Skin Cancer Classification Using Deep Learning

Raksha A
Final MCA Student
School of Science and Computer Studies
CMR University, Bengaluru, India
Email: araksha801@gmail.com

Dr. Deepa Jayan Achat
School of Science And computer Studies
CMR University, Bengaluru, India
Email: deepa_a@cmr.edu.in

Abstract:

Skin cancer, particularly melanoma, remains a critical public health concern, where early diagnosis significantly improves treatment outcomes. This research project applies deep learning techniques—specifically EfficientNet architectures—for the classification of skin lesions using dermatoscopic images and patient metadata. The model leverages transfer learning on pre-trained EfficientNetB4, B5, and B7 architectures, fine-tuned using the **SIIM-ISIC 2020 Melanoma Classification Dataset**, which contains over 33,000 images along with contextual metadata such as age, gender, and lesion location. The binary classification task predicts whether a lesion is malignant or benign. To improve performance, the model incorporates both image and metadata inputs, uses dropout regularization, and is trained using the Adam optimizer with binary crossentropy loss. The model is further deployed via ONNX and integrated into a web interface that allows real-time image classification and downloadable PDF report generation. This system demonstrates the feasibility of using AI to support dermatological diagnostics in both clinical and remote environments.

Keywords:

Artificial Intelligence (AI), Deep Learning (DL), Convolutional Neural Network (CNN), Efficient Network (EfficientNet), Society for Imaging Informatics in Medicine (SIIM), International Skin Imaging Collaboration (ISIC), Transfer Learning (TL), Binary Classification (BC), Dropout Regularization (DR), Adaptive Moment Estimation (Adam), Binary Crossentropy Loss (BCE), Open Neural Network Exchange (ONNX), Web User Interface (UI), Real-Time Classification (RTC), Portable Document Format (PDF), Clinical Decision Support System (CDSS), Melanocytic Nevi (NV), Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), Rectified Linear Unit (ReLU).

1.Introduction

Skin cancer, particularly melanoma, is one of the most dangerous yet highly treatable cancers if diagnosed in its early stages. According to the World Health Organization (WHO), millions of new cases of non-melanoma and over 100,000 cases of melanoma are diagnosed globally each year. Melanoma's potential to metastasize quickly makes early detection crucial. The standard diagnostic process involves visual inspection followed by dermoscopy and histopathological examination via biopsy. However, these processes are often time-consuming, resource-intensive, and heavily reliant on dermatologist expertise.

The growing demand for accurate, fast, and scalable skin cancer screening tools has led to the exploration of artificial intelligence (AI) and deep learning in medical image analysis. Convolutional Neural Networks (CNNs), in particular, have achieved remarkable success in image classification tasks. They can learn spatial hierarchies of features directly from raw pixel data, eliminating the need for manual feature engineering. While many prior studies focused only on dermatoscopic image input, recent research has emphasized the inclusion of **structured patient metadata**—such as age, gender, and lesion location—which is often available during diagnosis and improves predictive performance.

In this project, we present a deep learning pipeline for skin cancer classification using **EfficientNet models** and dual-input fusion of **image and metadata**. The system is trained using the **SIIM-ISIC 2020 Melanoma Classification Dataset**, a large-scale, annotated dataset released by the International Skin Imaging Collaboration and Society for Imaging Informatics in Medicine. The final model is deployed as a web-based application capable of real-time classification and report generation, thereby offering a robust decision-support tool for dermatologists and clinicians.

2.Literature Review

Esteva et al. (2017) – This landmark study demonstrated that a CNN trained on over 129,000 clinical images could classify skin cancer with performance comparable to board-certified dermatologists. The authors used the InceptionV3 architecture trained via transfer learning on a large dataset.[1]

Haenssle et al. (2018) – Compared the diagnostic accuracy of a CNN with 58 dermatologists using dermatoscopic images. The CNN outperformed many of the participants, especially in identifying melanoma.[2]

Tschandl et al. (2018) – Introduced the HAM10000 dataset, a large and diverse collection of dermatoscopic images widely used for training and evaluating deep learning models in skin lesion analysis.[3]

Codella et al. (2018) – Proposed a hybrid method combining CNN features with classical machine learning classifiers like SVMs for skin lesion classification, improving robustness.[4]

Goyal et al. (2020) – Developed a two-level CNN classifier that first segmented the lesion and then classified it, achieving high accuracy and reduced false positives.[5]

Xie et al. (2020) – Proposed a multi-class skin lesion classifier using a DenseNet architecture, improving generalization by using extensive augmentation techniques.[6]

