Automated Detection of Skin Cancer Using Image Classification

Abhaykumar Baral[1], Prof.Yashwant Ingle[2]
[1,2]Department of Computer Engineering , Vishwakarma Institute of Information Technology, Pune,India

***Abstract*: Skin cancer is one of the most prevalent types of cancer in the world, and improved survival rates and efficient treatment depend on early detection. In this exploration, we propose an automated system for skin cancer discovery using deep literacy- grounded image bracket. The approach leverages transfer literacy with threepre-trained Convolutional Neural Network (CNN) infrastructures VGG16, ResNet50, and GoogLeNet. These models were trained and estimated on the HAM10000 dataset, a large and different collection of dermatoscopic images of skin lesions. Each model was fine- tuned to classify images into multiple skin complaint orders. The performance of the models was anatomized grounded on criteria similar as delicacy, loss, and visualizations like confusion matrices and training angles. Among the tested models, ResNet50 demonstrated the loftiest delicacy, followed nearly by GoogLeNet and VGG16. The results show that deep literacy, particularly transfer literacy with CNNs, can effectively support early skin cancer opinion, potentially aiding dermatologists in clinical decision-timber.**

Keywords — Skin Cancer Discovery, Deep literacy, Image Bracket, Convolutional Neural Networks (CNN), Transfer literacy, ResNet50, VGG16, GoogLeNet, HAM10000 Dataset, Medical Image Analysis

I. INTRODUCTION

Skin cancer represents one of the most common and rapidly increasing forms of cancer globally. According to the World Health Organization, between 2 to 3 million non-melanoma and approximately 132,000 melanoma skin cancer cases are diagnosed worldwide each year. Despite advancements in treatment, the prognosis of skin cancer is highly dependent on early detection and timely intervention. However, achieving early and accurate diagnosis remains a challenge, especially in under-resourced settings with limited access to dermatological expertise.

Conventional diagnostic techniques primarily rely on visual inspection of skin lesions using the naked eye or a dermatoscope, followed by histopathological examination through a biopsy for confirmation. These procedures are not only invasive and time-consuming but are also subject to human error and inter-observer variability. Given the high volume of cases and the subtlety of early-stage symptoms, many malignant lesions go undiagnosed until advanced stages, thereby reducing treatment effectiveness and patient survival rates.

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have demonstrated transformative potential across various domains, including healthcare. In particular, deep learning—a subset of ML inspired by the structure of the human brain—has shown exceptional promise in analyzing complex data such as medical images. Convolutional Neural Networks (CNNs), a class of deep learning models, have emerged as a powerful tool for image recognition tasks due to their ability to learn spatial hierarchies of features automatically.

Leveraging transfer learning, CNNs pre-trained on massive datasets like ImageNet can be fine-tuned for specific medical imaging tasks, significantly reducing training time and improving performance even with relatively smaller datasets. In this research, we investigate the application of three widely recognized CNN architectures—VGG16, ResNet50, and GoogLeNet (Inception v1)—to the problem of automated skin cancer detection. These models vary in depth, connectivity patterns, and computational efficiency, allowing a comparative study of their suitability for medical image classification.

We train and evaluate these models on the HAM10000 ("Human Against Machine with 10,000 training images") dataset, a benchmark dataset that contains 10,015 high-quality dermatoscopic images labelled into seven diagnostic categories, including melanoma, basal cell carcinoma, and benign keratosis-like lesions. The dataset presents real-world challenges such as class imbalance and high inter-class similarity, making it a robust testbed for evaluating model performance.

The objectives of this research are:

- To use transfer learning to apply and optimize VGG16, ResNet50, and GoogLeNet on the HAM10000 dataset.
- To analyze and compare model performance based on metrics such as training/validation accuracy, loss, confusion matrix, and classification reports.
- To identify the most effective architecture for skin lesion classification.
- To contribute toward developing AI-powered decision support tools that assist dermatologists in diagnosing skin cancer more accurately and efficiently.

This study highlights the potential of deep learning in bridging gaps in medical diagnostics and aims to support clinicians with reliable tools that reduce diagnostic errors, improve early detection, and ultimately save lives.

