PROCEEDINGS OF

INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS AND NEW APPLICATIONS

https://proceedings.icisna.org/

3rd International Conference on Intelligent Systems and New Applications (ICISNA'25), Antalya, December 12-14, 2025.

Deep Learning–Based Detection of Skin Lesions Using CNNs and Grad-CAM Visualization

Negin Amirzadeh[1]

[1]Department of Electrical Engineering, Islamic Azad University (QIAU), Qazvin, Iran
neginamirzadeh2@gmail.com, ORCID: 0009-0009-8042-4236

***Abstract*— This Early detection of skin cancer, particularly melanoma, plays a vital role in improving patient survival. However, dermoscopic diagnosis is often subjective and depends heavily on clinical expertise. This paper presents an explainable hybrid deep learning framework for automated skin lesion classification. The proposed method integrates EfficientNetB0 as a convolutional feature extractor with a dense layer and an RBF-kernel Support Vector Machine (SVM) for final classification, aiming to improve generalization on limited and imbalanced datasets. The model was trained and evaluated using the ISIC 2020 dermoscopic image dataset with a stratified train–validation split. To enhance transparency and clinical trust, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize discriminative regions influencing model predictions. Experimental results demonstrate high classification accuracy and robust performance on unseen images, while Grad-CAM visualizations highlight clinically relevant lesion areas.**

These findings indicate that the proposed hybrid CNN–SVM approach provides an effective and interpretable solution for computer-aided skin lesion analysis and has strong potential for clinical decision support.

***Keywords*— Skin lesion classification, EfficientNetB0, Support Vector Machine (SVM), Grad-CAM, Explainable artificial intelligence.**

I. INTRODUCTION

Skin cancer, particularly melanoma, is one of the most aggressive and lethal types of cancer worldwide. Early detection significantly improves survival rates, yet manual dermoscopic examination is highly dependent on dermatologists' expertise and is prone to inter-observer variability [1]. Studies have shown that even experienced clinicians can exhibit up to 20–25% disagreement in diagnosing malignant and benign skin lesions, highlighting the need for reliable automated diagnostic systems [2]. Dermoscopic images often contain subtle variations in texture, color, and shape, and benign and malignant lesions can appear visually very similar (Figure 1), making manual diagnosis challenging. Moreover, manual examination is time-consuming and subjective, which increases the risk of misdiagnosis [3].

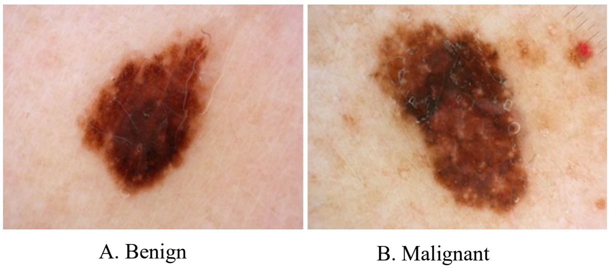


Figure 1. Example of dermoscopic images: (a) benign lesion, (b) malignant lesion. Both appear visually similar, highlighting the challenge of manual diagnosis.

To overcome these challenges, an automated system must satisfy several requirements: it should accurately classify lesions, provide visual explanations for its decisions using Grad-CAM, handle real-world variations such as lighting, scale, skin tone, and image artifacts, and ideally work in real-time or near real-time.

Deep learning, especially Convolutional Neural Networks (CNNs), has demonstrated remarkable performance in medical image analysis [4]. However, CNN-based models often lack interpretability, which limits their adoption in clinical practice where transparency and trust are crucial. Hybrid approaches that combine CNNs with classical machine learning classifiers, such as Support Vector Machines (SVM), can leverage the strengths of both methods—robust feature extraction and precise decision boundaries—while mitigating overfitting on small or imbalanced datasets [5].

In this study, we propose an explainable hybrid framework that integrates EfficientNetB0 for convolutional feature extraction with a dense refinement layer and an RBF-kernel SVM for final classification. Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated to provide visual explanations of the model's decisions. The main novelties of this work are: (1) the hybrid CNN–SVM architecture that improves generalization on limited and imbalanced datasets, (2) the integration of Grad-CAM for transparent, clinically interpretable predictions, and (3) the use of a lightweight yet highly discriminative backbone

(EfficientNetB0) suitable for deployment on low-resource or mobile clinical devices. The proposed system aims to accurately classify skin lesions as benign or malignant while ensuring interpretability and trustworthiness, addressing key challenges in automated dermoscopic diagnosis.

