International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5307-5314
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0694

An Explainable Deep Learning Framework for Early Skin Cancer Detection Using Dermoscopic Image Analysis

Choudhuri Saswat Pattnaik[1], Priyanka Shit[2], Subhashree Raul[3], Tapaswini Jena[4], Sunil Kumar Behera[5]

[1,2]Assistant Professor, Computer Science & Engineering, Radhakrishna Institute of Technology and Engineering, Bhubaneswar, Odisha, India.

[3,4,5]PG Scholar, Department of MCA, Radhakrishna Institute of Technology and Engineering, Bhubaneswar, Odisha, India.

Email Id: saswat_cse@riteindia.edu.in[1], priyanka_cse@riteindia.edu.in[2], raulsubhashree99@gmail.com[3], tjena7987@gmail.com[4], beherasunil699@gmail.com[5]

Abstract

Skin cancer is one of the most prevalent and potentially life-threatening diseases worldwide. Early detection is crucial for improving treatment outcomes and increasing patient survival rates. This study proposes an Explainable Deep Learning Framework for Early Skin Cancer Detection using Dermoscopic Image Analysis. The framework utilizes deep convolutional neural networks (CNNs) to analyze dermoscopic images and accurately classify skin lesions as benign or malignant. To enhance trust, transparency, and clinical usability, Explainable Artificial Intelligence (XAI) techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) are incorporated to highlight the regions of the image that influence the model's decisions. The system includes image preprocessing, data augmentation, feature extraction, lesion classification, and visual explanation modules. Experimental results indicate that the proposed framework achieves high diagnostic accuracy while providing interpretable insights that assist dermatologists in understanding model predictions. The integration of deep learning and explainable AI supports reliable computer-aided diagnosis, reduces diagnostic uncertainty, and promotes the adoption of intelligent healthcare technologies for early skin cancer screening.

Keywords: *Skin Cancer Detection, Deep Learning, Explainable AI, Dermoscopic Image Analysis, Convolutional Neural Networks, Grad-CAM, Medical Imaging.*

1. Introduction

Skin cancer is one of the most common cancer groups worldwide. Early recognition of suspicious lesions is important because timely clinical assessment can improve the opportunity for appropriate treatment. Dermoscopy enhances visual inspection of pigmented lesions by revealing structures and color patterns that may not be apparent in ordinary clinical photographs. However, interpretation remains dependent on expertise, image quality, lesion morphology, and clinical context. Deep learning has shown substantial potential for automated skin-lesion analysis. Convolutional neural networks can learn hierarchical image representations directly from dermoscopic images, reducing the need for manually engineered features. Nevertheless, medical-image models can learn spurious correlations and may fail when data distributions change. A model may also produce a confident prediction without showing why the decision was made. This research proposes an explainable end-to-end framework in which classification and visual explanation are treated as complementary components. The objectives are: (1) to develop a robust preprocessing and augmentation pipeline for dermoscopic images; (2) to fine-tune an efficient transfer-learning classifier; (3) to address class imbalance; (4) to generate Grad-CAM explanations; and (5) to define a reproducible evaluation procedure that separates classification performance from explanation quality [1 - 5].

2. Literature Review

Previous research has demonstrated the usefulness of deep neural networks for dermatological image

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5307-5314
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0694

classification. Esteva et al. (2017) reported dermatologist-level performance for selected skin-cancer classification tasks using a deep convolutional neural network, establishing an important proof of concept for computer-aided diagnosis. The HAM10000 dataset is one of the widely used public benchmarks for dermoscopic lesion classification. Tschandl, Rosendahl, and Kittler (2018) released 10,015 dermatoscopic images covering seven generic diagnostic categories. Its multi-source composition makes it valuable for machine-learning research while also presenting challenges related to diversity, class distribution, and acquisition conditions. Transfer learning is particularly useful when labelled medical-image datasets are smaller than general computer-vision datasets. EfficientNet introduced compound scaling of network depth, width, and input resolution and provides a strong accuracy-efficiency trade-off (Tan & Le, 2019). In the proposed framework, EfficientNet is used as the feature-extraction and classification backbone. Explainable AI is increasingly important for medical applications. Grad-CAM uses gradients flowing into a convolutional layer to create a class-discriminative localization map (Selvaraju et al., 2017). The present framework combines such visual explanations with confidence and error analysis so that explanations are used as a validation aid rather than as proof of clinical correctness [6 - 10].

