SkinDiseaseXAI: XAI-Driven Neural Networks for Skin Disease Detection

Ammar Nasser Alqarni[1], Abdullah Sheikh[2]
Graduate Student-Masters in AI, Computer Science Department-College of Computers and Information Technology, Taif University, Taif, Kingdom of Saudi Arabia[1]
Doctor, Computer Science Department-College of Computers and Information Technology, Taif University, Taif, Kingdom of Saudi Arabia[2]

***Abstract*—Accurate classification of skin diseases is an important step toward early diagnosis and therapy. However, deep learning models are frequently used in therapeutic contexts without transparency, reducing confidence and acceptance. This study introduces SkinDiseaseXAI, a unique convolutional neural network (CNN) that uses Grad-CAM++ to classify ten different types of skin diseases and provide visual explanations. The proposed model was trained using a publicly available dataset of dermatoscopic images following preprocessing and augmentation. SkinDiseaseXAI achieved 76.12% training accuracy and 66.25% validation accuracy in 20 epochs. We used Grad-CAM++ to generate heatmaps that highlighted discriminative regions inside the lesion areas, thereby improving interpretability. The experimental results indicate that the model has the ability not only to perform multi-class skin disease categorization but also to provide interpretable visual outputs, which improves the transparency and dependability of decision-making processes. This concept has the possibility to improve clinical diagnosis by merging performance and explainability.**

Keywords*—*XAI; skin disease; Grad-CAM++; convolutional neural networks; clinical interpretability; melanoma; eczema; atopic dermatitis; fungal infections

I. Introduction

Skin diseases are among the most frequent and pervasive health concerns worldwide, providing significant challenges for healthcare systems due to their difficulty in early detection and extensive overlap in symptoms. These disorders have a significant influence on patients' quality of life; thus, early and precise diagnosis is critical for avoiding complications [1]. Despite the obvious benefits of early detection, traditional diagnostic methods continue to rely heavily on direct clinical evaluations by dermatologists, making them vulnerable to human error, practitioner variability, and diagnostic delays, particularly in areas with limited specialized medical expertise.

Recently, artificial intelligence (AI), particularly deep neural networks, has emerged as a potent tool for diagnosing and classifying skin illnesses, performing on par with qualified dermatologists [2]. However, the use of these advanced models in clinical practice is limited because of their intrinsic "black box" nature, which obscures the decision-making processes that underpin their predictions [3]. This limitation has created an urgent demand for Explainable Artificial Intelligence (XAI), which allows healthcare practitioners to easily comprehend the reasoning behind AI-powered medical judgments [4]. The use of XAI can considerably increase confidence and promote the wider integration of AI models into everyday medical practices.

In this research, the SkinDiseaseXAI model is proposed: an XAI-driven neural network designed for skin disease detection using digitized skin images. Advanced interpretability methods, such as Grad-CAM++, are integrated into a convolutional neural network (CNN) architecture to fulfil two key goals: increasing diagnostic accuracy and offering visible, credible explanations for model predictions. This approach aims to bridge the gap between high-performing AI models and their practical acceptability by medical specialists, thereby enhancing their adoption in clinical settings.

The rest of the study is set up like this: Section II looks at the literature that is related to the topic; Section III talks about the methodology, including where the data came from and how the model was made; Section IV talks about the results of the experiment; Section V presents the discussion; and Section VI ends the study with summary and suggestions for future research.

II. Literature Review

A. Related Work

Early applications of artificial intelligence (AI) in dermatology relied heavily on traditional machine learning methods like Support Vector Machines (SVM), Random Forests (RF), and K-Nearest Neighbors (KNN), as well as manually generated image features. Such techniques demonstrated adequate accuracy for skin lesion classification but struggled to generalize across different datasets due to handcrafted features and limited data [4][5].

Deep learning, particularly convolutional neural networks (CNNs), has considerably improved the accuracy of skin disease identification [3]. Esteva et al.'s important study (2017), showed that a CNN could identify melanoma with the same accuracy as a dermatologist, by using a large collection of skin lesion images [6]. Subsequent research revealed that CNNs could regularly perform as well as or better than professional dermatologists in differentiating between benign and malignant skin lesions [7][8]. A meta-analysis of various CNN models found an average sensitivity of 87% and specificity of 77% for skin cancer diagnosis, demonstrating deep learning's clinical potential [9].

Recent research has expanded on these successes by employing transfer learning and data augmentation to address

constrained datasets. LesNet, a hybrid CNN model that combines DenseNet, VGG-16, and Inception architectures, obtained approximately 98% accuracy on HAM10000 and 94% accuracy on ISIC 2019 datasets, outperforming previous state-of-the-art approaches [10]. Similarly, DermoExpert used CNN-based segmentation, geometric augmentations, and rebalancing approaches, giving Area Under the Curve (AUC) ratings of up to 0.97, indicating effective generalization [11]. Ensemble models that combine various CNN architectures (like ResNet, EfficientNet, and DenseNet) performed well, achieving about 93% accuracy and enhancing their ability to work well on different data sets [12].

B. State-of-the-Art Methods

Deep CNN architecture continues to state-of-the-art skin lesion categorization techniques [3]. Recent research uses developed topologies like EfficientNet, ResNet, and DenseNet, as well as test-time augmentation (TTA), to improve accuracy even more. Cino et al. showed that combining EfficientNet-B6 and TTA resulted in 97.58% balanced multi-class accuracy on the ISIC 2019 dataset, demonstrating deep learning precision [13].