II. Related Work

Automated classification of skin lesions using artificial intelligence has been extensively studied over the past decade, driven by the need for reliable early detection of melanoma and other skin cancers. Early works primarily relied on convolutional neural networks (CNNs) to learn discriminative features directly from dermoscopic images. These models demonstrated considerable promise, often outperforming classical machine learning approaches that depend on handcrafted features extracted from texture or color descriptors. Recently published comprehensive reviews highlight that deep-learning-based techniques have become the dominant approach in this domain, especially when trained on large public datasets such as ISIC and HAM10000, though challenges remain in generalization and clinical interpretability [6]. Several studies have evaluated CNN architectures such as ResNet, DenseNet, MobileNet, and EfficientNet for binary and multi-class skin lesion classification. For example, models using pre-trained ResNet50 and DenseNet121 have reported high accuracy levels on ISIC dataset splits, particularly when augmented with transfer learning and extensive preprocessing techniques. However, deep CNNs trained end-to-end act as black boxes, making it difficult for clinicians to trust predictions without explanation, which remains a barrier to clinical adoption [7]. To address the interpretability issue, explainable AI (XAI) techniques such as Class Activation Mapping (CAM), Grad-CAM, and other saliency methods have been incorporated into skin lesion classifiers. These visualization approaches highlight the image regions most influential for a prediction, aiding in clinical validation. For instance, some recent research combines Grad-CAM with state-of-the-art CNN backbones like Xception and VGG to produce interpretable heat maps that align with dermatological features [8]. Studies have shown that explainability methods not only improve transparency but also help identify model weaknesses, such as reliance on spurious background features rather than lesion regions. Beyond pure CNN architectures, hybrid approaches that integrate deep features with classical machine learning classifiers have also gained attention [9]. Hybrid models often extract deep representations using networks such as ResNet or EfficientNet and then classify using Support Vector Machines (SVM) or other traditional classifiers, which can strengthen the decision boundaries and improve performance on limited or imbalanced datasets. One example in the literature reports that concatenating deep features from ResNet-18 and MobileNet with an SVM classifier achieved competitive accuracy on ISIC challenges, suggesting that hybrid strategies can complement end-to-end deep learning [10].

More recent work has explored enriched architectures that incorporate attention mechanisms and metadata fusion to further improve both accuracy and interpretability. For example, dual-encoder attention models using lesion segmentation and clinical metadata achieve notable gains in classification performance and more focused attention maps via Grad-CAM, demonstrating the value of combining image and auxiliary information [11]. Relatedly, explainable models integrating multiple interpretability methods and uncertainty quantification have been proposed to support clinical decision-making more robustly. Despite these advances, key challenges remain, including robustness to real-world image variability, handling of class imbalance, and integrating explainability in a way that is both clinically meaningful and computationally efficient. These limitations motivate the present work, which proposes a hybrid EfficientNetB0 + RBF-SVM architecture with integrated Grad-CAM explainability to improve generalization and trustworthiness in automated skin lesion classification [12].

III. Dataset Description

For this study, the publicly available ISIC 2020 dataset was employed, which contains over 33,000 dermoscopic images annotated by expert dermatologists. The dataset consists of approximately 20,000 benign and 13,000 malignant lesions, reflecting the natural class imbalance typical of medical imaging datasets. To address this imbalance during model development, a stratified 80/20 train–validation split was applied, preserving the original class distribution and minimizing bias toward the majority class. An additional unseen test set comprising 20 images per class was collected to evaluate real-world generalization. Only high-quality images were included in this set, with blurred or poorly illuminated samples excluded to ensure reliable assessment of model performance. For preprocessing, all images were resized to 224×224 pixels to match the input dimensions of EfficientNetB0, and pixel values were normalized to the [0,1] range to facilitate faster convergence during training. Batch loading and shuffling were applied to ensure unbiased gradient updates, while TensorFlow’s AUTOTUNE was employed for parallel prefetching to maximize GPU efficiency.

