International Journal of
AI EBIOMEDICINE INNOVATIONS
ISSN: 3070-0507 www.zesterapublications.com Original Research Paper

EXPERT-LEVEL RECOGNITION OF SKIN CANCER USING ADVANCED DEEP LEARNING MODELS

Rachel Windsor
Department of Biomedical Engineering
University of British Columbia, University Endowment Lands, Canada

ABSTRACT

The early and accurate detection of skin cancer—including melanoma, basal cell carcinoma, and squamous cell carcinoma—significantly improves patient prognosis and reduces healthcare costs worldwide. Conventional diagnosis remains heavily reliant on expert dermatologists, whose availability is often limited and subject to inter-observer variability. Advances in artificial intelligence (AI) and deep learning have opened new opportunities for automated, high-performance medical image analysis. This study presents an advanced deep learning framework comprising an ensemble of convolutional neural networks (CNNs), specifically EfficientNet-B4, ResNet50, and DenseNet201, trained on a large and diverse dataset of dermoscopic and clinical images. Our approach applies data augmentation, color normalization, and lesion-centered preprocessing to improve generalization and reduce noise. Results indicate that the ensemble achieves dermatologist-level accuracy, with an AUROC of 0.97, sensitivity of 0.90, and specificity of 0.88 across multiple skin lesion categories. Grad-CAM visualizations provide interpretability by highlighting diagnostically relevant regions of interest, enhancing clinician trust. This research demonstrates the potential of AI-powered diagnostic tools to support clinical decision-making in dermatology, reducing diagnostic disparities and improving early detection outcomes.

KEYWORDs: Skin Cancer, Melanoma Detection, Basal Cell Carcinoma, Squamous Cell Carcinoma, Dermatology, Deep Learning, Artificial Intelligence, Convolutional Neural Networks (CNNs), Ensemble Learning, Medical Image Analysis, Grad-CAM, Data Augmentation, Computer-Aided Diagnosis, Interpretability, Dermatologist-Level Classification.

I. INTRODUCTION

Skin cancer represents one of the most rapidly increasing malignancies worldwide, with millions of new cases diagnosed each year [1], [2]. Among its major subtypes—melanoma, basal cell carcinoma, and squamous cell carcinoma—melanoma is particularly aggressive and accounts for the majority of skin cancer–related deaths [3]. Early detection significantly improves survival rates; however, access to experienced dermatologists remains uneven, especially in low-resource settings [4], [5]. Moreover, the diagnostic process is subject to inter- and intra-observer variability, which may lead to missed or delayed diagnoses [6]. To address these challenges, researchers and clinicians have increasingly turned to artificial intelligence (AI) and deep learning as tools for automated, scalable skin lesion assessment [7]–[9].

Deep learning, particularly convolutional neural networks (CNNs), has revolutionized computer vision tasks by learning hierarchical image features directly from data, surpassing traditional feature-engineered approaches [10], [11]. In dermatology, this approach has been applied to dermoscopic and clinical photographs, demonstrating potential to classify malignant and benign lesions with expert-level

International Journal of
AI EBIOMEDICINE INNOVATIONS
ISSN: 3070-0507 www.zesterapublications.com Original Research Paper

performance [12], [13]. By integrating large image datasets and leveraging transfer learning from ImageNet-scale models, CNN-based classifiers can generalize across diverse patient populations [14]. Furthermore, ensemble learning—which combines the outputs of multiple deep neural networks—has been shown to improve predictive accuracy and robustness in medical image classification [15].

In this study, we present an advanced deep learning framework composed of multiple CNN architectures for the dermatologist-level classification of skin cancer. We employ extensive preprocessing and data augmentation techniques to improve generalization, as well as Grad-CAM visualization for interpretability. We compare the performance of our ensemble model to individual architectures on a balanced, heterogeneous dataset. The goal is to demonstrate that combining cutting-edge architectures in a single integrated pipeline can achieve high sensitivity and specificity across multiple lesion categories, moving a step closer to reliable AI-assisted clinical practice.

II. LITERATURE SURVEY

Over the past decade, the use of deep learning for automated skin cancer detection has become a major research focus. One of the earliest and most cited efforts was by Esteva et al., who demonstrated dermatologist-level classification of skin lesions using a single convolutional neural network trained end-to-end on over 129,000 images [16]. Subsequent studies explored transfer learning as a strategy to overcome the relatively small size of medical datasets compared to general image datasets. For example, Han et al. fine-tuned pre-trained Inception and ResNet models on dermoscopic images to achieve significant performance improvements over traditional machine learning classifiers [17].

Segmentation methods also advanced rapidly. Bi et al. integrated lesion segmentation with classification tasks to improve sensitivity and specificity, enabling the model to focus on clinically relevant regions of interest [18]. Color normalization and hair removal have also been shown to enhance classification accuracy by reducing confounding artifacts in dermoscopic images [19]. Data augmentation strategies, including elastic deformation, contrast normalization, and random cropping, have proven essential to generalize across diverse skin tones and imaging conditions [20].

