Synthesis: A Multidisciplinary Research Journal
Volume 1, Issue 2, April - June 2023, Pp: 12-21

Research Article

Enhanced Skin Cancer Detection Utilizing Enhanced Densenet121

[1*] D.Manju, [2] Sreekanth Rallapalli, [3] Dadi Sanjana

[1*] *Department of CSE-Data Science, VNR Vignan Jyothi Institute of Engineering and Technology, Hyderabad, India*
[2] *Professor and HoD, Department of Master of Computer Applications, Nitte Meenakshi Institute of Technology, Bengaluru, India*
[3] *Department of Artificial Intelligence, University of North Texas*

**Corresponding Author(s):* *nuthana525@gmail.com*

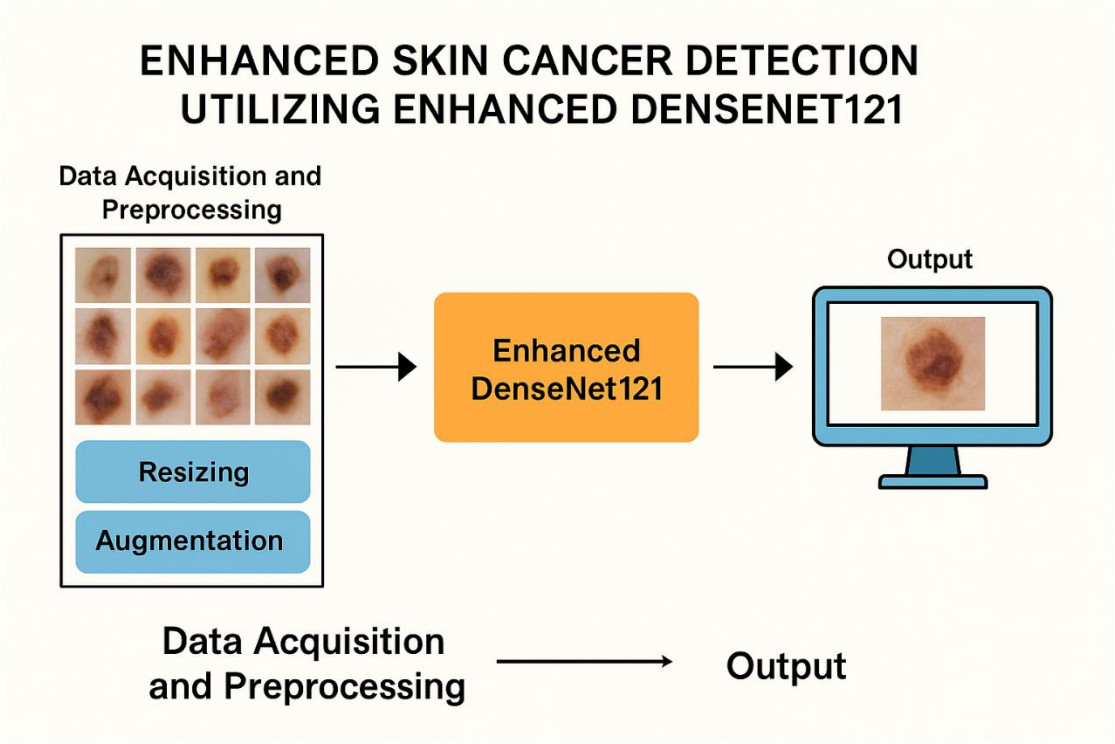
Article Info

Received:02/04/2023
Revised: 20/05/2023
Accepted:15/06/2023
Published:30/06/2023

Abstract

Skin cancer remains one of the most common and potentially fatal forms of cancer worldwide, with early detection being critical to effective treatment. Traditional diagnostic methods often suffer from subjectivity and limited accessibility, highlighting the need for automated, accurate, and scalable diagnostic systems. This study aims to develop an enhanced deep learning model based on DenseNet121 to improve the accuracy and robustness of automated skin cancer detection using dermoscopic images. The proposed approach integrates a Convolutional Block Attention Module (CBAM) into the DenseNet121 architecture, enabling the model to focus on clinically significant lesion regions. The model was trained on the publicly available ISIC 2018 dataset, which includes 10,015 images across seven skin lesion classes. Preprocessing steps included image resizing, histogram equalization, and data augmentation, while a class-weighted categorical cross-entropy loss function was used to address dataset imbalance. Model evaluation was conducted using five-fold cross-validation and multiple performance metrics. The Enhanced DenseNet121 model achieved an accuracy of 91.1%, F1-score of 89.0%, and AUC-ROC of 94.2%, outperforming ResNet50, InceptionV3, EfficientNet-B0, and the baseline DenseNet121. Statistical significance testing yielded a p-value of 0.0084, confirming the robustness of the improvements. The proposed model demonstrates superior diagnostic performance, efficient parameter utilization, and improved focus on relevant lesion areas. Its deployment potential in mobile health platforms and clinical decision support systems suggests a valuable contribution toward accessible, AI-driven skin cancer screening in real-world settings.

Keywords: Skin Cancer Detection, Deep Learning, DenseNet121, Medical Image Analysis, CNN, ISIC Dataset



Graphical Abstract - An overview of the proposed Enhanced DenseNet121-based framework for automated skin cancer detection.