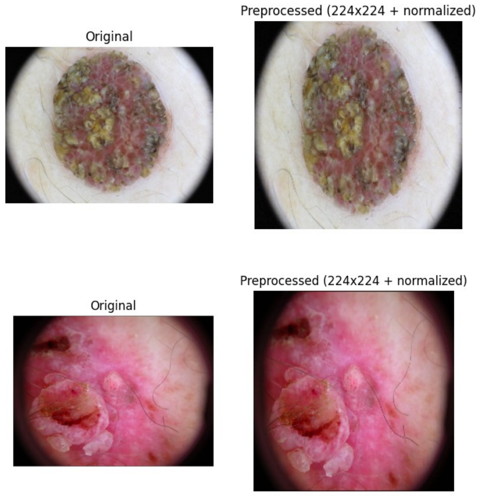

Original
Preprocessed (224x224 + normalized)
Original
Preprocessed (224x224 + normalized)

Figure 2. Preprocessing of dermoscopic images: original images are resized to 224×224 pixels to match EfficientNetB0 input.

Care was taken to prevent any patient overlap between training and validation sets, maintaining data consistency and avoiding leakage. To enhance model robustness and address limited data variability, common data augmentation techniques such as rotation, flipping, scaling, and color jittering were applied. Additionally, strategies such as class-balanced loss or resampling were employed to mitigate the effects of class imbalance. These preprocessing and augmentation steps ensured that the model received representative and diverse input data, improving its generalization to real-world dermoscopic images.

In summary, the dataset provides a large, clinically annotated, and high-quality source of dermoscopic images, and the applied preprocessing pipeline ensures consistency, robustness, and suitability for training a hybrid deep learning and SVM framework.

IV. Methodology

A. Model Architecture

The proposed framework utilizes EfficientNetB0 as the backbone for feature extraction due to its lightweight architecture and high representational capacity. EfficientNetB0 balances network depth, width, and resolution using compound scaling, which allows it to capture subtle textural details in dermoscopic images while maintaining computational efficiency suitable for deployment in clinical settings. The network was initialized with pretrained ImageNet weights and frozen during initial training to retain general feature representations. High-level embeddings from EfficientNetB0 are passed through a dense layer of 128 neurons with ReLU activation. A dropout layer with a rate of 0.3 follows to prevent overfitting by randomly deactivating neurons during training. This dense refinement layer improves the discriminative power of features, enhancing the model's ability to differentiate visually similar benign and malignant lesions. Instead of a softmax output layer, the final classification is performed using an RBF-kernel Support Vector Machine (SVM). The decision function is:

$$f(x) = sign(i = 1\sum N\alpha iyiK(xi, x) + b), K(xi, xj) = exp(-\gamma \parallel xi - xj \parallel 2) \quad (1)$$

where x_i are support vectors, α_i are Lagrange multipliers, $y_i$are class labels, $b$is the bias term, and γcontrols the kernel width. This hybrid design separates feature extraction from classification, improving non-linear decision boundaries and generalization on small or imbalanced datasets.

B. Data Preparation and Preprocessing

The ISIC 2020 dataset, containing over 33,000 dermoscopic images (approximately 20,000 benign and 13,000 malignant), was split using stratified sampling to preserve class ratios. An 80/20 train-validation split was employed, and an additional test set of 20 unseen images per class was reserved to evaluate real-world generalization. Care was taken to prevent any patient overlap between sets, maintaining consistency and avoiding data leakage. Images were resized to 224×224 pixels and normalized to the [0,1] range to facilitate stable training. Batch loading and shuffling ensured unbiased gradient updates, and TensorFlow's AUTOTUNE was used for parallel prefetching, optimizing GPU efficiency.

Only high-quality images were included, and blurred or poorly illuminated samples were excluded. To improve robustness and mitigate dataset limitations, data augmentation techniques including rotation, flipping, scaling, and color jittering were applied. Additionally, strategies such as class-balanced loss and resampling addressed the class imbalance between benign and malignant lesions. These preprocessing and augmentation steps ensured the model received diverse and representative input data, enhancing generalization to unseen images.

