CNN Applications in Skin Disease Prediction and Detection

Joseph Richardson
School of Data Science & Analytics
Kennesaw State University
Kennesaw, GA, USA
jrich256@students.kennesaw.edu

Sharanya D V
School of Data Science & Analytics
Kennesaw State University
Kennesaw, GA, USA
sdv1@students.kennesaw.edu

Venkata Abhiram Chitty
School of Data Science & Analytics
Kennesaw State University
Kennesaw, GA, USA
vchitty@students.kennesaw.edu

***Abstract*—** Skin cancer remains one of the most prevalent and deadly forms of cancer globally, necessitating advancements in early detection and accurate diagnosis to improve patient outcomes. This report explores the application of Convolutional Neural Networks (CNNs) for skin disease prediction and detection, utilizing the HAM10000 dataset—a comprehensive collection of dermatoscopic images of various skin lesions. Our research aims to develop a sophisticated model that enhances identifying potentially malignant conditions through deep learning techniques. We use ResNet-50, VGG16, and Inception-ResNet-V2 architectures to use data augmentation techniques to rectify class imbalances and enhance model resilience. Every model is painstakingly adjusted and assessed for efficiency and classification accuracy; the results show 87% accuracy across various architectures. This study underscores the transformative potential of CNNs in dermatology, paving the way for automated diagnostic tools that can assist healthcare professionals in making informed decisions, ultimately leading to improved patient care in the fight against skin cancer.

Keywords: Skin Cancer, Convolutional Neural Networks (CNNs), Deep Learning, HAM10000 Dataset, Transfer Learning and Medical Imaging.

I. INTRODUCTION

One of the most common and potentially fatal types of cancer in the world is still skin cancer, and improving patient outcomes and survival rates requires early identification and precise diagnosis. The use of deep learning methods, especially Convolutional Neural Networks (CNNs), has demonstrated great potential in recent years for improving and automating the categorization of skin lesions. This research aims to use the HAM10000 dataset to create a sophisticated skin cancer detection model by utilizing CNNs. By harnessing the power of deep learning, we aim to create a robust tool that can assist dermatologists in identifying potentially malignant skin conditions early. The significance of this project lies in its potential to improve early detection rates, streamline diagnostic processes, and ultimately contribute to better patient outcomes in the fight against skin cancer.

Recent studies have demonstrated the effectiveness of CNNs in skin cancer classification. Through careful preprocessing, innovative model architecture design, and rigorous evaluation, we seek to address the challenges inherent in medical image classification, including class imbalance and image variability. Our approach will involve data augmentation, normalization, and resizing during preprocessing to enhance the model's performance. We will also explore CNN architectures, including InceptionV3, ResNet50, and others, to determine the most effective model for our specific task. This study lays the groundwork for investigating how CNNs may be used to classify skin lesions. This critical position has the potential to revolutionize dermatological practice and pave the way for more accurate, automated diagnostic support systems in the healthcare industry. By developing a reliable and automated approach to skin lesion classification, we want to assist physicians in making more precise diagnoses and assisting in the early detection of malignancies. This will eventually enhance patient outcomes in the worldwide battle against skin cancer.

II. Related Works

Deep learning has revolutionized the field of medical image classification, particularly in skin lesion analysis, by leveraging the power of CNNs and transfer learning. Recent advancements focus on balancing accuracy, efficiency, and accessibility, with innovative approaches tailored to specific medical challenges.

Compact models like MobileNet V2, optimized for mobile devices due to their low computational demand and precise feature extraction capabilities, have been successfully integrated with advanced architectures such as LSTMs [1]. This combination enables real-time diagnosis of skin diseases in resource-limited settings, demonstrating the potential for practical deployment in underserved regions. Similarly, CNNs have proven invaluable for tasks such as melanoma classification from dermoscopic images, significantly outperforming traditional classifiers like decision trees [2]. These studies underscore the importance of early and accurate detection in improving patient outcomes and highlight the potential of automated diagnostic tools in dermatology.

The introduction of MedNet, a pre-trained CNN specifically designed for medical imaging, has addressed the limitations of models pre-trained on natural images like ImageNet [3]. By training on millions of grayscale and color medical images, MedNet generalizes well across various tasks, including classification, segmentation, and disease diagnosis. This tailored approach exemplifies how domain-specific pretraining can enhance model performance in medical imaging.

