Original Article

MobileNetV2 and EfficientNetB3 Ensemble for Scalable Dermoscopic Skin Lesion Classification

Allied Medical Research Journal
Vol:3 No.4: 97-104

Muhammad Faris[1], Ramsha Qayyum[1], Mir Farooq Ali[2], Muhammad Mansoor Mughal[1], Tariq Javed[1]

[1]Hamdard University, Karachi, Pakistan
[2]Marche Polytechnic University, Ancona, Italy

ABSTRACT

Background: Skin diseases, ranging from mild benign lesions to life-threatening malignant conditions, remain a major global health concern. Early and accurate diagnosis is critical to avoid complications, yet this is often limited by the shortage of dermatologists and the subjective nature of visual inspections, particularly in low-resource settings. To address this challenge, this study proposes an automated deep learning framework for skin disease classification using dermoscopic images.

Methods: The framework employs a hybrid learning approach by integrating transfer learning and ensemble learning techniques. Specifically, MobileNetV2 and EfficientNetB3 models were combined to leverage their unique strengths, thereby enhancing generalization and predictive accuracy. The system was trained on a well-annotated dataset of 22,177 dermoscopic images, representing eight diagnostic categories that include benign, malignant, and pre-cancerous skin conditions.

Results: Experimental results demonstrated strong classification performance, achieving a training accuracy of 96.81%, validation accuracy of 87.66% (loss of 0.455), and test accuracy of 86%. To improve clinical trust and interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to highlight the image regions that contributed most to the model's decisions. In addition, a user-friendly diagnostic interface was developed, enabling real-time image input, automated analysis, and clear interpretive guidance. This makes the system accessible not only to healthcare providers but also to non-specialists, bridging gaps in dermatological care.

Conclusion: The proposed solution offers a reliable, interpretable, and scalable application of artificial intelligence for skin disease screening, with significant implications for tele-dermatology and seamless integration into clinical workflows.

Keywords: *Deep Learning, Dermoscopic Images, Transfer Learning, Tele-dermatology, Skin Disease Classification.*

Received: July 14, 2025**; Revised:** September 24, 2025**; Accepted:** October 14, 2025
Corresponding Email: ramsha.qayyum@hamdard.edu.pk
DOI: https://doi.org/10.59564/amrj/03.04/014

INTRODUCTION

Skin diseases, benign and malignant, are among the most widespread health issues worldwide and a significant part of primary care consultation, especially in the low- and middle-income countries (LMICs).[1] It is crucial to have an early diagnosis, more so in the case of melanoma, which can be fatal when it is diagnosed late.[2] Nonetheless, diagnosis has conventionally been based on the skills of dermatologists and dermoscopic imaging, which is inaccessible in most of the rural and underserved regions.[3] Also, traditional clinical methods can be prone to observer bias and misunderstanding.[4]

Artificial intelligence (AI), and, more specifically, deep learning (DL), demonstrates the potential of enhanced diagnostic accuracy, scalability, and efficiency.[5,6] Convolutional neural networks (CNNs) have shown extraordinary performance in the feature extraction as well as image classification.[7] Our work is based on the idea that we introduce a new framework of transfer learning and ensemble learning by combining MobileNetV2 and EfficientNetB3 to optimize the outcome. The transfer learning facilitates adaption of large-scale pretrained models to medical imaging and the ensemble methods enhance generalizability.[8,9]

We trained our model and tested it on a randomized sample of 22,177 dermoscopic images of 8 diagnostic classes. The findings proved 96.81% and high training accuracy and classification potential, which reduces chances of misdiagnosis in actual clinical practices. Gradient-weighted Class Activation Mapping (Grad-CAM) was also used to improve the interpretability, which proved that the model would always focus on clinically relevant regions.[10] Also, a user-friendly web interface has been created, as it allows real time predictions and visual feedback to dermatologists, general practitioners, and telehealth providers.

