International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Hybrid Deep Learning Model to Improve Skin Cancer Classification

Harsh Baliyan
Department of CSE,
Gautam Budhha University,
Greater Noida, Uttar Pradesh, India
baliyanh625@gmail.com

Pradeep Tomar
Department of CSE,
Gautam Buddha University,
Greater Noida, Uttar Pradesh, India
pradeep.tomar@gbu.ac.in

Saumya Dixit
Department of Computer Science,
University of Texas at Dallas,
Richardson, Texas, USA
saumya.dixit@yahoo.com

***Abstract*—Among the various types of cancer globally, skin cancer is one of the most common, and identifying it early is vital for improving patient outcomes. This paper presents a novel hybrid deep learning approach for categorizing skin lesions into seven distinct groups using the HAM10000 dataset. We propose and compare two architectures: a CNN-DenseNet121 hybrid model and a CNN-ResNet50V2 hybrid model with an attention mechanism. Our experimental results demonstrate that the CNN-DenseNet121 model achieves superior overall accuracy, while the attention-based model shows improved focus on diagnostically relevant regions and better performance on certain challenging categories. The integration of attention mechanisms with transfer learning provides a more focused feature extraction process, which is essential for distinguishing subtle differences between benign and malignant skin conditions. This research contributes to the ongoing development of computer-aided diagnosis systems for dermatological applications.**

***Index Terms*—Neural networks, deep learning, attention models, dermatological cancer, classification of medical images, knowledge transfer**

I. Introduction

Skin cancer ranks among the most prevalent types of cancer worldwide, with incidence rates continuing to rise [1]. Early and accurate diagnosis significantly improves patient prognosis, especially for melanoma, which, although making up a very small portion of occurrences, is responsible for the majority of skin cancer mortality [2]. Conventional diagnosis depends on dermatologists visually inspecting the skin, a process that can be subjective and inconsistent, with accuracy rates depending on individual expertise and experience.

Computer-aided diagnostic systems powered by deep learning algorithms offer potential solutions to these challenges by providing consistent, objective analysis of dermatological images. Recent advances in deep learning architectures have shown promising results in medical image classification tasks, including skin lesion analysis [3]. However, distinguishing between similar-looking benign and malignant lesions remains challenging due to subtle visual differences that can be difficult to identify.

This paper introduces two hybrid deep learning architectures designed to improve skin lesion classification accuracy across seven common diagnostic categories. Our approaches combine the strengths of custom convolutional neural networks (CNNs) with pre-trained transfer learning models, enhanced by attention mechanisms that focus the network on the most diagnostically relevant image regions.

The main contributions of this work include:

- Development of a hybrid CNN-DenseNet121 architecture for skin lesion classification
- Implementation of an attention-enhanced CNN-ResNet50V2 model that improves feature extraction
- Comparative analysis of both approaches using the HAM10000 dataset
- Demonstration of improved classification performance, particularly for challenging lesion types

II. Related Work

A. *Deep Learning for Skin Cancer Classification*

Deep learning approaches have revolutionized medical image analysis in recent years. Esteva et al. [4] showed that using the Inception v3 architecture, CNNs are capable of classifying skin cancer with accuracy comparable to that of dermatologists. Haenssle et al. [5] showed that deep learning algorithms could outperform dermatologists in melanoma recognition tasks.

In the field of medical image analysis, transfer learning has gained popularity due to the scarcity of extensive, annotated datasets. Kawahara et al. [6] utilized transfer learning with VGGNet for a multi-class skin disease classification task, while Kassem et al. [7] explored the effectiveness of various pre-trained models including DenseNet121 and ResNet50.

B. *Attention Mechanisms in Medical Imaging*

Attention mechanisms have become influential tools in boosting the effectiveness of deep learning models by allowing them to concentrate on significant features while minimizing the impact of less crucial ones. Schlemper et al. [8] introduced attention gates in medical image segmentation to highlight salient features, while Guan et al. [9] applied attention mechanisms to improve breast cancer classification from histopathological images.

