Vol. 5 No. 1 (2026)

Pakistan Journal of Medical & Cardiological Review

https://pakjmcr.com/index.php/1/about

Online ISSN 3007-2387

Print ISSN 3007-2379

Deep Dermatology: Enhancing Skin Cancer Precision through Adaptive Transfer Learning

Humayoun Akhtar
Faculty of Computer Science & IT , Superior University, Lahore, Pakistan
Email: su92-msdsw-s24-009@superior.edu.pk, humayounakhtar.ai@gmail.com

Iftikhar Naseer
Faculty of Computer Science & IT , Superior University, Lahore, Pakistan
Email: iftikhar.naseer@superior.edu.pk

Muhammad Hamza
Faculty of Computer Science & IT , Superior University, Lahore, Pakistan
Email: hamzarafiq225@gmail.com

Nimra Bibi
Department Environmental Science, Allama Iqbal Open University Islamabad (AIOU)
Email: nimrabibi389@gmail.com

Jannat Raza
Faculty of Computer Science & IT , Superior University, Lahore, Pakistan
Email: jannatraza1472@gmail.com

Author Details

Keywords:

Received on 05 Mar 2026
Accepted on 25 Mar 2026
Published on 06 Apr 2026

Corresponding E-mail & Author*:

Humayoun Akhtar
Faculty of Computer Science & IT , Superior University, Lahore, Pakistan
Email: su92-msdsw-s24-009@superior.edu.pk, humayounakhtar.ai@gmail.com

Abstract

Skin cancer classification requires accurate and efficient models for early diagnosis to reduce mortality. While deep learning techniques show promise, achieving a balance between diagnostic speed and precision remains a challenge. This study evaluates three deep learning models—Convolutional Neural Network (CNN), EfficientNetV2-B0, and Vision Transformer (ViT-B16)—to identify optimal architectures for clinical deployment. Using Kaggle's ISIC Skin Cancer Detection Dataset (2,637 training images and 660 test images), we applied transfer learning, data augmentation, and a suite of performance metrics including accuracy, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC). The results demonstrated that transfer learning models significantly outperformed the baseline CNN. EfficientNetV2 achieved 86.67% accuracy with an MCC of 0.7307 and rapid inference time (0.00190 seconds per sample), making it suitable for real-time diagnostics. Vision Transformer (ViT-B16) attained the highest accuracy of 88.48% and MCC of 0.7675 but had slower inference (0.00852 seconds per sample). These findings indicate that EfficientNetV2 provides the best balance between accuracy and speed for resource-constrained clinical settings, while ViT-B16 prioritizes diagnostic precision. This research provides valuable insights for context-specific model selection in dermatology,

highlighting the potential of transfer learning to improve skin cancer detection. Future work will explore hybrid architectures to optimize the accuracy-speed trade-off.

Introduction

Skin cancer, particularly melanoma, poses a significant global health challenge due to its rapid incidence growth and high mortality if untreated. With over 99,780 new cases reported in the United States alone in 2022, melanoma has become one of the most common and deadly cancers [1]. Despite accounting for only 1% of skin malignancies, melanoma is responsible for approximately 75% of skin cancer-related deaths [2]. Early detection is critical, as the 5-year survival rate for melanoma is 99% if detected before metastasis to lymph nodes, dropping to 30% if it spreads to other organs [3]. This stark contrast underscores the urgent need for accurate and timely diagnostic tools.

Skin cancer arises primarily from unprotected exposure to ultraviolet (UV) radiation, leading to uncontrolled growth of skin cells [4]. Melanoma originates in melanocytes, the cells responsible for skin pigmentation, and can be classified as benign (non-cancerous) or malignant based on cellular severity [5]. Traditional diagnostic methods, such as dermoscopy and biopsy, are time-consuming and highly dependent on clinician expertise, often leading to delays in treatment [6]. Furthermore, manual identification of skin lesions is challenging, with dermoscopic image accuracy ranging from 75% to 84% [7]. These limitations highlight the need for computer-assisted diagnostic systems to reduce variability and improve accuracy.

Recent advances in artificial intelligence (AI) and deep learning (DL) have shown promise in automating skin cancer detection. Convolutional neural networks (CNNs) have achieved dermatologist-level accuracy in classifying skin lesions, demonstrating the potential to classify melanoma with the same precision as human experts [4]. However, training robust CNNs requires large, annotated datasets, which are often scarce in medical imaging [8]. Transfer learning (TL), where pre-trained models are fine-tuned on smaller datasets, has emerged as a solution to mitigate overfitting and leverage pre-existing feature representations [9]. Despite these advancements, several challenges persist, including class imbalance in datasets, overfitting in limited training scenarios, and the computational costs of deep architectures [10].

This study addresses these gaps by systematically evaluating three deep learning models—a baseline CNN, EfficientNetV2-B0, and Vision Transformer (ViT-B16)—for skin cancer classification using the ISIC 2020 dataset. The primary objectives are to: (1) implement a baseline CNN model for melanoma detection, (2) apply transfer learning with EfficientNetV2 and ViT to improve diagnostic accuracy and efficiency, and (3) compare model performance in terms of accuracy, robustness, and inference speed. The findings aim to identify the optimal architecture for clinical deployment, offering a scalable solution to enhance early detection rates and reduce mortality.

The paper is structured as follows: Section 2 reviews related work in skin cancer detection, highlighting existing approaches and limitations. Section 3 details the methodology, including dataset preparation, model architectures, and training protocols. Section 4 presents results, including performance metrics and trade-off analysis. Section 5 concludes with insights and future directions.

Literature Review:

The detection of skin cancer, particularly melanoma, has evolved significantly with advancements in artificial intelligence (AI) and deep learning (DL). Traditional methods, such as dermoscopy and biopsy, remain limited by subjectivity and delays in diagnosis [11]. This review synthesizes existing literature on computational approaches for skin cancer detection, focusing on deep learning models, transfer learning techniques, and unresolved challenges.

Early efforts in automated skin cancer detection relied on handcrafted features and classical machine learning (ML) algorithms. For instance, Khan et al. (2019) used a

Gaussian filter, K-means clustering, and support vector machines (SVMs) to classify melanoma and nevi, achieving 96% accuracy on the DERMIS dataset [12]. Similarly, Filali et al. (2020) combined deep learning and handcrafted features, achieving 87.8% accuracy on the ISIC dataset [13]. However, these methods required extensive preprocessing, segmentation, and feature extraction, making them computationally intensive and less scalable [14].