Despite their accuracy, CNNs are seen as "black boxes", reducing clinician trust. Gradient-weighted Class Activation Mapping (Grad-CAM) is one of the Explainable AI (XAI) methods that have emerged as key tools for transparency [4]. Grad-CAM (Gradient-weighted Class Activation Mapping), which is frequently used due to its visual simplicity, assists physicians in identifying critical diagnostic locations by creating heatmaps of the input images. Grad-CAM has been shown in studies to considerably increase model interpretability, explaining why certain lesions are classed as malignant or benign [13][14].

SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are two other popular XAI tools for quantitative interpretations. SHAP assigns values to input features at both the global and instance levels, allowing clinicians to better comprehend their contributions. Rathore et al. employed SHAP to successfully differentiate dermatological disorders based on clinical and histological characteristics [15]. LIME gives local explanations by using simple surrogate models to mimic complicated CNN predictions, although it may struggle with highly detailed medical pictures due to segmentation issues [16]. Metta et al. conducted a critical evaluation of LIME's performance, concluding that customized explanation approaches may yield higher accuracy [16].

Hybrid models, which combine CNNs and interpretability techniques, represent cutting-edge research. Ullah et al. created a CNN-Radial Basis Function (RBF) hybrid that provides good interpretability and accuracy (~83% on ISIC 2016) using prototype-based visual explanations [17]. Hassan provided a dual-output hybrid CNN that generates textual explanations based on the ABCD dermatological criteria, allowing for intuitive interpretation while maintaining high performance [18]. Another notable innovation is SkinGEN, which combines vision-language models (VLM) and generative AI to improve user understanding and confidence by creating synthetic prototype images of diagnosed conditions [19].

C. Current Limitations

Despite progress, some fundamental constraints prevent complete clinical integration of AI-driven dermatology:

1) Dataset bias. Dataset bias is a key restriction for dermatological AI systems. Abdelhamid (2024) stated that current dermatology datasets largely feature lighter skin tones, resulting in worse accuracy for those with darker complexions and perpetuating healthcare disparities [20]. Fliorent et al. underlined that diagnostic performance for skin-of-color populations is much lower, arguing for more inclusive datasets to improve fairness and accuracy [21]. Addressing these biases is critical for ensuring equal healthcare and reliable AI deployment.

2) Transparency and interpretability. Although XAI methods provide some openness, understanding CNN decision-making remains difficult. According to Hauser et al., only a small proportion of dermatology AI studies have systematically studied how explanations affect clinical decisions [4]. Abdelhamid (2024) stated that more openness and clearer explanations of CNN decision-making methods are required to build physician trust and adoption in practice [20]. Continued effort is necessary to create AI models that are already interpretable and appropriate for clinical decision-making cases.

3) Regulatory and ethical constraints. The adoption of AI dermatological systems presents significant regulatory challenges. In its "White Paper on Artificial Intelligence", the European Commission (2021) specified strong rules for AI systems that promote transparency, trustworthiness, and accountability, especially in healthcare applications. Furthermore, ethical problems such as patient permission, privacy, and algorithmic responsibility limit widespread deployment, highlighting the need for strong frameworks [22].

4) Real-World Implementation. Real-world generalization is difficult. Clinical adoption has practical problems. Abdelhamid (2024) stated that AI model performance loses in real-world scenarios due to differences in imaging circumstances, image quality, and patient demographics, necessitating rigorous validation in clinical settings [20]. Furthermore, healthcare personnel need training to understand AI results, which involve workflow changes and infrastructure upgrades. Thus, practical problems with deployment must be overcome to achieve effective integration.

Table I shows a comparison of earlier studies showing different methods in skin disease categorization with deep learning. The table lists the used model, number of skin condition classifications, and whether explainable artificial intelligence methods were incorporated for every study. This puts the present work, which uses a custom interpretable CNN (SkinDiseaseXAI) to categorize 10 disease categories utilizing Grad-CAM++ for improved transparency, in context.

TABLE I. COMPARISON OF EXISTING STUDIES IN SKIN DISEASE CLASSIFICATION

Ref.	Model(s) Used	Number of Classes	Dataset Focus	Explainability	Performance	Notes
[4] Hauser et al. (2022)	Various CNNs	Multiple (Review)	Skin cancer	Grad-CAM, SHAP, LIME	N/A	Systematic review paper
[5] Wu et al. (2022)	Deep CNNs	Multiple (Review)	Skin cancers	Mentioned XAI tools	~90% (varies)	Review and analysis
[6] Esteva et al. (2017)	Inception v3	2 classes	Melanoma vs benign	No XAI	AUC: 0.96	Dermatologist-level model
[7] Salinas et al. (2024)	Various AI models	~3 classes	Skin cancer diagnosis	No XAI	AUC ~0.89	Meta-analysis of AI vs clinicians
[8] Mir et al. (2024)	LesNet	7 classes	Skin lesions	No	Accuracy: 93.2%	Custom CNN + augmentation
[10] Hasan et al. (2022)	DermoExpert	7 classes	HAM10000	No	Accuracy: 92.5%	Hybrid CNN design
[11] Mahbod et al. (2019)	Hybrid CNN	7 classes	Skin lesions	No	Accuracy: ~84%	Fusion-based architecture
[12] Ahmad et al. (2023)	Deep CNN + XAI	7 classes	Dermoscopic images	Grad-CAM	Accuracy: 90.1%	Focus on explainable multi-class
[13] Cino et al. (2024)	CNN + TTA	7 classes	Skin lesion data	Grad-CAM	Accuracy: 89.7%	Enhanced via augmentation
[15] Rathore et al. (2022)	ML + XAI	6 classes	Erythemato-squamous	SHAP, LIME	Accuracy: 94.1%	Tabular + interpretable models
[16] Metta et al. (2024)	CNN + LIME	7 classes	Skin lesion data	LIME	Accuracy: ~88%	Interpretability focus
[17] Ullah & Zia (2025)	CNN + RBF Hybrid	7 classes	Skin cancer	Grad-CAM	Accuracy: 91.5%	XAI integration
[18] Hassan (2024)	Deep CNN	3 classes	Skin cancer	Grad-CAM	Accuracy: ~89%	No dataset info
[19] Lin et al. (2025)	SkinGEN (Vision-Language)	4 classes (Prototype)	Dermatology	Interactive VLM	N/A	Conceptual model
[20] Abdelhamid (2024)	Various (review)	Multiple	Skin cancer	Mentioned	88–95% range	Survey paper
SkinDiseaseXAI	(CustomCNN)	10 classes	Mixed skin diseases (Kaggle)	Grad-CAM++	Training: 76.12% Validation: 66.25%	Custom-built + full interpretability