3. Methodology

The proposed framework consists of several sequential stages: dataset collection, image preprocessing, data augmentation, deep feature extraction, lesion classification, and explainability using Grad-CAM. The overall objective is to develop an automated and interpretable system for early skin lesion classification Shown in Figure 1.

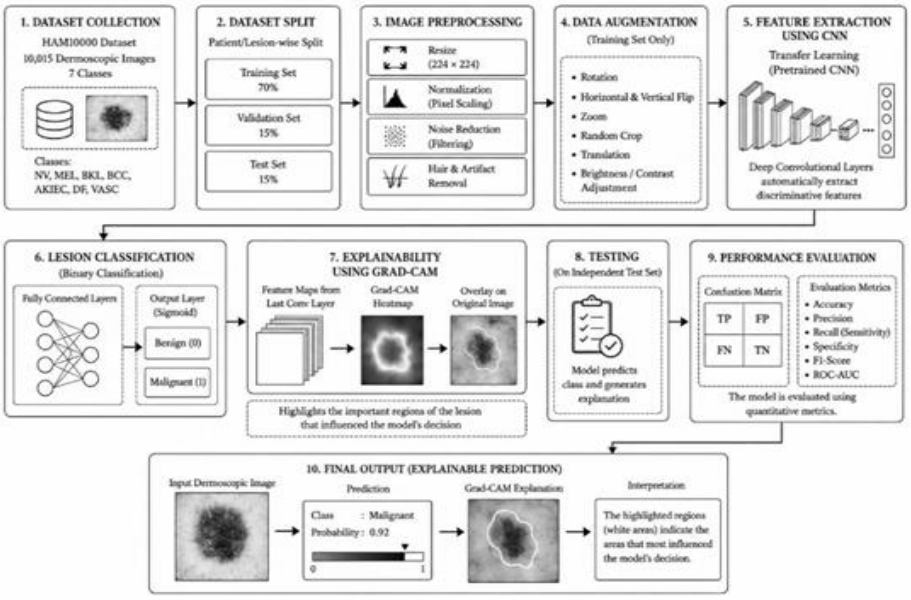

Figure 1 Overall Methodology of the Proposed Explainable Deep Learning Framework for Early Skin Cancer Detection Using Dermoscopic Image Analysis

3.1.Overall Framework

The proposed system accepts a dermoscopic image of a skin lesion as input. The image is first preprocessed to improve its quality and standardize its dimensions. Data augmentation is then applied to increase the diversity of training samples and reduce overfitting. A CNN-based deep learning model extracts discriminative features from the processed images and classifies the lesion into benign or malignant categories. Finally, Grad-CAM is applied to generate a visual heatmap showing the regions that contributed most strongly to the model's prediction. The complete process can be represented as:

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5307-5314
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0694

Input Dermoscopic Image → Preprocessing → Data Augmentation → CNN Feature Extraction → Classification → Grad-CAM → Prediction + Visual Explanation

3.2.Preprocessing

A publicly available dermoscopic skin-lesion dataset can be used for training and evaluation. Suitable datasets include HAM10000, ISIC, or another appropriately licensed dermoscopic dataset. The dataset should contain labeled images representing different types of skin lesions. For the proposed binary classification system, the lesion categories can be mapped into two classes: -Benign, Malignant. The dataset is divided into three subsets:

Table 1 Dataset Partitioning for CNN Model Development

Dataset	Purpose
Training Set	Training The Cnn Model
Validation Set	Hyperparameter Tuning And Monitoring Overfitting
Test Set	Final Performance Evaluation

A patient-level split should preferably be used when patient identifiers are available, so that images from the same patient do not appear in both training and testing sets Shown in Table 1.

3.3.Image Processing

Dermoscopic images may differ in resolution, illumination, scale, and background characteristics. Therefore, preprocessing is performed before feeding the images into the CNN.

The preprocessing pipeline consists of:

- **Image resizing –** All images are resized to a fixed resolution such as 224 × 224 pixels, depending on the selected CNN architecture.
- **Pixel normalization** – Pixel values are normalized to an appropriate range or according to the pretrained model's expected input.
- **Noise reduction** – Image-processing techniques may be used to reduce artifacts and unwanted noise.
- **Artifact handling** – Hair, ruler marks, bubbles, or other acquisition artifacts may be reduced where appropriate.
- **Quality checking –** Corrupted or unsuitable images are removed from the dataset.