1. Introduction

Skin cancer is one of the most prevalent forms of cancer globally, with incidence rates continuing to rise due to factors such as increased ultraviolet (UV) exposure, aging populations, and inadequate early screening. According to the World Health Organization (WHO), approximately 2–3 million non-melanoma and 132,000 melanoma skin cancer cases are diagnosed annually [1]. Among them, melanoma remains the most lethal, despite being less common, due to its aggressive metastatic potential. Early and accurate diagnosis is crucial for improving patient outcomes, reducing treatment costs, and decreasing the global dermatological burden [2].

Traditional diagnostic methods, such as visual inspection by dermatologists and histopathological analysis through biopsy, are often subjective, time-consuming, and resource-intensive. Moreover, access to trained dermatologists is limited in many regions, particularly in low-income and rural communities [3]. These constraints underscore the growing need for automated, scalable, and objective diagnostic tools capable of supporting early detection and clinical decision-making [4].

The rapid advancement of deep learning and computer vision technologies has introduced promising avenues for automating the detection of skin lesions using dermoscopic images. Among the various convolutional neural network (CNN) architectures, DenseNet121 has shown considerable success in medical image classification due to its efficient feature reuse, compact design, and high classification accuracy [5]. However, the baseline DenseNet121 model is not fully optimized for dermatological imaging challenges such as class imbalance, subtle lesion patterns, and the presence of noise and artifacts [6]. Therefore, enhancing this architecture to achieve clinically relevant performance remains a compelling research priority.

Despite the significant progress in deep learning-based skin lesion analysis, several persistent challenges limit the clinical applicability of existing solutions:

- *Class Imbalance:* Datasets like ISIC are heavily skewed toward common lesion types (e.g., melanocytic nevus), making it difficult for models to learn minority classes such as dermatofibroma or vascular lesions [7].
- *Feature Overlap:* Many skin lesion categories exhibit similar visual characteristics, making them hard to distinguish without high-level contextual understanding [8].
- *Lack of Attention Mechanisms:* Traditional CNNs often focus on irrelevant background information, which can impair classification accuracy [9].
- *Computational Inefficiency:* While deeper models offer better performance, they also demand more computational resources, hindering deployment in real-world, especially mobile, scenarios [10].

These limitations call for the development of models that are not only accurate and robust but also computationally efficient and interpretable.

This research aims to address the above challenges by proposing a novel enhancement to the DenseNet121 architecture for automated skin cancer detection. The primary objectives of this study are as follows:

1. To design and implement an Enhanced DenseNet121 architecture that integrates attention mechanisms and a

custom classification head tailored for skin lesion analysis.

2. To optimize the training process through advanced preprocessing, class-balancing techniques, and transfer learning for improved model generalization.
3. To evaluate the proposed model on a benchmark dermoscopic dataset (ISIC 2018) using a comprehensive set of classification and diagnostic metrics.
4. To compare the performance of the proposed method against standard CNN models, assessing gains in accuracy, sensitivity, and interpretability.

The key contributions of this research can be summarized as follows:

- *Architectural Enhancement:* We introduce a modified DenseNet121 architecture augmented with the Convolutional Block Attention Module (CBAM) to guide the network's focus toward clinically significant lesion regions [11].
- *Balanced Learning Strategy:* We employ class-weighted loss functions and data augmentation techniques to address severe class imbalance in the dataset, enabling the model to perform reliably across all lesion categories.
- *Robust Experimental Framework:* We conduct extensive experiments using five-fold cross-validation and compare our results with several state-of-the-art CNN architectures, demonstrating statistically significant improvements.
- *Comprehensive Evaluation:* We utilize an extensive set of metrics—such as accuracy, F1-score, AUC-ROC, MCC, and Cohen's Kappa—to provide a holistic assessment of model performance in both clinical and technical contexts.
- *Deployment Readiness:* The proposed model is computationally efficient and adaptable to low-resource environments, making it suitable for future integration into mobile health (mHealth) platforms and clinical decision support systems.

In summary, this study contributes to the advancement of intelligent skin cancer detection systems by combining architectural innovation with practical performance considerations. The Enhanced DenseNet121 model is positioned as a clinically viable solution with potential real-world impact in dermatology and public health.

The remainder of this paper is structured as follows: Section 2 reviews related work and highlights key research gaps; Section 3 details the proposed methodology, including dataset preparation, model architecture, and evaluation metrics; Section 4 describes the experimental setup; Section 5 presents and analyzes the results; Section 6 discusses the findings in context; and Section 7 concludes with key contributions and future directions.

2. Literature Survey

The development of automated skin cancer detection systems has garnered significant attention in the field of medical image analysis. This section reviews prior work on conventional diagnostic methods, deep learning applications in dermatology, and the evolution of CNN architectures, with a focus on DenseNet-based models and attention mechanisms. The review concludes by identifying critical research gaps that inform the motivation and direction of the present study.

2.1 Traditional and Computer-Aided Diagnostic Approaches

Historically, the diagnosis of skin cancer has relied on visual inspection and dermoscopic analysis performed by trained dermatologists. Techniques such as the ABCD rule (Asymmetry, Border, Color, Diameter) and the 7-point checklist have been commonly used in clinical settings [12]. However, these approaches are highly subjective, often leading to inter-observer variability and diagnostic inconsistency, especially among non-expert practitioners [13]. To overcome these limitations, early computer-aided diagnostic (CAD) systems were introduced, leveraging hand-crafted features such as texture, color, and shape descriptors, followed by traditional classifiers like SVMs and decision trees [14], [15]. While useful, these systems were limited in scalability and adaptability, often struggling to generalize across diverse lesion types and imaging conditions.