Overall, this research provides a scalable, easy-to-understand, and high-performing system for automatic skin disease classification, which has strong potential to be used in clinical decision-support systems and teledermatology settings.The variety of skin diseases is vast, and these can be both mild infections and serious cancers that can lead to death like melanoma. Their spread is on the rise all over the world because of various reasons such as urbanization, pollution, climate change and changes in lifestyles.[11] Melanoma and non-melanoma skin cancers represent a large fraction of cancer in the world and melanoma is highly metastatic and the most fatal of the two forms.[12] Diagnosis should therefore be done at an early stage since early intervention significantly enhances better prognosis and survival in malignant lesions of the skin.

Historical practice in diagnosis is based on the visual examination and dermoscopy by the dermatologist. Nevertheless, these approaches are subjective, they can also be inter-observer variants, and they are restricted due to the lack of dermatological skills, especially with low resources and rural locations[13]. These constraints underscore the importance of automated, scalable, and precise diagnostic instruments to help clinicians and increase access to underserved communities. The recent progress in the analysis of medical images, especially deep learning (DL) and computer vision has revolutionized dermatological studies. Convolutional Neural Networks (CNNs) have demonstrated good results in a range of classification, segmentation, and feature extraction and can often compete with dermatologists.[14] CNNs unlike the traditional methods learn hierarchical features directly on the raw pixels making them more efficient and accurate. Transfer learning has made additional contributions to this area by modifying existing pretrained architectures (e.g., EfficientNet and MobileNetV2) to medical imaging.[15] EfficientNet applies depth, width and resolution scaling in an optimized manner[16,17] and MobileNetV2 is small and can be used in a mobile setting or an embedded system.[18,19]

The other essential clinical adoption factor is interpretability. Gradient-weighted Class Activation Mapping (Grad-CAM) allows one to visualize the parts of an image that affect the predictions made by a model, which in turn promotes the development of trust and validation in clinicians.[20] Likewise, ensemble learning, which is a composite of multiple models, to increase the strength, decrease misclassification and promote generalizability is especially useful due to the intra-class similarities of dermatological lesions.[21,22] Training and validation of strong DL models have been achieved due to the presence of huge dermoscopic data sets, such as HAM10000, ISIC Archive, and DermNet23. Nonetheless, class imbalance is always an issue, and sometimes augmentation or resampling techniques are necessary.

The main objective of this study is to develop a hybrid deep learning model that combines EfficientNetB3 and MobileNetV2 architectures for accurate and interpretable classification of skin diseases. The proposed model aims to enhance diagnostic performance across eight dermatological categories using a carefully curated dataset. Additionally, the research seeks to design a user-friendly teledermatology interface that allows both clinicians and patients to easily access and utilize the system in clinical and remote healthcare settings.

METHODOLOGY

Data Preparation for Model Training

The extraction of the dataset was an important step in assuring the validity and the performance of the proposed model. The dataset utilized in the present study is the HAM10000 dataset of the International Skin Imaging Collaboration (ISIC), which comprises 11,720 dermoscopic images, which were assigned labels by a group of dermatologists, based on eight different categories: Actinic Keratosis, Basal Cell Carcinoma, Dermatofibroma, Melanoma, Nevus, Pigmented Benign Keratosis, Squamous Cell Carcinoma, and Vascular Lesions. These images were also made to have 256x256 pixels and preprocesses included shuffling, batching (32) and label coding. The images were also normalized in the 0-1 range in order to boost model learning and also became augmented by means of medically suitable approaches like rotations, flips, and zooms due to the lesser data and shows in Table I. This served to balance the dataset and minimize the risk of overfitting and finally increased the dataset to 22,177 images. The more rare classes were increased more, such as Vascular Lesions and Squamous Cell Carcinoma but the most common Nevus class did not change. WeB3D divided the dataset into training (70%), validation (20%), and test (10%) sets, which presented both a balanced training set and monitoring set and an unbiased evaluation set.