Class imbalance, a persistent issue in medical datasets, poses a unique challenge to deep learning models [4]. Imbalanced datasets can lead to skewed evaluation metrics and misleading performance assessments. Research has shown that analyzing individual class performance and adopting data balancing strategies, such as oversampling or data augmentation, can mitigate these effects, ensuring more reliable and fair outcomes. These insights informed our decision to augment our dataset to address class imbalances effectively.

The importance of model selection and preprocessing has also been emphasized in recent studies [5]. Comparisons of architectures such as DenseNet-201, Inception-ResNet-V2, and Inception-V3 have demonstrated that careful hyperparameter tuning and data augmentation significantly enhance the classification of benign and malignant skin lesions. Simplified CNN architectures, as demonstrated in lung image classification tasks, also highlight the potential to optimize accuracy while avoiding overfitting, even with limited training data [6].

Transfer learning has emerged as a critical tool in medical imaging, offering improved accuracy and efficiency when datasets are small [7]. Specialized architectures like VGG-16 and VGG-19 have been successfully applied to tasks such as Alzheimer's disease classification, achieving high performance across metrics like accuracy, precision, and AUC [8]. Additionally, ensemble learning strategies, which combine models such as ResNet-50 and Inception V3, have demonstrated improved classification accuracy by leveraging complementary strengths. Ensemble approaches have proven particularly effective for challenging cases, such as distinguishing between visually similar skin lesions [9].

Innovative models like AIDDA, built on EfficientNet-B4, further highlight the growing sophistication of deep learning in dermatology [10]. With accuracy rates surpassing 95%, these models are pushing the boundaries of what is possible in automated diagnosis. However, challenges remain, including the interpretability of deep learning models and the need for more diverse datasets that capture the full spectrum of medical conditions and imaging modalities.

This body of work demonstrates the rapid advancements in medical image classification and provides a solid foundation for our research, which builds on these techniques to tackle the specific challenges of skin lesion classification. By leveraging insights from prior studies, particularly in addressing class imbalance and adopting optimized CNN architectures, our approach aims to advance the state of the art in this critical area of medical diagnostics.

III. Dataset

The Dataset used in the project is HAM10000 (Human Against Machine with 10000) is an extensive collection of multi-sources dermatoscopic images of common pigmented skin lesions. The Dataset consists of 10015 images in it. There are a total of seven categories of images in this Dataset:

0. Melanocytic nevi (nv)
1. Melanoma (mel)
2. Benign Keratosis-like lesions (bkl)
3. Basal cell carcinoma (bcc)
4. Actinic Keratoses and Intraepithelial carcinoma (akiec)
5. Vascular lesions (vasc)
6. Dermatofibroma (df)

The metadata file has information like lesion_id: Identifier for the lesion, image_id: Unique name of the image file, dx: Diagnosis label, dx_type: Type of diagnosis confirmation, age: Age of the patient, sex: Gender of the patient and localization: Anatomical site of the lesion.

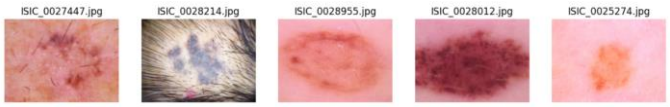

Fig. 1: Sample Data Images

IV. Methods

To enhance the diversity and representativeness of our dataset, we employed data augmentation techniques, motivated by the need to improve the model's robustness and generalization capabilities. Initially, our dataset comprised 10,000 images, which we determined was insufficient to capture the variability required for effective training, particularly given the imbalances in class distributions.