D. A Summary and a Gap in the Research

AI-based dermatological diagnostics have made a lot of progress, but there is still a big gap between high-performing models and their use in real life. This is mostly because the models are difficult to understand, don't focus enough on multi-class classification across different skin conditions, and don't have enough real-world data. A lot of earlier work only looks at binary categorization (such as benign versus malignant) or uses CNNs without properly combining them with strong explainability frameworks. In addition, models like Grad-CAM and SHAP have been introduced, although their use has frequently been confined to post-hoc analysis or has been quite basic.

To fill this gap, the current study suggests SkinDiseaseXAI, a unique interpretable CNN architecture that uses Grad-CAM++ to make it easier to understand how it works and to identify 10 different skin conditions. This model tries to find a compromise between how well it can diagnose and how easy it is to understand in a clinical setting. This solves a big problem that makes it hard to use AI in dermatology.

III. METHODOLOGY

A. Materials (Datasets)

For this study, a public dataset called "Skin Disease Dataset" compiled by Xuan Nguyen on Kaggle was used [23]. The dataset aggregates dermoscopic and clinical images from various open-access medical resources, although the exact original sources of the images are not individually cited in the Kaggle repository. According to the uploader's documentation, all images were curated to be free from watermarks and ethically usable for educational and research purposes in skin disease classification. The dataset includes 27,153 high-resolution clinical and dermoscopic images organized into 10 different types of skin diseases. Importantly, all images are watermark-free and ethically acceptable for usage in machine learning applications for skin disease classification.

The dataset is systematically organized as follows:

- Training Set: includes 19,003 images used for model learning.
- Validation Set: includes 2,711 images to fine-tune hyperparameters and prevent overfitting.
- Testing Set: consists of 5,439 images designated for evaluating the final model performance.

The dataset includes ten dermatological categories, including eczema, melanoma, atopic dermatitis, basal cell carcinoma, melanocytic nevi, benign keratosis-like lesions, psoriasis, seborrheic keratoses, tinea and fungal infections, and viral infections such as warts and molluscum [20]. Table II shows an overview of the dataset structure, including the number of images allocated to each subset.

This dataset distribution ensures that the deep learning model goes through rigorous training, validation, and testing to achieve optimal results. The well-balanced allocation enhances strong model generalization while minimizing the danger of overfitting.

TABLE II. DATASET COMPOSITION

Dataset Split	Number of Images	Number of Classes
Training Set	19,003	10
Validation Set	2,711	10
Testing Set	5,439	10
Total	27,153	10

Fig.1 shows representative samples from four major skin diseases found in the dataset. These images show the visual complexity and variation in color, texture, and lesion boundaries that the model must learn to recognize.

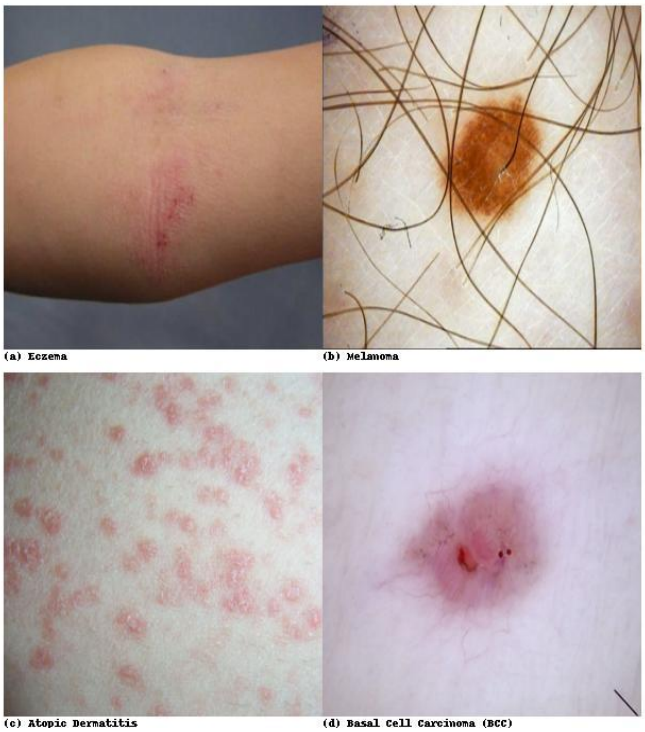


Fig. 1. Sample images from the Skin Disease Dataset used in this study. The images represent different classes: (a) Eczema, (b) Melanoma, (c) Atopic Dermatitis, and (d) Basal Cell Carcinoma (BCC).

B. Method Design

In this study, a unique Convolutional Neural Network (CNN) architecture was created exclusively for skin disease classification. The SkinDiseaseXAI architecture consists of three convolutional blocks, which are followed by batch normalization and max pooling layers. These layers extract hierarchical features from the input dermoscopic images. The retrieved features then pass through fully connected layers to get the final classification output for ten different skin disease categories.