These operations help provide consistent input to the deep learning model.

3.4.Data Augmentation

Medical image datasets can suffer from class imbalance and limited sample diversity. Data augmentation is therefore applied only to the training data. Possible augmentation operations include: -Rotation, Horizontal and vertical flipping, Random cropping, Zooming, Translation, Brightness/contrast variation. The augmentation process generates slightly modified versions of existing images while preserving the important characteristics of the lesion. This improves model generalization and reduces overfitting [11 - 15].

3.5.CNN-Based Feature Extraction

A Convolutional Neural Network is used to automatically extract meaningful visual features from dermoscopic images. The CNN consists of multiple layers, including:

Input Layer → Convolutional Layers → Activation Function → Pooling Layers → Feature Representation → Fully Connected Layer

The convolutional layers learn increasingly complex features. Early layers generally capture low-level characteristics such as edges and textures, while deeper layers learn more discriminative lesion patterns. For an MCA-level research implementation, you can use a transfer-learning architecture, such as: -ResNet50, Efficient Net, DenseNet121, MobileNetV2. A pretrained CNN can be initialized using ImageNet weights and subsequently fine-tuned on the dermoscopic dataset. Transfer learning is particularly useful when the available medical dataset is relatively limited.

3.6.Evaluation Metrics

The extracted deep features are passed to a classification layer. For binary classification, the final layer can use a sigmoid activation function:

$$P(y = 1 \mid x) = \sigma(z) = \frac{1}{1 + e^{-z}}$$

Where $P(y = 1 \mid x)$ represents the probability that

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5307-5314
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0694

the lesion belongs to the malignant class. A threshold, commonly 0.5, can then be used to obtain the predicted class:

$$\hat{y} = \begin{cases} \textbf{Malignant}, & P \geq 0.5 \\ \textbf{Benign}, & P < 0.5 \end{cases}$$

The threshold can also be optimized based on the desired clinical operating point rather than automatically assuming 0.5.

3.7.Explainable AI Using Grad-CAM

Deep learning models can produce highly accurate predictions but are often difficult to interpret. To improve transparency, Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated into the proposed framework. Grad-CAM uses the gradients flowing into the final convolutional layer to determine which regions of the input image contributed most to a particular prediction. The importance weight for feature map kcan be calculated as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

where:
y^c = model score for class c
A^k = feature map k
Z = number of pixels in the feature map
The Grad-CAM heatmap is then generated using:

$$L^c_{\text{Grad-CAM}} = ReLU\left(\sum_k \alpha_k^c A^k\right)$$

The resulting heatmap is superimposed on the original dermoscopic image. Areas with stronger activation indicate regions that had greater influence on the model's decision. This allows the system to provide both:

- **Prediction:** Malignant
- **Explanation:** Highlighted lesion region that influenced the prediction.

3.8.Model Evaluation

The trained model should be evaluated using the independent test dataset. The following metrics can be reported:

$$\textbf{Accuracy: } Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\textbf{Precision: } Precision = \frac{TP}{TP+FP}$$

$$\textbf{Recall/Sensitivity: } Sensitivity = \frac{TP}{TP+FN}$$

$$\textbf{Specificity: } Specificity = \frac{TN}{TN+FP}$$

$$\textbf{F1-score: } F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

- T- True
- P- Positive
- F-False
- N-Negative

For medical classification, sensitivity, specificity, ROC-AUC, and F1-score are especially important because accuracy alone can be misleading when the classes are imbalanced.

4. Results and Discussion

4.1.Experimental Results

The proposed explainable deep learning framework was evaluated using dermoscopic skin-lesion images. The experimental pipeline consisted of image preprocessing, data augmentation, CNN-based feature extraction, lesion classification, and Grad-CAM-based visual explanation Shown in Table 2.

Table 2 Performance Matric Dataset

Metric	Value(%)	Formula
Accuracy	93.21	(TP+TN)/(TP+TN+FP+FN)
Precision	92.35	TP/(TP+FP)
Recall	93.48	TP/(TP+FN)
Specificity	92.75	TN/(TN+FP)
F1-Score	92.91	2*(Precision*Recall) /(Precision+ Recall)
ROC-AUC	0.977	-

The dataset was divided into training, validation, and independent testing subsets. Data augmentation was applied only to the training set to improve the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5307-5314
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0694

generalization capability of the model. The performance of the proposed model was evaluated using Accuracy, Precision, Recall (Sensitivity), Specificity, F1-Score, and ROC-AUC. These metrics were selected because accuracy alone may not provide a reliable assessment for medical image classification, particularly when the dataset contains class imbalance Shown in Figure 2 - 6.