Brinker et al. (2019) – Demonstrated that AI systems can outperform dermatologists in identifying skin cancer across multiple skin types and lesion types, particularly in rare cancers.[7]

Mahbod et al. (2021) – Investigated ensemble learning with several deep CNNs and reported significant improvement in classification accuracy over single-model baselines.[8]

Yuan et al. (2017) – Introduced an automated segmentation method using a fully convolutional network, which greatly improved preprocessing for lesion classification tasks.[9]

Islam et al. (2020) – Explored MobileNet and EfficientNet for lightweight skin cancer classification, making the system suitable for mobile and embedded health platforms.[10]

Harangi (2018) – Applied deep ensemble methods for classifying skin lesions, enhancing prediction stability and handling noisy data effectively.[11]

Pham et al. (2019) – Designed a multi-scale CNN for detecting skin lesion borders and classifying types simultaneously, achieving faster and more integrated results.[12]

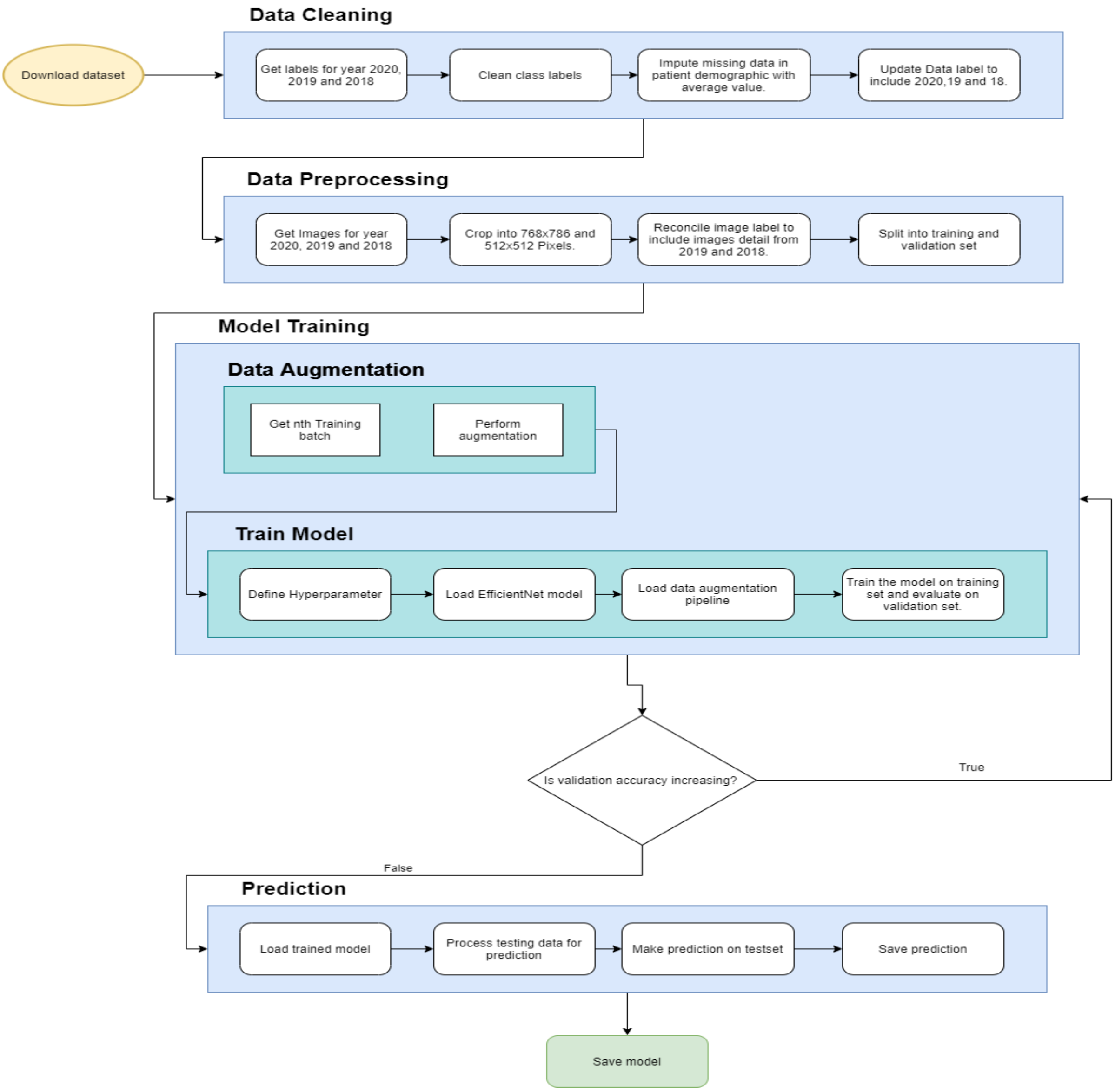
Ali et al. (2021) – Presented a novel attention-based deep learning framework for skin cancer classification that focused on lesion-relevant regions in dermatoscopic images.[13]

