Enhancing Early Detection of Skin Cancer in Clinical Practice with Hybrid Deep Learning Models

Azzedine El Mrabet

Laboratory of Advanced Systems Engineering, Ibn Tofail University, Kenitra, Morocco
azzedine.elmrabet@uit.ac.ma (corresponding author)

Mohamed Benaly

Faculty of Sciences, Laboratory of Electronic Systems, Information Processing, Mechanics and Energetics, Ibn Tofail University, Morocco
mohamed.benaly@uit.ac.ma

Imam Alihamidi

Laboratory of Advanced Systems Engineering, Ibn Tofail University, Kenitra, Morocco
imam.alihamidi@uit.ac.ma

Bouchra Kouach

Laboratory of Advanced Systems Engineering, Ibn Tofail University, Kenitra, Morocco
bouchra.kouach@uit.ac.ma

Laamari Hlou

Faculty of Sciences, Laboratory of Electronic Systems, Information Processing, Mechanics and Energetics, Ibn Tofail University, Kenitra, Morocco
laamari.hlou@uit.ac.ma

Rachid El Gouri

Laboratory of Advanced Systems Engineering, Ibn Tofail University, Kenitra, Morocco
rachid.elgouri@uit.ac.ma

ABSTRACT

Skin cancer is a significant global health issue where early detection is essential to improve outcomes. This study evaluates hybrid deep learning models that combine CNN architectures (MobileNetV2, ResNet-18, EfficientNet-B0, and others) with metadata (age, lesion localization) for classification using the SLICE-3D subset of the ISIC 2024 dataset. MobileNetV2 achieved a recall of 99.2% and an accuracy of 97.7%, while EfficientNet-B0 demonstrated a recall of 98.5% and an accuracy of 97.2%, making them ideal for telemedicine in resource-limited settings due to their low computational demands. ResNet-18 and DenseNet-121, with recalls of 99.0% and 98.7%, respectively, excelled in clinical applications but required greater computational resources. These hybrid models show great potential as accessible and accurate tools for improving skin cancer detection. Future work should validate these findings on diverse datasets and optimize preprocessing to further enhance sensitivity and early diagnostic accuracy.

Keywords-skin cancer detection; hybrid deep learning models; telemedicine; early cancer detection; ISIC 2024 dataset; machine learning in healthcare; melanoma detection; clinical applications of AI; medical images

I. INTRODUCTION

Skin cancer is a major global health problem, and early detection is essential to improve survival rates and reduce deaths. Melanoma, the deadliest form of skin cancer, is especially dangerous due to its fast ability to spread [1]. The early diagnosis of melanoma and other malignant skin lesions can significantly improve treatment outcomes but often requires access to specialized clinics and high-quality imaging tools, which are not widely available in rural or underserved areas [2]. Traditional methods to detect skin cancer rely on visual checks by dermatologists, followed by a biopsy to confirm the diagnosis [3]. Recent advances in Artificial Intelligence (AI), especially deep learning, have introduced new effective methods for detecting skin cancer by analyzing images of skin lesions [4].

Implementing AI-driven diagnostic tools outside specialist clinics presents unique challenges. Skin cancer screening based on images typically relies on high-quality standardized images taken with specialized equipment, which may not be available in primary care or telemedicine settings. In many cases, image quality varies significantly, reducing diagnostic accuracy. The rise of telemedicine, accelerated by the COVID-19 pandemic, has increased the need for reliable diagnostic tools that work well with diverse image formats, including cellphone photos [5]. AI systems capable of adapting to lower-quality images could improve access to skin cancer screening and early treatment, particularly for remote and underserved communities.