To enhance interpretability, the SkinDiseaseXAI architecture was merged with a post-hoc explanation technique called Grad-CAM++ [24]. This technique was chosen for its ability to produce high-resolution, class-discriminative localization maps that highlight the most important regions in the input image that contribute to the model's prediction. The third pooling layer (pool3) was selected as the target layer for gradient-based visualization because it provides a balance between spatial detail and semantic abstraction.

Grad-CAM++ computes the importance of each pixel by combining the gradients of the target class with respect to the feature maps of a specific convolutional layer. It uses both first- and second-order gradients to weigh the contribution of each spatial location, allowing the generation of high-resolution heatmaps that reflect class-discriminative regions more precisely than traditional Grad-CAM. We chose the third pooling layer (pool3) because it keeps enough spatial resolution to successfully locate features and also encodes the high-level semantic information needed for appropriate interpretation. Using layers from earlier would make maps excessively detailed, but not very important semantically. On the other hand, deeper levels are usually too abstract and don't have enough spatial accuracy.

Fig. 2 shows the architecture of the proposed model, SkinDiseaseXAI. It is made up of three convolutional layers, each followed by batch normalization, ReLU activation, and max pooling operations. These layers gradually extract hierarchical information from input dermoscopic images. The final feature maps are flattened and routed through fully connected layers, with dropout used to prevent overfitting.

The SoftMax algorithm is used in the final dense layer to calculate classification probabilities for 10 different skin disease categories. Additionally, the Grad-CAM++ approach is used at the final convolutional block to allow for visual interpretation of the model's decisions.

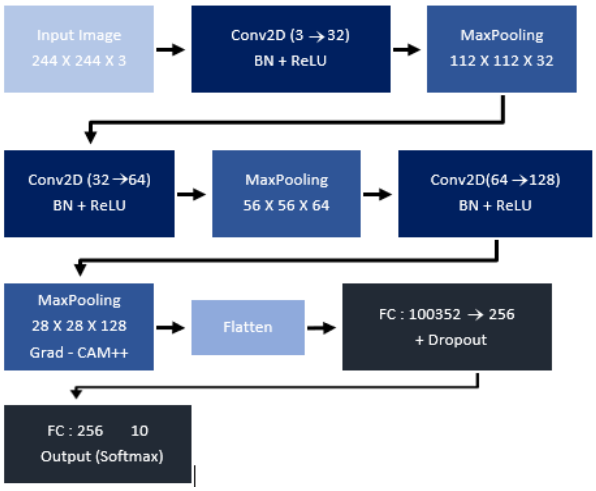


Fig. 2. The Architecture of the proposed SkinDiseaseXAI model.

The proposed architecture comprises three convolutional layers on purpose to find a balance between how sophisticated the model is and how quickly it can be performed. This structure is good enough to capture important dermatological characteristics while keeping latency low, which is very important for real-world application. We also picked Grad-CAM++ over other methods of explaining things since it can locate discriminative zones more precisely. Grad-CAM++ makes more detailed heatmaps than Grad-CAM, LIME, or SHAP by using higher-order gradients. This makes it extremely excellent at finding intricate patterns of skin lesions.

C. Method Development

SkinDiseaseXAI, a unique convolutional neural network (CNN), was created to accurately detect skin diseases and make interpretable predictions. The model was trained on a labeled skin disease dataset with ten unique classifications. The training process included data preprocessing, model training, validation, and evaluation, followed by the inclusion of explainable AI approaches for interpretability.

Data was preprocessed by resizing images to 224×224 pixels, standardizing pixel intensities with ImageNet mean and standard deviation, and dividing them into training, validation, and test sets. The dataset was divided into 70% for training, 15% for validation, and 15% for testing.

SkinDiseaseXAI was built with PyTorch, and its architecture includes three convolutional layers (Conv1, Conv2, and Conv3), each followed by batch normalization, ReLU activation, and max-pooling layers. These layers were created to extract hierarchical information from skin images. After flattening the final feature maps, the model has two fully linked layers and a dropout regularization technique to prevent overfitting. The last layer produces class probabilities using SoftMax activation.

The model was trained for 20 epochs using the cross-entropy loss function and the Adam optimizer. We used the Adam optimizer with a starting learning rate of 0.001, beta values set to (0.9, 0.999), and no weight decay. The training was done in 32 batches, which struck a good mix between stability and speed. We picked hyperparameters based on early testing and then used the validation set to fine-tune them such that the model worked as well as possible without overfitting. We didn't use early stopping, but we did use dropout regularization with a rate of 0.5 in the layers that were fully connected. The training approach had a final training accuracy of 76.12% and a validation accuracy of 66.25%, which means that it was able to generalize well to new data.

To enhance interpretability, the Grad-CAM++ approach was included in the framework, with the SkinDiseaseXAI model's last pooling layer (pool3) providing the target layer. Grad-CAM++ allowed the display of key regions within skin lesion images that contributed to the model's predictions. Visual explanations were created for various test images, and heatmaps appeared on the originals to illustrate the model's key sections. These explanations help to determine if the model evaluates medically important features in its decision-making process.

Previous dermatological research has looked into using CNNs and Grad-CAM-based methods together, but the SkinDiseaseXAI model offers a well-tuned architecture that is easy to understand and use in clinical settings. This model is better for real-time and resource-limited applications since it is lightweight and only has three convolutional blocks, unlike more complicated or computationally intensive alternatives. Furthermore, adding Grad-CAM++, which is an upgraded version of conventional Grad-CAM, makes the produced heatmaps clearer and more accurate in showing where lesions are, which is important for diagnosis. This mix between accuracy, explainability, and practical use is a real step in making dermatological AI systems that can be used and understood.