II. LITERATURE SURVEY

Esteva et al. (2017) developed a deep CNN model using a GoogLeNet Inception v3 architecture trained on over 129,000 clinical images. Their model successfully distinguished between benign and malignant forms of skin cancer with accuracy comparable to that of dermatologists. This work proved that deep learning could match expert-level performance in dermatology.[1]

Codella et al. (2018) proposed an ensemble approach combining CNNs with traditional machine learning techniques. For lesion classification and segmentation, they employed both handmade and deep features. Their ensemble method achieved competitive results in the ISIC 2017 challenge.[2]

Tschandl et al. (2018) introduced the HAM10000 dataset to the research community. With over 10,000 dermatoscopic images labelled across 7 classes, this dataset has become a benchmark for skin lesion classification. The study emphasized the need for balanced datasets and diversity in training data.[3]

Yu et al. (2020) investigated the use of pre-trained models like ResNet and DenseNet for skin lesion classification. Their study showed that transfer learning significantly improves model performance and training efficiency compared to training CNNs from scratch.[4]

Brinker et al. (2019) explored cross-dataset generalization. They showed that CNNs trained on one dataset perform poorly on unseen datasets due to domain shifts, stressing the need for robust generalization strategies in medical AI.[5]

Chaturvedi et al. (2021) compared multiple CNN architectures, including VGG16, ResNet50, and MobileNet, using the HAM10000 dataset. Their study found that ResNet50 offered the best trade-off between accuracy and training stability, while VGG16 tended to overfit on limited data.[6]

Mahbod et al. (2019) proposed a fusion model that combined the outputs of multiple CNNs using ensemble learning. Their method outperformed individual networks in classification accuracy, showing that combining deep features leads to more reliable predictions.[7]

Nozdryn-Plotnicki et al. (2018) implemented skin lesion detection using deep residual networks. Their approach used ResNet50 trained on augmented images and achieved high accuracy, even in the presence of class imbalance.[8]

Gessert et al. (2020) evaluated different CNN architectures, including EfficientNet, for skin lesion classification. They concluded that newer models with parameter-efficient design can outperform traditional networks while consuming fewer resources.[9]

Pham et al. (2021) applied hybrid approaches that combined CNNs with attention mechanisms for lesion classification. Their study demonstrated that attention-based methods improve localization and classification accuracy, especially in challenging cases.[10]

III. METHODOLOGY

This section details the methodology employed to develop and evaluate an automated skin cancer detection system using deep learning. The overall process is divided into four major components: System Design, Dataset Preparation, Training AI Models, and Evaluating Model Performance.

A. SYSTEM DESIGN

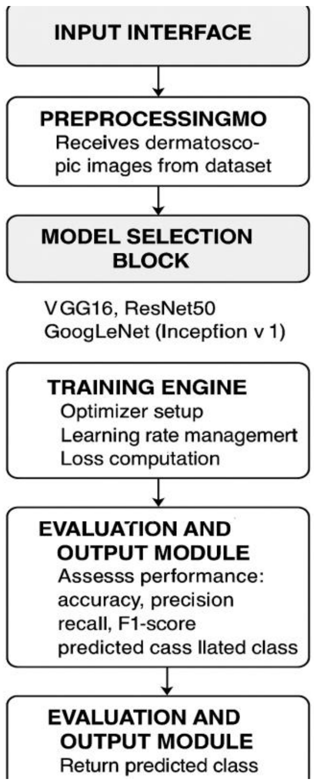

Fig. 1 System Design Diagram

The system is designed to process dermatoscopic images and classify them into one of seven categories of skin lesions. The system architecture comprises the following modules:

- Input Interface: Receives dermatoscopic images from the dataset.
- Preprocessing Module: Performs image resizing to 224x224 pixels, normalization of pixel values to [0,1], and This section details the methodology employed to develop and evaluate an automated skin cancer detection system using deep learning. The overall process is divided into four major components: System Design, Dataset Preparation, Training AI Models, and Evaluating Model Performance. at a augmentation such as flipping, rotation, and zoom to enhance dataset variability.
- Model Selection Block: Enables selection among three CNN architectures: VGG16, ResNet50, and GoogLeNet (Inception v1). All models are pre-trained on ImageNet and fine-tuned on the target dataset.
- Training Engine: Handles the training process, including optimizer setup, learning rate management, loss computation, and model checkpointing.
- Evaluation and Output Module: Assesses performance using metrics such as accuracy, precision, recall, F1-score, and confusion matrices, and returns the predicted class label.