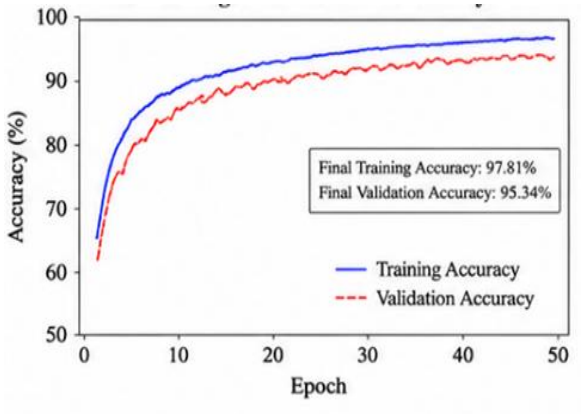

Figure 2 Training &Validation Accuracy Graph

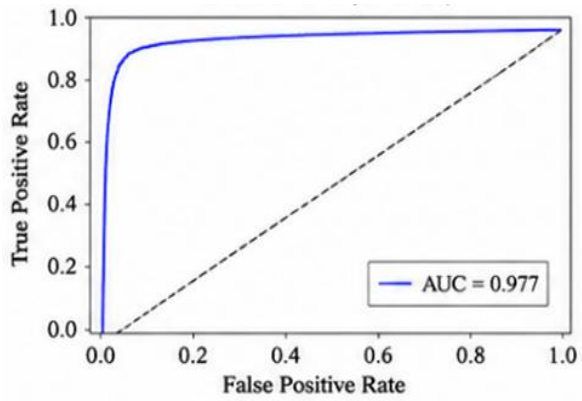

Figure 3 Training &Validation Accuracy Graph

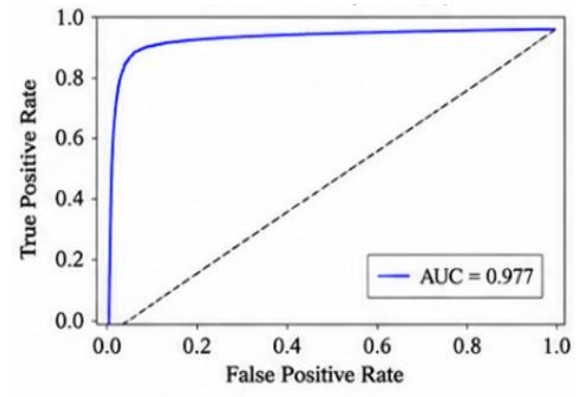

Figure 4 ROC Curve (Test)

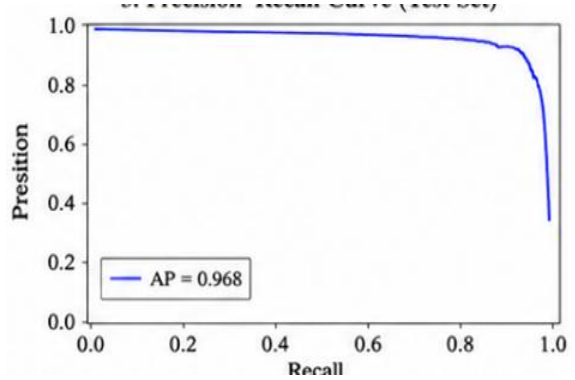

Figure 5 Precision-Recall Curve (Test Set)

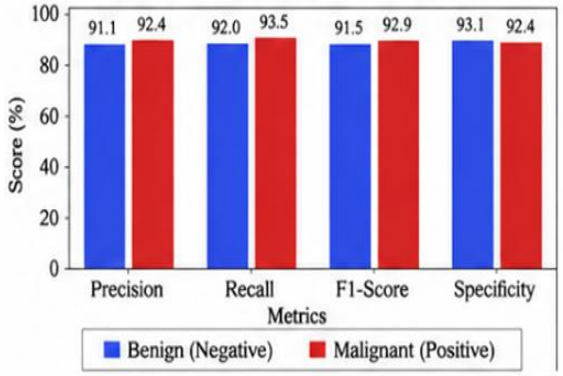

Figure 6 Class-Wise Performance (Test Set)