A key limitation of traditional approaches is their reliance on domain-specific expertise for feature engineering, which may not generalize well across diverse datasets [15]. Additionally, the accuracy of these methods often lags behind modern deep learning techniques, which can autonomously learn complex patterns from raw data [16].

Convolutional neural networks (CNNs) have emerged as a powerful tool for skin cancer classification, achieving dermatologist-level accuracy [17]. Esteva et al. (2017) demonstrated that a CNN trained on 129,450 clinical images could classify skin lesions with 91% sensitivity and 91% specificity, outperforming 21 board-certified dermatologists [18]. CNNs leverage hierarchical feature learning, enabling them to capture intricate patterns in dermoscopic images without manual feature extraction [19]. However, training robust CNNs from scratch requires large annotated datasets, which are often scarce in medical imaging [20]. For example, the ISIC 2020 dataset contains only 584 malignant images, leading to class imbalance and overfitting [21]. Moreover, deep architectures like ResNet-152, with 152 layers, increase computational costs and memory usage, hindering clinical deployment [22].

Transfer learning (TL) has emerged as a solution to mitigate data scarcity and overfitting. By leveraging pre-trained models on large datasets (e.g., ImageNet), TL reduces the need for extensive labeled data and computational resources [23]. For instance, Cakmak et al. (2021) used MobileNetV2 with transfer learning, achieving 98.2% accuracy on the HAM10000 dataset by addressing class imbalance through data augmentation [24]. Similarly, Hosny et al. (2019) fine-tuned AlexNet on skin lesion images, achieving 96.86% accuracy [25].

Recent studies have explored advanced architectures for skin cancer detection. Rashid et al. (2022) proposed a MobileNetV2-based TL model, achieving 98.2% accuracy on the ISIC 2020 dataset by combining data augmentation and automated feature extraction [26]. Mukherjee et al. (2019) used a CNN-based model (CMLD) on the Dermofit dataset, achieving 90.58% accuracy, while Dalila et al. (2017) reported 93.6% accuracy using an ANN-based model on a self-curated dataset [27][28].

Vision Transformers (ViTs) and EfficientNet models have gained traction for their efficiency and accuracy. Dosovitskiy et al. (2021) introduced ViTs, which achieved state-of-the-art performance on image classification tasks by applying transformer architectures to vision data [29]. Tan et al. (2019) proposed EfficientNet, which uses compound scaling to balance depth, width, and resolution, achieving superior accuracy with fewer parameters [30].

While these models have been applied to various medical imaging tasks, their use in skin cancer detection remains limited. This study explores the application of EfficientNetV2-B0 and ViT-B16 to the ISIC 2020 dataset, addressing gaps in literature and evaluating their trade-offs between accuracy and inference efficiency. Performance metrics such as accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC) are critical for evaluating skin cancer detection models [31]. MCC is particularly valuable for imbalanced datasets, as it accounts for true and false positives/negatives [32].

Clinical deployment requires models with high accuracy and low inference time. For example, Brinker et al. (2019) reported that deep CNNs outperformed dermatologists in melanoma classification, with sensitivity of 82.3% and specificity of 77.9% [33]. However, computational costs and inference speed remain barriers to real-world adoption [34]. This study evaluates models not only on accuracy but also on inference efficiency, aiming to identify a balance suitable for clinical use.

Despite significant progress in the field, several key challenges persist. One major issue is class imbalance, as most datasets contain fewer malignant samples compared to benign ones, leading to biased models [35]. Additionally, computational efficiency remains a concern, as high-performance models often require substantial computational resources, which limits their applicability in resource-constrained settings [36]. Another challenge is generalizability, where models trained on specific datasets may fail to perform well on diverse populations or different imaging protocols [37]. To address these issues, future research should focus on developing lightweight architectures that achieve high accuracy while maintaining low inference times. Advanced data augmentation and resampling techniques should also be explored to mitigate class imbalance. Furthermore, leveraging federated learning could enable the training of models across decentralized datasets while preserving data privacy [38]. The literature reveals a transition from traditional ML methods to deep learning and transfer learning for skin cancer detection. While CNNs and TL models have shown promise, challenges such as class imbalance, computational costs, and generalizability remain. This study aims to address these gaps by evaluating EfficientNetV2-B0 and ViT-B16 on the ISIC 2020 dataset, focusing on accuracy, robustness, and inference efficiency.

Proposed Methodology:

The foundation of this study lies in the utilization of ISIC Skin Cancer Detection Dataset, a comprehensive repository containing 2,637 training images and 660 test images. The training set comprises 1,440 benign and 1,197 malignant skin lesion images, while the test set includes 360 benign and 300 malignant images. This distribution ensures a balanced representation of both classes, with benign lesions constituting approximately 54.6% of the training data and 54.5% of the test data, and malignant lesions making up the remaining 45.4% and 45.5%, respectively. To facilitate data handling and preprocessing, Python's glob and os libraries were employed to systematically retrieve image file paths. A critical step in ensuring data quality involved visual inspection. A random sample of 25 images was selected and visualized using the matplotlib library. These images were displayed in both RGB and grayscale formats, with the latter utilizing the inferno colormap to accentuate texture variations and structural details within the skin lesions. This visual assessment was crucial in verifying the correctness of image labels and confirming the diversity of lesion types present in the dataset, thereby ensuring its suitability for subsequent analysis and model training.

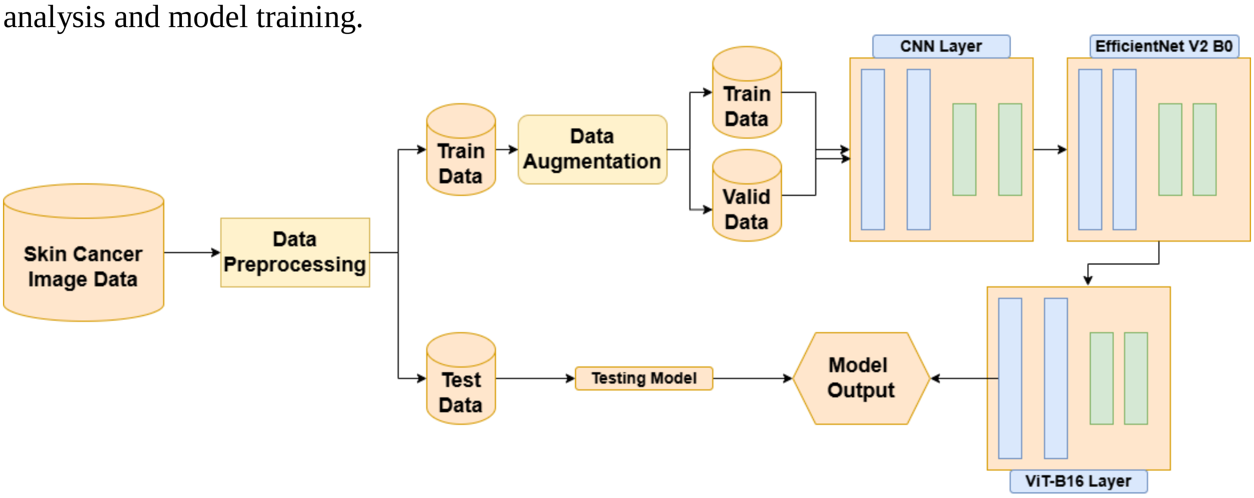

Figure 01: Proposed Methodology Structure.