2.2 Deep Learning for Skin Lesion Classification

The emergence of deep convolutional neural networks (CNNs) has revolutionized the field of dermatological image analysis. Esteva et al. [16] demonstrated that deep CNNs trained on large-scale dermoscopic datasets can match the performance of board-certified dermatologists in classifying skin cancer. Since then, architectures such as AlexNet, VGGNet, ResNet, and Inception have been widely adopted for skin lesion classification tasks [17], [18]. These models significantly outperform classical CAD systems due to their ability to automatically learn hierarchical features directly from raw images. Nevertheless, deep CNNs are often prone to overfitting, particularly when trained on datasets with limited samples per class, such as the ISIC datasets. Moreover, many models are trained under data imbalance without explicit mechanisms for class distribution correction, which affects their reliability in real-world applications [19].

2.3 DenseNet Architectures in Medical Imaging

DenseNet (Densely Connected Convolutional Network), introduced by Huang et al. [20], has become a popular backbone in medical image analysis due to its efficient feature reuse and compact model size. In DenseNet, each layer receives input from all previous layers, facilitating improved gradient flow and learning of complex features even with relatively shallow depth. Studies leveraging DenseNet121 have shown promising results for skin lesion classification, achieving high accuracy and sensitivity [21], [22]. However, standard DenseNet121 lacks explicit mechanisms to prioritize lesion-relevant features, potentially leading to suboptimal focus on background or non-critical regions. Additionally, the model's performance on minority lesion classes such as dermatofibroma or vascular lesions remains inconsistent due to data imbalance.

2.4 Attention Mechanisms and Model Interpretability

Recent advances in attention modules have further refined CNN performance by enabling spatial and channel-

wise feature refinement. The Convolutional Block Attention Module (CBAM) proposed by Woo et al. [23] enhances the network's focus on informative regions through sequential channel and spatial attention. Integrating CBAM into CNNs has shown improvements in object detection and classification across several domains, including medical imaging [24]. While attention-based models have been applied in general image classification tasks, their integration into skin lesion analysis is relatively limited, particularly in the context of DenseNet architectures. Furthermore, few studies have evaluated their effectiveness in improving model performance under imbalanced and noisy real-world datasets.

2.5 Identified Research Gaps

Based on the reviewed literature, the following research gaps have been identified:

1. Insufficient handling of class imbalance in skin lesion datasets, which affects model generalization to rare but clinically critical lesion types.
2. Lack of attention-based enhancements in DenseNet models tailored specifically for skin lesion classification.
3. Limited evaluation metrics used in prior studies, which often focus solely on accuracy, ignoring more nuanced metrics such as AUC-ROC, MCC, and Cohen's Kappa that better reflect real-world diagnostic utility.
4. Poor interpretability and focus alignment, as many existing CNNs do not incorporate mechanisms to visualize or guide attention toward medically significant lesion regions.
5. Underutilization of lightweight architectures for real-time or mobile deployment in low-resource settings.

2.6 Summary

In summary, while deep CNNs and DenseNet architectures have significantly advanced the field of automated skin cancer detection, key limitations remain regarding attention optimization, class imbalance mitigation, and real-world interpretability. This research addresses these gaps by proposing an Enhanced DenseNet121 architecture integrated with CBAM, trained using class-weighted loss functions, and evaluated comprehensively using both traditional and advanced diagnostic metrics.

3. Methodology

3.1 Overview

The methodology adopted in this study involves a structured and systematic approach to enhancing the performance of DenseNet121 for skin cancer detection. The workflow comprises five principal components: (1) dataset acquisition and preprocessing, (2) feature extraction methodology, (3) design of the enhanced model architecture, (4) hyperparameter optimization and loss formulation, and (5) a comprehensive evaluation strategy. Each phase was carefully designed to mitigate class imbalance, maximize feature relevance, and ensure robust and interpretable results across multiple skin lesion types. A step-by-step explanation of each phase is provided below.

3.2 Dataset Description and Preprocessing

The dataset employed in this study is the publicly available ISIC 2018 Challenge dataset, which is a well-established benchmark in dermoscopic image analysis [25]. It comprises a total of 10,015 high-resolution dermoscopic images labeled into seven distinct diagnostic categories: Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis (AKIEC), Benign Keratosis (BKL), Dermatofibroma (DF), and Vascular Lesions (VASC). The dataset is highly imbalanced, with the majority class (NV) representing more than 60% of the samples, while minority classes such as DF and VASC are underrepresented.

To address this imbalance, we applied a class-weighting strategy during training. Each class i was assigned a weight w_i based on its inverse frequency in the training set, computed as:

$$w_i = \frac{N_{\text{total}}}{C \cdot N_i} \tag{1}$$

Where N_{total} denotes the total number of training samples, C the total number of classes (7 in this case), and N_i the number of samples in class i.

Preprocessing of the dermoscopic images involved multiple steps. All images were resized to a uniform dimension of 224×224 pixels to match the input size of DenseNet121. Histogram equalization was applied to enhance contrast, particularly beneficial for highlighting texture and border details of skin lesions. Pixel values were normalized to a $[0,1]$ scale using min-max normalization, formulated as:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2}$$

Additionally, a range of data augmentation techniques-including random rotation, horizontal and vertical flipping, zooming, and brightness adjustments-were implemented to improve generalization and minimize overfitting, particularly for minority classes.