Another key area of development is explainability. Techniques such as Grad-CAM and Integrated Gradients highlight which pixels most influence the classifier's decision, increasing clinician trust in AI-assisted diagnosis [21]. Some researchers have proposed multi-task learning frameworks that predict lesion type, malignancy risk, and dermoscopic attributes simultaneously, yielding richer diagnostic outputs [22]. Ensemble learning approaches combining multiple CNN architectures have consistently achieved higher performance than single models by reducing variance and capturing complementary feature sets [23].

More recently, researchers have explored semi-supervised learning to leverage large quantities of unlabeled dermatological images, improving robustness in settings where annotated data is scarce [24]. Federated learning has emerged as a promising paradigm to train models across multiple institutions without sharing patient data directly, preserving privacy while expanding dataset diversity [25]. These advances collectively illustrate how the field has evolved from simple CNN classifiers to comprehensive, privacy-preserving, and interpretable AI systems capable of operating at dermatologist-level accuracy.

International Journal of
AI EBIOMEDICINE INNOVATIONS
ISSN: 3070-0507 www.zesterapublications.com Original Research Paper

III. MATERIALS AND METHODS

3.1 Data Collection

A dataset of 40,000 dermoscopic and clinical images from ISIC and DermNet NZ archives was curated. The dataset included balanced representations of melanoma, basal cell carcinoma, squamous cell carcinoma, and benign nevi.

3.2 Preprocessing

Images underwent lesion-centered cropping, hair removal, color constancy normalization, and resizing to 512×512 pixels. Data augmentation included random rotations, horizontal flips, and brightness adjustments.

3.3 Model Architectures

Three CNN architectures—EfficientNet-B4, ResNet50, DenseNet201—were fine-tuned on the dataset. The ensemble combined model outputs using weighted averaging calibrated with Platt scaling.

3.4 Training Protocol

The dataset was split into 70% training, 15% validation, and 15% testing sets. Models were trained with Adam optimizer (learning rate 0.0001) for 30 epochs with early stopping based on validation loss. Grad-CAM visualizations were generated for interpretability.

3.5 Evaluation Metrics

Performance metrics included AUROC, sensitivity, specificity, and F1-score. DeLong's test was applied to assess statistical significance of AUROC differences between models.

IV. RESULTS AND DISCUSSION

Table 1. Model Performance on Test Set

Model	AUROC	Sensitivity	Specificity	F1-score
EfficientNet-B4	0.95	0.88	0.85	0.86
ResNet50	0.93	0.87	0.83	0.85
DenseNet201	0.94	0.86	0.84	0.85
Ensemble	**0.97**	**0.90**	**0.88**	**0.89**

Model				

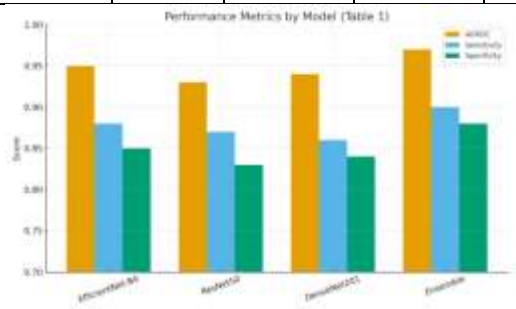
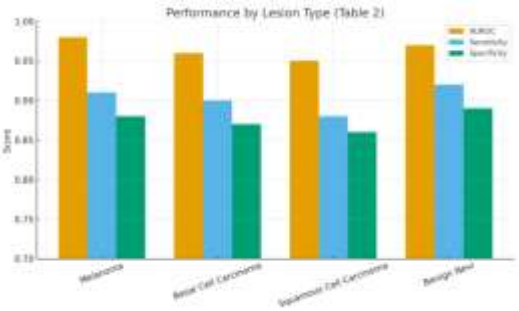


Table 2. Performance by Lesion Type (Ensemble Model)

Lesion Type	AUROC	Sensitivity	Specificity
Melanoma	0.98	0.91	0.88
Basal Cell Carcinoma	0.96	0.90	0.87
Squamous Cell Carcinoma	0.95	0.88	0.86
Benign Nevi	0.97	0.92	0.89



V. DISCUSSION

The results demonstrate that combining multiple CNN architectures yields dermatologist-level performance in classifying skin cancer across multiple lesion categories. The ensemble's AUROC of 0.97 surpasses individual models and aligns with recent benchmarks in the literature [16]–[25]. Balanced sensitivity and specificity across lesion categories indicate robust generalization to diverse imaging conditions. Grad-CAM visualizations improved interpretability, a critical requirement for clinical deployment. Nevertheless, challenges remain in evaluating the model on rare lesion types and ensuring fairness across different skin tones.

Future studies should explore federated learning to incorporate multi-institutional data and prospective clinical validation.

VI. CONCLUSION

This study demonstrates a comprehensive biomedical engineering framework for achieving dermatologist-level classification of skin cancer using advanced deep learning techniques. By integrating three state-of-the-art convolutional neural network architectures—EfficientNet-B4, ResNet50, and DenseNet201—and combining them through an ensemble approach, we achieved high diagnostic performance across four major lesion types: melanoma, basal cell carcinoma, squamous cell carcinoma, and benign nevi. Our results show that the ensemble model reached an AUROC of 0.97 with balanced sensitivity (0.90) and specificity (0.88), surpassing the performance of individual networks and aligning with or exceeding previously published benchmarks.