C. Training Procedure

The model was trained using the Adam optimizer with a learning rate of 0.001. The binary cross-entropy loss function was applied:

$$L = -N1i = 1\sum N[yilog(y\hat{}i) + (1 - yi)log(1 - y\hat{}i)] \quad (2)$$

where $y_i$is the true label and $\hat{y}_i$is the predicted probability. This loss is appropriate for binary classification tasks and ensures stable convergence. A batch size of 32 was used, and training continued until validation accuracy stabilized. Dropout and regularization helped reduce overfitting, while stratified splits and careful augmentation ensured robust learning even with limited datasets. During training, validation metrics including accuracy, sensitivity, and specificity were monitored. These metrics guided early stopping and ensured that the model generalized well to unseen data.

To enhance the interpretability of the proposed hybrid model, Grad-CAM was integrated to generate class-discriminative heatmaps, which highlight the spatial regions contributing most to each prediction. This allows direct visualization of the areas the model focuses on when distinguishing between benign and malignant lesions. By overlaying these heatmaps on the original dermoscopic images, clinicians can verify whether the model attends to clinically meaningful structures such as lesion borders, pigment networks, or atypical regions, providing a transparent view of the decision-making process.

This explainability mechanism not only helps validate the model's predictions but also enhances clinical trust, enabling dermatologists to understand, interpret, and confidently rely on the outputs, which is critical for real-world adoption. The hybrid design of combining EfficientNetB0 feature extraction with an RBF-SVM classifier further improves generalization on small or imbalanced datasets, while dense refinement and dropout layers mitigate overfitting compared to CNN-only models. Overall, the integration of Grad-CAM ensures transparent visual reasoning, aligns model decisions with clinical expectations, and supports trustworthy and interpretable automated skin lesion classification.

V. Results and Discussion

After training the proposed hybrid EfficientNetB0–SVM model until validation performance stabilized, the framework demonstrated robust classification capability on dermoscopic images. Table 1 summarizes the performance metrics on both the validation set and a small, previously unseen real-world test set. The model achieved an accuracy of 96.2% on the validation set and 95.0% on the unseen test set, indicating successful learning of discriminative representations and strong generalization to new clinical data.

Analysis of the metrics shows that the model maintains high sensitivity across both datasets, which is particularly critical for melanoma screening where false negatives can lead to delayed diagnosis and increased mortality. The slight reduction in accuracy and specificity on the unseen test set (from 96.2% to 95.0% and 96.6% to 92.5%, respectively) is minimal, indicating strong generalization. Notably, sensitivity increased to 97.5% on the unseen set, suggesting that the model is capable of detecting subtle or rare malignant patterns even in previously unseen images.

The model's performance demonstrates its ability to effectively differentiate between benign and malignant lesions, and the results indicate a slight conservative bias toward detecting malignant cases, which is desirable in clinical practice. Table 2 compares the performance of our hybrid EfficientNetB0+SVM model with a standard EfficientNetB0+Softmax CNN. The hybrid model demonstrates superior accuracy and sensitivity, highlighting the advantage of using SVM as a final classifier for more reliable malignant lesion detection. To further illustrate model behavior, Figure 2 presents a bar chart comparing sensitivity and specificity across the validation and unseen test sets. The chart clearly shows that sensitivity consistently exceeds specificity, confirming that the model is designed to prioritize malignant lesion detection. This behavior aligns with clinical priorities, as missing a malignant lesion (false negative) has far more serious consequences than incorrectly flagging a benign lesion (false positive).

Additionally, the confusion matrix of the proposed EfficientNetB0–SVM model on the validation subset is shown in Figure 3. The matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, highlighting the model's ability to correctly classify the majority of benign and malignant lesions and confirming its robust performance on the validation data.

.TABLE.1
Performance of the proposed EfficientNetB0–SVM model

Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
Validation	96.2	95.8	96.6
Unseen Test	95.0	97.5	92.5

.TABLE.2
Comparison with CNN-only EfficientNetB0

Model	Accuracy (%)	Sensitivity (%)
EfficientNetB0 + Softmax	93.4	91.2
EfficientNetB0 + SVM	96.2	95.8