Table-1 Distribution of original and augmented images across eight skin disease classes

Classes	No. of original images	No. of Augmented images	Total No. of images
Basal Cell Carcinoma	619	2,319	2,938
Melanoma	1,305	1,695	2,861
Squamous Cell Carcinoma	229	2,387	2,616
Vascular Lesion	180	2,475	2,655
Pigmented Benign Keratosis	1,338	1,662	2,871
Actinic Keratosis	149	2,458	2,607
Dermatofibroma	160	2,469	2,629
Nevus	7,737	-	3,000

Model Building and Architectural Discovery

Four convolutional neural network (CNN) models were constructed and evaluated against each other:

- **Model A (Hybrid- ResNet50 + VGG16)**
  Hybrid deep residual learning with high spatial representation was not sparse in parameters.

- **Model B (Lightweight Hybrid-EfficientNetB3 + MobileNetV2)**
  Known as the efficiency-oriented one, it used the concept of compound scaling and depth-wise separable convolution to make the computation less expensive with efficiency and feature retrieval intact.

- **Model C (Regularized Hybrid)**
  Model B with dropout, batch normalization, and global average pooling to stabilize the training process and minimize overfitting.

- **Model D (Task-Tuned Hybrid)**
  The final selected design is shown in Fig. 1. It was based on EfficientNetB3 and MobileNetV2 backbones and custom convolutional layers, ReLU activation, and MaxPooling to learn dermatology specific patterns. A global average pooling layer maintained spatial detail, batch normalization enhanced training, dropout (0.5) enhanced regularization to enhance training. The last classification layer applied the softmax activation with eight classes.

  Model D was shown to have the most accurate, efficient and clinical relevant balance, having the benefit of transfer learning and with dermatology specific refinements.

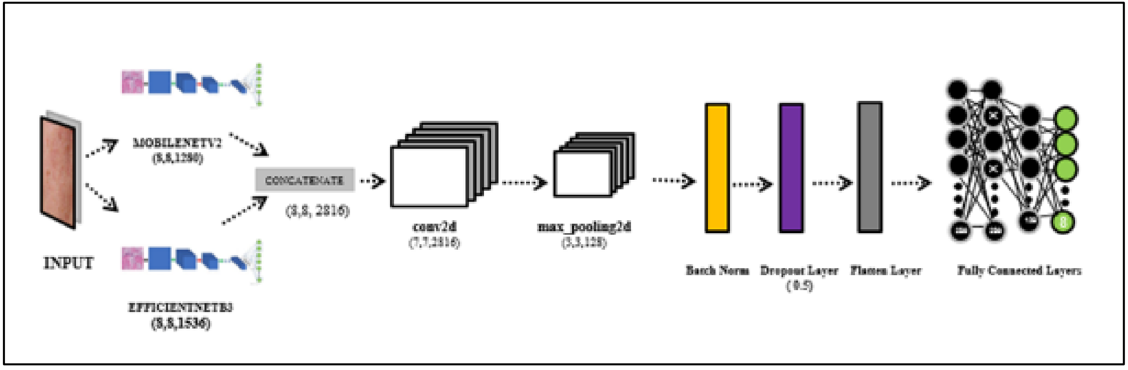

Fig.1 Architecture of task tuned Hybrid model

Training Configuration

The Adam optimizer used and sparse categorical cross-entropy loss were used to train the model. An early-stopping (not more than 15 epochs) and a learning rate scheduler avoided overfitting. The training was implemented on Kaggle Tesla T4 GPU (CUDA 12.6) based on the metric of accuracy.

Model Evaluation

Testing on the test set used accuracy, precision, recall, and F1-score and the macro and weighted averages were reported to take into account the imbalance in the classes. Findings affirmed that the model had good generalizability and had a strong predictive accuracy in various skin conditions.

Interpretability with Grad-CAM.

Gradient-weighted Class Activation Mapping (Grad-CAM) was then used to enhance the transparency of the last convolutional layer. Heatmaps were used to identify the areas that made the greatest contribution to predictions, which validated that the model was paying attention to clinically relevant factors and not a background noise. This was a step of interpretation that enhanced confidence in the reliability of the model to be used medically.