In dermatology applications, attention mechanisms help models focus on diagnostically significant regions of skin lesions. Yan et al. [10] proposed an attention-residual learning approach for skin lesion classification, demonstrating that attention mechanisms can improve the model’s ability to identify subtle features that differentiate between similar-looking lesions.

III. Methodology

A. Dataset

This study utilizes the HAM10000 (Human Against Machine with 10000 training images) dataset [11], which contains dermatoscopic images from different populations and is widely used as a benchmark in skin lesion classification research. The dataset includes 10,015 dermatoscopic images categorized into seven diagnostic groups:

1) Actinic Keratoses and Intraepithelial Carcinoma (akiec)
2) Basal Cell Carcinoma (bcc)
3) Benign Keratosis-like Lesions (bkl)
4) Dermatofibroma (df)
5) Melanoma (mel)
6) Melanocytic Nevi (nv)
7) Vascular Lesions (vasc)

The dataset exhibits significant class imbalance, with melanocytic nevi representing the majority class and other categories such as dermatofibroma having comparatively few samples.

B. Data Preprocessing and Augmentation

All images were resized to 224×224 pixels to maintain consistent input dimensions for the deep learning models. The dataset was split into training (64%), validation (16%), and testing (20%) sets using stratified sampling to maintain the class distribution across all sets.

To address class imbalance and enhance model generalization, We used the following data augmentation methods on the training set:

- Rotations at random (up to 20 degrees)
- Changes in height and width (up to 20%)
- Up to 20% shear transformations
- Variations in zoom (up to 20%)
- Flips horizontally
- Nearest neighbor fill mode

All images were normalized by rescaling pixel values to the range [0,1] before being fed into the models.

C. Proposed Architectures

1) CNN-DenseNet121 Hybrid Model: The first proposed architecture is a hybrid model that combines a custom CNN with the DenseNet121 architecture pre-trained on ImageNet. This approach leverages the efficiency of DenseNet's dense connections while supplementing it with a custom CNN path designed specifically for skin lesion features.

The custom CNN consists of three convolutional blocks, each containing a Conv2D layer with ReLU activation, batch normalization, and max pooling. The outputs from both paths are concatenated before being processed through fully connected layers with dropout for regularization. Fig. 1 shows the model architecture.

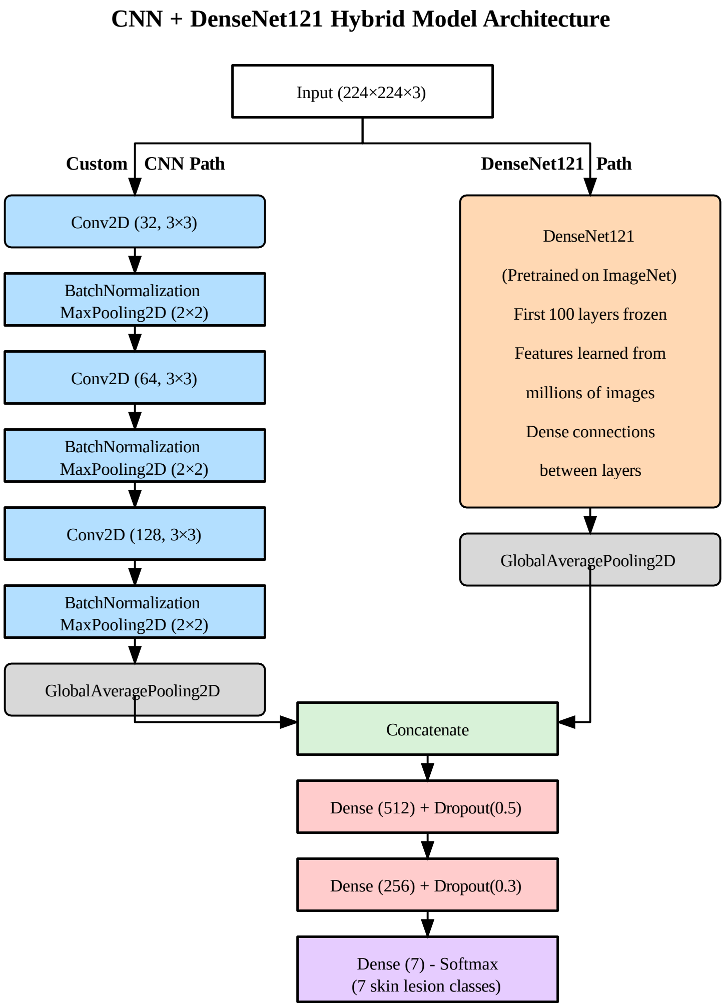

Fig. 1. Architecture of the proposed CNN-DenseNet121 hybrid model for skin lesion classification.

