# Skin Cancer Detection: Image Classification Using CNN Architectures with CLAHE

MSc Research Project
Data Analytics

Syed Munazir Shajahan

School of Computing
National College of Ireland

Supervisor: Prof. Jaswinder Singh

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

**Student Name:** Syed Munazir Shajahan

**Student ID:**

**Programme:** MSc in Data Analytics **Year:** 2023 - 2024

**Module:** MSc Research Project

**Supervisor:** Prof. Jaswinder Singh

**Submission Due Date:** 29/01/2025

**Project Title:** Skin Cancer Detection: Image Classification Using CNN Architectures with CLAHE

**Word Count:** **7241** **Page Count:** 22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** Syed Munazir Shajahan

**Date:** 29/01/2025

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Skin Cancer Detection: Image Classification Using CNN Architectures with CLAHE

Syed Munazir Shajahan
x23103914

**Abstract**

Skin cancer is a significant health problem that is common and can be improved if recognized early and avoided. In this paper, the authors examine Contrast Limited Adaptive Histogram Equalization (CLAHE), which is applicable for changing the images to enhance their contrast, to improve the efficiency of various Convolutional Neural Networks (CNNs) in detecting various skin lesions. They used the HAM10000 dataset with 10,015 skin images for testing the three implemented CNNs ResNet, DenseNet, and EfficientNet with raw and CLAHE applied images. The first case results showed that CLAHE works well with models in detecting skin features and increases classification accuracy. The DenseNet turned out to be the one with the highest performance of 77% accuracy and recall when given CLAHE-enhanced images, whereas ResNet came up to 75% accuracy; however, the precision revealed a poorer performance compared to the others for EfficientNet. The study highlights the good use of CLAHE as the best means of improving the detection of skin cancer by means of CNN models, namely DenseNet for most effective results. New image enhancement methods and larger datasets will be discussed in future work.

**Keywords:** Skin cancer, CNN, CLAHE Technique, Image Classification, Deep Learning

# 1 Introduction

## 1.1 Background and Motivation

Skin cancer is one most dangerous cancer around the world and it has different varieties, such as melanoma, basal cell carcinoma, squamous cell carcinoma, etc,. It is one of the most prevalent types of cancers worldwide. The early detection of cancer improves outcomes and decreases costs associated with late-stage treatments. Despite these ongoing advances, skin cancer diagnosis often requires purely visual examinations combined with biopsies, which may be further constrained by subjectivity, accessibility, and time consumption. The growing demand for precision in early diagnosis creates a high relevance of using technological advancements in medical imaging and the analysis of data. CNNs have remarkable potential in medical image classification, since their core role is to analyze visual data at high accuracy (Georgios Kourounis et al., 2023). Through processing image data in multi-tiered architectures, CNNs can extract patterns and features in images that may be quite hard to observe by eye. CNN architectures such as ResNet, DenseNet, and EfficientNet are known for their ability to analyze large datasets with high levels of detail efficiently while minimizing overfitting, hence being important for tasks in medical imaging (Alzubaidi et al., 2021)

In this project, the implementation of CLAHE, an image enhancement technique, is investigated together with CNN architectures to check whether improved image contrast indeed enhances the classification results. CLAHE adapts local image contrast adaptively and may enhance the visibility of skin lesions, thus helping the CNN models identify relevant features for classification. Given the significant role that input quality plays in model performance, the motivation for this study lies in assessing whether CLAHE, as a preprocessing technique, can optimize CNN classification model evaluation (accuracy, recall, precision, F1 score) in skin cancer detection and streamline the diagnostic process.[1]

## 1.2 Research Question and Objectives

The proposed research question:
"How does CLAHE (image enhancement technique) impacts the performance of the CNN models (ResNet, DenseNet, EfficientNet) in skin cancer detection, and which approach - using raw images or CLAHE gives better results in terms of recall, precision, accuracy, and F1-score?"