Deep learning models, particularly Convolutional Neural Networks (CNNs), are highly effective in medical image processing, achieving high sensitivity and specificity in tasks such as skin lesion classification [6, 7]. CNN architectures such as ResNet, MobileNet, and VGG extract detailed visual features efficiently but often overlook patient metadata (e.g., age, sex, lesion location) that can enhance diagnostic accuracy. Integrating metadata with image analysis in hybrid CNN models improves efficacy by utilizing both visual and contextual information [8]. Recent advances in deep learning offer significant potential for identifying skin cancer, addressing limitations of conventional methods such as skilled visual inspection, dermoscopic evaluation, and biopsy, which are resource-intensive and less accessible in underserved areas. CNNs provide automated and scalable solutions by analyzing skin lesion images [9]. These models excel at recognizing patterns, distinguishing malignant from benign lesions, and achieving diagnostic performance that often rivals or surpasses that of dermatologists. For instance, in [10], CNNs outperformed dermatologists in multiclass skin lesion classification. In [9], it was shown that a market-approved CNN matched dermatologist-level performance across various types of lesion, highlighting the scalability of CNNs in clinical applications.

Hybrid deep learning frameworks that combine CNN-based image analysis with patient metadata (e.g., age, lesion location, gender) have shown promise in improving diagnostic precision. By integrating visual and contextual data, these models achieve greater accuracy than image-only approaches [11]. For instance, in [8], incorporating metadata into CNNs significantly improved the interpretability and accuracy of melanoma diagnosis. In [12], the use of an Inception-ResNet-v2 model with metadata achieved a 5% improvement in accuracy, highlighting the clinical relevance of such methods, especially in telemedicine settings with limited clinical data.

This study aims to bridge specialist dermatologic care and general clinical use by evaluating hybrid CNN models with 3D Total Body Photography (3D-TBP) data. The ISIC 2024 dataset, which features high-resolution lesion images from 3D body captures, mimics the quality of smartphone photos common in telemedicine. Comprehensive metadata for each lesion allows hybrid models to combine visual and tabular data for more accurate predictions.

II. METHODOLOGY

A. Dataset Description

The SLICE-3D subset of the International Skin Imaging Collaboration (ISIC) 2024 dataset [13] was used, which contains approximately 500,000 images of skin lesions derived from 3D TBP. The dataset integrates high-resolution images with comprehensive metadata, allowing the development of hybrid diagnostic models that combine visual and tabular data. Each image is linked to a unique identifier (isic_id) and paired with demographic and anatomical details, such as age, sex, and lesion location. These features are supplemented with binary diagnostic labels, malignant (1) or benign (0), that serve as ground truth for model training and evaluation. The image data consists of 15×15 mm cropped regions stored in JPEG format, optimized in an HDF5 file for efficient loading during training. Metadata includes lesion characteristics (e.g., size, perimeter, color irregularity, contrast) and patient-specific attributes, providing critical contextual information to enhance diagnostic precision. To address the class imbalance, a common challenge in medical datasets due to the lower prevalence of malignant cases, random downsampling of benign samples and upsampling of malignant samples were applied to create a balanced training set. Additionally, a class-weighted loss function was used to assign higher penalties for misclassifying malignant lesions, ensuring that the model prioritizes these critical cases during training.

B. Data Preprocessing

Preprocessing was used to ensure compatibility with deep learning models and to improve robustness against variations in clinical settings. Images were resized to 128×128 pixels to reduce computational demands while maintaining sufficient resolution for feature extraction. Data augmentation techniques, including random horizontal and vertical flips, rotations, and brightness adjustments, were applied to enhance model generalization and performance across diverse imaging conditions. Metadata attributes, such as sex and lesion location, were encoded numerically using LabelEncoder, while continuous variables, such as age and lesion size, were normalized with StandardScaler to standardize their range and facilitate convergence during training. The processed metadata was then integrated with image features to create a hybrid dataset, combining both visual and contextual information for enhanced diagnostic accuracy.

C. Model Architecture

This study assessed six CNN architectures: MobileNet, ResNet, EfficientNet, VGG, DenseNet, and GoogLeNet. Each backbone possesses distinct architectural features that affect its performance and computing demands.