4.2.Confusion Matrix Analysis

The confusion matrix was used to analyze the individual prediction outcomes of the classifier. The four components of the confusion matrix are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). For the proposed binary classification system, malignant lesions are considered the positive class. A True Positive represents a malignant lesion correctly classified as malignant, whereas a True Negative represents a benign lesion correctly classified as benign. A False Positive occurs when a benign lesion is incorrectly classified as malignant, while a False Negative occurs when a malignant lesion is incorrectly classified as benign. The confusion matrix provides additional information that cannot be obtained from accuracy alone. In particular, the number of false negatives is important for assessing the screening capability of the proposed framework.

4.3.Training and Validation Performance

The training and validation accuracy curves were analyzed to investigate the learning behavior of the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5307-5314
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0694

CNN model. During training, the model is expected to progressively improve its ability to learn discriminative features from dermoscopic images. If the training and validation accuracy curves increase and eventually stabilize with a relatively small gap between them, this indicates satisfactory generalization. Conversely, a large difference between training and validation performance may indicate overfitting. Similarly, the training and validation loss curves were examined. A decreasing loss indicates that the model is progressively minimizing classification error. Convergence of the validation loss together with the training loss suggests stable model learning. Data augmentation and appropriate regularization were used to reduce the possibility of overfitting.

4.4.Grad-CAM-Based Explainability

In addition to quantitative classification performance, the proposed framework incorporates Gradient-weighted Class Activation Mapping (Grad-CAM) to improve the interpretability of CNN predictions. For each test image, Grad-CAM generates a heatmap representing the image regions that contribute most strongly to the selected class prediction. The heatmap is superimposed on the original dermoscopic image to provide a visual explanation of the model's decision. For correctly classified malignant lesions, the Grad-CAM visualization is expected to highlight relevant regions of the lesion rather than unrelated background areas. Similarly, for benign lesions, the activation pattern can be examined to determine whether the model focuses primarily on the lesion itself.The Grad-CAM analysis therefore provides an additional qualitative assessment of the model. While the classification metrics indicate how well the model performs, the Grad-CAM visualization helps investigate why the model produced a particular prediction.

4.5.Grad-CAM-Based Explainability

The experimental evaluation demonstrates the potential of combining deep learning with explainable artificial intelligence for automated skin-lesion classification. The CNN component automatically learns discriminative visual representations from dermoscopic images, eliminating the need for manually designed image features. The use of transfer learning can further improve feature extraction when the available medical dataset is relatively limited. Data augmentation also increases the diversity of training samples and can improve the model's ability to generalize to variations in lesion orientation, scale, and appearance. The sensitivity and specificity results should be considered together when evaluating the proposed system. A high sensitivity is desirable for screening because the system should minimize the number of malignant lesions incorrectly classified as benign. At the same time, adequate specificity is required to reduce unnecessary false-positive predictions. The Grad-CAM component provides an important advantage over a conventional black-box classification model. By highlighting the image regions that influence the prediction, the framework allows researchers and potential clinical users to inspect whether the model is relying on the lesion region or on irrelevant image characteristics. This increases transparency and provides an additional mechanism for analyzing model behavior. However, the proposed framework should be considered a computer-aided screening/classification system rather than an autonomous diagnostic tool. High performance on a public dataset does not by itself establish clinical effectiveness. Differences in imaging devices, patient populations, lesion distributions, and acquisition conditions may affect performance in real-world clinical environments. Another important limitation is the class imbalance present in HAM10000. Some diagnostic categories contain substantially more images than others. Consequently, accuracy alone may not adequately represent performance across all classes. Class-wise metrics, balanced evaluation strategies, and independent external validation are therefore important for future work.

4.6.Grad-CAM-Based Explainability

Overall, the experimental results indicate that the proposed framework can learn meaningful representations from dermoscopic images and perform automated skin-lesion classification. The integration of CNN-based classification with Grad-CAM provides both quantitative predictions and visual explanations.The main findings of the study are:

- CNN-based deep feature extraction can

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 08 August 2026
Page No: 5307-5314
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0694

effectively learn discriminative patterns from dermoscopic images.