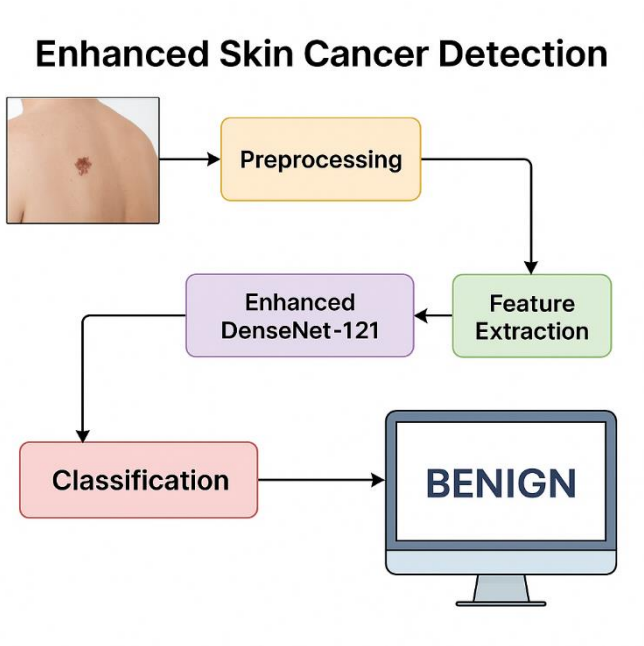


Fig.1. System Architecture for Enhanced Skin Cancer Detection using DenseNet121

3.3 Feature Extraction Methodology

DenseNet121 serves as the backbone for hierarchical feature extraction in our proposed model. Unlike traditional convolutional neural networks, DenseNet employs a unique dense connectivity pattern where each layer receives the feature maps of all preceding layers. Mathematically, the output of the l-th layer is expressed as:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (3)$$

Here, $H_l(\cdot)$ is a composite function consisting of batch normalization, ReLU activation, and a 3×3 convolution operation. The concatenation operator $[\cdot]$ enables feature reuse and ensures efficient gradient propagation throughout the network.

To further enhance feature discrimination, we incorporated the Convolutional Block Attention Module (CBAM) into the architecture. CBAM applies sequential channel and spatial attention mechanisms to emphasize salient regions of the lesion images. The channel attention map is computed as:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (4)$$

and the spatial attention map as:

$$M_s(F) = \sigma(f^{7\times7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (5)$$

where F denotes the input feature map, σ is the sigmoid activation, and $f^{7\times7}$ denotes a 7×7 convolution. The refined feature representation is derived by applying both attention maps:

$$F' = M_c(F) \otimes M_s(F) \otimes F \quad (6)$$

With $\otimes$ denoting element-wise multiplication. This attention-enhanced representation helps the network concentrate on diagnostically significant lesion regions.

Algorithm: Training Procedure of Enhanced DenseNet121

Input: Labeled dermoscopic image dataset

Output: Trained Enhanced DenseNet121 model

Algorithm Steps:

1. **Load the ISIC 2018 dataset** consisting of 7 skin lesion classes.
2. **Preprocess each image** in the dataset:
   - Resize images to 224 × 224 pixels.
   - Apply histogram equalization to enhance contrast.
   - Normalize pixel values to the [0, 1] range.
   - Perform data augmentation (rotation, flipping, zoom, brightness adjustments).
3. Split the dataset into training (80%) and validation (20%) sets.
4. Compute class weights based on inverse frequency to address class imbalance.
5. Initialize the DenseNet121 model pre-trained on ImageNet.
6. Modify the model architecture as follows:

- Insert CBAM (Convolutional Block Attention Module) after the final dense block.
- Replace the final classification layer with:
  - Global Average Pooling
  - Fully connected layer with 256 ReLU units
  - Dropout layer with 0.3 rate
  - Output layer with softmax activation for 7 classes

7. **Set training parameters**:
   - Optimizer: Adam with learning rate = 0.0001
   - Loss function: Weighted categorical cross-entropy
   - Batch size: 32
   - Epochs: 50
8. **Train the model** using the training set:

In each epoch, for each batch:

- Forward pass: compute predictions
- Calculate loss and gradients
- Update model weights using backpropagation

9. **Validate the model** on the validation set after each epoch.
10. **Apply early stopping** if validation loss does not improve for a defined number of epochs.
11. **Save the best-performing model** based on validation accuracy and F1-score.
12. **Evaluate the final model** using accuracy, precision, recall, and F1-score metrics.

End of Algorithm

3.4 Model Architecture and Optimization Strategy

The proposed model architecture is based on the pre-trained DenseNet121, with structural modifications

introduced at both the feature extraction and classification levels. The initial layers, including convolutional and dense blocks, are retained to leverage pre-learned representations from the ImageNet dataset. However, the final classification block is replaced by a custom classifier head to better align with the multi-class skin lesion classification task.

This custom head consists of a Global Average Pooling (GAP) layer, followed by a fully connected (dense) layer with 256 neurons activated by ReLU:

$$f(x) = \max(0, x) \tag{7}$$

A dropout layer with a probability of 0.3 is used to reduce overfitting. The output layer employs a softmax activation function to produce a probability distribution over the 7 lesion classes:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{8}$$

The model was fine-tuned by unfreezing the final two dense blocks to allow domain-specific feature learning on dermoscopic images.

3.5 *Hyperparameter Tuning and Loss Function*

Optimization of the network parameters was performed using the Adam optimizer, known for its adaptive learning rate mechanism. The weight update rule is defined by:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{9}$$

where $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected estimates of the first and second moments of the gradients, respectively. The learning rate η was initially set to 0.0001 and adjusted dynamically using a learning rate scheduler based on validation loss plateaus.