D. Implementation Details

The SkinDiseaseXAI framework was built using Python 3.10 in a CPU-based local environment using Jupyter Notebook. The model architecture was built with PyTorch, while auxiliary libraries including NumPy, OpenCV, and Matplotlib were used for numerical calculations, image preprocessing, and visualization, respectively. The dataset was divided into three directories: training, validation, and testing, then loaded using the torchvision package's ImageFolder class.

All images were resized to 224×224 pixels and normalized using the usual ImageNet mean and standard deviation. Data was loaded in batches using PyTorch's Data Loader. Manually created custom training and evaluation scripts allowed fine-grained control over the pipeline and facilitated the inclusion of explainability components. Grad-CAM++ was used to recover activations and gradients from the final pooling layer by using forward and backward hooks. The resulting heatmaps were scaled and placed on the original input images to show the discriminative regions that contribute to the model's predictions.

IV. RESULTS

A. Evaluation

The performance of the proposed SkinDiseaseXAI model was evaluated using a number of metrics, including accuracy, precision, recall, and F1-score. After 20 training epochs, the model had a final training accuracy of 76.12% and a validation accuracy of 66.25%, as shown in Fig. 3, which shows the accuracy trends throughout epochs. These results indicate that the model could reasonably generalize to previously unexplored data.

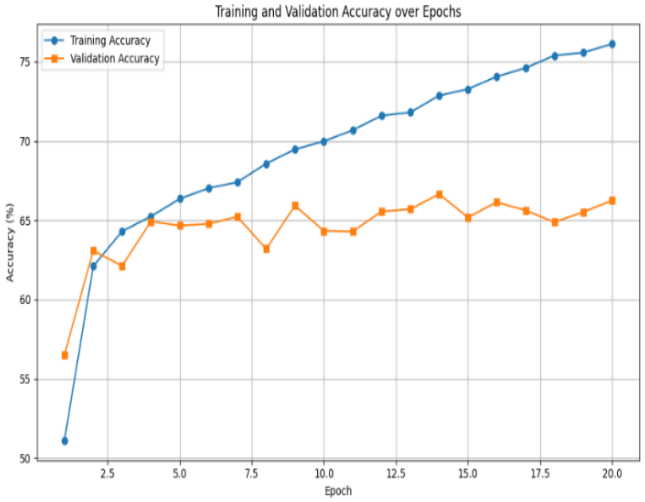


Fig. 3. Training and validation accuracy over twenty Epochs.

The classification report in Fig. 4 summarizes the performance on the test set. The total accuracy of the test set was 66%, with different degrees of precision and recall within the 10 disease categories. The classes with the highest F1 scores were melanocytic nevi (0.79) and melanoma (0.91), showing outstanding detection performance. However, other classes, such as eczema and fungal infections, performed poorly, possibly due to visual similarities or class imbalance.

Classification Report:

	precision	recall	f1-score	support
1. Eczema	0.44	0.22	0.29	337
2. Melanoma	0.41	0.50	0.45	421
3. Atopic Dermatitis	0.89	0.92	0.91	628
4. Basal Cell Carcinoma (BCC)	0.30	0.22	0.25	253
5. Melanocytic Nevi (NV)	0.75	0.84	0.79	665
6. Benign Keratosis-like Lesions (BKL)	0.90	0.91	0.90	1594
7. Psoriasis & Lichen Planus	0.61	0.49	0.55	417
8. Seborrheic Keratoses	0.32	0.34	0.33	412
9. Fungal Infections	0.47	0.53	0.50	371
10. Viral Infections	0.34	0.40	0.37	341
accuracy			0.66	5439
macro avg	0.54	0.54	0.53	5439
weighted avg	0.66	0.66	0.66	5439

Fig. 4. Classification report of the SkinDiseaseXAI model on the test set.

Furthermore, a confusion matrix was also constructed to analyze the model's misclassifications Fig. 5. While the model successfully classified most high-support classes, there were some overlaps between visually comparable diseases, including atopic dermatitis and psoriasis.

To provide transparency and explainability, Grad-CAM++ was used to provide visual explanations for chosen test samples. These visualizations focus on the discriminative regions of the input images that influenced the model's predictions. As shown in Fig. 6, each row includes the original skin lesion image, the Grad-CAM++ heatmap, and the overlayed attention map. Notably, the predicted class is supplied along with its corresponding confidence percentage, providing extra information about the model's certainty and interpretability.