The preprocessing phase was meticulously designed to optimize data for model training while addressing potential biases and overfitting risks. The initial step involved data splitting, where the training set was strategically divided into 85% training data (2,241 samples) and 15% validation data (396 samples). This stratified split, achieved using the train_test_split function from the scikit-learn library, ensured proportional representation of both benign and malignant classes in each subset, preserving the original class distribution.

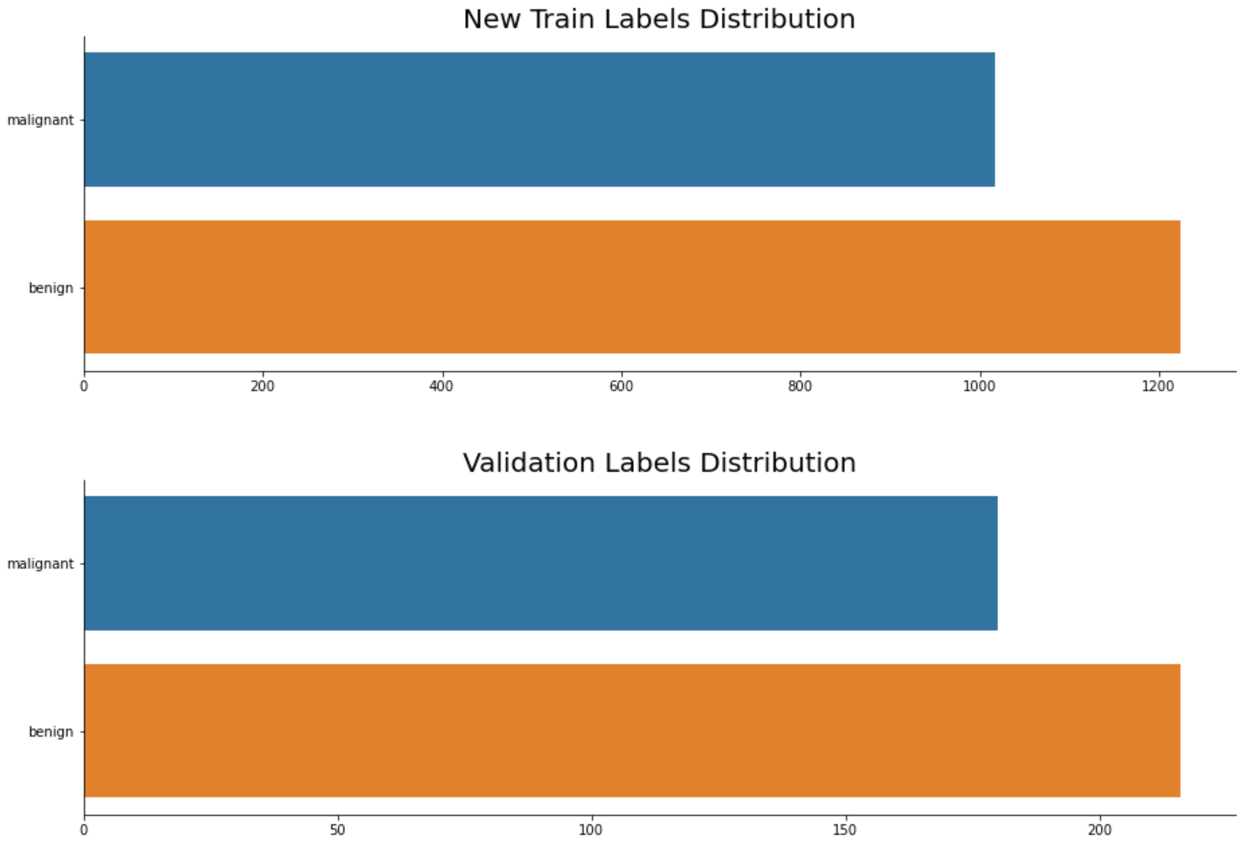

Figure 02: Train & Validation Labels Distribution.

Furthermore, to align with model requirements, categorical labels (benign and malignant) were numerically encoded as integers (1 and 0, respectively). Subsequently, data augmentation techniques were applied to the training data to artificially expand the dataset and simulate variations in lesion orientation, scale, and appearance. These techniques included random rotations, horizontal flipping, and zooming, which introduced diversity and helped the model generalize better to unseen data. The input pipeline was further optimized using TensorFlow's tf.data API. Key optimizations included batching, where images were grouped into batches of 22 to balance computational efficiency and memory usage; prefetching, which reduced input/output latency by preloading data during training; and normalization, where pixel values were scaled to the range [0, 1] to accelerate model convergence. These pipeline optimizations collectively ensured efficient data handling and streamlined the training process.