Overall Performance Using Custom Metrics

Even though the most commonly used evaluation indicators are accuracy and loss, they may not be adequate to evaluate a model performance in multi-class and imbalanced classification problems. To have a better analysis of the final model, the Task-Tuned Hybrid Model was tested on based on the precision, recall and F1-score metrics.

Evaluation of the Final Model

Based on the comparative results, Model D was selected as the final architecture for further evaluation. The following sections present an in-depth analysis of its performance using additional metrics and visualization techniques.

Interface Development

A user-friendly interface was developed with Streamlit to be used in a practical deployment. It allows users to upload pictures, get predictions and real-time Grad-CAM overlays, without its technical knowledge. With the combination of TensorFlow and OpenCV, the system facilitates effective and convenient application in the clinical practice especially in areas with limited resources.

Summary of Model Performance

Table-2 shows the values of training, validation, and test accuracy, and loss of the four developed models. It accentuates the development of the performance of the first hybrid model to the last optimized task specific model.

Class-Wise Performance Analysis

Although standard measures of model evaluation are accuracy and loss, which do not necessarily represent performance in multi-class or unbalanced classification problems. To give a more detailed evaluation Task-Tuned Hybrid Model was also tested in terms of precision, recall, and F1-score.

Table-2. Comparative Performance Metrics of All Models

Models	Training Accuracy	Training loss	Validation Accuracy	Validation loss	Test accuracy	Test Loss
Model A	62.08%	1.034	40.69%	2.271	56%	1.293
Model B	90%	0.2712	75%	0.9362	73%	0.79
Model C	81.95%	0.0950	83.79%	0.464	83%	0.460
Model D	96.81%	0.490	87.66%	0.455	86%	0.411

The model was found to be accurate to the extent of 0.97 which means that 97 percent of the positive predictions were true. It had a recall of 0.96 indicating good identification of 96% of true positive, and F1-score of 0.98, which indicates the perfect mix of precision and recall as illustrated in Table-3. These uniform high scores indicate that model has strong reliability and balanced predictive potential that can be potentially useful in terms of real and interpretable classification of skin diseases across various categories.

Table-3 Class-Wise Distribution

Class	Class Names	Precision	Recall	F1-Score
0	Nevus	0.70	0.74	0.72
1	Actinic Keratosis	0.82	0.88	0.85
2	Basal Cell Carcinoma	0.82	0.69	0.75
3	Dermatofibroma	0.85	0.89	0.87
4	Melanoma	0.50	0.58	0.54
5	Pigmented Benign Keratosis	0.58	0.45	0.51
6	Squamous Cell Carcinoma	0.65	0.73	0.68
7	Vascular Lesion	0.97	0.95	0.96

Interpretability through Grad-CAM Visualizations

To understand the model predictions and what the model focused on during classification, Gradient-weighted Class Activation Mapping (Grad-CAM) was used. Grad-CAM produces heatmaps, which are used to show the most significant regions of an image that led to the model decision. Fig. 2 makes side-by-side comparisons of original test images and their Grad-CAM overlaid images, showing the areas where the model focused during prediction.

Such visualizations are evidence of the model capability to localize clinically relevant features, which demonstrates the fact that its decisions are based on dermatological patterns. This interpretability does not only confirm the reliability of the model decision-making process but also enhances its possibility to be used in the diagnosis of skin diseases with the trustworthiness of clinical use.

DISCUSSION

This paper has shown that the classification of multi-class skin disease can be significantly enhanced by an effective architectural polishing instead of merely adding complexity to the models. Having substituted heavy and less compatible backbone networks with lightweight, pretrained models, including EfficientNetB3 and MobileNetV2, and introducing several other improvements, including the use of more convolutional layers, dropout, batch normalization, and global average pooling, the proposed hybrid model found an effective balance between the accuracy, generalizability, and interpretability.