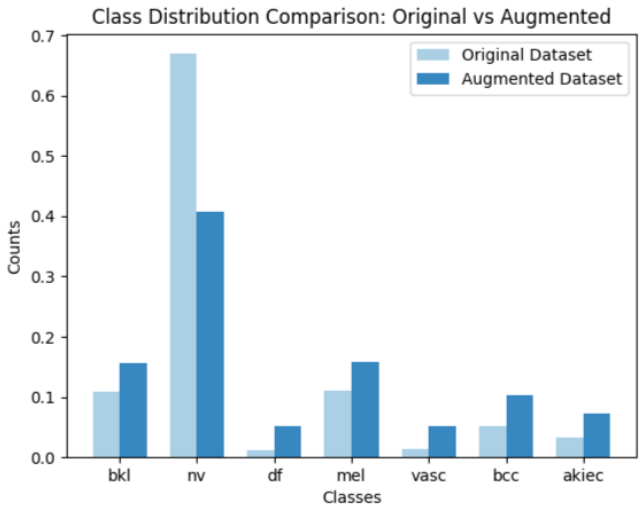

Fig. 2: Comparison of class distributions before vs after data augmentation and oversampling.

Drawing on insights from the paper Unraveling the Impact of Class Imbalance on Deep Learning Models for Medical Image Classification, we recognized the importance of addressing class imbalances to mitigate bias and ensure similar performance across all classes. This understanding informed our decision to expand the dataset to 50,000 images through augmentation, with careful attention to preserving the integrity of the training and testing sets.

To prevent data leakage and maintain a rigorous evaluation process, the dataset was first split into training and testing subsets before augmentation to prevent data leakage and maintain a rigorous evaluation process. Once this division was established, augmentation was applied independently to the training and testing sets. For the training data, augmentation involved transformations such as random rotations, flipping, zooming, and adjustments to brightness and contrast, all designed to simulate real-world variability and enhance the model's ability to generalize. This process increased the training set size and helped address identified class imbalances by generating additional examples for underrepresented categories. Table 1 shows the distribution of classes before and after data augmentation.

The testing dataset underwent a similar augmentation process, conducted independently to ensure that no overlap occurred between the training and testing sets. This approach was critical for

preserving the integrity of our model evaluation, ensuring that the test results would accurately reflect the model's performance on unseen data.

A validation set was derived from the augmented training data, comprising 10% of the total training images to facilitate hyperparameter tuning and monitor performance during training, a validation set was derived from the augmented training data, comprising 10% of the total training images. Through this carefully structured augmentation pipeline, we effectively increased the size and diversity of our dataset while addressing class imbalances, enabling the development of a more robust and fairer model for medical image classification.

Table 1: Class Distributions Before and After Augmentation

Class	Before - Frequency	Before - Percent	After - Frequency	After - Percent
bkl	1099	11.0%	7693	15.6%
nv	6705	66.9%	20115	40.7%
df	115	1.1%	2518	5.1%
mel	1113	11.1%	7791	15.8%
vasc	142	1.4%	2556	5.2%
bcc	514	5.1%	5140	10.4%
akiec	327	3.3%	3597	7.3%

Table 1: Class distributions before vs after data augmentation and oversampling.

V. Experiments

To evaluate the performance of various CNN architectures in classifying skin lesion images, we conducted a series of experiments using pre-trained models fine-tuned for our dataset. Our goal was to compare these models based on their classification accuracy, robustness to class imbalances, and computational efficiency. The experiments were designed to explore the capabilities of state-of-the-art architectures, leveraging transfer learning and data augmentation to maximize performance.

Each model was initialized with weights pre-trained on ImageNet to benefit from features learned on a large-scale dataset. The architectures were adapted to suit the specific requirements of skin lesion classification, including resizing input image shapes to (3, 224, 224), adding custom fully connected layers, and employing strategies like batch normalization and dropout to mitigate overfitting. The training process included early stopping and learning rate adjustments to ensure convergence while avoiding unnecessary computations.

We assessed the performance of three prominent CNN architectures: ResNet-50, VGG16, and Inception-ResNet-V2. Each model was fine-tuned on the augmented training dataset, with hyperparameters optimized for this task. Validation and testing were performed on independent datasets to ensure the reliability of the results. The subsequent subsections detail the setup, modifications, and outcomes for each model.

A. Model 1: ResNet-50

Initially, the ResNet50 architecture, pre-trained on the ImageNet dataset, serves as the backbone of the model for feature extraction. This pretrained model's top layers are removed, allowing adaptation to the skin lesion classification task. To fine-tune the model, a custom head is added, consisting of global average pooling, a dense layer with 1024 neurons using ReLU activation, and an output layer with softmax activation corresponding to the number of classes in the dataset. All layers of the ResNet50 model are unfrozen during training, allowing the network to adjust pre-trained weights to the specific characteristics of skin lesion images. The model is compiled with the Adam optimizer, a learning rate of 0.00001, categorical cross-entropy loss, and accuracy as the evaluation metric.