- MobileNetV2 [14] is a lightweight architecture with reduced parameters, optimized for efficient inference, rendering it suitable for resource-limited settings.

- ResNet-18 [15] uses residual connections to facilitate deep feature extraction and alleviate vanishing gradient problems.

- EfficientNetB0 [16] attains an equilibrium among model depth, width, and resolution, delivering robust performance with a reduced number of parameters.

- VGG-11 [17] is a profound, sequential design that facilitates intricate feature extraction, although with increased memory consumption.

- DenseNet-121 [18] incorporates dense connections that promote feature reutilization and optimize gradient flow, thus increasing performance on intricate classification problems.

- GoogLeNet [19] utilizes inception modules to capture multi-scale characteristics, enabling the analysis of patterns at various resolutions.

The model integrates image and tabular data in a hybrid architecture, as shown in Figure 1. The CNN backbone processes image inputs to extract feature vectors, while tabular data is processed through fully connected layers. The outputs from both branches are concatenated into a unified feature vector and passed through additional dense layers for binary classification. This architecture combines visual and contextual information to enhance diagnostic accuracy.

D. Training and Validation

1) Train-Validation Split

The dataset was partitioned between the training and validation sets using an 80:20 split, ensuring that both sets were stratified by the target label to preserve class balance. Stratification is crucial to address the intrinsic class imbalance in skin cancer datasets, where benign cases generally exceed malignant ones.

2) Training Parameters and Optimization

- Loss function: Binary cross-entropy loss is used as the objective function, suitable for binary classification tasks.

- Optimizer: The Adam optimizer is employed with a learning rate of 1e-4 and weight decay to prevent overfitting. A ReduceLROnPlateau scheduler further reduces the learning rate when validation performance plateaus, aiding in stable convergence.

- Early stopping: An early stopping mechanism with patience of 50 epochs is used to halt training if validation loss does not improve, preserving computational resources and avoiding overfitting.

- Model checkpoints: Checkpoints save the best-performing model based on validation ROC-AUC scores, ensuring the optimal model is used for testing.

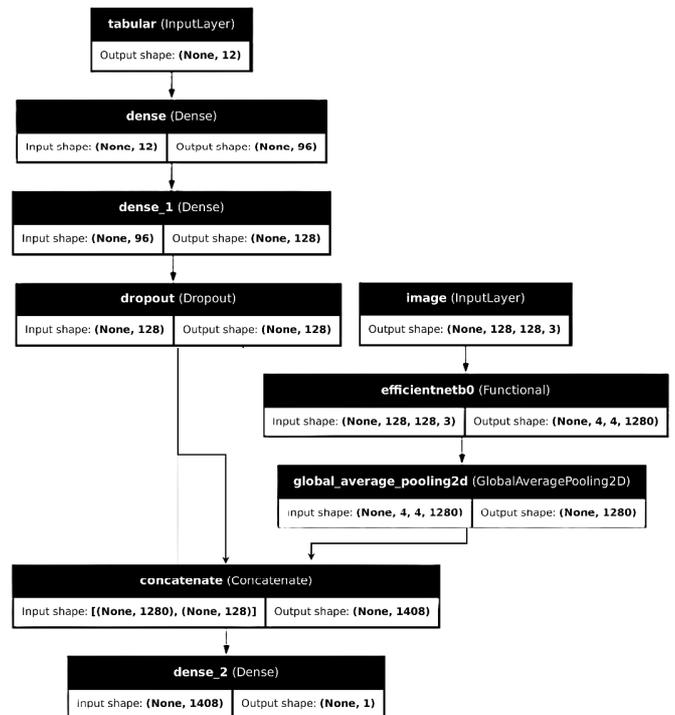Training was performed on a Tesla P100-PCIE-16GB GPU, supplied by Kaggle.

Fig. 1.   Hybrid model architecture with a network for image features and metadata.

3) Evaluation Metrics

The following metrics were employed to comprehensively evaluate model performance.