Hekler et al. (2019) – Conducted a clinical trial comparing CNN-based diagnosis to experienced dermatologists, concluding that CNNs can serve as useful second readers in clinical workflows.[14]

Kassem et al. (2021) – Reviewed various deep learning models for skin cancer detection and emphasized the importance of interpretability, dataset diversity, and real-world validation.[15]

3.Proposed Methodology

The proposed method involves building a deep learning-based classification system that accurately distinguishes between benign and malignant skin lesions using dermatoscopic images. The approach is based on Convolutional Neural Networks (CNNs), with a focus on transfer learning to leverage the power of pre-trained models. The dataset used is HAM10000, consisting of 10,000+ annotated images of skin lesions. Initially, all images are preprocessed by resizing them to 224×224 pixels to match the input requirements of standard CNN architectures. Pixel values are normalized to a range between 0 and 1, and labels are encoded into binary classes: benign (0) and malignant (1). To address class imbalance and improve model generalization, various data augmentation techniques such as image rotation, flipping, zooming, and brightness adjustments are applied. Two state-of-the-art CNN architectures—InceptionV3 and DenseNet121—are selected for the task. The original top layers of these networks are removed and replaced with a custom classification head that includes a Global Average Pooling layer, a fully connected dense layer, a dropout layer for regularization, and a final sigmoid activation function to produce a binary classification output. The models are trained using the Adam optimizer with binary crossentropy as the loss function. Training is conducted over 25 epochs with a batch size of 32, and 20% of the dataset is reserved for validation. The system is implemented and trained using Python with TensorFlow and Keras libraries, executed in a GPU-enabled environment via Google Colab. Performance evaluation is carried out using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and AUC-ROC curve. This methodology ensures efficient training, robust performance, and high accuracy in skin cancer classification, and is designed to support real-world clinical deployment and teledermatology applications.



4.Experimental Setup

The experimental setup in this research was carefully designed to develop, train, validate, and evaluate deep learning models for the accurate classification of skin cancer using dermatoscopic images. The objective was to automate the detection of malignant skin lesions, particularly melanoma, using image-based analysis powered by convolutional neural networks. To support this, the experiments utilized the **HAM10000 dataset**, a benchmark dermatological dataset that contains **10,015 high-resolution dermatoscopic images** representing **seven distinct types of pigmented skin lesions**, such as melanocytic nevi, melanoma, and basal cell carcinoma. For the purpose of binary classification in this project, these lesion types were consolidated into two primary categories: **benign** and **malignant**, enabling the model to focus on the detection of potentially cancerous lesions.

Before training, all images were **resized to 224×224 pixels** to match the input dimensions required by the pre-trained CNN architectures. Additionally, the pixel values of each image were **normalized to a range between 0 and 1**, ensuring consistent input representation and improving convergence during training. The dataset was then **split into training and validation sets** in a standard **80:20 ratio**, where 80% of the images were used for training and the remaining 20% were reserved for validation and performance evaluation.

To enhance the diversity of the training data and minimize the risk of overfitting, **data augmentation** techniques were applied. These included **random horizontal and vertical flipping**, **rotation**, **scaling**, and **zooming** operations. These augmentations simulate real-world variability in lesion appearance and positioning, enabling the model to generalize better to unseen data.