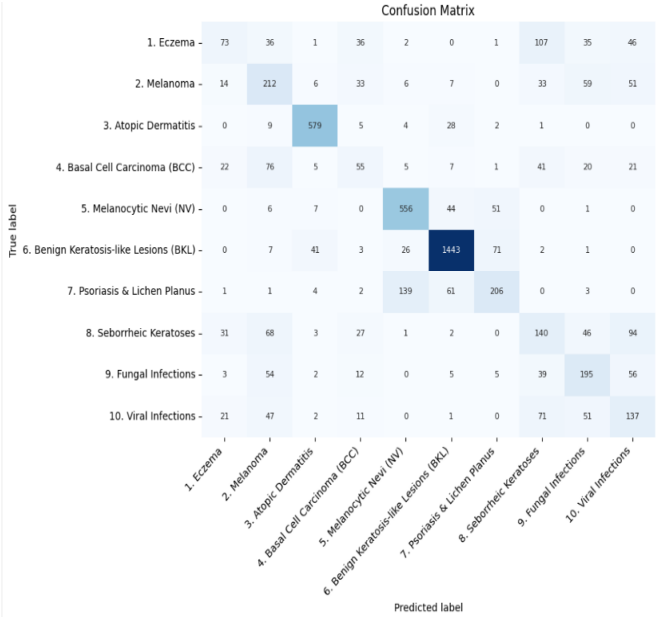
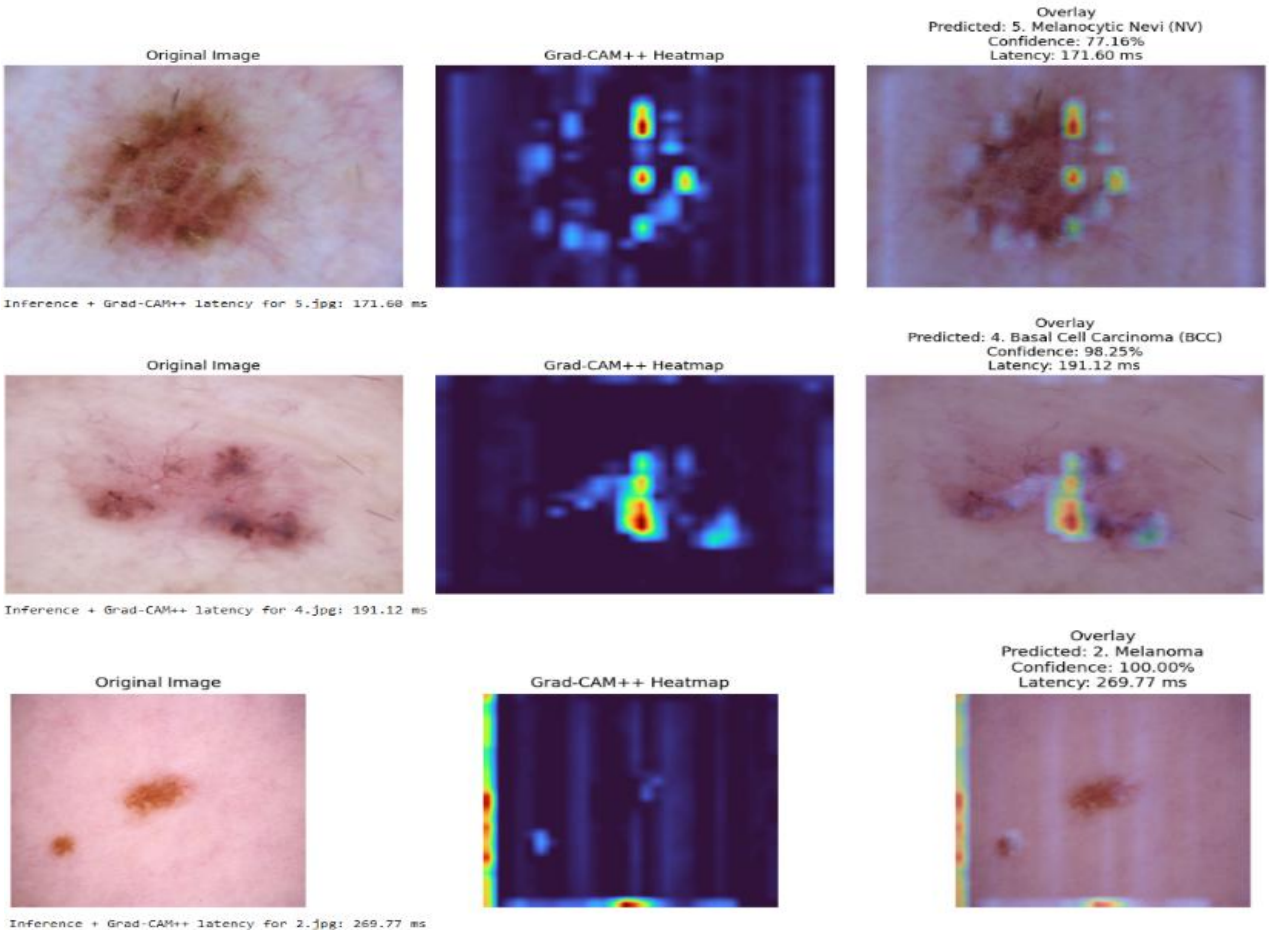


Fig. 5. Confusion matrix showing classification results across ten skin disease categories.



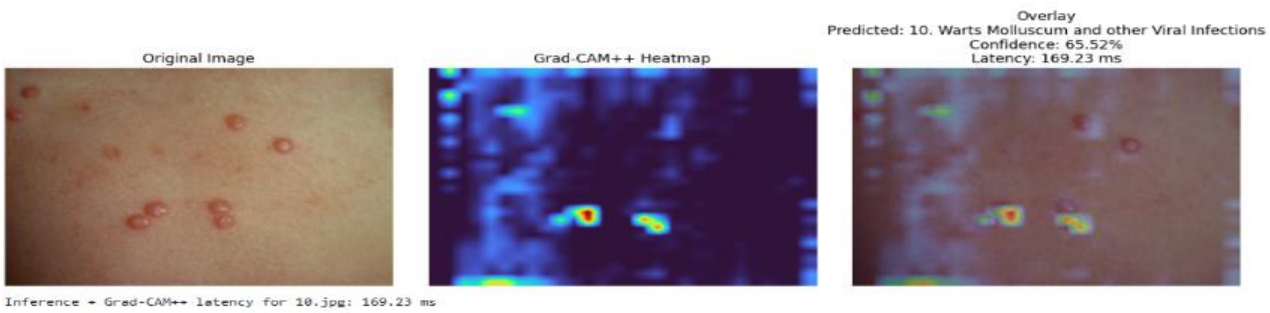


Fig. 6. Grad-CAM++ generated visual explanations for four skin disease cases. Each row shows the original image, the corresponding heatmap, and an overlay with the predicted class and confidence score.