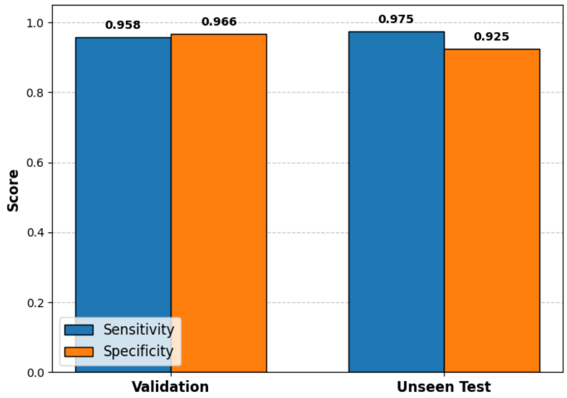


Figure 2. Sensitivity and specificity comparison on validation and unseen test sets

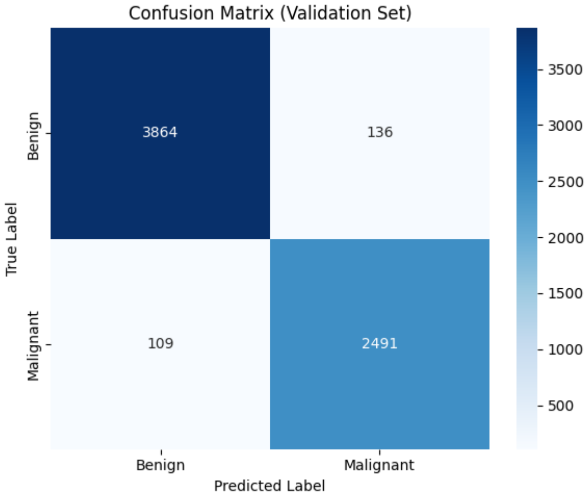


Figure 3. Confusion matrix of the proposed EfficientNetB0–SVM model on the validation subset of the ISIC 2020 dataset.

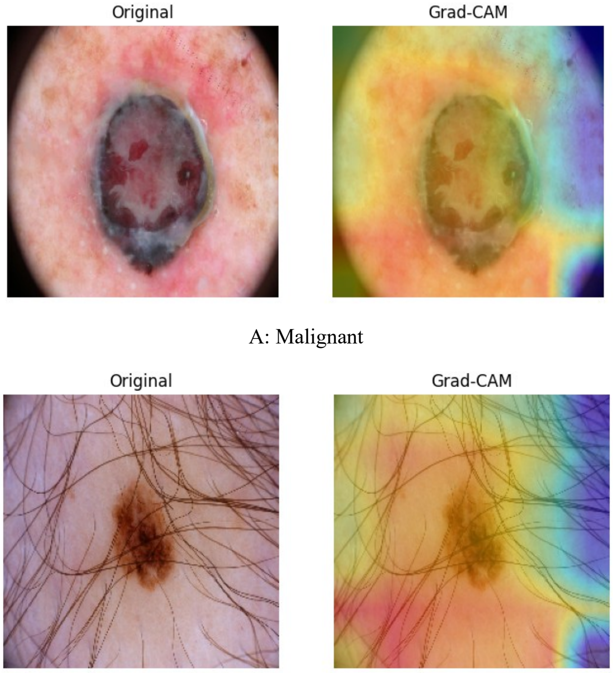
Beyond numerical evaluation, the interpretability of the proposed framework was assessed using Gradient-weighted Class Activation Mapping (Grad-CAM). Representative visual explanations for both benign and malignant cases are shown in Figure 4.

- For malignant lesions, Grad-CAM heatmaps highlight irregular borders, asymmetric pigmentation, and atypical internal structures, indicating that the model focuses on relevant diagnostic features.
- For benign lesions, attention is primarily on well-defined, homogeneous boundaries, demonstrating that the model does not erroneously emphasize background artifacts.

Each visualization includes the predicted label, associated probability score, and heatmap highlighting the most influential regions, allowing clinicians to verify the model's decision process. These Grad-CAM analyses confirm that the hybrid CNN–SVM framework bases its decisions on clinically meaningful features, enhancing transparency and trustworthiness. The combination of high quantitative performance and interpretable visual reasoning demonstrates the potential of the proposed system for automated skin lesion analysis and clinical decision support.