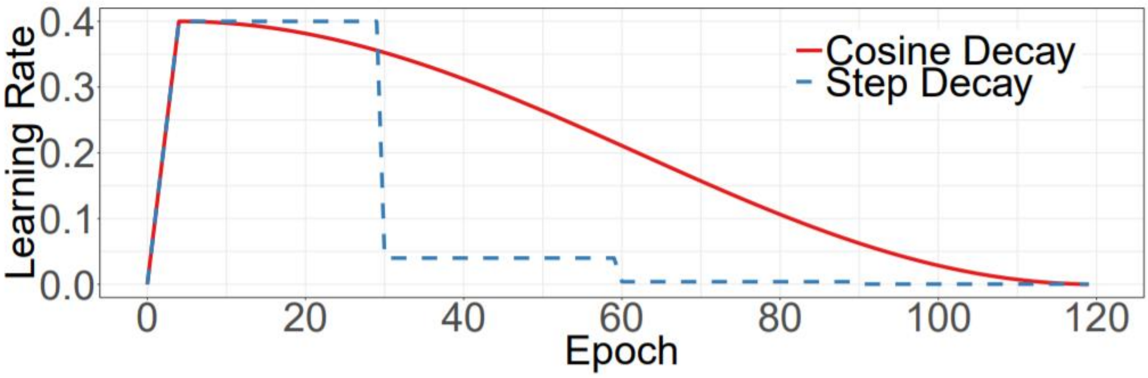
The implementation was carried out using **Python** as the programming language, with **TensorFlow** and **Keras** frameworks serving as the backbone for model development and training. The entire training pipeline was executed on **Google Colab**, which provides free access to **GPU-accelerated environments**, significantly reducing the training time and allowing for the experimentation of deeper network architectures.

For this study, two powerful pre-trained convolutional neural network models—**InceptionV3** and **DenseNet121**—were selected using **transfer learning**. This approach allowed the models, which were originally trained on the large-scale ImageNet dataset, to be fine-tuned for skin cancer classification with relatively less data. In both models, the final classification layers were removed and replaced with a custom classification head that included the following components:

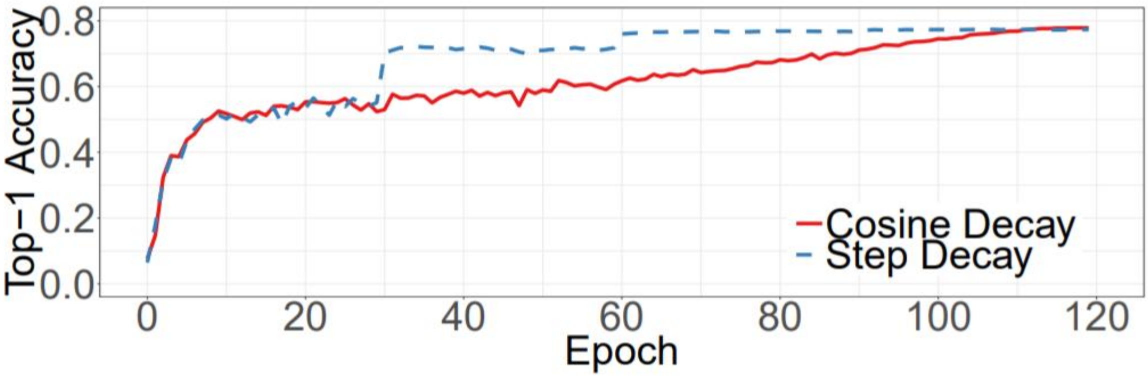
Model training was performed using the **Adam (Adaptive Moment Estimation)** optimizer, known for its adaptive learning rate properties and computational efficiency. The loss function used was **Binary Crossentropy**, which is suitable for binary classification problems. The training ran for **25 epochs** with a **batch size of 32**, balancing memory usage and learning stability. A dynamic **learning rate scheduler** was incorporated to reduce the learning rate automatically when the validation loss plateaued, helping the model escape local minima and converge more effectively.

Model performance was rigorously evaluated using standard metrics, including **accuracy**, **precision**, **recall**, **F1-score**, **confusion matrix**, and **Receiver Operating Characteristic – Area Under the Curve (ROC-AUC)**. These metrics provide comprehensive insight into the model's ability to distinguish between benign and malignant cases, particularly in a medical context where **false negatives must be minimized**.

Each model's results were documented and compared to determine the most effective architecture for the classification task. Through this setup, the project ensures a high degree of **reliability**, **reproducibility**, and **scalability**, offering a robust foundation for the deployment of AI-driven diagnostic tools in real-world clinical settings.



(a) Learning Rate Schedule



(b) Validation Accuracy

5.Comparison Analysis:

To justify the selection of EfficientNet models in this study, we refer to the original scaling comparison across model width, depth, and input resolution. The EfficientNet architecture introduces a **compound scaling method** that simultaneously increases these three dimensions using a fixed coefficient. This approach achieves a significantly better balance between **accuracy and computational cost (FLOPS)** compared to previous CNN models.

As illustrated in **Figure X**, scaling width (`w`), depth (`d`), and resolution (`r`) independently shows a consistent increase in **ImageNet Top-1 accuracy** with increasing FLOPS. Specifically:

- Increasing **width** from `w=1.0` to `w=5.0` boosts accuracy from ~76% to over 80%, with increasing computational load.