- Accuracy: Measures the proportion of correctly classified samples out of the total samples.

- Precision and Recall: Precision indicates the fraction of true positive predictions among all positive predictions, while recall captures the fraction of actual positives correctly identified. These metrics are crucial for assessing the model's performance in distinguishing malignant lesions.

- F1-score: The harmonic means of precision and recall, offering a balanced assessment between them.

- ROC-AUC (Receiver Operating Characteristic Area Under Curve): Indicates the model's capacity to differentiate between classes at various thresholds, with elevated values signifying superior performance. ROC-AUC is important for model selection due to its significance in medical diagnosis.

- Confusion Matrix: Confusion matrices were used for each model to evaluate true positives, false positives, false negatives, and true negatives, offering insights into

diagnostic inaccuracies and emphasizing each model's proficiency in accurately classifying malignant lesions.

Alongside accuracy and classification metrics, computational efficiency was assessed using memory consumption, parameter quantity, and training duration. These parameters are especially pertinent for assessing the viability of implementing each CNN backbone in resource-constrained clinical environments.

III. RESULTS

1) Model Performance

The performance of each CNN backbone (MobileNet, GoogLeNet, ResNet, EfficientNet, VGG, and DenseNet) was evaluated on the validation set, and the results are shown in Table I.

TABLE I. PERFORMANCE METRICS OF CNN BACKBONES ON VALIDATION SET

Backbone	Accuracy	Precision	Recall	F1 score	ROC AUC
MobileNet	0.977	0.942	0.992	0.967	0.995
ResNet	0.962	0.903	0.990	0.944	0.996
GoogLeNet	0.95	0.87	0.99	0.94	0.99
EfficientNet	0.972	0.933	0.985	0.958	0.994
VGG	0.976	0.940	0.990	0.964	0.993
DenseNet	0.973	0.935	0.987	0.960	0.995

MobileNet demonstrated exceptional performance in identifying cancerous lesions, achieving the highest accuracy of 0.977 and an almost perfect recall of 0.992, underscoring its remarkable sensitivity. Its F1 score of 0.967 further highlights its robust overall performance in both benign and malignant cases. ResNet excelled with the highest ROC-AUC of 0.996, indicating superior classification efficacy, and maintained a high recall of 0.990, although its accuracy and precision were slightly lower than MobileNet. Meanwhile, VGG and DenseNet also exhibited strong performance, with VGG achieving a high precision of 0.940 and DenseNet striking an effective balance with a recall of 0.987 and an F1 score of 0.960.

B. Training and Validation Metrics Across CNN Backbones

Figure 2 illustrates the training and validation loss, as well as the AUC metrics, for various CNN backbones over 35 epochs. The plots for train and validation losses show that MobileNet and EfficientNet converge faster than the other backbones, reaching their minimum validation loss within the first 10 epochs. This indicates efficient learning, suggesting that these models perform well on the dataset with fewer training steps. In contrast, VGG and GoogLeNet show more pronounced fluctuations in validation loss, pointing to potential instability in the training process or increased sensitivity to the learning rate. The train and validation AUC and plots reveal a sharp initial gain in AUC for all models during the first few epochs, followed by slower convergence. MobileNet and DenseNet-121 achieved near-optimal AUC values quickly, demonstrating strong learning capabilities. MobileNet, in particular, achieved fast convergence and high stability during training, making it especially suitable for telemedicine applications that demand both efficiency and reliability.
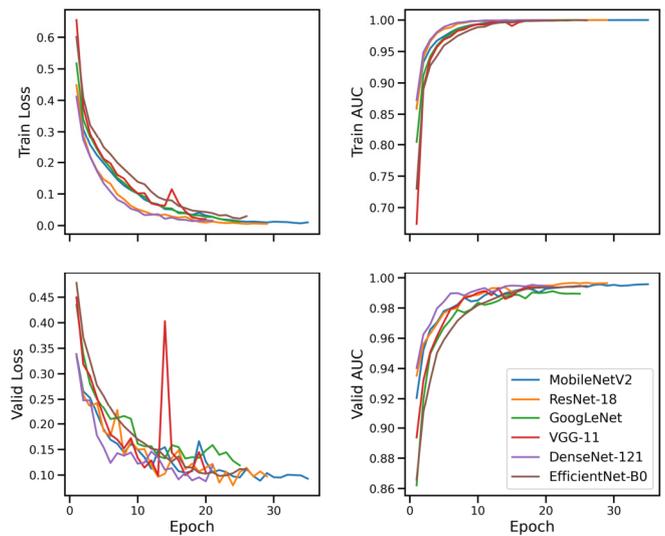