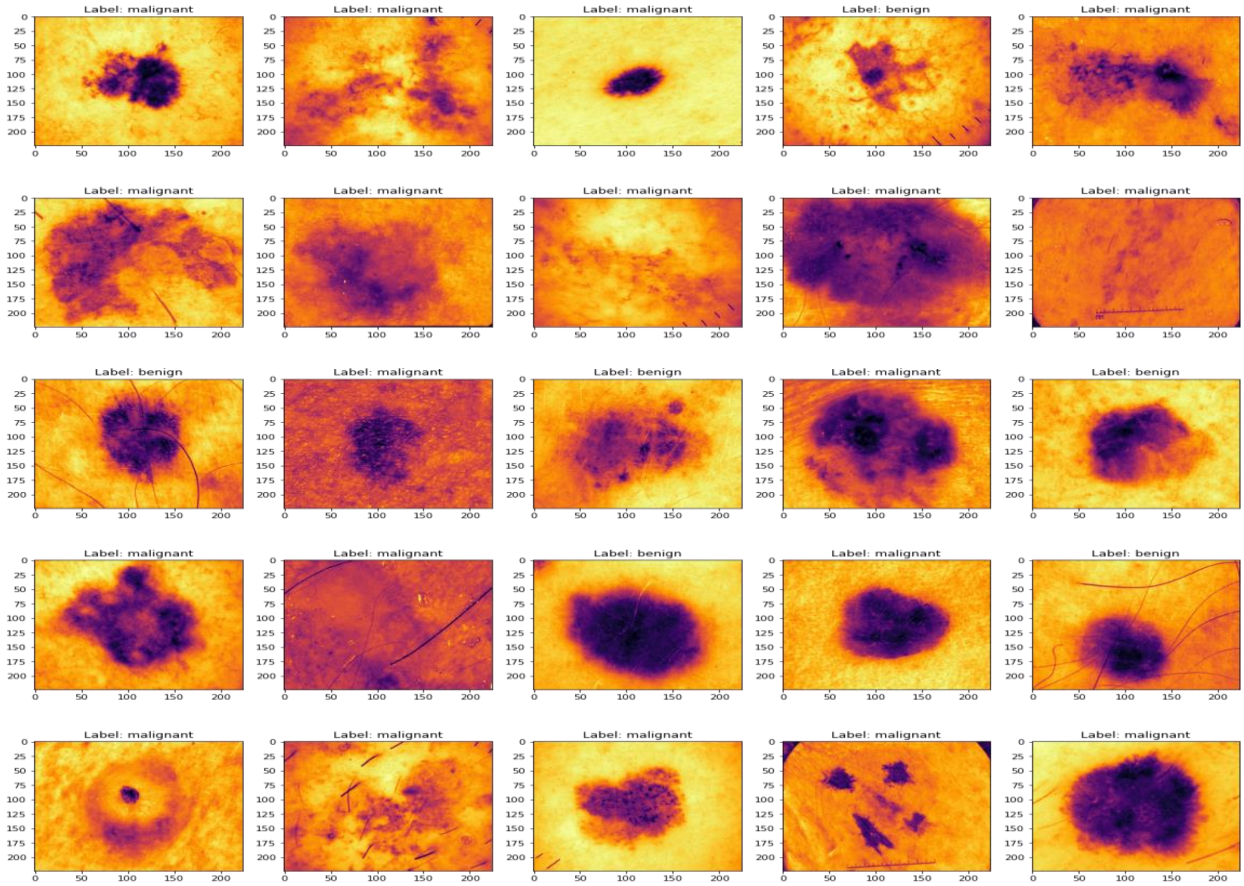

Figure 03: Multiple random samples image of data.

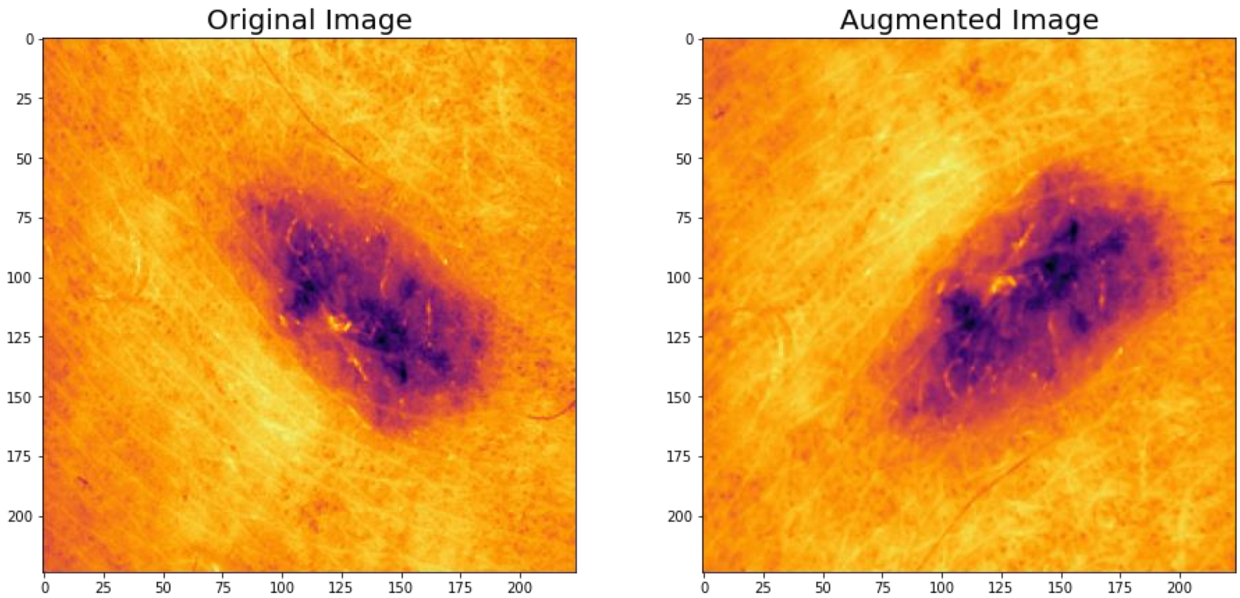

Figure 04: Image Data Augmentation Layer.

This study implemented three distinct models to evaluate their efficacy in skin cancer classification: a baseline Convolutional Neural Network (CNN), EfficientNetV2-B0, and Vision Transformer (ViT-B16). The baseline CNN was custom-designed with a sequential architecture comprising four primary layers. The initial layer consisted of 16 convolutional filters with a 3×3 kernel size and ReLU activation, followed by a 2×2 max-pooling layer to reduce spatial dimensions. The second convolutional block employed 8 filters, also with a 3×3 kernel and ReLU activation. Subsequently, the feature maps were flattened and passed through a dense layer with 128 units and ReLU activation. To mitigate overfitting, a dropout layer with a rate of 20% was incorporated.

