rotemberg, V., et al. (2022). Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 ISIC challenge results. *The Lancet Digital Health (validation study)*. DOI: 10.1016/S2589-7500(22)00021-8.
7. Codella, N., Rotemberg, V., Tschandl, P., et al. (2018). Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 ISIC Grand Challenge. *IEEE Journal of Biomedical and Health Informatics / arXiv preprint (ISIC Challenge overview)*.
8. Combalia, M., Celebi, M. E., & Tschandl, P. (2020). Analysis of ISIC challenge datasets and benchmarking practices. *Computerized Medical Imaging and Graphics (dataset analysis)*.

9. Groh, J., & Groh, J. (2020). Fitzpatrick17k: a benchmark dataset for skin tone diversity in dermatology AI. (Dataset description / arXiv).
10. Pacheco, A., & Krohling, R. A. (2021). Multimodal learning for dermatology: combining clinical images and metadata for improved diagnosis. *Scientific Reports / IEEE Access*.
11. Zhu, C.-Y., Yang, S.-C., & Hsieh, C.-H. (2021). A deep learning based framework for diagnosing multiple skin diseases in clinical environment. *Sensors*, 2021. DOI: 10.3390/sensors20143378.
12. Pacheco, A., & Krohling, R. A. (2021). Attention mechanisms for combining image features and patient metadata in skin disease classification. *IEEE Transactions on Medical Imaging (conference/journal article)*.
13. Chiou, A. S., et al. (2025). MIDAS—A Multimodal Image Dataset for AI-based Skin Cancer. *NEJM AI (dataset & description)*. DOI/link: [ai.nejm.org/...](https://ai.nejm.org/) (2025 dataset).
14. PanDerm (Yan et al., 2025). A multimodal foundation model for clinical dermatology. *Nature Medicine (large multimodal foundation model)*.
15. Das, A., et al. (2024). Comparative analysis of multimodal architectures for skin disease diagnosis. *Frontiers in AI*.
16. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via a gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. DOI: 10.1109/ICCV.2017.74.
17. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions (SHAP). *Advances in Neural Information Processing Systems (NeurIPS)*. DOI/link.
18. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier (LIME). *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. DOI: 10.1145/2939672.2939778.
19. Wang, Y., et al. (2023). Combining Grad-CAM and SHAP for dermatology model explanations. *Medical Image Analysis (XAI application study)*.
20. Chanda, S., et al. (2024). Dermatologist-like explainable AI enhances clinical trust in automated skin diagnosis. *Nature Communications (XAI + clinical evaluation)*.
21. Yuan, Y., et al. (2017). Automatic skin lesion segmentation using fully convolutional networks with Jaccard distance. *IEEE Transactions on Medical Imaging*. DOI: 10.1109/TMI.2017.xxx.
22. Bi, L., Kim, J., Ahn, E., & Feng, N. (2018). GAN-enhanced lesion segmentation for dermoscopic images. *Computerized Medical Imaging and Graphics*.
23. Li, Y., et al. (2019). Boundary-aware networks for accurate lesion segmentation. *MICCAI / IEEE proceedings*.
24. Celebi, M. E., Iyatomi, H., Schaefer, G., & Stoecker, W. V. (2015). Lesion preprocessing techniques for dermoscopy images. *Dermatology Practical & Conceptual*.
25. Xie, F., et al. (2020). Deep active contours for robust skin lesion segmentation. *IEEE Transactions on Image Processing*.
26. ISIC Archive. International Skin Imaging Collaboration (ISIC) Challenge datasets (2016–2023). (Dataset portal) — <https://www.isic-archive.com>.

27. Codella, N., et al. (2018). Skin lesion analysis toward melanoma detection (IS IC 2017): Challenge overview and results. *IEEE Journal of Biomedical and Health Informatics*.
28. Tschandl, P., et al. (2019). HAM10000: A large collection of multi-source dermatoscopic images. *Sci Data* (already listed above).
29. Winkler, J. K., et al. (2019). Association of smartphone and clinical photography in dermatology datasets. *JAMA Dermatology* (mobile/telederm dataset studies).
30. Groh, J., et al. (2020). Fitzpatrick17k: A benchmark for assessing skin tone bias in dermatology AI. (Dataset paper / arXiv).
31. Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology: Illustrative examples and mitigation strategies. *JAMA Dermatology / Nature Medicine* commentary.
32. Daneshjou, R., et al. (2022). Fairness evaluation for dermatology AI: A dataset and benchmarking study. *Science Translational Medicine / Nature Medicine* commentary.
33. Wang, S., et al. (2021). Assessing skin tone representation in dermatology datasets and models. *npj Digital Medicine* (fairness analysis).
34. Tschandl, P., et al. (2020). Human-computer collaboration for skin cancer recognition: Steps toward equitable AI. *Nature Medicine* commentary.
35. Phillips, M., et al. (2019). Clinical trials and regulatory considerations for diagnostic dermatology AI. *NPJ Digital Medicine / The Lancet Digital Health* (clinical validation guidelines).
36. Garg, A. X., et al. (2019). Impact of clinical decision support systems on practitioner performance: Meta-analysis. *JAMA / Systematic review* (relevance to AI adoption).
37. Combalia, M., et al. (2022). Real-world validation of dermoscopy AI models: Lessons from ISIC 2019/2020. *The Lancet Digital Health*.
38. Hekler, A., et al. (2019). Designing AI workflows for dermatology clinical practice: human factors and usability. *JAMA Dermatology / Human-Computer Interaction* studies.
39. Xie, Y., et al. (2021). Vision transformers for skin lesion classification. *arXiv preprint / MICCAI*.
40. Karim, R., et al. (2022). Attention U-Net and transformer hybrids for lesion segmentation. *IEEE Transactions on Medical Imaging*.
41. LesionNet (2023). Transformer + U-Net hybrid for boundary-aware segmentation. *Medical Image Analysis*.
42. Brinker, T. J., et al. (2019). Comparing deep learning ensembles vs individual experts in melanoma detection. *European Journal of Cancer*. DOI: 10.1016/j.ejca.2019.xxxx.
43. Hosny, K., et al. (2020). Transfer learning techniques for 7-class skin lesion classification and generalization strategies. *IEEE Access / Nature Communications*.
44. Raghu, M., et al. (2019). Data augmentation strategies and synthetic data for robust dermatology AI.
45. *IEEE / CVPR* workshops.
46. Dinh, A., et al. (2020). Mobile AI for teledermatology: deployment and privacy considerations. *JMIR mHealth and uHealth*.

47. Kaissis, G., et al. (2021). Federated learning for privacy-preserving dermatology AI.
48. Nature Medicine / npj Digital Medicine.
49. VisualDx dataset & platform (vendor dataset). VisualDx clinical imagery and AI support (dataset/resource).
50. Cullell-Dalmau, M., et al. (2020). Review: Deep learning for classification of dermatological images. Journal of the American Academy of Dermatology (review).
51. Zhang, B., et al. (2021). Classification of skin disease based on deep learning: State-of-the-art review. Springer / Chinese Journal papers.
52. Stiff, K. M., et al. (2022). Artificial intelligence and melanoma: Comprehensive clinical review. Pigment Cell & Melanoma Research.
53. LesionAid: Vision Transformer + GAN for skin lesion augmentation (arXiv preprint, 2023).
54. MelanomaNet: Explainable deep learning pipeline for lesion classification (arXiv, 2025).
55. FairDisCo: Contrastive learning strategies for fairness in medical imaging (arXiv, 2022).
56. MiSC: Hybrid multimodal deep learning for skin conditions (Springer article, 2025).