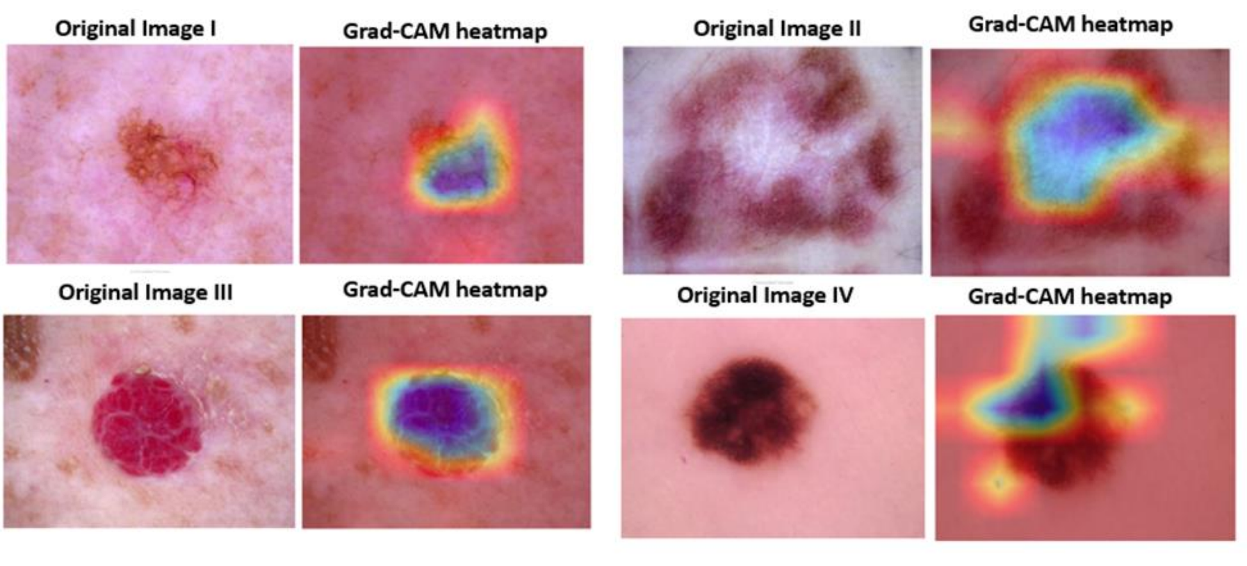

Fig.2 Comparison of original and Grad-CAM heatmap visualizations for the model's predictions

The test accuracy of 86 and high precision and recall rates show that the model has a high level of reliability in categorizing skin diseases with low chances of misclassifying. These results correspond to the recent literature to the point that too much complexity may result in overfitting and poor practical utility.

Grad-CAM visualizations also confirmed the interpretability of the model clinically as they indicated that the model predictions were guided by medically sensitive parts of the skin images. This enhances the confidence in the model decision-making process and promotes its The other limitation is related to the composition of the dataset, as it was not as diverse in the skin tone and demographic composition. These biases may influence the generalization of models on other populations and geographical areas. Moreover, the model architecture is computationally less expensive than most deep networks, but a more thorough analysis of the computational demands and scalability on low-resource hardware (e.g. mobile devices or rural clinic systems) is needed before scaling to large scale. Lastly, the paper fails to exhaust the practical implementation of this model in clinical practice. Practical implementation would involve clinical validation, user education, ethics, and EHR integrability. The consideration of these points in the workplace will be important in ensuring safe and effective implementation of the potential application to the use of the model as a decision-support tool in clinical and teledermatology practice. Nonetheless, there are a number of critical drawbacks that should be noted. The model demonstrated relatively low performance in detecting melanoma, which is one of the most difficult conditions of the skin to detect because of its visual similarity to benign lesions and inter-class variability. Also, the cases of failure showed that sometimes the model was not able to work with low-quality or darkened images, which should be improved, and the data should be further preprocessed through image enhancement.

AI-based dermatological tool in the daily health establishment practice.

To conclude, although the proposed model is performing promisingly and has interpretability, future studies must emphasize on the perfection of melanoma detection, improve the diversity of datasets, determine the scalability of the model, and create deployment options that are compliant with clinical workflow realities.