For data augmentation, the ImageDataGenerator performs random transformations on the images, such as rotations, shifts, and flips, which increases the diversity of the training data and helps prevent overfitting. The training process is carried out over 50 epochs, with the model evaluated on a 80-20 dataset split. Each batch is of size 32, and steps per epoch are calculated based on the total number of training samples. The model is trained with both training and validation data fed through the generators, which allows for real-time augmentation during training. The model's performance is assessed on the test set, providing critical metrics like test accuracy and loss.

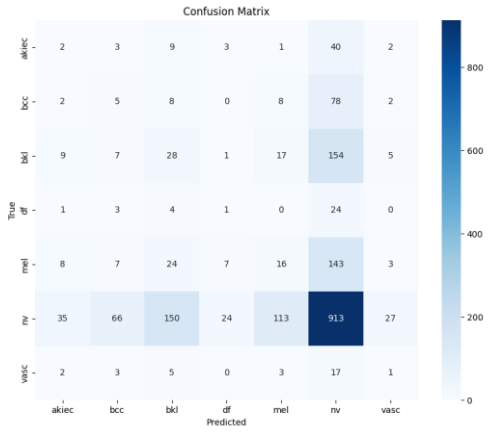

Fig. 3: Confusion matrix for ResNet-50 model.

The model achieved an accuracy of 87% and a classification report includes the F1 scores for each class and precision and recall assessing the model further. These metrics provide information about the model's accuracy in classifying skin lesions.

B. Model 2: VGG16

The implemented model for skin disease classification leverages the power of VGG16, a pretrained convolutional neural network designed initially for ImageNet. The architecture begins with the base VGG16 model, where the fully connected layers are removed to allow custom adaptation. The high-dimensional feature maps from the convolutional layers are reduced using a GlobalAveragePooling2D layer, which efficiently condenses the learned features. A dense layer with 1024 units and ReLU activation is added to learn dataset-specific representations, followed by a final softmax-activated layer for multi-class classification. The number of output neurons corresponds to the number of unique disease classes in the dataset. To enhance learning, all layers, including those of the VGG16 base model, are set as trainable, allowing for fine-tuning of pretrained weights better to suit the unique characteristics of the skin disease dataset.

The training pipeline is meticulously designed to ensure robustness and generalization. The input images are resized to 224x224 pixels and preprocessed for compatibility with the VGG16 architecture. A comprehensive data augmentation strategy is applied using ImageDataGenerator, which includes random rotations, flips, and shifts to introduce variability into the training data, thus reducing overfitting and enabling the model to generalize better.

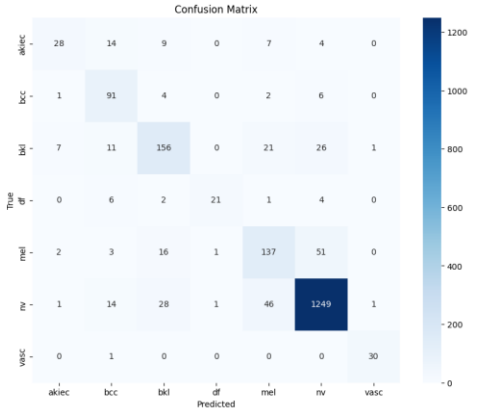

Fig. 4: Confusion matrix for VGG16 model.

Training is conducted using the Adam optimizer with a fine-tuned learning rate of 1e-05. The categorical cross-entropy loss function and accuracy metrics are used to evaluate performance during training. The model is trained for 50 epochs with a batch size of 32, and the validation set is used to monitor performance in real time, ensuring that the model is neither underfitting nor overfitting.

The model achieved an accuracy of 85.5%. and a classification report offers a detailed insight of model performance by providing accuracy, recall, and F1 scores for every class. The model shows its efficacy in identifying skin disorders by achieving competitive accuracy on the test set.