Fig. 2. Training and validation metrics across different backbones.

C. Training Efficiency and Computational Resource Usage

Training time, memory usage, and parameter count are critical factors for evaluating the deployability of the model in clinical settings, particularly those with limited computational resources. MobileNet required the least memory (8.54 MB) and completed training in 287 seconds, making it ideal for resource-constrained environments. ResNet offered a balanced approach with moderate resource usage and high diagnostic precision. VGG, despite its accuracy, consumed the most memory (427.26 MB), making it more suitable for resource-rich settings. DenseNet struck an effective balance between performance and efficiency, requiring relatively low resources and delivering strong results. The comparisons in Table II highlight MobileNet and EfficientNet as highly viable options for real-time applications on limited hardware.

TABLE II. TRAINING EFFICIENCY AND RESOURCE USAGE

Backbone	Training time (s)	Total memory (MB)	Trainable parameters
MobileNet	287	8.54	2,238,561
ResNet	231	42.69	11,190,433
GoogLeNet	210	21.42	5614337
EfficientNet	237	15.34	4,022,237
VGG	235	427.26	112,002,529
DenseNet	258	26.58	6,968,289

D. Confusion Matrices

Figure 3 highlights the strengths and weaknesses of each CNN backbone in distinguishing benign from malignant lesions. MobileNetV2 demonstrated high recall with minimal false negatives, making it ideal for early cancer detection in telemedicine, although moderate false positives in low-contrast lesions suggest preprocessing improvements. ResNet-18 and DenseNet-121 showed strong sensitivity with few false negatives but moderate false positives in benign cases with atypical textures. GoogLeNet, although highly sensitive, produced more false positives due to overestimating malignancy. VGG-11 balanced false positive and negative rates

but required significant memory, limiting its use in resource-constrained settings. EfficientNet-B0 offered a balanced performance, with low false negatives and controlled false positives, making it well-suited for real-time, resource-limited applications. MobileNetV2 and EfficientNet-B0 stand out for telemedicine, while ResNet-18 and DenseNet-121 are more suited for clinical environments.
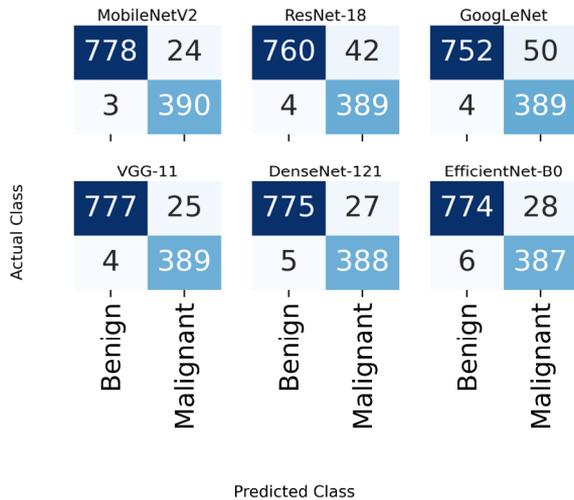

Fig. 3.   Confusion matrices for different backbones on the validation set.