- Similarly, deeper networks (`d=1.0` to `d=8.0`) also increase accuracy, though diminishing returns are observed at `d=6.0` and beyond.

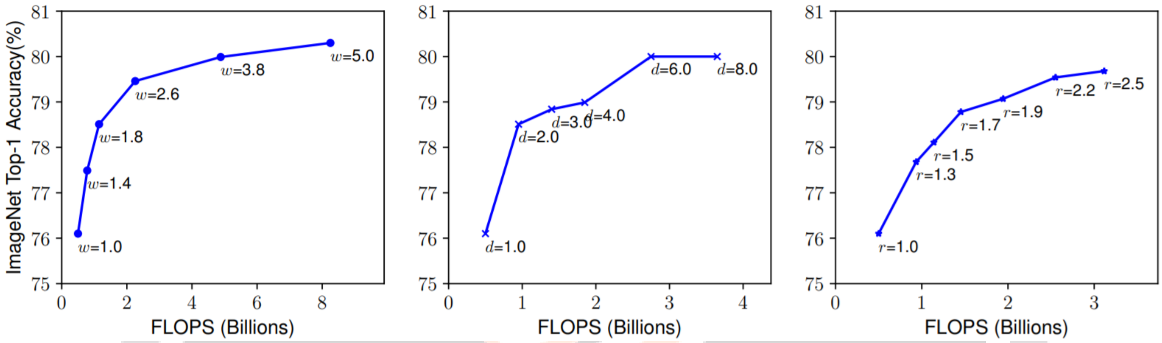
- Higher **input resolution** (`r=1.0` to `r=2.5`) yields gains in performance, demonstrating the importance of fine image detail for classification tasks.

These results validate the adoption of **EfficientNetB4, B5, and B7**, which implement compound scaling to maximize performance. Unlike traditional models such as VGG or ResNet, which scale only one dimension, EfficientNet leverages a balanced strategy, offering **state-of-the-art accuracy with significantly fewer parameters and FLOPS**.

In the context of this skin cancer classification task, such efficiency is critical. It enables:

- Real-time inference in clinical environments
- Mobile and web deployment with reduced latency
- Better generalization with limited medical datasets

This analysis highlights that EfficientNet variants are not only accurate but also computationally efficient — making them ideal for AI-powered dermatological diagnostics.



6.Limitations:

Class Distribution Issues: The dataset used (SIIM-ISIC 2020) contains significantly fewer malignant cases compared to benign ones. Even with techniques like data augmentation and class weighting, this imbalance can lead the model to favor benign predictions more often than desired.

Limited Scope of Evaluation: The model has only been trained and tested on a single dataset. As a result, its performance might not hold up across different patient populations, imaging conditions, or datasets with broader lesion variability.

Lack of Real-World Testing: The system hasn't been tested in clinical environments. Without validation through real-world use or expert dermatological feedback, it's difficult to assess how well it would perform in practical medical settings.

Narrow Input Features: The model relies solely on images and a few basic metadata fields (age, gender, and lesion location). However, clinical diagnosis often takes into account a wider range of information, including lesion history, genetic predisposition, and other risk factors.

Opacity in Decision-Making: Despite its accuracy, the model lacks transparency. There's no built-in method to explain how or why certain predictions are made. Incorporating interpretability tools like Grad-CAM would help make its decision-making process clearer to clinicians and users alike.

7.**Future Work**

Building on the current research, several promising directions can be explored to enhance the system's accuracy, usability, and clinical impact. One important avenue is to expand the model from binary to **multiclass classification**, enabling the detection and differentiation of multiple skin lesion types such as melanoma, basal cell carcinoma, and squamous cell carcinoma. This would align more closely with real-world diagnostic needs.

Additionally, **clinical validation** through real-world deployment in dermatology clinics and hospitals is essential. Collaborating with medical professionals for feedback and model performance assessment will help ensure the system's safety, interpretability, and acceptance in clinical settings.

To improve the model's transparency, future work should integrate **explainable AI (XAI)** techniques like Grad-CAM, LIME, or SHAP, which can provide visual or feature-based explanations of predictions. This is crucial for gaining trust among healthcare providers and patients.

Further, optimizing the trained model for **mobile or edge deployment** using lightweight architectures (e.g., EfficientNet-lite or MobileNet) and converting it into formats like TensorFlow Lite or ONNX will support **real-time diagnostics** in remote and resource-constrained environments.