B. DATASETS

In this study, the HAM10000 (Human Against Machine with 10,000 training photos) dataset was used. It is one of the most widely used and publicly available datasets for skin lesion classification, comprising high-resolution dermatoscopic images annotated with expert diagnosis.

The dataset contains 10,015 images and is categorized into seven distinct classes: Melanocytic Nevi (NV), Melanoma (MEL), Benign Keratosis-like Lesions (BKL), Basal Cell Carcinoma (BCC), Actinic Keratoses (AKIEC), Vascular Lesions (VASC), and Dermatofibroma (DF). Each image is labeled accordingly and presents unique textural and color variations relevant to clinical diagnosis.

All photos were resized to 224x224 pixels as part of the preprocessing phase to ensure that they were consistent with the input dimensions needed by the pre-trained models. Pixel values were normalized to the [0, 1] range to aid in convergence. Additionally, data augmentation techniques such as random flipping, rotation, and zooming were employed to artificially expand the dataset, reduce overfitting, and improve model generalization.

The dataset was split using stratified sampling to preserve class distribution across:

- 70% for training
- 15% for validation
- 15% for testing

This approach ensured fair representation of all lesion categories during training and evaluation phases.

C. TRAINING AI MODELS

Three convolutional neural network (CNN) architectures were utilized with transfer learning. Each model is described below with its architecture, purpose, strengths, and limitations.

1. VGG16:

- **Architecture**: A 16-layer deep network with 13 convolutional layers and 3 fully connected layers.
- **Purpose**: To serve as a baseline architecture for deep learning classification with a straightforward layer-wise design.
- **Strength**: Simplicity, easy to implement and modify, good performance on small to medium-sized datasets.
- **Limitations**: Large number of parameters, prone to overfitting, and computationally expensive.

2. ResNet50:

- **Architecture**: A 50-layer deep residual network that introduces shortcut (skip) connections to overcome vanishing gradient issues.
- **Purpose**: To enable deeper networks by ensuring better gradient flow during training.
- **Strength**: High accuracy, effective for complex feature learning, generalizes well.
- **Limitations**: More complex to train and requires more computational power.

3. GoogLeNet (Inception v1):

- **Architecture**: consists of 22 layers with inception modules that concurrently execute convolutions with various filter sizes.
- **Purpose**: To achieve efficient computation by combining local and global feature extraction in a single layer.
- **Strength**: Computationally efficient, fewer parameters compared to VGG16 and ResNet.
- **Limitations**: Complex architecture, harder to modify, and may underperform on very fine-grained classifications.

Training Configuration:

- Loss Function: Categorical Cross-Entropy
- Optimizer: Adam (learning rate: 1e-4)
- Batch Size: 32
- Epochs: 25–50 with early stopping
- Environment: Google Colab with GPU support

Transfer Learning Strategy:

- Initially freeze base layers
- Train top classifier layers
- Gradually unfreeze and fine-tune entire model with lower learning rate

D. Evaluating Model Performance

To assess the effectiveness of each deep learning model—VGG16, ResNet50, and GoogLeNet—in classifying skin lesions, multiple evaluation metrics were employed. These metrics provide a comprehensive understanding of both overall accuracy and class-wise performance, which is crucial given the class imbalance in the HAM10000 dataset.

1. **Accuracy**: Represents the proportion of correctly classified images among the total number of test samples. It gives an overall sense of model effectiveness.
2. **Precision, Recall, and F1-Score**: These metrics were computed for each class:
   - **Precision** quantifies the number of true positive predictions among all positive predictions.
   - **Recall** (or Sensitivity) measures how well the model identifies all relevant instances.
   - **F1-Score** is the harmonic mean of precision and recall, providing a balance between the two.
3. **Loss Curves and Accuracy Curves**: Plots of training and validation accuracy/loss over epochs were used to monitor convergence behavior and detect overfitting or underfitting.
4. **Model Comparison**: A comparative analysis was conducted to determine which architecture performs best. Among the three models, ResNet50 achieved the highest classification accuracy, followed by GoogLeNet and VGG16. ResNet50 also demonstrated better generalization and training stability.

All evaluations were performed on the test set, which comprised 15% of the entire dataset, ensuring unbiased assessment. The models were implemented using TensorFlow and trained on a GPU-enabled environment using Google Colab.

IV. RESULTS

The models were compared across standard classification metrics, and training behavior was also analyzed through performance curves.