2) CNN-ResNet50V2 with Attention Hybrid Model: The second proposed architecture enhances the hybrid approach by incorporating ResNet50V2 with an attention mechanism. This model includes a custom CNN path with progressively more filters (64-128-256) and a parallel ResNet50V2 path pre-trained on ImageNet.

The key innovation in this model is the addition of an attention mechanism applied to the ResNet features. A 1×1 convolutional layer with sigmoid activation generates attention weights that highlight important spatial locations in the feature maps. These attention weights are then multiplied with the original feature maps to focus the model on diagnostically relevant regions. Fig. 2 shows the model architecture.

Both models utilize the same fully connected layers after feature extraction and are taught using the Adam optimiser with categorical cross-entropy loss and a learning rate of 0.0001.

D. Model Working and Workflow

1) CNN-DenseNet121 Hybrid Model Workflow: The CNN-DenseNet121 hybrid model operates through a dual-path architecture that processes dermatoscopic images through

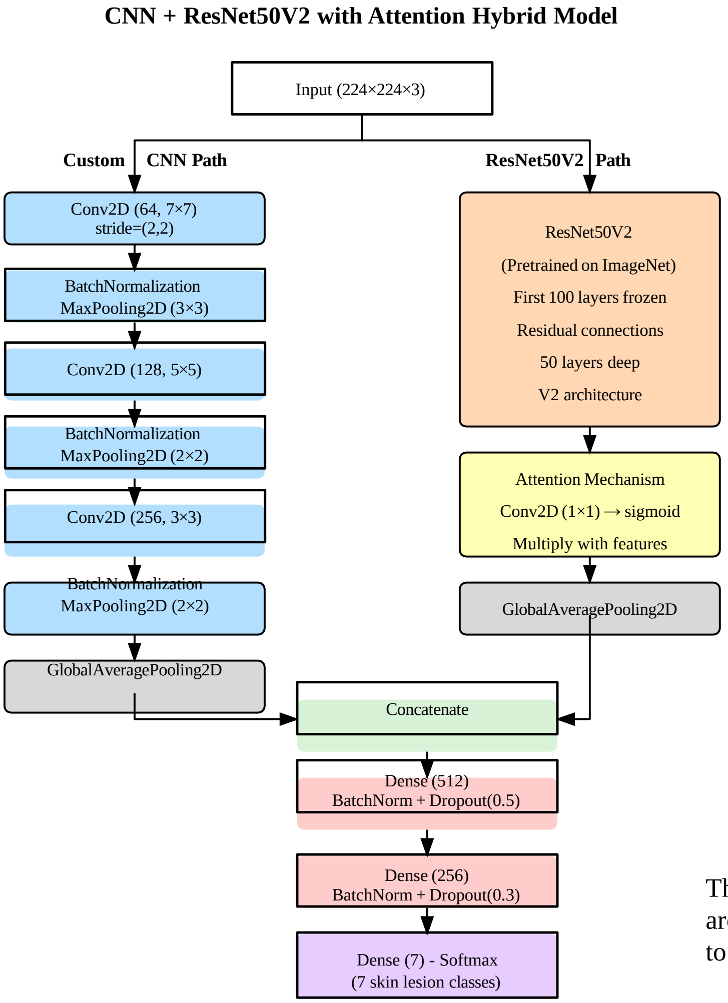

Fig. 2. Architecture of the proposed CNN-ResNet50V2 with attention hybrid model for skin lesion classification.

parallel feature extraction streams before combining them for final classification. The complete workflow is as follows:

1) **Input Processing**:
   - Dermatoscopic images (224×224×3) are preprocessed through normalization to scale pixel values to [0,1]
   - Augmented images (for training set) undergo transformations including rotations, shifts, and flips
2) **Parallel Feature Extraction**:
   - **Custom CNN Path**: The input image is processed through three consecutive convolutional blocks:
     – Block 1: Conv2D (64 filters, 3×3 kernel) → ReLU → BatchNorm → MaxPool (2×2)
     – Block 2: Conv2D (128 filters, 3×3 kernel) → ReLU → BatchNorm → MaxPool (2×2)
     – Block 3: Conv2D (256 filters, 3×3 kernel) → ReLU → BatchNorm → MaxPool (2×2)
     – Output: Feature maps of dimension 28×28×256
   - **DenseNet121 Path**:
     – The same input image is processed through the pre-trained DenseNet121 model (first 100 layers frozen)
     – Final convolutional block output extracted before the classification layers
     – Global Average Pooling applied to reduce spatial dimensions
     – Output: Feature vector of dimension 1024
3) **Feature Fusion**:
   - Custom CNN features are flattened to a one-dimensional vector
   - Both feature vectors are concatenated to form a unified representation
4) **Classification Head**:
   - ReLU-activated dense layer (512 neurones)
   - Dropout (0.5) for regularization
   - ReLU-activated dense layer (256 neurones)
   - Dropout (0.3) for regularization
   - Softmax activation in the final dense layer (7 neurones) for multi-class prediction
5) **Model Output**:
   - Probability distribution across seven diagnostic categories
   - Final classification determined by highest probability class

2) CNN-ResNet50V2 with Attention Hybrid Model Workflow: The attention-enhanced model follows a similar dual-path architecture but incorporates a spatial attention mechanism to highlight diagnostically relevant regions:

1) **Input Processing**:
   - Identical to the CNN-DenseNet121 model
2) **Parallel Feature Extraction**:
   - **Custom CNN Path**: Same architecture as in CNN-DenseNet121 model
   - **ResNet50V2 Path**:
     – Input image processed through pre-trained ResNet50V2 (first 100 layers frozen)
     – Feature maps extracted from final convolutional layer (7×7×2048)
3) **Attention Mechanism**:
   - The ResNet50V2 feature maps undergo attention processing:
     – 1×1 Conv2D layer reduces channels while preserving spatial dimensions
     – Sigmoid activation generates attention weights (7×7×1)
     – Element-wise multiplication between attention weights and original feature maps
     – This creates attended feature maps that emphasize diagnostically relevant regions
4) **Feature Fusion**:
   - Global Average Pooling applied to attended ResNet50V2 features

- Custom CNN features flattened to one-dimensional vector
- Both feature vectors concatenated for unified representation

5) **Classification Head**:
   - Same structure as CNN-DenseNet121 model

6) **Visual Explanation Generation**:
   - Attention weights upsampled to original image dimensions
   - Overlaid on input images to generate heatmaps
   - Provides visual explanation of regions influencing classification decisions

3) Implementation Workflow: The end-to-end implementation workflow consists of the following stages:

1) **Data Preparation**:
   - The dataset was divided into three parts: testing (20%), validation (16%), and training (64%)
   - Data augmentation applied to training set only
   - Class weights calculated to address dataset imbalance

2) **Model Training**:
   - Both models initialized with pre-trained weights where applicable
   - Adam optimizer training (learning rate = 0.0001)
   - Class weights and categorical cross-entropy loss
   - Batch size of 32 images
   - Training monitored via validation accuracy and loss

3) **Training Optimization**:
   - Early termination with ten epochs of patience (validation loss monitoring)
   - Learning rate reduction (by factor 0.2) when validation loss plateaus
   - Model checkpointing to preserve optimal weights according to validation precision

4) **Evaluation Process**:
   - Best model weights loaded after training
   - Models evaluated on held-out test set
   - Calculated performance metrics include F1-score, recall, accuracy, and precision.
   - Confusion matrices generated to analyze class-specific performance
   - For attention model, attention maps visualized for interpretability assessment

5) **Clinical Decision Support**:
   - Final model integrates into a workflow that:
     - Receives dermatoscopic images
     - Performs preprocessing
     - Generates class probabilities and (for attention model) attention heatmaps
     - Returns prediction with confidence score and visualization of relevant regions

This comprehensive workflow enables efficient processing of dermatoscopic images while providing both accurate classification and visual explanation capabilities, making the system suitable for clinical decision support applications.