Limitations

Although these are promising results, there are a number of limitations that need to be mentioned. The model showed relatively worse results in recognition of melanoma, which is a category that is still especially tricky because of its unobtrusive and changeable appearance. Also, the data used was filtered and low in diversity that can limit the applicability of the model into the clinical setting where the quality of images, light, and skin tones differ dramatically. Thus, the arguments related to the reduction of false positives and false negatives should be perceived with caution, because the performance may vary depending on the types of diseases and population groups.

Recommendations

The future work will be aimed at a number of directions. First, it will be possible to increase the size of the dataset to cover various populations, imaging devices, and clinical settings to decrease bias and increase the robustness of the models. Second, melanoma detection will be enhanced by targeting augmentation of data, ensemble learning, and attention-based mechanisms. Third, full real-world validation in partnership with dermatologists will be undertaken to determine the value of the system diagnostics and usefulness during the clinical processes. Lastly, consideration of computational efficiency and deployment suitability in healthcare organizations with limited resources will be given a high priority to make the model accessible and viable in the international application of teledermatology.

CONCLUSION

This paper presented an advanced deep learning architecture in the classification of multi-class skin diseases based on lightweight pretrained backbones, with a set of extra convolutional layers and regularization methods, including dropout, batch normalization, and global average pooling. The hybrid model was observed to produce significant results, and the training, validation, and test accuracy were 96.81, 87.66 and 86, respectively. The evaluation metrics such as its precision (0.97), and recall (0.96) show that the model is consistent in thedifferentiating between the various dermatological classes. Moreover, Grad-CAM visualizations were also very easy to interpret, and confirmed the model focus to be clinically relevant, which increased confidence in the model prediction.

Acknowledgments

None.

Author Contributions

Muhammad Faris contributed to the study conception, design, and data collection. **Ramsha Qayyum** assisted in data analysis and manuscript drafting. **Mir Farooq Ali** supervised the research process and provided critical revisions. **Muhammad Masoor Mughal** contributed to data interpretation and statistical analysis. **Tariq Javed** assisted in literature review and final editing of the manuscript. All authors read and approved the final version of the manuscript.

Ethical Approval

Not applicable.

Grant Support and Funding Disclosure

None.

Conflict of Interests

None.

REFERENCES

1. Yakupu A, Aimaier R, Yuan B, Chen B, Cheng J, Zhao Y, Peng Y, Dong J, Lu S. The burden of skin and subcutaneous diseases: findings from the Global Burden of Disease Study 2019. *Front Public Health.* 2023;11:1145513.
DOI: https://doi.org/10.3389/fpubh.2023.1145513
2. Helgadottir H, Mikiver R, Schultz K, Nielsen K, Portelli F, Lapins J, Puig S, Isaksson K. Melanoma incidence and mortality trends among patients aged 59 years or younger in Sweden. *JAMA Dermatol.* 2024;160(11):1201–10.
DOI: https://doi.org/10.1001/jamadermatol.2024.3514
3. Urban K, Chu S, Giesey RL, Mehrmal S, Uppal P, Delost ME, Delost GR. Burden of skin disease and associated socioeconomic status in Asia: a cross-sectional analysis from the Global Burden of Disease Study 1990–2017. *JAAD Int.* 2021;2:40–50.
DOI: https://doi.org/10.1016/j.jdin.2020.10.006
4. Marghoob AA, Swindle LD, Moricz CZ, Sanchez Negron FA, Slue B, Halpern AC. Instruments and new technologies for the in vivo diagnosis of melanoma. *J Am Acad Dermatol.* 2003;49(5):777–97.
DOI: https://doi.org/10.1067/S0190-9622(03)02183-6
5. Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R. The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules.* 2020;10(8):1123.
DOI: https://doi.org/10.3390/biom10081123
6. Waheed SR, Saadi MS, Rahim MSM, Suaib NM, Najjar FH, Adnan MM, Salim AA. Melanoma skin cancer classification based on CNN deep learning algorithms. *Malays J Fundam Appl Sci.* 2023;19(3):299–305.
DOI: https://doi.org/10.11113/mjfas.v19n3.2900
7. Mienye ID, Swart TG, Obaido G, Jordan M, Ilono P. Deep convolutional neural networks in medical image analysis: a review. *Information.* 2025;16(3):195.
DOI: https://doi.org/10.3390/info16030195
8. Tan M, Le Q. EfficientNetV2: smaller models and faster training. *Proc Int Conf Mach Learn (ICML).* 2021;10096–10106. Available from: https://proceedings.mlr.press/v139/tan21a.html
9. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks.