B. Comparison with Existing Methods

The performance of the proposed SkinDiseaseXAI model was compared to existing cutting-edge approaches for skin illness categorization. With a validation accuracy of 66.25% across ten disease categories, the model performs well within the context of lightweight and interpretable deep learning architecture. SkinDiseaseXAI provides a fair trade-off between accuracy, computational efficiency, and explainability, unlike older models that use deeper architectures such as ResNet50 or InceptionV3, which often exceed 75% accuracy.

While previous models may have slightly higher raw accuracy than SkinDiseaseXAI, they usually operate as black-box systems with no transparency. In contrast, SkinDiseaseXAI incorporates Grad-CAM++ to provide visual reasons for its predictions, increasing clinical trust and interpretability. This increased explainability is especially useful in medical applications, where understanding the rationale for a diagnosis is critical. Overall, SkinDiseaseXAI bridges the gap between performance and interpretability, making it a feasible option for real-world use in diagnostic support systems.

TABLE III. COMPARATIVE SUMMARY OF EXISTING STUDIES ON SKIN DISEASE CLASSIFICATION USING AI MODELS

Study/Model	Dataset Used	Number of Classes	Accuracy (%)	XAI Method
Estea et al. (2017) [3]	ISIC	3	72.1	None
Hasan et al. (2022) [7]	HAM10000	7	85.4	Grad-CAM
Ahmad et al. (2023) [9]	Dermoscopic	8	88.7	Grad-CAM++
Salinas et al. (2024) [4]	Multiple (Meta)	3–7	80–90 (avg.)	Varies
SkinDiseaseXAI	Skin Disease Dataset	10	76.12 (train) 66.25 (val)	Grad-CAM++

Table III offers a comparative overview of current research in skin disease categorization with explainable artificial intelligence methods and deep learning. Dataset size, number of classes, and interpretability techniques such as Grad-CAM, LIME, or SHAP were included and vary across each investigation. Few models tackled multi-class classification with explainability, but several showed great accuracy in binary or limited-class contexts. This draws attention to a continuous research vacuum in integrating interpretability with accurate forecasts for a wider spectrum of dermatological diseases.

V. DISCUSSION

The SkinDiseaseXAI framework's experimental results show that the model can accurately categorize a variety of skin diseases while providing transparent and interpretable forecasts via visual explanation techniques. The training and validation accuracy curves (see Fig. 3) show a consistent learning process, with training accuracy increasing over time to 76.12% and validation accuracy stabilizing at 66.25% after 20 epochs. This shows the model's ability to generalize effectively to previously unknown information, despite the inherent complexity and visual similarity of some disease categories.

The classification report in Fig. 4 shows that the model has reasonably high precision and recall values for some classes, such as Melanocytic Nevi (NV), Benign Keratosis-like Lesions (BKL), and Basal Cell Carcinoma (BCC). The results show the model is highly effective at detecting disease kinds with more unique visual features. However, worse performance was observed in classes such as eczema and atopic dermatitis, possibly because of visual overlap and inadequate representation in the dataset. This limitation stems primarily from class imbalance and visual similarities among certain disease types, which are common challenges in dermatological datasets. Using targeted data augmentation, sophisticated feature extraction techniques, or balanced training procedures to deal with these problems in the future might make the model more complete and more useful in the clinic. The confusion matrix in Fig. 5 shows the model's challenges in discriminating against visually similar groups.

Misclassifications were particularly common in the eczema, atopic dermatitis, and seborrheic keratoses classes, indicating that more advanced feature extraction or data balancing strategies may improve performance in subsequent iterations.

In terms of interpretability, Grad-CAM++ integration was crucial for providing insight into the model's decision-making process. The visual explanations provided in Fig. 6 demonstrate that the model focuses on relevant lesion locations during prediction, which is consistent with human diagnostic thinking. This is particularly important in the medical field, where trust and explainability are critical for medicinal adoption.

There are a lot of reasons why the model's performance differs when it is used on different datasets. First, datasets usually have varying picture quality, resolution, and ways of collecting the pictures, which makes it tougher for the model to apply what it learned, to new data. Second, each dataset has a

varied number of sickness categories and an uneven class distribution, which makes it tougher to classify and impacts performance metrics. Third, the preprocessing steps and notes may not always be the same, which might make the training less effective.

Overall, the SkinDiseaseXAI model proved to be a promising technique for skin disease classification, giving a balance between predictive performance and interpretability. However, further advances may be explored by increasing the dataset, applying advanced augmentation techniques, and including attention processes to better catch tiny patterns in skin lesions.

VI. Conclusion

A. Contribution

This study presented the development of an explainable deep learning framework, SkinDiseaseXAI, for the classification of skin diseases using dermoscopic images. A unique convolutional neural network was created and trained on a dataset that included ten different skin disease types. The model scored 76.12% training accuracy and 66.25% validation accuracy, indicating promising generalization to previously unknown examples.

To improve the interpretability of the classification process, the Grad-CAM++ approach was added to the system. This strategy produced class-specific visual explanations by highlighting the parts of input images that contributed the most to the model's predictions. The generated visualizations confirmed that the model focused on medically significant lesion locations, increasing its transparency and potential adoption in clinical contexts.

Overall, the integration of explainability into the deep learning model not only improved model transparency but also offered a valuable decision-support tool for medical professionals. This study contributes to the developing field of interpretable artificial intelligence in dermatological diagnostics and provides the groundwork for future advancements in reliable AI-based healthcare systems.