E. Model Predictions

Figure 4 illustrates prediction probabilities from different CNN backbones for selected test images. GoogLeNet displayed higher confidence, assigning a probability of 0.9775 to ISIC_0015657, compared to near-zero predictions from MobileNet, ResNet, VGG, DenseNet, and EfficientNet. This overconfidence increases the risk of false positives in ambiguous cases. In contrast, ResNet and DenseNet adopted a more cautious approach, with lower probabilities under uncertain conditions, while EfficientNet showed moderate sensitivity, as seen in ISIC_0015729. These results highlight the trade-off between GoogLeNet's assertiveness, which may lead to false positives, and the conservative yet reliable predictions of other models, such as ResNet and DenseNet.

IV.   DISCUSSION

A. Key Findings and Interpretation

The hybrid model architecture demonstrated strong diagnostic accuracy and computational efficiency, particularly with MobileNetV2 and EfficientNet-B0, making them ideal for telemedicine applications. MobileNetV2's low false-negative rate highlights its suitability for early cancer detection in resource-limited settings, while EfficientNet-B0 balances recall and precision, reducing unnecessary follow-ups. In clinical environments where computational resources are less restricted, ResNet-18 and DenseNet-121 excel due to their high sensitivity and feature extraction capabilities. Conversely, GoogLeNet's tendency to produce false positives highlights its potential for minimizing missed malignancies but also underscores the need for cautious deployment to avoid excessive follow-ups.
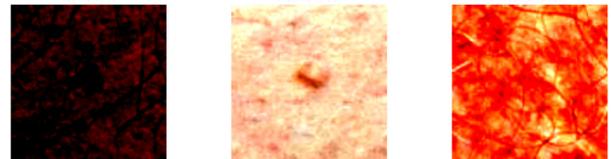

Fig. 4.   Predictions from CNN backbones on test images.

B. Comparison with Existing Literature

Recent studies confirm that combining image data with metadata improves skin cancer classification accuracy. A systematic review in [20] highlighted that integrating patient data such as age, sex, and lesion location into CNN-based skin cancer classifiers consistently improved diagnostic accuracy across various studies. In [21], 92.34% balanced accuracy was achieved with a CNN+ANN hybrid model compared to 73.69% for image-only CNNs, using the ISIC 2019 dataset, a benchmark dataset with 25,331 dermoscopic images labeled for multiple skin lesion types and metadata. In [22], skin lesion classification achieved 95.2% accuracy using GAN-based image augmentation applied to the ISIC 2018 dataset, which contains 10,015 dermoscopic images annotated for diagnostic tasks. In [23], a hybrid AI framework was proposed, leveraging diverse datasets, including Derm7pt, a dermatologist-annotated dataset with metadata, achieving 93.04% accuracy, 92.0% recall, and 93.0% precision. In [24], the ISIC 2019 and ISIC 2020 datasets were employed, achieving accuracies of up to 96% with a hybrid VGG19+SVM model and demonstrating the effectiveness of generative AI for data augmentation. This study used the ISIC 2024 dataset, specifically the SLICE-3D subset with 500,000 high-resolution images and comprehensive metadata (e.g., patient age, sex, lesion localization). MobileNetV2 achieved 99.2% recall and 97.7% accuracy, while EfficientNet-B0 achieved 98.5% recall and 97.2% accuracy, demonstrating their suitability for telemedicine applications. ResNet-18 and DenseNet-121 achieved 99.0% and 98.7% recall, respectively, with strong performance in clinical settings. These findings further validate hybrid models as accessible and accurate tools to improve skin cancer detection in various contexts (Table III).