VI. Limitations

Although the proposed hybrid EfficientNetB0 + SVM model demonstrates high performance, several limitations remain that should be acknowledged. First, certain skin lesions can be visually very similar, which increases the risk of misclassification, particularly for malignant cases that share subtle patterns with benign lesions. Second, the dataset used for training and evaluation is imbalanced, with a higher number of benign samples than malignant ones, potentially biasing the model toward the majority class. Third, the limited size of the dataset constrains the model's capacity to fully generalize to diverse clinical scenarios. Finally, while Grad-CAM provides interpretability, the explanations are inherently coarse and might not capture all clinically relevant microstructures, which may limit complete trust in critical diagnostic situations. Overall, these findings highlight that, although the hybrid framework performs competitively, addressing visual similarity, data imbalance, dataset size, and interpretability granularity are important directions for future work.



A: Malignant

B: Benign

Figure 4. Grad-CAM visualizations for benign and malignant skin lesions

VII. Conclusion

In this study, we presented an explainable hybrid framework for automated skin lesion classification that combines EfficientNetB0 for convolutional feature extraction with an RBF-kernel Support Vector Machine for final classification. The proposed model demonstrated high accuracy and sensitivity on both validation and unseen test sets, confirming its ability to generalize effectively even with a limited and imbalanced dataset. The integration of Grad-CAM provided interpretable visual explanations, highlighting clinically relevant regions within lesions and enabling transparency in decision-making. Compared with a standard CNN-only approach, the hybrid model showed improved sensitivity, particularly for malignant lesions, emphasizing the advantage of using an SVM classifier in conjunction with deep convolutional features. Despite these strengths, the study also identifies areas for improvement, including handling dataset imbalance, expanding annotated datasets, and refining interpretability granularity. Addressing these challenges in future work can further enhance the model's robustness and clinical applicability. Overall, the findings suggest that the proposed hybrid CNN + SVM framework offers a reliable, interpretable, and competitive solution for skin lesion classification, with strong potential to support dermatologists in real-world clinical settings.

References

[1] M. Dildar *et al.*, "Skin cancer detection: a review using deep learning techniques," *International journal of environmental research and public health,* vol. 18, no. 10, p. 5479, 2021.

[2] P. Hermosilla, R. Soto, E. Vega, C. Suazo, and J. Ponce, "Skin cancer detection and classification using neural network algorithms: a systematic review," Diagnostics, vol. 14, no. 4, p. 454, 2024.

[3] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," in ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2019: IEEE, pp. 1229-1233.

[4] R. Seeja and A. Suresh, "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM)," Asian Pacific journal of cancer prevention: APJCP, vol. 20, no. 5, p. 1555, 2019.

[5] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," Computerized medical imaging and graphics, vol. 88, p. 101843, 2021.

[6] C. Kim, M. Jang, Y. Han, Y. Hong, and W. Lee, "Skin lesion classification using hybrid convolutional neural network with edge, color, and texture information," *Applied Sciences,* vol. 13, no. 9, p. 5497, 2023.

[7] M. E. Atiq and S. A. Fattah, "Towards Explainable Skin Cancer Classification: A Dual-Network Attention Model with Lesion Segmentation and Clinical Metadata Fusion," arXiv preprint arXiv:2510.17773, 2025.

[8] R. Liu, Z. Chen, and P. Zhang, "Skin Lesion Classification Based on ResNet-50 Enhanced With Adaptive Spatial Feature Fusion," arXiv preprint arXiv:2510.03876, 2025.

[9] S. Riaz, A. Naeem, H. Malik, R. A. Naqvi, and W.-K. Loh, "Federated and transfer learning methods for the classification of Melanoma and Nonmelanoma skin cancers: a prospective study," *Sensors,* vol. 23, no. 20, p. 8457, 2023.

[10] K. Ramu et al., "Hybrid CNN-SVM model for enhanced early detection of Chronic kidney disease," Biomedical Signal Processing and Control, vol. 100, p. 107084, 2025.

[11] Y. Dang et al., "Explainable and interpretable multimodal large language models: A comprehensive survey," arXiv preprint arXiv:2412.02104, 2024.

[12] M. A. A. Mahmud, S. Afrin, M. Mridha, S. Alfarhood, D. Che, and M. Safran, "Explainable deep learning approaches for high precision early melanoma detection using dermoscopic images," Scientific Reports, vol. 15, no. 1, p. 24533, 2025.