C. Model 3: Inception-ResNet-V2

The Inception-ResNet-V2 architecture was employed for our third experiment, leveraging its pre-trained weights on the ImageNet dataset. The model was downloaded through the Keras library in Python and adapted to our specific task of skin lesion classification. On top of the base layer, we incorporated a global average pooling 2D layer, followed by a 2D convolutional layer with 128 filters of size 3, using ReLU activation. To improve stability and mitigate overfitting, a batch normalization layer and a dropout layer with a rate of 0.5 were added. A second global average pooling

layer followed, feeding into the final classification output, which consisted of seven classes and employed a softmax activation function.

For optimization, we utilized early stopping to halt training when validation performance plateaued, alongside a ReduceLROnPlateau callback to dynamically adjust the learning rate as improvements slowed. The model was initially trained with a batch size of 250, a learning rate of 1e-05, and a maximum of 90 epochs. Following this initial phase, fine-tuning was performed by unfreezing 50 layers of the base model and retraining with a lower learning rate of 1e-06 to further refine the feature extraction for our specific dataset.

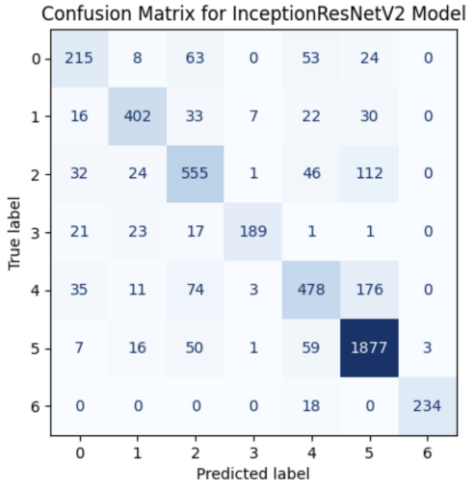

Fig. 5: Confusion matrix for InceptionResNetV2 model.

The model achieved a test accuracy of 80.0%, demonstrating robust performance in distinguishing between the seven skin lesion classes. To further evaluate the model's effectiveness, we analyzed its confusion matrix, providing insights into its classification capabilities and areas for improvement. These results highlight the potential of the Inception-ResNet-V2 architecture when fine-tuned for medical imaging tasks.

The model performs exceptionally well for Class 6 (high precision, recall, and F1-score) and Class 5 (high recall and F1-score), likely because these classes dominate the dataset or have distinguishing features. Performance is weaker for Class 0 and Class 4, as indicated by lower precision, recall, and F1-scores. The overall accuracy and weighted averages suggest balanced performance but highlight the need for improvement in underrepresented or harder-to-classify classes.

VI. Conclusion

In this study, we evaluated the performance of three convolutional neural network architectures—ResNet-50, VGG-16, and Inception ResNet-V2—on a skin lesion classification task. The results, summarized in Table 2, show trade-offs between test accuracy and other metrics like precision, recall, and F1 scores.

Model	Train Accuracy	Test Accuracy	Precision	Recall	F1-score
Resnet-50	0.9934	0.8785	0.48	0.49	0.48
VGG16	0.9881	0.8547	0.86	0.85	0.75
Inception-ResNet-V2	0.9998	0.800081	0.80	0.80	0.80

Fig. 1. Table 2: Experiments results.

ResNet-50 achieved the highest test accuracy at 87.85%, showing strong overall performance. However, its precision, recall, and F1 scores were relatively low at 48, 49, and 48, indicating it struggled to balance false positives and false negatives. VGG-16 had the highest precision and recall scores, with a slightly lower test accuracy of 85.47%, making it more reliable for minimizing misclassifications. Inception ResNet-V2 had the lowest test accuracy at about 80% but the highest F1 score at 0.8, showing it managed the trade-off between precision and recall better.

Precision and recall are generally considered more important than overall accuracy for skin lesion classification. High precision reduces false positives, avoiding unnecessary stress and follow-ups, while high recall ensures malignant cases are identified, allowing for timely treatment. While ResNet-50 had the best accuracy, the better precision, recall, and F1 scores of VGG-16 and Inception ResNet-V2 suggest they are better suited for this application.