TABLE III.   COMPARISON WITH EXISTING STUDIES

Study	Year	Backbone used	Dataset	Accuracy (%)
[21]	2021	CNN + ANN hybrid	ISIC 2019	92.34
[22]	2020	GAN-enhanced CNN	ISIC 2018	95.2
[23]	2024	Hybrid AI Framework	ISIC 2019	93.04
[24]	2023	VGG19+SVM Hybrid	ISIC 2019 & 2020	96
This study	2024	MobileNetV2, ResNet, GoogLeNet, EfficientNet, VGG, DenseNet	ISIC 2024	97.7 96.2 95 97.2 97.6 97.3

C. Clinical Implications

The high recall rates of MobileNetV2 and EfficientNet-B0, along with their computational efficiency, make them suitable for telemedicine, particularly in underserved areas where early detection can improve outcomes. These models could be integrated into mobile platforms for primary care or telehealth professionals to efficiently identify high-risk cases. For specialized clinical settings, ResNet-18 and DenseNet-121 provide greater accuracy and sensitivity, critical where diagnostic precision is paramount. The incorporation of metadata into hybrid models further enhances clinical decision-making by combining visual and contextual features for comprehensive analysis.

D. Limitations and Future Directions

The reliance on synthetic class balancing may introduce biases, potentially overfitting to malignant features. Future research should explore alternative balancing techniques and include more diverse datasets to enhance generalizability. Further studies should evaluate the impact of specific metadata, such as patient demographics or lesion history, on model performance. Additionally, external datasets from mobile devices and diverse clinical environments should be used to validate findings. Advanced techniques such as transfer learning and domain adaptation can further optimize model performance across varied imaging conditions, ensuring robust functionality in both clinical and telemedicine contexts.

V. CONCLUSION

This study highlights the potential of hybrid deep-learning models that integrate image data and metadata for early skin cancer detection. MobileNetV2 and EfficientNet-B0 were particularly effective for telemedicine applications, offering high recall, low false-negative rates, and computational efficiency, making them suitable for resource-constrained settings. These models provide affordable and reliable preliminary screening tools, crucial to improve outcomes in underserved populations. In clinical environments where accuracy is critical, ResNet-18 and DenseNet-121 performed exceptionally well, with low false-negative rates, making them strong alternatives where computational resources are available. Integration of metadata with image analysis improves model precision and allows for more comprehensive assessments by leveraging both visual and contextual features.

Despite these promising results, limitations such as the reliance on synthetic class balancing and controlled datasets need to be addressed. Future research should validate these models using diverse real-world datasets, particularly images from mobile devices in varied clinical environments. Optimizing preprocessing steps and incorporating more complex metadata could further improve model specificity and diagnostic utility.

In summary, hybrid deep learning frameworks bridge the gap between specialized dermatology and general telemedicine, offering reliable diagnostic tools that enhance early detection and reduce the global burden of skin cancer, especially in underserved areas.

REFERENCES

[1] *The Skin Cancer Foundation*. https://www.skincancer.org/.

[2] V. A. Rajendran and S. Shanmugam, "Automated Skin Cancer Detection and Classification using Cat Swarm Optimization with a Deep Learning Model," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12734–12739, Feb. 2024, https://doi.org/10.48084/etasr.6681.

[3] H. Vega-Huerta *et al.*, "Convolutional Neural Networks on Assembling Classification Models to Detect Melanoma Skin Cancer," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 14, pp. 59–76, Nov. 2022, https://doi.org/10.3991/ijoe.v18i14.34435.

[4] A. A. Adegun and S. Viriri, "Deep Learning-Based System for Automatic Melanoma Detection," *IEEE Access*, vol. 8, pp. 7160–7172, 2020, https://doi.org/10.1109/ACCESS.2019.2962812.

[5] K. Vodrahalli *et al.*, "Development and Clinical Evaluation of an Artificial Intelligence Support Tool for Improving Telemedicine Photo Quality," *JAMA Dermatology*, vol. 159, no. 5, May 2023, Art. no. 496, https://doi.org/10.1001/jamadermatol.2023.0091.

[6] A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani, and A. Almuhaimeed, "Skin Cancer Detection Using Combined Decision of Deep Learners," *IEEE Access*, vol. 10, pp. 118198–118212, 2022, https://doi.org/10.1109/ACCESS.2022.3220329.