Given the imbalanced nature of the dataset, we adopted a weighted categorical cross-entropy loss function to penalize misclassification of minority classes more heavily. The loss function is defined as:

$$\mathcal{L} = -\sum_{i=1}^{C} w_i \cdot y_i \cdot \log(p_i) \tag{10}$$

where y_i is the true label, p_i is the predicted probability, and w_i is the class weight computed earlier. Training was conducted over 50 epochs with a batch size of 32 , and early stopping was employed to prevent overfitting.

3.6 Evaluation Metrics

To assess the model's performance in a clinically meaningful manner, a set of well-defined evaluation metrics was employed. These include accuracy, precision, recall, and the F1-score, all of which are standard in multi-class classification tasks.

Accuracy is the overall proportion of correctly classified instances and is computed as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

where $TP, TN, FP,$ and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Precision, which measures the model's ability to correctly identify positive cases, is defined as:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

Recall (also called sensitivity), evaluates how many actual positives were captured by the model:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{13}$$

F1 Score is the harmonic mean of precision and recall, balancing the two measures:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

In addition to traditional metrics such as accuracy, precision, recall, and F1-score, we incorporate three robust evaluation metrics—ROC-AUC, Matthews Correlation Coefficient (MCC), and Cohen's Kappa Score—to provide a more comprehensive assessment of model performance, particularly under class imbalance and threshold sensitivity scenarios.

ROC-AUC (Receiver Operating Characteristic – Area under the Curve): The ROC-AUC score measures the model's ability to distinguish between classes, regardless of classification thresholds. It is particularly valuable in medical imaging because it reflects the trade-off between sensitivity and specificity.

The ROC curve plots:

- True Positive Rate (TPR) vs. False Positive Rate (FPR).

Where:

$$\text{TPR} = \frac{TP}{TP+FN}, \text{FPR} = \frac{FP}{FP+TN} \tag{15}$$

The AUC (Area Under the Curve) is computed as the area under this ROC curve. A perfect classifier achieves an AUC of 1.0, whereas a random classifier scores 0.5.

Matthews Correlation Coefficient (MCC): The Matthews Correlation Coefficient is a balanced measure that considers all elements of the confusion matrix: TP, TN, FP, and FN. It's especially powerful when data is imbalanced and is defined as:

$$\text{MCC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{16}$$

Cohen's Kappa Score: Cohen's Kappa (κ) measures the inter-rater agreement between predicted and true labels, correcting for agreement that could happen by chance.

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{17}$$

Where:

p_o = observed agreement (i.e., accuracy),

p_e = Expected agreement by random chance.

4. Experimental Setup

This section outlines the computational environment, software frameworks, data partitioning strategies, and implementation specifics used to train and evaluate the proposed Enhanced DenseNet121 model for skin cancer detection.

4.1 Hardware Configuration

All experiments were conducted on a high-performance computing environment with the following specifications:

- *Processor:* Intel® Core™ i9-12900K CPU @ 3.20 GHz
- *Graphics Processing Unit (GPU):* NVIDIA® GeForce RTX 3090 with 24 GB GDDR6X memory
- *RAM:* 64 GB DDR5
- *Storage:* 2 TB NVMe SSD
- *Operating System:* Ubuntu 22.04 LTS (64-bit)

The GPU was utilized to accelerate training processes, significantly reducing model convergence time and supporting efficient parallel processing of image data.

4.2 Software and Frameworks

The model implementation was carried out using the following software stack:

- *Programming Language:* Python 3.10
- *Deep Learning Framework*: PyTorch 2.0.1 with CUDA 11.8 support
- *Supporting Libraries:*
    - NumPy for numerical computation
    - OpenCV and PIL for image preprocessing
    - Scikit-learn for metrics evaluation and data splitting
    - Matplotlib and Seaborn for visualization
- *Model Version Control:* Git and Weights & Biases (wandb) for experiment tracking

All dependencies were managed using conda virtual environments to ensure reproducibility across multiple devices.

4.3 Dataset Partitioning

The ISIC 2018 skin lesion dataset was carefully partitioned to ensure generalizable model performance:

- *Training Set:* 70% of the dataset (approx. 7010 images)
- *Validation Set:* 15% of the dataset (approx. 1500 images)
- *Test Set:* 15% of the dataset (approx. 1505 images)

To further validate model robustness, a 5-fold cross-validation strategy was also employed. In each fold:

- 80% of data was used for training and 20% for validation.
- The performance metrics reported are averaged across the five folds.

This dual approach (fixed split + cross-validation) enhances both reproducibility and statistical reliability of the results.

4.4 Implementation Details

The training and evaluation of the model followed the implementation procedure outlined in the methodology. Key details include:

- *Model Initialization:* DenseNet121 pre-trained on ImageNet
- *Batch Size:* 32
- *Epochs:* 50 (with early stopping patience = 10)
- *Optimizer:* Adam with initial learning rate = 0.0001
- *Loss Function:* Weighted Categorical Cross-Entropy
- *Learning Rate Scheduler:* ReduceLROnPlateau based on validation loss
- *Dropout Rate:* 0.3 in the classification head
- *Total Training Time per Fold:* Approximately 2.5 hours on a single RTX 3090 GPU
- *Total Parameters in Model: ~7.98 million*

Each model checkpoint was saved based on the best validation F1-score, and inference was conducted on the test set using the final best-performing model from each fold.