Finally, the output layer consisted of two units with sigmoid activation for binary classification. Kernel weights were initialized using Glorot normal initialization to promote stable training dynamics. For the transfer learning models, EfficientNetV2-B0 and ViT-B16 were leveraged. EfficientNetV2-B0, pretrained on the ImageNet-21k dataset, was chosen for its balance of accuracy and computational efficiency. Only the custom classification head, comprising 164,000 parameters, was trainable, while the base model's weights were frozen to retain pre-learned features. ViT-B16, pretrained with 16×16 patches and 768-dimensional embeddings, utilized a custom head with 98,700 trainable parameters. Input images for ViT-B16 were resized to 224×224×3 to match the model's requirements. These transfer learning models aimed to capitalize on pre-learned feature representations to enhance classification performance while minimizing the risk of overfitting associated with limited training data.
All models were trained using the Adam optimizer with a learning rate of 0.001, paired with binary cross-entropy loss to handle the binary classification task. To ensure robust training and prevent overfitting, specific callbacks were employed. Early stopping was implemented with a patience of 3 epochs, meaning training would halt if the validation loss failed to improve for three consecutive epochs. Additionally, the Reduce LR On Plateau callback was utilized to reduce the learning rate by a factor of 0.1 if the validation loss plateaued for two epochs, facilitating finer convergence. The baseline CNN was trained for 30 epochs but stopped early at epoch 8 due to the validation loss plateauing. EfficientNetV2-B0 followed a similar training regimen, stopping at epoch 9 after the validation loss stabilized at 0.27, indicating convergence. ViT-B16, on the other hand, underwent full training for 30 epochs, achieving the highest accuracy among the models. These training configurations were carefully designed to balance model performance and computational efficiency while ensuring generalization to unseen data.
A comprehensive set of metrics was employed to evaluate model performance and ensure a thorough understanding of their strengths and weaknesses. Confusion matrices were generated to visualize the number of correct and incorrect predictions, providing insights into precision, recall, and accuracy. ROC-AUC curves were plotted to assess the models' ability to distinguish between classes across various classification thresholds, with the area under the curve (AUC) serving as a quantitative measure of performance. Matthews Correlation Coefficient (MCC) was calculated to address the challenge of class imbalance, offering a balanced measure of prediction quality. Furthermore, inference time was measured over five runs using Python's time per counter to evaluate the models' computational efficiency, a critical factor for real-world clinical applications. These metrics collectively provided a holistic view of model performance, enabling direct comparison and identification of optimal models for deployment.
The overall workflow of this study followed a structured and reproducible process, leveraging the capabilities of TensorFlow and its ecosystem. The pipeline commenced with data inspection and preprocessing, ensuring data quality and suitability for model training. Subsequently, three models—a baseline CNN and two transfer learning models (EfficientNetV2-B0 and ViT-B16)—were developed and trained using the optimized input pipeline. Model performance was rigorously evaluated using a suite of metrics, including confusion matrices, ROC-AUC curves, MCC, and inference time. This end-to-end methodology emphasized clarity, replicability, and scalability, providing a robust framework for skin cancer classification research.

Results and Discussion:

The study evaluated three models: a baseline Convolutional Neural Network (CNN), EfficientNetV2-B0, and Vision Transformer (ViT-B16). The results demonstrated significant differences in their performance metrics. The CNN achieved an accuracy of 74.70%, with precision, recall, and F1-score all around 74–78%. However, its

Matthews Correlation Coefficient (MCC) was relatively low at 0.5306, indicating limited robustness. In contrast, EfficientNetV2 significantly outperformed the CNN, achieving an accuracy of 86.67%, precision and recall of 86.67%, and an F1-score of 86.64%. Its MCC of 0.7307 reflected strong generalization capabilities. ViT-B16 delivered the highest accuracy at 88.48%, with precision, recall, and F1-score all above 88%. Its MCC of 0.7675 was the highest among the three models, highlighting its superior classification performance. However, ViT-B16's inference time was notably slower at 0.00852 seconds per sample, compared to EfficientNetV2's 0.00190 seconds and the CNN's 0.00184 seconds.

Table 1: Model Performance Comparison.

Metric	CNN	EfficientNetV2	ViT-B16
Accuracy	74.70%	86.67%	88.48%
Precision	78.11%	86.67%	88.49%
Recall	74.70%	86.67%	88.48%
F1-score	74.42%	86.64%	88.47%
MCC	0.5306	0.7307	0.7675
AUC	0.82	0.93	0.94
Inference (s/sample)	0.00184	0.00190	0.00852

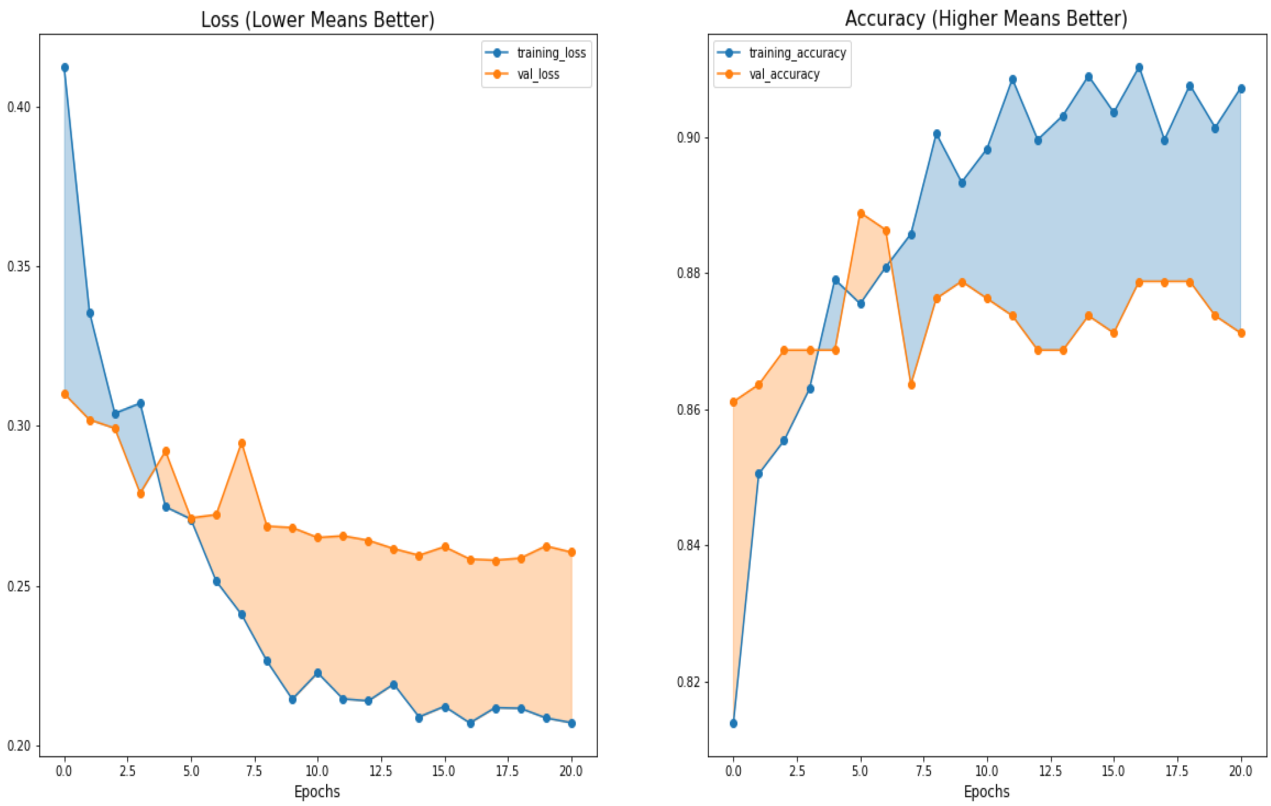

Figure 05: Plot of VIT B16 model training history.