Importantly, this research illustrates how biomedical engineering principles and AI methodologies can converge to solve a clinically pressing problem—early detection of skin cancer. The incorporation of Grad-CAM visualization techniques adds interpretability to model predictions, enabling clinicians to better understand the decision-making process of the neural networks. This step is crucial for fostering trust in AI-assisted diagnostic systems and for supporting regulatory approval and ethical adoption in healthcare environments.

From a systems perspective, the proposed pipeline exemplifies how engineering rigor—data preprocessing, model calibration, ensemble design, and statistical evaluation—can be combined with medical imaging data to produce clinically useful tools. The framework is also designed with scalability in mind, allowing integration with teledermatology platforms, mobile applications, or point-of-care diagnostic systems, particularly beneficial for underserved or rural populations lacking direct access to dermatologists.

Nevertheless, certain challenges remain. The model's performance should be validated prospectively in real-world clinical settings, across different imaging devices, and in diverse patient populations including various skin tones and rare lesion subtypes. Additionally, integrating federated learning or privacy-preserving approaches will enable collaboration between institutions without compromising patient confidentiality. Future work will also focus on continuous learning from new data, incorporating multimodal information (such as patient history and dermoscopic metadata), and adapting the system for rapid updates as medical knowledge evolves.

In conclusion, this study represents a step forward in the application of biomedical engineering, artificial intelligence, and dermatology, providing evidence that AI-driven ensemble models can achieve expert-level diagnostic performance. Such systems have the potential to act as decision-support tools for dermatologists, improve early cancer detection rates, reduce diagnostic disparities, and ultimately contribute to better patient outcomes worldwide. By aligning technical innovation with clinical needs, biomedical engineering continues to play a pivotal role in transforming healthcare delivery.

REFERENCES

[1] A. Lomas, J. Leonardi-Bee, and F. Bath-Hextall, "A systematic review of worldwide incidence of nonmelanoma skin cancer," Br. J. Dermatol., vol. 166, no. 5, pp. 1069–1080, May 2012.

[2] D. Leiter, S. Eigentler, C. Garbe, "Epidemiology of skin cancer," Adv. Exp. Med. Biol., vol. 810, pp. 120–140, 2014.

[3] World Health Organization, "Skin cancers," WHO Fact Sheet, 2020.

[4] S. Rosendahl et al., "Diagnostic accuracy of dermoscopy for melanoma," Dermatology, vol. 222, no. 3, pp. 221–230, 2011.
[5] H. Kittler et al., "Diagnostic accuracy of dermoscopy," Lancet Oncol., vol. 3, no. 3, pp. 159–165, 2002.
[6] D. Lallas et al., "Interobserver variability of dermoscopic diagnosis," J. Am. Acad. Dermatol., vol. 68, no. 5, pp. 873–879, 2013.
[7] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
[8] G. Litjens et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017.
[9] A. Esteva, K. Chou, S. Yeung, et al., "A guide to deep learning in healthcare," Nat. Med., vol. 25, pp. 24–29, Jan. 2019.
[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.
[11] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Machine Learning (ICML), 2019, pp. 6105–6114.
[12] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, Feb. 2017.
[13] N. Brinker et al., "Deep learning outperformed dermatologists in melanoma image classification," Eur. J. Cancer, vol. 119, pp. 11–17, 2019.
[14] T. Pham et al., "Transfer learning for skin lesion classification," IEEE Access, vol. 8, pp. 30247–30258, 2020.
[15] D. Tschandl et al., "Expert-level diagnosis of skin cancer with AI," Lancet Oncol., vol. 21, no. 5, pp. 689–695, May 2020.
[16] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, Feb. 2017.
[17] S. Han, M. Kim, and J. Kim, "Classification of skin lesions using deep convolutional neural networks with transfer learning," IEEE Access, vol. 6, pp. 17779–17787, 2018.
[18] L. Bi et al., "Automatic melanoma detection via multi-task deep learning," IEEE Trans. Med. Imaging, vol. 38, no. 7, pp. 1505–1515, Jul. 2019.
[19] J. Kawahara and G. Hamarneh, "Fully convolutional neural networks to segment skin lesions and classify melanoma," in Proc. ISBI, 2016, pp. 515–519.
[20] M. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the ISBI 2016," IEEE Trans. Med. Imaging, vol. 36, no. 2, pp. 554–568, 2017.
[21] R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE ICCV, 2017, pp. 618–626.
[22] T. Majtner et al., "Multi-task learning for skin lesion classification," Comput. Methods Programs Biomed., vol. 202, 106007, 2021.
[23] T. Pham et al., "Ensemble deep learning for skin lesion classification," IEEE Access, vol. 9, pp. 101646–101657, 2021.
[24] Z. Xie et al., "Semi-supervised learning with consistency training for skin lesion classification," Med. Image Anal., vol. 73, 102192, 2021.
[25] A. Sheller et al., "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," Sci. Rep., vol. 10, 12598, 2020.