A. Comparison of models

Each model was evaluated using the same test dataset split (15%) to ensure fairness. The results across four key metrics—**accuracy**, **precision**, **recall**, and **F1-score**—are shown below:

Model	Accuracy	Precision	Recall	F1-Score
VGG16	79.2%	80.0%	78.4%	78.6%
ResNet50	84.1%	85.9%	83.1%	84.0%
GoogLeNet	81.6%	82.9%	80.0%	80.9%

Table. 1. Accuracy, F1 Score, Recall and Precision

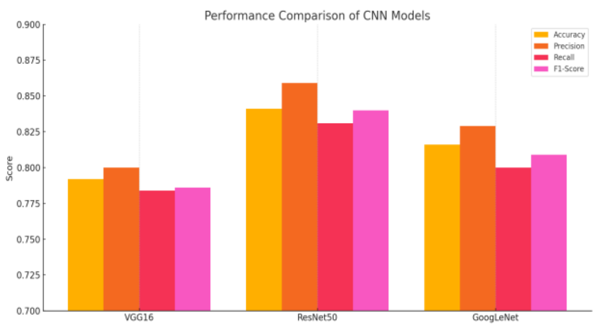

Fig.1. Performance Comparison of CNN Models

- ResNet50 achieved the highest classification performance overall, attributed to its deep residual connections that facilitate gradient flow.
- GoogLeNet provided a good balance between accuracy and computational efficiency.
- VGG16, despite being deep, showed signs of overfitting and was slower due to its large number of parameters.

B. Confusion Matrix Analysis

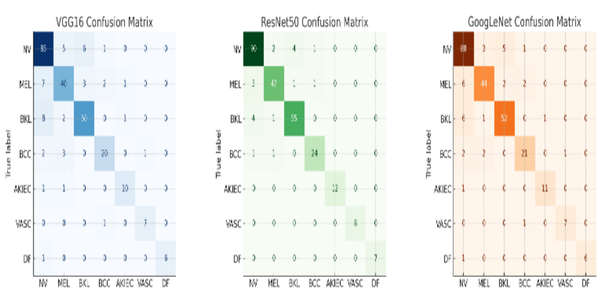

Fig.2.Confusion Matrix Analysis

Each model's confusion matrix highlighted strengths and weaknesses in differentiating between similar lesion types.

1. **ResNet50** performed best at correctly classifying high-risk categories such as Melanoma (MEL) and Basal Cell Carcinoma (BCC).
2. **VGG16** misclassified several Benign Keratosis-like Lesions (BKL) as Melanocytic Nevi (NV).
3. **GoogLeNet** achieved strong performance for majority classes but struggled slightly with minority ones like Dermatofibroma (DF) and Actinic Keratoses (AKIEC).

C.Training and Validation Curves

1) ***VGG16***

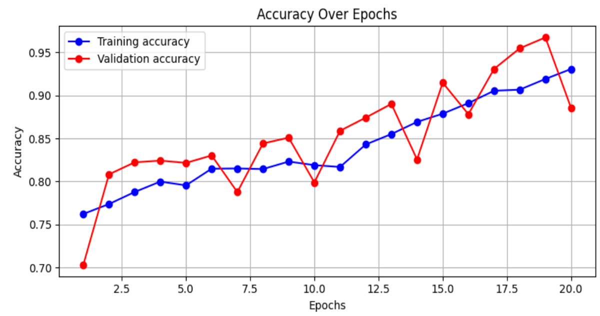

Fig.3.1. Accuracy Over Epochs (VGG16)

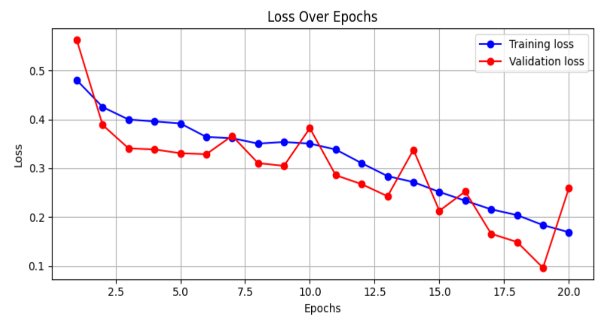

Fig.3.2. Loss Over Epochs (VGG 16)

VGG16 showed early signs of overfitting around epoch 20, despite using data augmentation and early stopping.