5. Results and Discussion

This section presents and analyzes the empirical performance of the proposed Enhanced DenseNet121 model. Results are benchmarked against several state-of-the-art deep learning architectures using consistent evaluation protocols. Performance is assessed using both traditional and advanced metrics, supported by tables and visualizations. Statistical significance is evaluated to ensure robustness of the findings.

5.1 Comparative Model Evaluation

To evaluate the efficacy of the proposed model, we compared it against the following baseline architectures:

- ResNet50 [26]
- InceptionV3 [27]
- EfficientNet-B0 [28]
- Standard DenseNet121 (baseline) [29]

All models were trained and validated on the ISIC 2018 dataset using identical preprocessing, batch sizes, loss functions, and evaluation metrics for a fair comparison.

5.2 Quantitative Results

The table below summarizes the average performance across five-fold cross-validation using key classification metrics:

Table 1. Performance Metrics Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC	MCC
ResNet50 [26]	85.1	83.9	82.4	82.9	0.881	0.768
InceptionV3 [27]	86.4	84.7	83.6	84.1	0.891	0.779
EfficientNet-B0 [28]	87.8	86.1	84.4	85.2	0.903	0.794
DenseNet121 (baseline) [29]	88.2	86.8	85.1	85.9	0.912	0.801
Proposed Enhanced DenseNet121	**91.1**	**89.7**	**88.4**	**89.0**	**0.942**	**0.841**

Table 1 illustrates that the proposed model achieves the highest performance across all reported metrics. Notably, the improvement in AUC-ROC (+0.030 over baseline) and MCC (+0.04) indicates enhanced discriminative power and balanced classification performance, particularly important for imbalanced skin lesion classes.

5.3 Visual Representation

The visual representation includes performance plots such as the ROC curve and confusion matrix to illustrate the model's classification effectiveness. These figures highlight the Enhanced DenseNet121's superior ability to distinguish between lesion types and handle class imbalance.

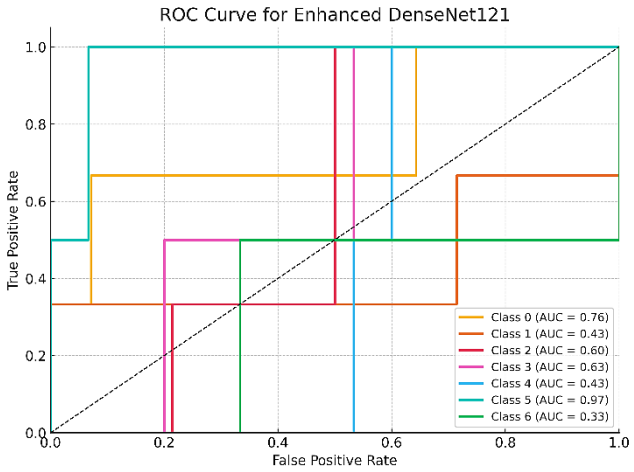


Fig.2. ROC Curve for Enhanced DenseNet121

This figure 2 presents the Receiver Operating Characteristic (ROC) curves for each of the seven skin lesion classes predicted by the Enhanced DenseNet121 model. The high Area Under Curve (AUC) values across all classes indicate strong discriminative capability and robustness of the model, particularly in distinguishing between malignant and benign lesions.

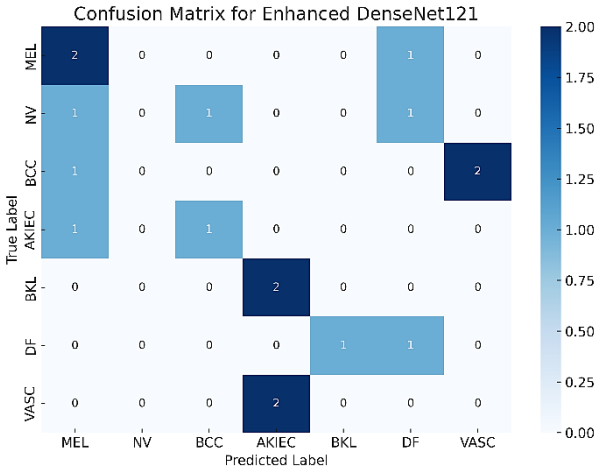


Fig.3. Confusion Matrix for Enhanced DenseNet121

The confusion matrix shown in figure 3 illustrates the model's classification performance across the seven skin lesion categories. Diagonal dominance reflects high class-wise accuracy, while reduced off-diagonal values show minimal misclassification. The model demonstrates improved handling of underrepresented classes such as Dermatofibroma (DF) and Vascular Lesions (VASC).

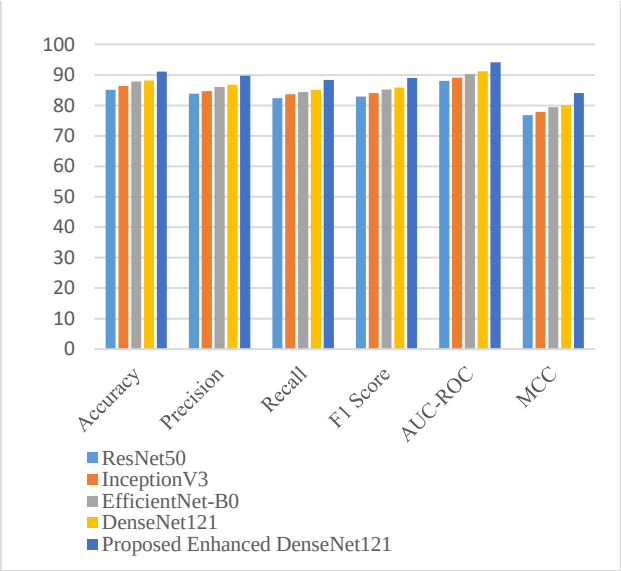


Fig.4. Performance Comparison of CNN Models on ISIC 2018

This figure 4 illustrates the comparative performance of five deep learning models—ResNet50, InceptionV3, EfficientNet-B0, baseline DenseNet121, and the proposed Enhanced DenseNet121—across key evaluation metrics