VII. Future Works

Future work includes investigating the overfitting observed in the Inception ResNet-V2 model. Analyzing its behavior during training and validation could help identify factors contributing to its high

training accuracy but lower generalization performance. Exploring alternative hyperparameter combinations, such as learning rates, dropout rates, and batch sizes, may mitigate overfitting and improve overall effectiveness. Fine-tuning additional layers or employing regularization techniques like weight decay are potential strategies for enhancing generalizability.

Another focus is expanding the training dataset to include more diverse medical imaging data. Training on larger and more varied datasets can increase robustness and improve effectiveness in real-world applications. Incorporating additional skin lesion datasets, as well as data from other medical imaging domains, may enable the models to learn more generalized features and perform better across different imaging conditions.

Finally, exploring ensemble learning techniques to combine the strengths of multiple models is a promising direction. Leveraging the complementary advantages of architectures like ResNet-50, VGG-16, and Inception ResNet-V2 could result in improved classification accuracy and balance between precision and recall.

VIII. Target Conference Venue

We'd like to submit our work to the Medical Imaging with Deep Learning (MIDL) 2025 conference. MIDL focuses on the intersection of medical imaging and deep learning, making it a suitable platform for our research. This conference encourages submissions from both seasoned and emerging researchers, fostering a supportive environment for early-stage work. The submission deadline is 24 January, 2025.

IX. References

[1] Srinivasu, Parvathaneni Naga, et al. "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM." Sensors 21.8 (2021): 2852. https://www.mdpi.com/1424-8220/21/8/2852.

[2] Allugunti, Viswanatha Reddy. "A machine learning model for skin disease classification using convolution neural network." International Journal of Computing, Programming and Database Management 3.1 (2022): 141-147. https://www.researchgate.net/publication/361228242_A_machine_learning_model_for_skin_disease_classification_using_convolution_neural_network.

[3] Alzubaidi, Laith, et al. "MedNet: pre-trained convolutional neural network model for the medical imaging tasks." arXiv preprint arXiv:2110.06512 (2021). https://arxiv.org/abs/2110.06512.

[4] Hellín, Carlos J., et al. "Unraveling the impact of class imbalance on deep-learning models for medical image classification." Applied Sciences 14.8 (2024): 3419. https://doi.org/10.3390/app14083419.

[5] Villa-Pulgarin, Juan Pablo, et al. "Optimized Convolutional Neural Network Models for Skin Lesion Classification." Computers, Materials & Continua, vol. 70, no. 2, 2022, pp. 2131-2148. https://doi.org/10.32604/cmc.2022.019529.

[6] Li, Qing, et al. "Medical Image Classification with Convolutional Neural Network." Proceedings of the 2014 13th International Conference on Control, Automation, Robotics & Vision (ICARCV), Singapore, 10–12 December 2014, pp. 844–848. https://doi.org/10.1109/ICARCV.2014.7064414

[7] Salehi, Ahmad Waleed, et al. "A study of CNN and transfer learning in medical imaging: advantages, challenges, future scope." Sustainability 15.7 (2023): 5930. https://www.mdpi.com/2071-1050/15/7/5930.

[8] Ajagbe, Sunday Adeola, et al. "Multi-classification of Alzheimer Disease on Magnetic Resonance Images (MRI) Using Deep Convolutional Neural Network (DCNN) Approaches." International Journal of Advanced Computer Research, vol. 11, no. 53, 2021, pp. 51–60. https://doi.org/10.19101/IJACR.2021.1152001.

[9] Wu, Haijing, et al. "A Deep Learning, Image-Based Approach for Automated Diagnosis for Inflammatory Skin Diseases." Annals of

Translational Medicine, vol. 8, no. 9, 2020, p. 581. https://doi.org/10.21037/atm.2020.04.39.

[10] Shahin, Ahmed H., et al. "Deep Ensemble Learning for Skin Lesion Classification from Dermoscopic Images." Proceedings of CIBEC, Cairo, Egypt, 2018, pp. 150–153. https://doi.org/10.1109/CIBEC.2018.8641706.

X. APPENDIX

	Joseph Richardson	Sharanya DV	Venkata Abhiram Chitty
Role	Leader / Modeler	Modeler	Modeler
Meetings Attended	11	11	11
Percent	100%	100%	100%

Table 3: Meeting attendance and group members contribution.