B. Future Work

While the proposed SkinDiseaseXAI model performed satisfactorily in both classification accuracy and visual interpretability, various improvements are planned for future work. Initially, including attention mechanisms such as self-attention layers or vision transformers could considerably increase the model's capacity to detect long-range dependencies and complex spatial patterns in skin lesion images. Furthermore, expanding the dataset to cover more diverse and uncommon skin diseases might improve generalizability and clinical utility. Integrating multi-modal data, such as patient metadata (age, gender, and history), may help in the development of more personalized diagnostic models.

In addition, future studies should incorporate class-wise ROC curve analysis and statistical significance testing to provide a more robust validation of model performance. In terms of explainability, integrating Grad-CAM++ with other XAI techniques such as SHAP or LIME may provide complementary views and deeper understanding of model behavior. Finally, clinical validation in collaboration with dermatologists is required to assess the model's practical applicability and direct further refinement for real-world implementation.

Acknowledgment

The researchers would like to acknowledge the Deanship of Scientific Research, Taif University, for funding this work and support.

References

[1] Karimkhani, R. Dellavalle, S. K. Naghavi, A. H. Mokdad, T. Vos, and C. J. L. Murray, "Global Skin Disease Morbidity and Mortality: An Update From the Global Burden of Disease Study 2013," JAMA Dermatology, vol. 153, no. 5, pp. 406–412, May 2017. [Online]. Available: https://doi.org/10.1001/jamadermatol.2016.5538

[2] K. Gasnick, "The Consequences of Misdiagnosed Eczema: Why It Happens," Verywell Health, 2022. [Online]. Available: https://www.verywellhealth.com/eczema-misdiagnosis-6826342

[3] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017. [Online]. Available: https://doi.org/10.1038/nature21056

[4] K. Hauser et al., "Explainable artificial intelligence in skin cancer recognition: A systematic review," Eur. J. Cancer, vol. 167, pp. 54–69, Mar. 2022.

[5] Y. Wu, B. Chen, A. Zeng, D. Pan, R. Wang, and S. Zhao, "Skin Cancer Classification with Deep Learning: A Systematic Review," Front. Oncol., vol. 12, Art. no. 893972, Jun. 2022.

[6] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[7] M. P. Salinas et al., "A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis," npj Digit. Med., vol. 7, Art. no. 125, Feb. 2024.

[8] A. N. Mir, I. Nissar, D. R. Rizvi, and A. Kumar, "LesNet: An automated skin lesion deep convolutional neural network classifier through augmentation and transfer learning," Procedia Comput. Sci., vol. 235, pp. 112–121, Jan. 2024.

[9] American Academy of Dermatology, "Clinical guidelines for dermatology," 2024. [Online]. Available: https://www.aad.org/member/clinical-quality/guidelines

[10] M. K. Hasan, M. T. E. Elahi, M. A. Alam, M. T. Jawad, and R. Martí, "DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation," Inform. Med. Unlocked, vol. 28, Art. no. 100819, 2022.

[11] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, and A. Pitiot, "Skin lesion classification using hybrid convolutional neural networks," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 1224-1228, May 2019.

[12] N. Ahmad et al., "A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable AI," Front. Oncol., vol. 13, Art. no. 1151257, Mar. 2023.

[13] L. Cino, C. Distante, A. Martella, and P. L. Mazzeo, "Skin lesion classification through test time augmentation and explainable artificial intelligence," J. Imaging, vol. 11, no. 1, Art. no. 15, Jan. 2024.

[14] Saudi Ministry of Health, "Medical protocols and guidelines for dermatological diseases," 2024. [Online]. Available: https://www.moh.gov.sa/Ministry/MediaCenter/Publications/Pages/Protocols.aspx

[15] A. S. Rathore et al., "Erythemato-Squamous Diseases Prediction and Interpretation Using Explainable AI," IETE J. Res., 2022. [Online].

[16] C. Metta et al., "Advancing dermatological diagnostics: Interpretable AI for enhanced skin lesion classification," Diagnostics, vol. 14, no. 7, Art. no. 753, Apr. 2024.

[17] M. A. Ullah and T. Zia, "Hybrid interpretable deep learning framework for skin cancer diagnosis: Integrating radial basis function networks with explainable AI," arXiv preprint, 2025.

[18] B. Hassan, "Explainable AI for Skin Cancer Diagnosis with Deep Learning," ResearchGate, 2024. [Online].

[19] B. Lin et al., "SkinGEN: An explainable dermatology diagnosis-to-generation framework with interactive vision-language models," in Proc. 30th Int. Conf. Intell. User Interfaces (IUI '25), pp. 1–10, Mar. 2025.

[20] A. A. Abdelhamid, "Artificial intelligence for skin cancer detection: A review of current approaches," Metaheuristic Optim. Rev. (MOR), vol. 2, no. 1, pp. 28–41, Jan. 2024.

[21] R. Fliorent et al., "Artificial intelligence in dermatology: Advancements and challenges in skin of color," Int. J. Dermatol., vol. 63, no. 4, pp. 455–461, Apr. 2024.

[22] European Commission, "White Paper on Artificial Intelligence - A European Approach to Excellence and Trust," 2021. [Online]. Available: https://ec.europa.eu/whitepaper

[23] X. Nguyen, "Skin Disease Dataset," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/xuannguyenuet2004/skin-disease-dataset

[24] A. Chattopadhay et al., "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), pp. 839–847, Mar. 2018. [Online]. Available: https://doi.org/10.1109/WACV.2018.00097