Proc IEEE Conf Comput Vis Pattern Recognit. 2018:4510–20.
DOI: https://doi.org/10.1109/CVPR.2018.00474

10. Wang H, Wang Z, Du M, Yang F, Zhang Z, Ding S, Mardziel P, Hu X. Score-CAM: score-weighted visual explanations for convolutional neural networks. *Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops.* 2020;24–5. Available from: https://openaccess.thecvf.com/content_CVPRW_2020/html/W31/Wang_Score-CAM_Score-Weighted_Visual_Explanations_for_Convolutional_Neural_Networks_CVPRW_2020_paper.html
11. Karimkhani C, Dellavalle RP, Coffeng LE, Flohr C, Hay RJ, Langan SM, Nsoesie EO, et al. Global skin disease morbidity and mortality: an update from the Global Burden of Disease Study 2013. *JAMA Dermatol.* 2017;153(5):406–12.
DOI: https://doi.org/10.1001/jamadermatol.2016.5538
12. Helgadottir H, Mikiver R, Schultz K, Nielsen K, Portelli F, Lapins J, Puig S, Isaksson K. Melanoma incidence and mortality trends among patients aged 59 years or younger in Sweden. *JAMA Dermatol.* 2024;160(11):1201–10.
DOI: https://doi.org/10.1001/jamadermatol.2024.3514
13. Marghoob AA, Swindle LD, Moricz CZ, Sanchez Negron FA, Slue B, Halpern AC. Instruments and new technologies for the in vivo diagnosis of melanoma. *J Am Acad Dermatol.* 2003;49(5):777–97.
DOI: https://doi.org/10.1067/S0190-9622(03)02183-6
14. Nawaz K, Zanib A, Shabir I, Li J, Wang Y, Mahmood T, Rehman A. Skin cancer detection using dermoscopic images with convolutional neural network. *Sci Rep.* 2025;15(1):7252.
DOI: https://doi.org/10.1038/s41598-025-91446-6
15. Yu H, Yang LT, Zhang Q, Armstrong D, Deen MJ. Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing.* 2021;444:92–110.
DOI: https://doi.org/10.1016/j.neucom.2020.04.157
16. Li X, Shen L, Xie X, Chen F, Li Y. Deep learning in skin disease diagnosis: a review of current applications and future trends. *Comput Methods Programs Biomed.* 2022;226:107141.
DOI: https://doi.org/10.1016/j.cmpb.2022.107141
17. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proc Mach Learn Res.* 2019;97:6105–14. Available from: https://proceedings.mlr.press/v97/tan19a.html
18. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks. *Proc IEEE Conf Comput Vis Pattern Recognit.* 2018:4510–20.
DOI: https://doi.org/10.1109/CVPR.2018.00474
19. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella NCF, Halpern A, et al. Human–computer collaboration for skin cancer recognition. *Nat Med.* 2020;26(8):1229–34.
DOI: https://doi.org/10.1038/s41591-020-0942-0
20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis (ICCV).* 2017:618–26.
DOI: https://doi.org/10.1109/ICCV.2017.74
21. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. *Proc IEEE Conf Comput Vis Pattern Recognit.* 2009:248–55.
DOI: https://doi.org/10.1109/CVPR.2009.5206848
22. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115–8.
DOI: https://doi.org/10.1038/nature21056
23. Codella NC, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC). *arXiv preprint.* 2019. Available from: https://arxiv.org/abs/1902.03368