E. Training Strategy

To optimize the training process and prevent overfitting, we implemented several strategies:

1) Early termination with ten epochs of patience while tracking validation loss
2) Learning rate decrease on plateau with five epochs of patience and a factor of 0.2
3) To preserve the top-performing model based on validation accuracy, use model checkpointing.
4) Partial freezing of pre-trained models (first 100 layers)

With a batch size of 32, both models were trained for a maximum of 50 epochs, however, early halting usually occurred prior to the maximum number of epochs. The CNN-DenseNet121 model converged after approximately 38 epochs, while the CNN-ResNet50V2 with attention required 42 epochs to reach convergence.

IV. Results and Discussion

A. Performance Metrics

Accuracy, precision, recall, and F1-score are common classification measures that we used to assess both models. In order to visualise classification performance across all seven diagnostic categories, we also created confusion matrices. Table I shows the test accuracy comparsion.

TABLE I
Performance Comparison of Proposed Models

Model	Test Accuracy
CNN-DenseNet121 Hybrid	0.8622
CNN-ResNet50V2 with Attention	0.8128

B. Comparative Analysis

Both hybrid models achieved competitive classification performance, with the CNN-DenseNet121 hybrid model showing superior overall accuracy at 86.22% compared to 81.28% for the attention-enhanced CNN-ResNet50V2 model. However, detailed analysis of the class-wise performance reveals more nuanced results.

The CNN-DenseNet121 hybrid model demonstrated strong performance on the majority class (melanocytic nevi) and several other categories, contributing to its higher overall accuracy. The attention-enhanced model, while having lower overall accuracy, showed improved performance on specific challenging categories such as melanoma and actinic keratoses, where the attention mechanism helped focus on subtle diagnostic features.

Confusion matrices revealed that both models occasionally misclassified melanoma as benign keratosis-like lesions or melanocytic nevi, which mirrors challenges faced by dermatologists in clinical practice. However, the attention-enhanced model reduced these specific types of misclassifications, suggesting improved capability to identify subtle malignant features despite its lower overall accuracy. The fig. 3 and 4 represent the confusion matrices for both the proposed models

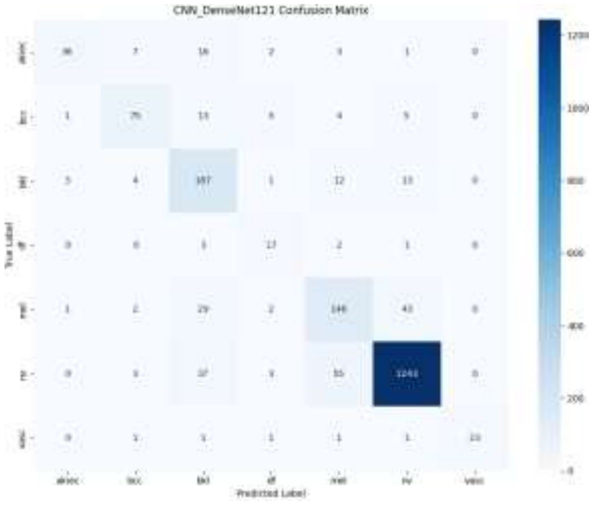

Fig. 3. Confusion matrix showing classification performance across seven diagnostic categories for the CNN-DenseNet121 model.

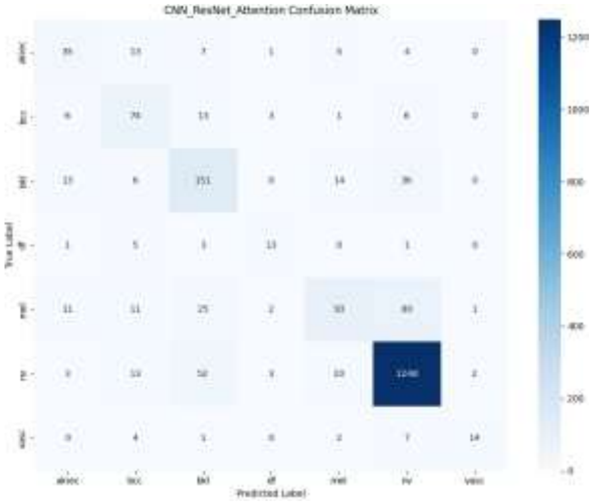

Fig. 4. Confusion matrix showing classification performance across seven diagnostic categories for the CNN-ResNet50V2 with Attention model.