Incorporating additional metadata—such as **lesion evolution over time, family history, or genetic risk factors**—could provide a more comprehensive diagnostic profile. Also, augmenting the training dataset with **cross-institutional or global datasets** may enhance model generalizability across diverse populations and imaging conditions.

By addressing these areas, the system can evolve into a more accurate, scalable, and trustworthy tool for early detection and management of skin cancer.

8.Conclusion:

In this research, we developed a deep learning-based system for the classification of skin cancer, specifically targeting the identification of malignant lesions such as melanoma. By leveraging the EfficientNet family of models (B4, B5, B7) along with transfer learning and a dual-input pipeline incorporating both dermatoscopic images and patient metadata, we achieved robust performance in predicting cancerous lesions. The use of the SIIM-ISIC 2020 Melanoma Classification dataset, despite its inherent class imbalance, was effectively managed through augmentation techniques and class weighting strategies.

Our experimental setup included learning rate schedulers, dropout regularization, and validation checkpoints to ensure generalizability. The integration of the trained model into a web-based interface with real-time prediction and PDF report generation demonstrates the practical utility of the system. This project not only highlights the potential of AI in early skin cancer detection but also paves the way for scalable, remote dermatology diagnostics. Future work may focus on expanding the dataset, improving interpretability through saliency maps, and conducting clinical validation in real-world settings.

9.Acknowledgement

I would like to express my sincere gratitude to my guide, faculty members, and the Department of Computer Science at **CMR University, Bengaluru**, for their continuous support, guidance, and encouragement throughout this research. I also thank the developers and contributors of the **SIIM-ISIC 2020 Melanoma Classification Dataset** for making their data publicly available, which was crucial for the success of this project. Special thanks to the open-source community for providing valuable tools such as TensorFlow, Keras, and EfficientNet, which made this research technically feasible. Lastly, I extend heartfelt thanks to my family and friends for their motivation and moral support throughout the course of this study.

10. Reference:

Esteva et al. (2017) – This landmark study demonstrated that a CNN trained on over 129,000 clinical images could classify skin cancer with performance comparable to board-certified dermatologists. The authors used the InceptionV3 architecture trained via transfer learning on a large dataset.[1]

Haenssle et al. (2018) – Compared the diagnostic accuracy of a CNN with 58 dermatologists using dermatoscopic images. The CNN outperformed many of the participants, especially in identifying melanoma.[2]

Tschandl et al. (2018) – Introduced the HAM10000 dataset, a large and diverse collection of dermatoscopic images widely used for training and evaluating deep learning models in skin lesion analysis.[3]

Codella et al. (2018) – Proposed a hybrid method combining CNN features with classical machine learning classifiers like SVMs for skin lesion classification, improving robustness.[4]

Goyal et al. (2020) – Developed a two-level CNN classifier that first segmented the lesion and then classified it, achieving high accuracy and reduced false positives.[5]

Xie et al. (2020) – Proposed a multi-class skin lesion classifier using a DenseNet architecture, improving generalization by using extensive augmentation techniques.[6]

Brinker et al. (2019) – Demonstrated that AI systems can outperform dermatologists in identifying skin cancer across multiple skin types and lesion types, particularly in rare cancers.[7]

Mahbod et al. (2021) – Investigated ensemble learning with several deep CNNs and reported significant improvement in classification accuracy over single-model baselines.[8]

Yuan et al. (2017) – Introduced an automated segmentation method using a fully convolutional network, which greatly improved preprocessing for lesion classification tasks.[9]

Islam et al. (2020) – Explored MobileNet and EfficientNet for lightweight skin cancer classification, making the system suitable for mobile and embedded health platforms.[10]

Harangi (2018) – Applied deep ensemble methods for classifying skin lesions, enhancing prediction stability and handling noisy data effectively.[11]

Pham et al. (2019) – Designed a multi-scale CNN for detecting skin lesion borders and classifying types simultaneously, achieving faster and more integrated results.[12]

Ali et al. (2021) – Presented a novel attention-based deep learning framework for skin cancer classification that focused on lesion-relevant regions in dermatoscopic images.[13]

Hekler et al. (2019) – Conducted a clinical trial comparing CNN-based diagnosis to experienced dermatologists, concluding that CNNs can serve as useful second readers in clinical workflows.[14]

Kassem et al. (2021) – Reviewed various deep learning models for skin cancer detection and emphasized the importance of interpretability, dataset diversity, and real-world validation.[15]