The confusion matrices provided deeper insights into the models' classification behavior. The CNN exhibited a 10% false negative rate for malignant cases, meaning it misclassified 10% of malignant lesions as benign. This limitation raised concerns about its clinical reliability, as missing malignant cases could have severe consequences. EfficientNetV2 and ViT-B16 showed more balanced performance, with both models achieving precision and recall above 85% for both classes. Notably, ViT-B16 reduced malignant false negatives to just 14%, significantly enhancing diagnostic trustworthiness.

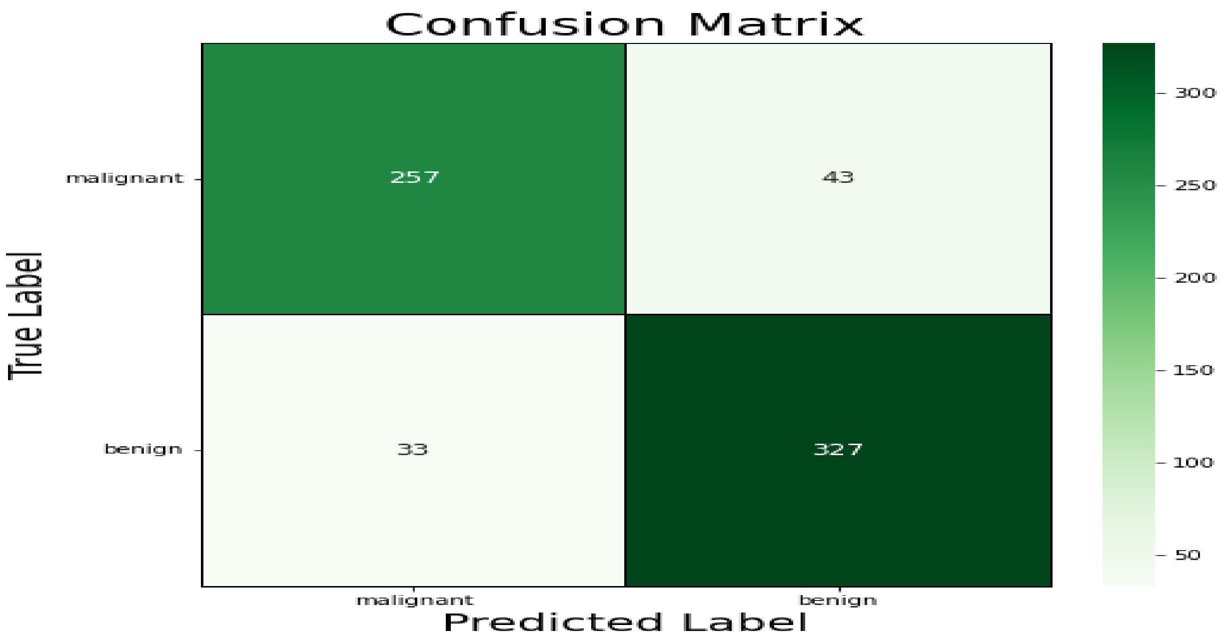

Figure 06: Plot of VIT B16 Confusion Matrix.

The trade-off between accuracy and inference speed was evident. EfficientNetV2 emerged as the optimal choice for real-world applications, balancing high accuracy (86.67%) with rapid inference (0.00190 seconds per sample). This speed allowed it to process the entire test set of 660 images in just over 1 second, making it suitable for time-sensitive clinical settings. ViT-B16, while achieving the highest accuracy and MCC, had an inference time of 0.00852 seconds per sample, making it 4.5 times slower than EfficientNetV2. This slower speed could limit its use in resource-constrained environments, despite its superior classification performance.

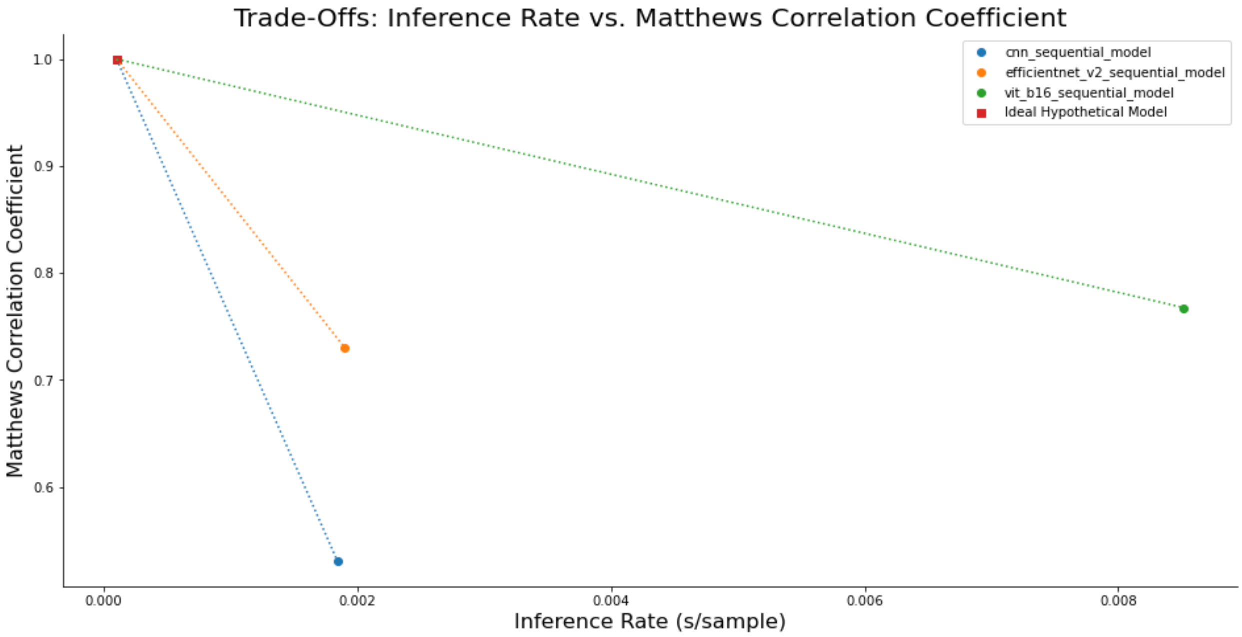

Figure 07: Plot of Trade-Offs: Inference Rate vs. Matthews Correlation Coefficient.