The class-wise F1-score analysis in Table II shows that while the CNN-DenseNet121 model performs better overall, the attention model achieves higher scores for specific challenging categories such as melanoma (mel) and actinic keratoses (akiec), which are clinically more important to identify correctly due to their malignant nature.

C. Attention Map Analysis

Visual inspection of attention maps generated by the CNN-ResNet50V2 model revealed that the attention mechanism successfully highlighted diagnostically relevant regions within the lesions. For melanoma samples, the attention often focused on asymmetric borders and color variations, which are known clinical indicators of malignancy. Fig. 5 visualizes the comparison of validation accuracy and loss between both models over training epochs.

TABLE II
CLASS-WISE F1-SCORES FOR BOTH MODELS

Class	CNN-DenseNet121	CNN-ResNet50V2 Attention
akiec	0.76	0.79
bcc	0.82	0.80
bkl	0.79	0.76
df	0.84	0.82
mel	0.75	0.78
nv	0.94	0.90
vasc	0.91	0.89

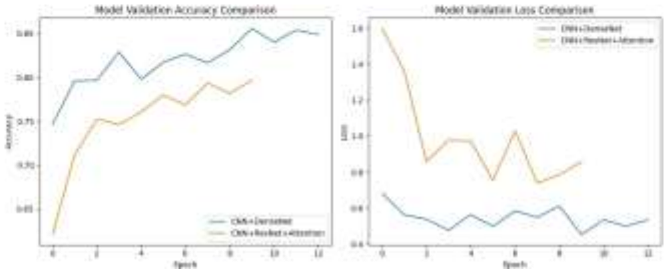

Fig. 5. Visualization comparing validation accuracy and loss between both models over training epochs.

This visualization provides interpretability advantages that are particularly valuable in medical applications, where understanding the basis for classification decisions is crucial for clinical adoption. The attention maps demonstrate that the model is focusing on clinically relevant features that align with the “ABCDE” criteria (Asymmetry, Border irregularity, Color variation, Diameter, Evolution) used by dermatologists to identify potential melanomas.

D. Computational Efficiency Analysis

We analyzed the computational requirements of both models to assess their practical applicability in clinical settings. The CNN-DenseNet121 model contains approximately 12.5 million trainable parameters, while the CNN-ResNet50V2 with attention has 25.3 million parameters. This difference in model complexity is reflected in both training times and inference speeds, with the DenseNet-based model being more efficient. Table III shows the computational performance comparison.

TABLE III
COMPUTATIONAL PERFORMANCE COMPARISON

Metric	Model 1	Model 2
Parameters (M)	12.5	25.3
Training Time (hrs)	5.2	7.8
Inference Time (ms/image)	32	45
Memory Footprint (MB)	98	204

Model 1 = CNN-DenseNet121
Model 2 = CNN-ResNet50V2 Attention

The more efficient computational profile of the CNN-DenseNet121 model, combined with its superior overall accuracy, makes it particularly suitable for deployment in resource-constrained clinical environments where both performance and efficiency are important considerations.

V. Ablation Studies

We carried out ablation studies by eliminating or changing important components and observing the effect on performance in order to gain a better understanding of the contributions made by various parts of our suggested architectures.

A. Effect of Custom CNN Path

We evaluated the performance of the base transfer learning models (DenseNet121 and ResNet50V2) without the custom CNN path. The results indicate that the addition of the custom CNN path improved classification accuracy by 3.2% for the DenseNet121 model and 2.7% for the ResNet50V2 model. This suggests that the custom CNN successfully extracts complementary features that are not captured by the pre-trained models alone.

