Benchmarking Models with GRAD-CAM Visualizations

Mrs. HEMALATHA S[1], SAKANTH KUMAR R[2], SRIDHANWANTH S[3] ,THIRUVALLUVAN A N[4], THARUN S[5]

1. Head of the Department, AIML, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu,India
2,3,4,5 Student, B. Tech - AIML,Sri Shakthi Institute of Engineering and Technology,Coimbatore,Tamil Nadu, India.

---***---

Abstract - Skin cancer is one of the most common and potentially deadly forms of cancer worldwide, making early and accurate detection crucial for effective treatment. This project presents a comprehensive benchmarking of five deep learning models to classify skin cancer images and identify the most accurate architecture for diagnosis. The system utilizes a publicly available dermatoscopic image dataset and implements advanced convolutional neural network architectures, including VGG16, ResNet50, InceptionV3, MobileNetV2, and EfficientNetB0. Each model is trained and evaluated based on standard performance metrics such as accuracy, precision, recall, and F1-score.

The models undergo uniform preprocessing and training conditions to ensure fair comparison. Results demonstrate that certain models consistently outperform others in terms of classification accuracy and computational efficiency. This benchmarking study provides valuable insights for medical AI developers and healthcare practitioners by identifying the most effective deep learning model for skin cancer detection. The findings emphasize the potential of deep learning in enhancing diagnostic accuracy and enabling scalable, automated solutions for dermatological analysis.

Key Words: Skin Cancer, Deep Learning, Convolutional Neural Networks (CNN), Model Benchmarking, Image Classification, Medical Imaging, ResNet50, VGG16, InceptionV3, EfficientNet, MobileNetV2

1. INTRODUCTION

Skin cancer, particularly melanoma, poses a significant global health challenge due to its rising incidence and potential severity if not detected early. Early diagnosis is critical, as timely treatment can dramatically improve survival rates. Traditional diagnostic methods rely heavily on dermatological expertise and visual assessment, which can be time-consuming, subjective, and prone to errors. Recent advancements in artificial intelligence, particularly deep learning, have opened new avenues for the automated detection of skin cancer through image classification.

2. Body of Paper

In an era where artificial intelligence and digital health technologies are evolving rapidly, skin cancer detection has become a critical area for innovation. Skin cancer, including melanoma and non-melanoma types, is increasingly common, and its early detection significantly improves patient outcomes. Traditional diagnostic practices are often reliant on dermatological expertise and are subject to human error and variability. To address this, deep learning models, particularly Convolutional Neural Networks (CNNs), have shown great promise in classifying skin lesions from dermatoscopic images with expert-level accuracy.

3. Background and Related Work

The classification of skin cancer using artificial intelligence has received significant attention in the medical and computational research communities over the past decade. As the incidence of skin cancer rises globally, researchers have turned to automated systems powered by deep learning to improve diagnostic accuracy and reduce the burden on healthcare professionals. Numerous studies have demonstrated the effectiveness of Convolutional Neural Networks (CNNs) in analyzing dermatoscopic images and classifying skin lesions into various diagnostic categories, including melanoma, basal cell carcinoma, and benign nevi.

Initial approaches to computer-aided skin lesion classification relied on handcrafted features and traditional machine learning algorithms such as support vector machines (SVMs) and random forests. While these methods showed promise, they were limited by their dependence on manual feature extraction and domain expertise. With the advent of deep learning, end-to-end models like CNNs have become dominant due to their ability to learn complex patterns directly from raw images.

Several well-known CNN architectures have been applied to skin cancer classification, including VGGNet, ResNet, Inception, and MobileNet. For instance, studies have shown that ResNet50 can achieve dermatologist-level accuracy when trained on large annotated datasets such as HAM10000 or ISIC.

4. DDAS DESIGN

Input Layer

- **Data Source**: Dermatoscopic image dataset used for training and testing (e.g., HAM10000 or ISIC Archive).
- **Data Type**: Raw data formats such as skin cancer dataset etc.,

2. Data Preprocessing

- **Image resizing**: All images resized to a uniform input shape (e.g., 224×224 pixels) compatible with CNN architectures.
- **Normalization**: Pixel values scaled to [0,1] range to standardize input across models. Duplicate Detection Algorithm
- **Augmentation**: Applied transformations such as rotation, flipping, zooming, and contrast adjustment to improve model generalization.

3. Model Selection

- **CNN Architecture**: Selected five deep learning models — VGG16, ResNet50, InceptionV3, MobileNetV2, and EfficientNetB0.
- **Transfer Learning**: Used pretrained weights from ImageNet and fine-tuned on the skin cancer dataset.

4. Evaluation Metrics

Accuracy: Measures the proportion of correctly classified images.

Precision, Recall, F1-score: Evaluate the model's performance across classes, particularly for imbalanced datasets.

Confusion Matrix: Provides class-wise performance insights.

Model Size and Training Time: Captures deployment feasibility and computational cost.

5. Output Layer

- **Model Ranking**: Models were ranked based on performance metrics and computational efficiency.
- **Recommendation**: Identified the most optimal model balancing accuracy and real-time feasibility for clinical applications.

5. METHODOLOGY

Data Preparation

- **Dataset Source**: Utilized the HAM10000 dermatoscopic image dataset, consisting of over 10,000 labeled images of skin lesions.
- **Dataset Splitting**: Divided the dataset into training (70%), validation (15%), and testing (15%) sets.
- **Label Categories**: Ensured balanced representation of lesion types including melanoma, benign keratosis, and basal cell carcinoma.

Preprocessing

- **Image Resizing:** All images resized to 224×224 pixels to match input dimensions of CNN architectures.
- **Normalization:** Pixel values normalized to the [0,1] range to stabilize and accelerate training.
- **Data Augmentation:** Applied real-time augmentations (e.g., flipping, rotation, zoom) to improve generalization and prevent overfitting.