[7] T. Imran, A. S. Alghamdi, and M. S. Alkatheiri, "Enhanced Skin Cancer Classification using Deep Learning and Nature-based Feature Optimization," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12702–12710, Feb. 2024, https://doi.org/10.48084/etasr.6604.

[8] S. Wang, Y. Yin, D. Wang, Y. Wang, and Y. Jin, "Interpretability-Based Multimodal Convolutional Neural Networks for Skin Lesion Diagnosis," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 12623–12637, Dec. 2022, https://doi.org/10.1109/TCYB.2021.3069920.

[9] R. C. Maron *et al.*, "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *European Journal of Cancer*, vol. 119, pp. 57–65, Sep. 2019, https://doi.org/10.1016/j.ejca.2019.06.013.

[10] H. A. Haenssle *et al.*, "Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions," *Annals of Oncology*, vol. 31, no. 1, pp. 137–143, Jan. 2020, https://doi.org/10.1016/j.annonc.2019.10.013.

[11] A. Kharbouche, Z. Madini, Y. Zouine, and N. El-Haryqy, "Signal demodulation with Deep Learning Methods for visible light communication," in *2023 9th International Conference on Optimization and Applications (ICOA)*, AbuDhabi, United Arab Emirates, Oct. 2023, pp. 1–5, https://doi.org/10.1109/ICOA58279.2023.10308822.

[12] R. A. Mehir and A. Ameri, "Skin Cancer Detection Based on Deep Learning," *Journal of Biomedical Physics and Engineering*, vol. 12, no. 6, Dec. 2022, https://doi.org/10.31661/jbpe.v0i0.2207-1517.

[13] N. R. Kurtansky *et al.*, "The SLICE-3D dataset: 400,000 skin lesion image crops extracted from 3D TBP for skin cancer detection," *Scientific Data*, vol. 11, no. 1, Aug. 2024, Art. no. 884, https://doi.org/10.1038/s41597-024-03743-w.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 4510–4520, https://doi.org/10.1109/CVPR.2018.00474.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

[16] B. Koonce, "EfficientNet," in *Convolutional Neural Networks with Swift for Tensorflow*, Berkeley, CA, USA: Apress, 2021, pp. 109–123.

[17] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, Apr. 10, 2015, https://doi.org/10.48550/arXiv.1409.1556.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.

[19] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594.

[20] J. Höhn *et al.*, "Integrating Patient Data Into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review," *Journal of Medical Internet Research*, vol. 23, no. 7, Jul. 2021, Art. no. e20708, https://doi.org/10.2196/20708.

[21] D. N. A. Ningrum *et al.*, "Deep Learning Classifier with Patient's Metadata of Dermoscopic Images in Malignant Melanoma Detection," *Journal of Multidisciplinary Healthcare*, vol. Volume 14, pp. 877–885, Apr. 2021, https://doi.org/10.2147/JMDH.S306284.

[22] Z. Qin, Z. Liu, P. Zhu, and Y. Xue, "A GAN-based image synthesis method for skin lesion classification," *Computer Methods and Programs in Biomedicine*, vol. 195, Oct. 2020, Art. no. 105568, https://doi.org/10.1016/j.cmpb.2020.105568.

[23] E. Farea, R. A. A. Saleh, H. AbuAlkebash, A. A. R. Farea, and M. A. Al-antari, "A hybrid deep learning skin cancer prediction framework," *Engineering Science and Technology, an International Journal*, vol. 57, Sep. 2024, Art. no. 101818, https://doi.org/10.1016/j.jestch.2024.101818.

[24] M. Saeed, A. Naseer, H. Masood, S. U. Rehman, and V. Gruhn, "The Power of Generative AI to Augment for Enhanced Skin Cancer Classification: A Deep Learning Approach," *IEEE Access*, vol. 11, pp. 130330–130344, 2023, https://doi.org/10.1109/ACCESS.2023.3332628.