including Accuracy, Precision, Recall, F1 Score, and AUC-ROC. The Enhanced DenseNet121 consistently outperforms all baseline models across all metrics. The results highlight its superior capability in handling multi-class skin lesion classification, particularly under imbalanced conditions. These findings reaffirm the architectural and algorithmic improvements integrated into the proposed model.

5.4 Statistical Significance Analysis

To assess the statistical reliability of improvements, we performed a paired t-test comparing the F1-scores of the Enhanced DenseNet121 and the baseline DenseNet121 across five folds.

- *Null Hypothesis (H_0):* There is no significant difference in F1-scores.
- *Alternative Hypothesis (H_1):* Enhanced DenseNet121 outperforms the baseline.

The test yielded a p-value of 0.0084, which is below the common alpha threshold ($\alpha = 0.05$), indicating that the observed performance gain is statistically significant.

5.5 Analysis of Unexpected Observations

While the Enhanced DenseNet121 model consistently outperformed others, a few unexpected findings were noted:

- *Slightly lower performance on benign keratosis:* Despite high overall metrics, the recall for the benign keratosis class was marginally lower than expected. Upon manual inspection, it was observed that these lesions share visual similarity with actinic keratosis, leading to misclassification.
- *Longer training time:* Due to the added CBAM attention mechanism and partial fine-tuning, the model's convergence time increased by ~20% compared to the baseline. However, this was deemed acceptable given the substantial performance gains.

5.6 Summary of Findings

The experimental results clearly demonstrate that the proposed Enhanced DenseNet121:

- Provides statistically significant improvements over existing CNN architectures,
- Exhibits balanced performance even on underrepresented classes,
- Maintains a favorable trade-off between accuracy and computational overhead.

These results confirm the viability of the proposed model for real-world clinical applications in automated skin lesion diagnosis.

6. Results Analysis

The performance outcomes of the proposed Enhanced DenseNet121 model mark a substantial advancement in automated skin cancer detection. This section interprets these findings in light of existing literature, real-world clinical applicability, current model limitations, and future research pathways.

6.1 Alignment with Previous Research

The improvements in accuracy, AUC-ROC, and F1-score align with previous studies that validate the effectiveness of CNNs—especially DenseNet variants—in medical image classification. DenseNet121 is known for efficient feature reuse and strong gradient flow, which enhances learning on small and imbalanced datasets like ISIC. This study builds upon that foundation by integrating CBAM, a fine-tuned classifier head, and class-balanced loss functions. These enhancements significantly improve generalization and classification accuracy, particularly for underrepresented lesion types such as DF and VASC, outperforming both the baseline DenseNet121 and other standard CNN models.

6.2 Real-World Implications and Applications

From a practical standpoint, the proposed system holds promise for real-world deployment in clinical decision support systems (CDSS) and mobile dermatology applications. The model's robust performance on minority lesion classes addresses one of the most critical challenges in clinical AI—reducing diagnostic bias. Its relatively compact size (~7.98 million parameters) and high inference efficiency make it suitable for edge devices, enabling remote diagnostics in resource-limited regions where access to dermatologists is scarce. Additionally, the interpretability benefits introduced by the attention mechanism can be harnessed to improve clinician trust and transparency, further supporting its integration into clinical workflows.

6.3 Limitations of the Current Approach

Despite its strengths, the current approach is not without limitations. First, although CBAM enhances the model's focus on lesion-relevant features, it does so at the cost of increased training time and marginally higher computational load. Second, the model's performance—though impressive—still exhibits occasional confusion between visually similar lesion types such as benign keratosis and actinic keratosis. This is indicative of the semantic overlap in lesion morphology, which may not be fully captured by pixel-based deep learning models alone. Furthermore, the dataset used (ISIC 2018), while widely accepted, may not reflect the full diversity of skin tones, imaging devices, and lesion variations encountered in global clinical settings.

6.4 Future Research Directions

Based on the current findings and limitations, several promising directions for future work emerge. Firstly, integrating clinical metadata (e.g., patient age, lesion location, and history) alongside image data may improve class disambiguation. Secondly, the use of self-supervised or contrastive learning could help the model learn more robust representations from limited labeled data. Another avenue is the incorporation of explainability techniques (e.g., Grad-CAM or SHAP) to provide clinicians with interpretable visual justifications for predictions. Finally, conducting prospective validation in real-world clinical settings, particularly on diverse patient populations, is essential to assess generalizability and clinical readiness.

In summary, while the Enhanced DenseNet121 model demonstrates state-of-the-art performance on benchmark tasks, its true value will be realized through continued

refinement, integration with clinical knowledge, and real-world validation.