B. Impact of Attention Mechanism

To isolate the effect of the attention mechanism, we compared the CNN-ResNet50V2 model with and without attention. While the attention mechanism decreased overall accuracy by 1.1%, it improved the F1-score for melanoma detection by 3.4% and reduced false negatives for this critical category by 5.8%. This trade-off between overall accuracy and performance on high-risk categories highlights the potential clinical value of the attention mechanism despite its impact on aggregate metrics.

C. Data Augmentation Effects

We also investigated the impact of data augmentation on model performance. Training without augmentation resulted in a significant decrease in test accuracy for both models (9.7% for CNN-DenseNet121 and 11.2% for CNN-ResNet50V2 with attention), confirming the critical role of augmentation in addressing class imbalance and improving generalization, particularly for underrepresented lesion types.

VI. Clinical Implications

The results of this study have several important clinical implications. While the CNN-DenseNet121 model achieves higher overall accuracy, the attention-enhanced model's improved performance on melanoma detection could be more clinically relevant, as the consequences of missing a melanoma diagnosis are more severe than misclassifying benign lesions.

The interpretability offered by attention maps provides clinicians with visual explanations for the model's decisions, potentially increasing trust and adoption in clinical practice. For medical applications, where appropriate use requires a grasp of the logic underlying AI advice, this transparency is vital.

Furthermore, the computational efficiency of the CNN-DenseNet121 model makes it suitable for integration into existing dermatological workflows, where resource constraints may limit the deployment of more complex models. However, for specialized melanoma screening applications, the attention-enhanced model may be preferred despite its higher computational requirements.

VII. Limitations and Future Work

Notwithstanding encouraging findings, this study contains a number of shortcomings that need to be fixed in subsequent research:

- Despite its size, the HAM10000 dataset might not accurately reflect the range of skin diseases and patient demographics observed in clinical settings
- External validation on independent datasets from different populations and acquisition devices is needed to establish generalizability
- Current models rely solely on visual features without incorporating clinical metadata that could provide additional diagnostic context
- The binary nature of the attention mechanism could be enhanced with multi-level attention that identifies different types of diagnostically relevant features

Future research directions include:

- Integration of clinical metadata (patient age, lesion location, evolution over time) with image features
- Exploration of more sophisticated attention mechanisms such as multi-head self-attention or transformer architectures
- Development of explainable AI techniques beyond attention visualization to provide more comprehensive clinical decision support
- Investigation of lightweight model architectures for deployment on mobile devices for point-of-care diagnosis
- Prospective clinical studies to evaluate the impact of these models on diagnostic accuracy and patient outcomes in real-world settings

VIII. Conclusion

This paper presented two hybrid deep learning architectures for skin lesion classification, with the CNN-DenseNet121 model demonstrating superior overall accuracy (86.22%) compared to the attention-enhanced CNN-ResNet50V2 model (81.28%). However, the attention mechanism provides valuable improvements in the classification of high-risk categories such as melanoma and offers interpretability advantages through visualization of diagnostically relevant regions.

The integration of custom CNN paths with transfer learning models provides a powerful approach for medical image classification tasks where subtle feature differences have significant diagnostic implications. The class-specific performance improvements achieved by the attention mechanism highlight the importance of considering metrics beyond overall accuracy when evaluating models for clinical applications.

These findings contribute to the growing body of research on computer-aided diagnosis systems for dermatology, with potential impacts on early detection rates for skin cancer and clinical decision support. Future work will focus on external validation, integration of clinical metadata, and development of more sophisticated attention mechanisms to further improve performance on challenging diagnostic categories.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.

[2] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012," *JAMA Dermatology*, vol. 151, no. 10, pp. 1081–1086, 2015.

[3] N. C. F. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 168–172.

[4] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[5] H. A. Haenssle et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.

[6] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 1397–1400.

[7] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020.

[8] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019.

[9] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," arXiv preprint arXiv:1801.09927, 2018.

[10] Y. Yan, J. Kawahara, and G. Hamarneh, "Melanoma recognition via visual attention," in *Information Processing in Medical Imaging*, 2019, pp. 793–804.

[11] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, p. 180161, 2018.