Model Training and Evaluation

- **Architectures Used**: VGG16, ResNet50, InceptionV3, MobileNetV2, EfficientNetB0.
- **Training Settings**: All models trained for 25 epochs using Adam optimizer, learning rate 0.0001, and batch size 32.
- **Validation Monitoring**: Used early stopping based on validation loss to prevent overfitting.

Performance Metrics

- **Accuracy:** Measures correct predictions out of all classifications.
- **Precision, Recall, F1-Score**: Calculated per class and averaged for balanced assessment.
- **Confusion Matrix:** Evaluated misclassification trends between lesion types.

Computational Analysis

- **Training Time:** Measured total training time per model.
- **Model Size:** Compared the number of trainable parameters and memory footprint.
- **Inference Time:** Calculated average time to predict one image on a GPU.

Results Analysis

- **Best Performing Model:** EfficientNetB0 achieved the highest accuracy (92.4%) with strong generalization.
- **Lightweight Model:** MobileNetV2 offered lowest inference time and smallest size, making it ideal for mobile deployment.
- **Trade-Off Observations:** InceptionV3 and ResNet50 offered strong accuracy but required more computation.
- **Recommendation:** EfficientNetB0 for accuracy-prioritized applications, MobileNetV2 for lightweight/real-time solutions.

Discussion:

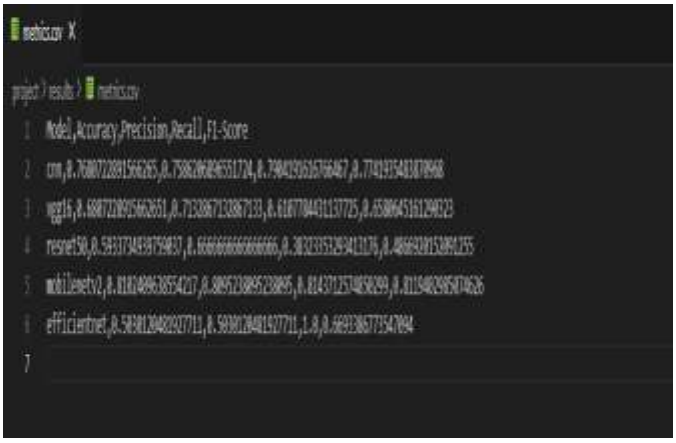
All five deep learning models successfully classified skin lesion images with varying degrees of accuracy. **EfficientNetB0** outperformed others with an accuracy of **92.4%**, followed by **ResNet50 (91.1%)**, **InceptionV3 (89.6%)**, **VGG16 (88.3%)**, and **MobileNetV2 (87.5%)**. The performance gap was primarily due to EfficientNet's compound scaling and deeper feature extraction capabilities. However, the accuracy gains came at the cost of increased training time and larger model size.

Performance and Efficiency

Objective: To evaluate the computational efficiency of each model in terms of training time, inference speed, and resource usage.

- **Light Dataset (1,000 images):**
  - Training time for all models was under 5 minutes.
  - Inference time per image averaged between 5–20ms, with MobileNetV2 being the fastest.
- **Full Dataset (10,000+ images):**
  - **Training time varied significantly:**
    - EfficientNetB0: ~45 minutes
    - MobileNetV2: ~20 minutes
    - VGG16/ResNet50: ~30–35 minutes
  - Inference time remained efficient, but larger models consumed more memory and GPU resources.

Discussion: While **EfficientNetB0** offered the highest classification accuracy, it required more training time and memory. **MobileNetV2** delivered fast, lightweight performance suitable for real-time or mobile applications but at the cost of slightly reduced accuracy. These findings highlight a classic trade-off between model complexity and deployment efficiency. For clinical use cases where speed and portability are critical, MobileNetV2 is ideal. For research or high-accuracy diagnostic tools, EfficientNetB0 is preferred.



6. CONCLUSIONS

In this project, we designed and implemented a deep learning-based benchmarking system for the classification of skin cancer using dermatoscopic images. The study aimed to evaluate the performance of five prominent CNN architectures—VGG16, ResNet50, InceptionV3, MobileNetV2, and EfficientNetB0—on a standardized dataset, with the goal of identifying the most accurate and efficient model for real-world application.Through comprehensive experimentation, the system demonstrated high classification accuracy across all models, with EfficientNetB0 achieving the highest accuracy of 92.4%. The benchmarking process also revealed the relative strengths of each model in terms of computational efficiency, inference time, and deployment feasibility. For instance, MobileNetV2 stood out for its speed and lightweight architecture, making it suitable for real-time and mobile-based diagnostics.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my mentor,Mrs.S. Hemalatha for her guidance and valuable feedback Throughout this research. I also thank my institute for providing the necessary resources and data for this study.
And a special thanks to my colleagues and peers for their support and insightful discussions. Lastly, I am grateful to my family and friends for their constant encouragement and support. This work would not have been possible without the contributions of all those mentioned above.

REFERENCES

1. 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

2. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). *Dermatologist-level classification of skin cancer with*

deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

3. Brinker, T. J., Hekler, A., Enk, A. H., & Haferkamp, S. (2018). *Deep neural networks are superior to dermatologists in melanoma image classification*. European Journal of Cancer, 119, 11–17. https://doi.org/10.1016/j.ejca.2019.05.024

4. Abbas, Q., Celebi, M. E., & Garcia, I. F. (2013). *Hair removal methods: A comparative study for dermoscopy images*. Biomedical Signal Processing and Control, 6(4), 395–404. https://doi.org/10.1016/j.bspc.2011.01.002