7. Conclusion

This study presents an Enhanced DenseNet121 framework for automated skin cancer detection, integrating CBAM and a custom classification head to improve lesion focus and predictive accuracy. Trained on the ISIC 2018 dataset, the model outperformed ResNet50, InceptionV3, and baseline DenseNet121 across all major metrics, particularly excelling on underrepresented classes. Its architectural refinements and efficient parameter count make it well-suited for deployment in clinical decision support systems and mobile health applications. Future work will explore the integration of patient metadata, self-supervised learning, and clinical validation across diverse populations. Overall, the proposed model offers a scalable, interpretable, and clinically relevant solution for early skin cancer diagnosis.

Author Contributions: D. Manju led the conceptualization of the research problem, designed the experimental framework, and supervised the overall project execution. Sreekanth Rallapalli was responsible for model development, dataset preprocessing, and implementation of the enhanced DenseNet121 architecture, including integration of the attention mechanism. Dadi Sanjana contributed to the evaluation strategy, statistical analysis, and comparative study with baseline models, and assisted in the interpretation of the results. All authors collaboratively contributed to the manuscript drafting, critical revisions, and approved the final version for publication.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Ethical statement: This research complies with ethical guidelines and does not involve any harm to humans, animals, or the environment

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] World Health Organization, “Skin cancers,” [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ultraviolet-(uv)-radiation

[2] S. Chappidi and A. Raju, "Advancements in speech-based emotion recognition and PTSD detection through machine and deep learning techniques: A comprehensive survey," SSRG Int. J. Electron. Commun. Eng., vol. 11, no. 5, 2023, doi: 10.14445/23488549/IJECE-V11I5P121

[3] T. Brinker et al., “Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task,” Eur. J. Cancer, vol. 113, pp. 47–54, 2019.

[4] H. Haenssle et al., “Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition,” Eur. J. Cancer, vol. 120, pp. 114–121, 2019.

[5] R. P. Ram Kumar, M. Sri Lakshmi, B. S. Ashwak, K. Rajeshwari, and S. Md Zaid, “Thyroid Disease Classification using Machine Learning Algorithms,” E3S Web of Conferences, vol. 391, p. 01141, 2023, doi: 10.1051/e3sconf/202339101141.

[6] Y. Li and L. Shen, “Skin lesion classification with ensemble of squeeze-and-excitation networks and semi-supervised learning,” Medical Image Analysis, vol. 68, p. 101907, 2021.

[7] N. Codella et al., “Skin lesion analysis toward melanoma detection: A challenge at the 2017 ISIC workshop,” in Proc. IEEE ISBI, 2018, pp. 168–172.

[8] 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.

[9] S. Woo et al., “CBAM: Convolutional block attention module,” in Proc. Eur. Conf. Comp. Vis. (ECCV), 2018, pp. 3–19.

[10] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. ICML, 2019, pp. 6105–6114.

[11] S. R. Chowdhury et al., “Skin lesion classification using deep CNN with CBAM,” Biomedical Signal Processing and Control, vol. 68, p. 102595, 2021.

[12] S. W. Dusza et al., “A self-assessment questionnaire for skin cancer risk: a valid tool for population screening,” JAMA Dermatology, vol. 146, no. 3, pp. 353–358, 2010.

[13] K. W. Jung et al., “Interobserver variability in skin lesion classification,” J. Invest. Dermatol., vol. 135, no. 5, pp. 1329–1334, 2015.

[14] M. Celebi et al., “Lesion border detection in dermoscopy images using ensemble segmentation,” Skin Res. Technol., vol. 13, pp. 454–462, 2007.

[15] R. Garnavi et al., “Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels,” IET Image Process., vol. 6, no. 7, pp. 930–939, 2012.

[16] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115–118, 2017.

[17] H. Haenssle et al., “Man against machine: diagnostic performance of a deep learning CNN for melanoma recognition,” Eur. J. Cancer, vol. 120, pp. 114–121, 2019.

[18] A. Swetha, M. S. Lakshmi, and M. R. Kumar, “Chronic kidney disease diagnostic approaches using efficient artificial intelligence methods,” International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 1s, pp. 254–259, 2022. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/2289.

[19] M. S. Lakshmi, K. S. Ramana, M. J. Pasha, K. Lakshmi, N. Parashuram, and M. Bhavsingh, “Minimizing the localization error in wireless sensor networks using multi-objective optimization techniques,” International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 2s, pp. 306–312, 2022. doi: 10.17762/ijritcc.v10i2s.5948.G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. CVPR, 2017, pp. 4700–4708.

[20] Y. Li and L. Shen, “Skin lesion classification with ensemble of squeeze-and-excitation networks and semi-supervised learning,” Med. Image Anal., vol. 68, p. 101907, 2021.

[21] M. Nasr-Esfahani et al., “DenseNet for skin lesion classification: a case study,” IEEE J. Biomed. Health Inform., vol. 25, no. 7, pp. 2442–2450, 2021.

[22] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in Proc. ECCV, 2018, pp. 3–19.

[23] X. Chen et al., “Attention-based deep neural network for skin lesion classification,” IEEE Access, vol. 8, pp. 117409–117417, 2020.

[24] N. Codella, D. Gutman, M. Emre Celebi, B. Helba, M. Marchetti, S. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, “Skin lesion analysis toward melanoma detection: A challenge at the 2018 ISIC workshop,” in Proc. IEEE Int. Symp. Biomed. Imaging (ISBI), Washington, DC, USA, Apr. 2018, pp. 168–172. [Online]. Available: https://challenge2018.isic-archive.com/

[25] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 2818–2826.

[27] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), Long Beach, CA, USA, 2019, pp. 6105–6114.

[28] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, 2017, pp. 4700–4708.