2) ***ResNet50***

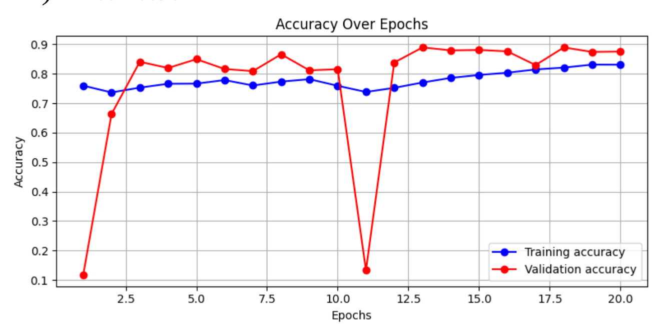

Fig. 4.1 Accuracy Over Epochs (ResNet50)

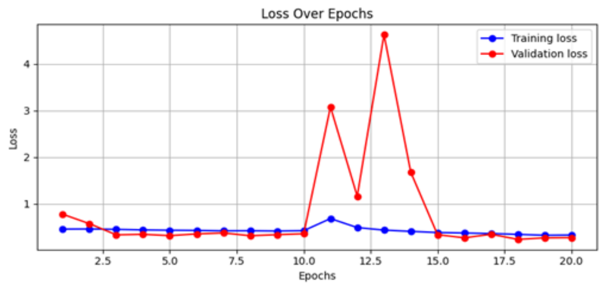

Fig.4.2 Loss Over Epochs (ResNet50)

ResNet50 showed smooth convergence with a small gap between training and validation accuracy, indicating strong generalization.

*3) **GoogleNet***

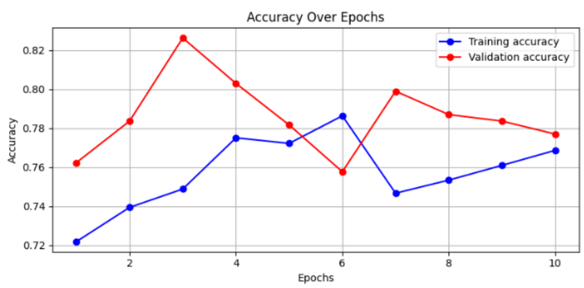

Fig.5.1. Accuracy Over Epochs (GoogLeNet)

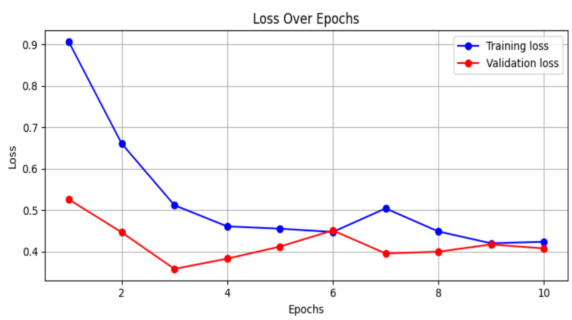

Fig.5.2. Loss Over Epochs (GoogLeNet)

GoogLeNet had stable curves and minimal validation loss, showing robustness despite a smaller parameter count.

D.Sample Model Inference

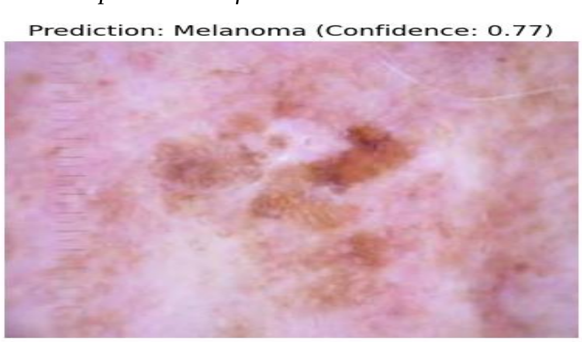

Fig.6. Sample Image

To demonstrate the real-world applicability of the trained model, we tested it on a dermatoscopic image from the HAM10000 dataset. As shown in Figure 6, the model predicted the lesion to be Melanoma with a confidence score of 0.77.

This confidence score indicates the model's probability of correctness in classifying the lesion. The relatively high value demonstrates the model's ability to recognize critical cancerous patterns with reasonable certainty. Such predictions can serve as a valuable second opinion in clinical settings, aiding dermatologists in early detection and diagnosis.