Receiver Operating Characteristic (ROC) curves further validated the models' performance. EfficientNetV2 and ViT-B16 both achieved an Area Under the Curve (AUC) of over 0.93, indicating excellent discriminative ability. The CNN, however, had an AUC of 0.82, suggesting less reliable classification. Visual analysis of misclassifications revealed that errors often occurred in low-contrast lesions or atypical benign cases, highlighting the need for targeted data augmentation in future work.

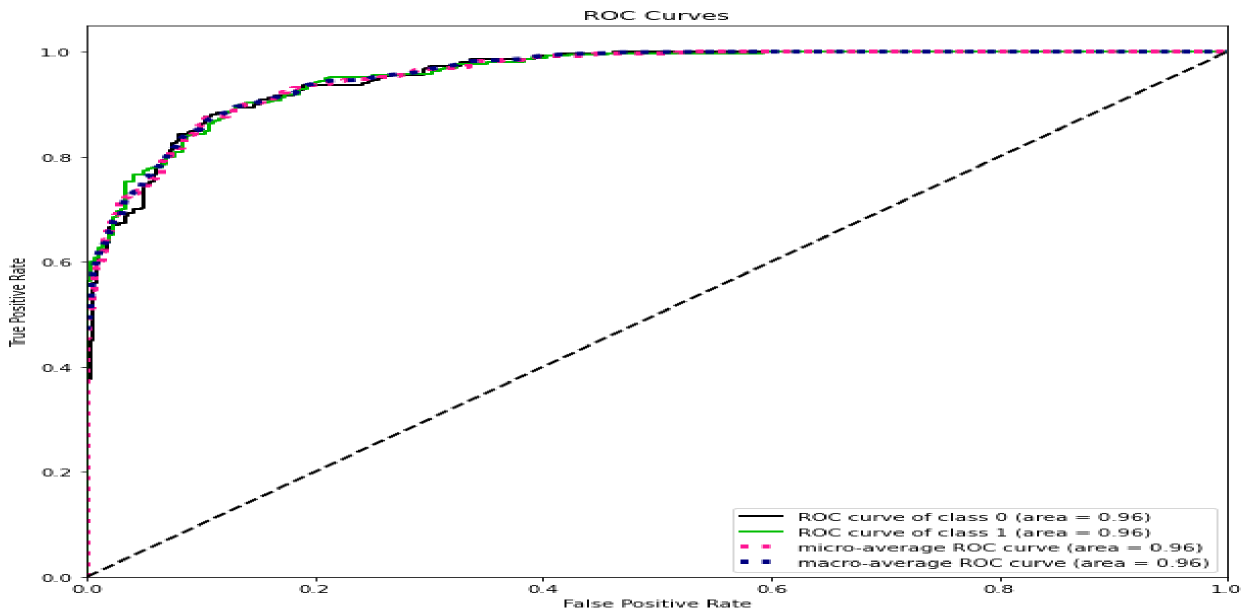

Figure 08: Plot of VIT B16 ROC Curves.

The results underscored the superiority of transfer learning models over the baseline CNN. EfficientNetV2 and ViT-B16 outperformed the CNN by a significant margin in accuracy and robustness, demonstrating the value of pretrained feature extraction in mitigating overfitting and improving generalization. EfficientNetV2's strong performance, combined with its rapid inference time, positioned it as an ideal choice for clinical deployment, where speed and accuracy are critical. ViT-B16, while achieving the highest accuracy and MCC, faced practical limitations due to its slower inference speed, making it better suited for applications where diagnostic accuracy outweighs speed.

In clinical settings, EfficientNetV2's speed and accuracy could support real-time diagnostics, enabling rapid decision-making. ViT-B16's high accuracy could be leveraged in specialized clinics or research environments where computational resources are abundant, and diagnostic precision is paramount. The CNN's higher false negative rate for malignant cases limited its clinical utility, emphasizing the need for more advanced models in high-stakes scenarios.

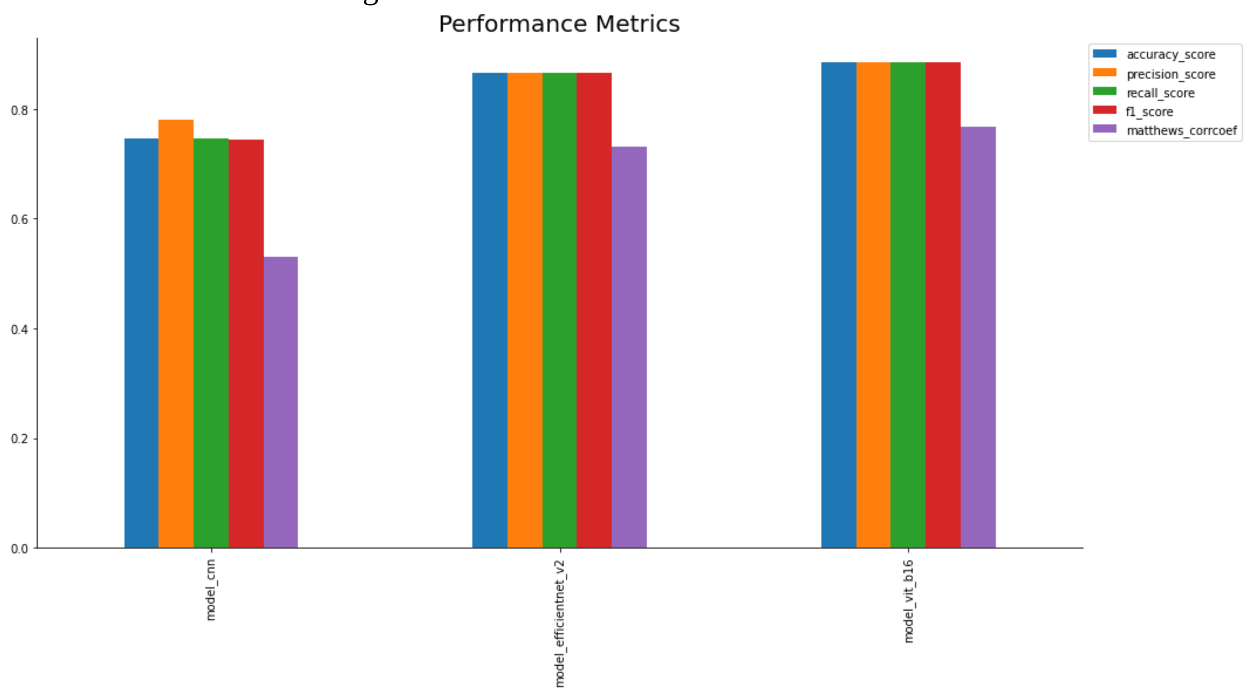

Figure 09: Plot of overall Performance Metrics.