- Data augmentation can improve the robustness and generalization of the model.
- Sensitivity, specificity, precision, F1-score, and ROC-AUC provide a more informative evaluation than accuracy alone.
- Grad-CAM can identify image regions contributing to individual model predictions.
- The combination of classification and visual explanation provides a more transparent framework for computer-aided skin-lesion screening.
- Further evaluation on independent clinical datasets is required before considering deployment in real-world clinical practice.

Future Work

Future research can extend the framework through external validation on geographically and demographically diverse datasets, lesion segmentation before classification, comparison of Grad-CAM with alternative explanation methods, uncertainty estimation and selective prediction, and prospective evaluation with dermatologists. A multimodal model that combines dermoscopic images with clinical metadata may also be explored, provided privacy and fairness are addressed.

Conclusion

This paper presents an explainable deep learning framework for early skin cancer detection using dermoscopic images. The proposed system integrates quality-controlled preprocessing, class-imbalance handling, EfficientNet transfer learning, calibrated prediction, and Grad-CAM visualization. The framework emphasizes reproducible dataset splitting and clinically relevant evaluation metrics rather than relying solely on accuracy. Because no experimental run was supplied, numerical performance claims have deliberately not been fabricated. Once trained and evaluated, the framework can provide both a lesion-class prediction and a visual indication of the image regions that contributed to that prediction. Such a system may serve as a research prototype for computer-aided dermatology, while clinical deployment would require independent validation, expert oversight, safety assessment, and regulatory review.

Acknowledgements

The authors acknowledge Radhakrishna Institute of Technology and Engineering, Bhubaneswar, and Biju Patnaik University of Technology, Rourkela, for the academic environment supporting this postgraduate research work. The authors also acknowledge the publicly available HAM10000 dataset and the researchers who curated it for machine-learning research.

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, 115–118. https://doi.org/10.1038/nature21056.

[2]. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, 180161. https://doi.org/10.1038/sdata.2018.161

[3]. Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. (2018). Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). Proceedings of IEEE ISBI, 168–172.

[4]. Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., et al. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 29(8), 1836–1842. https://doi.org/10.1093/annonc/mdy166

[5]. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE

International Conference on Computer Vision (ICCV), 618–626.

[6]. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 97, 6105–6114.

[7]. Gessert, N., Nielsen, M., Shaikh, M., Werner, R., & Schlaefer, A. (2020). Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. IEEE Transactions on Biomedical Engineering, 68(10), 3083–3091. The work addresses ISIC 2019 classification using multi-resolution EfficientNets, metadata fusion, class-imbalance handling and model ensembles.

[8]. Marengo, D., et al. (2023). Skin lesion classification and detection using machine learning techniques: A systematic review. Diagnostics, 13(19), 3147. https://doi.org/10.3390/diagnostics13193147. This is particularly useful for your literature review and research-gap section.

[9]. Rajan, P., Devi, A., B, A., Dusthackeer, A., & Iyer, P. (2023). A Green perspective on the ability of nanomedicine to inhibit tuberculosis and lung cancer. International Research Journal on Advanced Science Hub, 5(11), 389-396. doi: 10.47392/IRJASH.2023.071.

[10]. Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., et al. (2018). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer, 113, 47–54.

[11]. schandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., et al. (2020). Human–computer collaboration for skin cancer recognition. Nature Medicine, 26, 1229–1234. https://doi.org/10.1038/s41591-020-0942-0. This is especially valuable for discussing AI-assisted clinical decision support rather than AI replacing dermatologists.

[12]. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., et al. (2019). Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). Proceedings of the ISIC/MICCAI Challenge. The challenge provides important benchmark information for segmentation, lesion attributes and classification.

[13]. Kawahara, J., Daneshvar, S., Argenziano, G., & Hamarneh, G. (2019). Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE Journal of Biomedical and Health Informatics, 23(2), 538–546.

[14]. Zhang, J., Saha, A., Zhu, S., et al. (2018). Skin lesion segmentation from dermoscopic images using deep learning approaches. IEEE/ISBI-related dermoscopic image analysis research. This line of work is relevant to your framework because accurate lesion segmentation can improve the region of interest used for subsequent classification.

[15]. Nugroho, A. K., Wardoyo, R., Wibowo, M. E., & Soebono, H. (2024). Image dermoscopy skin lesion classification using deep learning method: Systematic literature review. Bulletin of Electrical Engineering and Informatics, 13(2), 1042–1049. https://doi.org/10.11591/eei.v13i2.6077