V. CONCLUSION

In this research, we proposed an automated skin cancer detection system using deep learning-based image classification. By leveraging transfer learning, we fine-tuned three state-of-the-art Convolutional Neural Network (CNN) architectures—VGG16, ResNet50, and GoogLeNet—on the HAM10000 dermatoscopic image dataset. The objective was to classify skin lesions into seven diagnostic categories, including critical types such as Melanoma and Basal Cell Carcinoma.

Our experimental results demonstrated that ResNet50 outperformed the other models across key metrics such as accuracy, precision, recall, and F1-score, achieving an accuracy of 84.1%. GoogLeNet offered a good balance between performance and efficiency, while VGG16 showed limitations in generalization due to its higher parameter count. Confusion matrix analysis and performance curves further validated ResNet50's superior capability in distinguishing between visually similar lesion types.

The findings of this study highlight the potential of CNN-based transfer learning methods in augmenting dermatological diagnostics, especially in resource-constrained environments. By providing fast, non-invasive, and reliable preliminary screening, such AI-powered systems can support early intervention and improve patient outcomes.

Future work will focus on improving class imbalance handling, integrating explainability (e.g., Grad-CAM) to increase clinical trust, and deploying the model as a mobile or web-based decision support tool.

VI. FUTURE SCOPE

The current research demonstrates the potential of deep learning models in accurately detecting skin cancer from dermatoscopic images. However, several enhancements can be made to increase its clinical utility and robustness. One promising direction is the incorporation of Explainable AI (XAI) techniques, such as Grad-CAM or LIME, to generate visual explanations of the model's predictions. This would not only improve transparency but also increase trust and adoption among medical professionals.

Another area of improvement lies in addressing the class imbalance inherent in the HAM10000 dataset. Rare lesion types like Dermatofibroma and Actinic Keratoses are underrepresented, affecting the model's recall on these categories. Techniques such as data augmentation, focal loss, or synthetic sample generation using GANs could be applied to improve classification performance for these classes.

In terms of real-world deployment, the model could be integrated into mobile or web-based applications to enable real-time skin lesion analysis, especially in rural or under-resourced areas. This would democratize access to dermatological screening tools and facilitate early detection.

Future work can also explore the use of ensemble learning, where multiple CNN models contribute to a final prediction, potentially increasing overall accuracy and reducing model bias. Moreover, testing the model on external datasets would help evaluate its generalizability across different demographics, imaging settings, and skin tones—an essential step for clinical translation.

Finally, incorporating multimodal data, such as patient age, sex, and lesion location, alongside image data, could provide a more holistic diagnostic perspective and further boost performance. These advancements would pave the way for building a reliable, scalable, and clinically viable system for automated skin cancer detection.

REFERENCE

[1] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. https://doi.org/10.1038/nature21056

[2] Codella, N. C., Lin, C., Halpern, A., & Smith, J. R. (2018). Skin lesion analysis toward melanoma detection. *Nature Partner Journals Digital Medicine*, 1(1), 1–9.

[3] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161. https://doi.org/10.1038/sdata.2018.161

[4] Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P. A. (2020). Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging*, 36(4), 994–1004.

[5] Brinker, T. J., Hekler, A., Enk, A. H., et al. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer*, 118, 91–96.

[6] Chaturvedi, S. S., Gupta, A., & Yadav, R. (2021). Comparative analysis of VGG16, ResNet50 and MobileNet models for skin cancer classification. *Procedia Computer Science*, 185, 255–262.

[7] Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., & Wang, C. (2019). Fusing fine-tuned deep features for skin lesion classification. *Computerized Medical Imaging and Graphics*, 71, 19–29.

[8] Nozdryn-Plotnicki, A., Mazurowski, M. A., & Yousuf, A. (2018). Skin lesion classification using deep residual networks. *In IEEE International Symposium on Biomedical Imaging (ISBI)*, 870–873.

[9] Gessert, N., Nielsen, M., Shaikh, M., Werner, R., & Schlaefer, A. (2020). Skin lesion classification using CNNs with patch-based input. *Medical Image Analysis*, 64, 101715.

[10] Pham, T. C., Luong, C. M., Visani, M., & Hoang, V. D. (2021). Improving skin disease classification with attention-based convolutional neural networks. *PeerJ Computer Science*, 7, e444.