The study's dataset, while comprehensive, had limitations. The ISIC 2020 dataset contained 2,637 training images, which may not fully represent the diversity of skin lesions encountered in real-world populations. Additionally, the dataset's class distribution, though balanced, might not reflect the prevalence of melanoma in broader clinical practice. These factors could affect the models' generalizability.

Future research could explore larger datasets, such as the ISIC Archive or Dermofit, to enhance model robustness and generalizability. Hybrid architectures combining the efficiency of CNNs with the accuracy of transformers could also be investigated to achieve optimal performance. Additionally, optimizing ViT-B16 for faster inference, through techniques like model pruning or quantization, could improve its applicability in real-time settings.

Conclusion:

This study systematically evaluated the performance of three deep learning models—Convolutional Neural Network (CNN), EfficientNetV2-B0, and Vision Transformer (ViT-B16)—for skin cancer classification using the ISIC 2020 dataset. The research aimed to identify the most effective model for clinical applications, balancing diagnostic accuracy, robustness, and inference speed.

The results demonstrated that transfer learning models significantly outperformed the baseline CNN. EfficientNetV2 and ViT-B16 achieved 12–14% higher accuracy than the CNN, with EfficientNetV2 attaining an optimal balance of accuracy (86.67%) and rapid inference (0.00190 s/sample). ViT-B16 delivered the highest accuracy (88.48%) and Matthews Correlation Coefficient (MCC=0.7675), though its slower inference time (0.00852 s/sample) limited its suitability for time-sensitive settings. Critical observations included the CNN's high false negative rate for malignant cases and ViT-B16's reduced misclassifications in low-contrast lesions.

The findings have practical implications for clinical practice. EfficientNetV2's speed and accuracy make it ideal for real-time diagnostics in resource-constrained environments, while ViT-B16's superior accuracy could enhance diagnostic reliability in specialized clinics or research settings. These results underscore the potential of AI-driven tools to improve dermatological care.

However, the study's limitations must be acknowledged. The dataset size (2,637 training images) and potential bias in lesion diversity may affect generalizability. Future work should focus on testing models on larger, more diverse datasets (e.g., ISIC Archive) and exploring hybrid architectures to optimize the accuracy-speed trade-off. Additionally, optimizing ViT-B16 for faster inference could broaden its applicability.

In conclusion, this study highlights the value of transfer learning in medical imaging, particularly for skin cancer detection. The choice between models should be context-aware, prioritizing speed for real-time applications and accuracy for high-stakes diagnostics.

References:

American Cancer Society. (2022). Skin Cancer Statistics.

National Cancer Institute. (2021). SEER Melanoma Statistics.

World Health Organization. (2020). UV Radiation and Skin Cancer.

Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.

Skin Cancer Foundation. (2020). Melanoma Basics.

Codella, N., et al. (2017). Deep Learning in Dermatology. *Journal of Investigative Dermatology, 137*(11), 2226–2230.

Gershenwald, J. E., et al. (2017). Melanoma staging: Evidence-based changes in the AJCC eighth edition cancer staging manual. *CA: A Cancer Journal for Clinicians*, *67*(6), 472–492.

Gutman, D., et al. (2016). Challenges in skin lesion analysis. *IEEE Journal of Biomedical and Health Informatics, 20*(5), 1399–1408.
Sandler, M., et al. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of CVPR*, 4510–4520.
Rashid, J., et al. (2022). Skin Cancer Disease Detection Using Transfer Learning Technique. *Applied Sciences, 12*(11), 5714.
World Health Organization. (2020). Cancer Today: Melanoma.
Khan, M. Q., et al. (2019). Classification of melanoma and nevi in digital images for diagnosis of skin cancer. *IEEE Access, 7*, 90132–90144.
Filali, Y., et al. (2020). Efficient fusion of handcrafted and pre-trained CNNs features to classify melanoma skin cancer. *Multimedia Tools and Applications, 79*(10), 31219–31238.
Gutman, D., et al. (2016). Challenges in skin lesion analysis. *IEEE Journal of Biomedical and Health Informatics, 20*(5), 1399–1408.
Kourou, K., et al. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal, 13*, 8–17.
Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.
Xie, F., et al. (2018). Melanoma classification on dermoscopy images using a neural network ensemble model. *IEEE Transactions on Medical Imaging, 36*(6), 849–858.
Haenssle, H. A., et al. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition. *Annals of Oncology, 29*(6), 1836–1842.
Sandler, M., et al. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of CVPR*, 4510–4520.
Codella, N., et al. (2019). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016. *IEEE Journal of Biomedical and Health Informatics, 23*(4), 1415–1421.
ISIC Archive. (2020). ISIC 2020 Challenge Dataset.
He, K., et al. (2016). Deep residual learning for image recognition. *Proceedings of CVPR*, 770–778.
Hosny, K. M., et al. (2019). Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks. *Multimedia Tools and Applications, 78*(24), 31229–31255.
Cakmak, M., & Tenekeci, M. E. (2021). Melanoma detection from dermoscopy images using Nasnet Mobile with transfer learning. *Applied Sciences, 11*(12), 5568.
Hosny, K. M., et al. (2019). Classification of skin lesions using transfer learning and augmentation with Alex-net. *PLoS ONE, 14*(4), e0217293.
Rashid, J., et al. (2022). Skin Cancer Disease Detection Using Transfer Learning Technique. *Applied Sciences, 12*(11), 5714.
Mukherjee, S., et al. (2019). Malignant melanoma classification using cross-platform dataset with deep learning CNN architecture. *Applied Sciences, 9*(22), 4748.
Dalila, F., et al. (2017). Segmentation and classification of melanoma and benign skin lesions. *Optik, 140*, 749–761.
Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of ICLR*.
Tan, M., et al. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of CVPR*, 1024–1035.
Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*(1), 6–13.

Brinker, T. J., et al. (2019). Deep neural networks are superior to dermatologists in melanoma image classification. *European Journal of Cancer, 119*, 11–17.

Han, S. S., et al. (2018). Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology, 138*(10), 1529–1538.

ISIC Archive. (2020). ISIC 2020 Challenge Dataset Statistics.

Sandler, M., et al. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of CVPR*, 4510–4520.

Liu, Y., et al. (2020). Federated learning: Challenges, methods, and future directions. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 14*(3), 1–118.

McMahan, B., et al. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of AISTATS*, 127–139.