The three main objectives of this project are,

- First objective, the dataset is more than 10,000 of skin lesions and about seven classes, it needs loading and feeding into CNN models with adding all the necessary image loading, data cleaning & transformations. Further, it's plan to build a code to run the model training, building and evaluation seamlessly.
- The second objective is the training and evaluation of ResNet, DenseNet, and EfficientNet using both - Raw Images and CLAHE Images (two approaches). This will provide a comparison to determine whether image enhancement offers measurable improvement in the classification outcomes and which architecture performs the best under each preprocessing condition.
- The third objective is to compare the performance of the models with respect to accuracy, precision, recall, and F1 score. Indeed, these metrics serve to clarify the capabilities of the models in relation to skin cancer classification. The study should identify any improvements in enhanced performance of the CNN architectures compared to the results produced by both raw and CLAHE data.

[1] https://academic.oup.com/pmj/article/99/1178/1287/7289070
2 https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8

## 1.3 Contribution

CNNs have been used widely for medical imaging, but few studies have addressed how specific preprocessing techniques, such as CLAHE, impact the classification of skin cancer detection. This research, by incorporating CLAHE as an enhancement step, assesses if contrast adjustments can optimize the performance of CONs in increasing diagnostic performance. It assesses three CNN models-ResNet, DenseNet, and EfficientNet on raw and CLAHE preprocessed images. The current methodology permits analyzing a deep interplay between image enhancement and CNN learning processes that estimate the quality of enhanced images based on whether they would allow more effective capturing by CNN models of necessary diagnostic features. The results would shed light on which of the available methods works better on the model evaluations-enhanced with CLAHE or not. These findings could serve the future research and practical uses of dermatology through exemplifying ways of ameliorating diagnostic accuracy related to automated skin cancer detection.

# 2 Related Work

This section explores key research contributions and technological advancements relevant to skin cancer detection using deep learning. It specifically focuses on the application of Convolutional Neural Networks (CNNs), the role of image enhancement techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE), and the comparative performance of CNN architectures in medical image classification. The project provides the critical insights into how preprocessing enhances feature extraction and how architectural differences influence model performance in skin cancer detection.

## 2.1 Skin Cancer Detection Using Deep Learning

The work "Skin Cancer Detection using Deep Learning" was presented by (R. S. Kumar et al., 2022) at International Conference on Electronics and Renewable Systems (ICEARS). It discussed the application of deep learning in detecting skin cancer, especially melanoma, using image processing. The authors have pointed to the effectiveness of CNNs concerning skin lesion classification and diagnosis over medical image datasets. Because of this, deep learning models like CNN are employed, which can hierarchically learn features from the raw images themselves. The paper discusses the integration of a number of machine learning algorithms, including Support Vector Machine classification and K Means clustering, in order to improve the accuracy of skin cancer detection. The models will be trained with large datasets of skin images to categorize various skin conditions; the prime focus will remain on melanoma, as it is one of the most dangerous forms of skin cancer. In the paper "Skin Cancer Classification and Detection Using VGG-19 and DenseNet," the authors (A. Barbadekar, et al., 2023) explore the application of advanced convolutional neural networks for accurate classification and detection of skin cancer. The research work presented at ICCINS makes use of the VGG-19 and DenseNet architectures to identify skin lesions with efficacy. In this work, the authors develop an enhanced model with computational modeling and image segmentation for better detection accuracy of malignant and benign lesions. The findings of this study highlight the potential of AI-driven tools in changing dermatology.

(Azadeh Noori Hoshyar et al., 2011) presented a comprehensive review of automatic early skin cancer detection techniques in their paper published at the 2011 International Conference on Computer Science and Service System (CSSS) in Nanjing, China. The paper reviews different computational approaches and methodologies that are directed toward improving early detection of skin cancer, a crucial factor for improving survival rates. By analyzing existing technologies and their limitations, the authors bring out the potential of automation in dermatological diagnostics. (Y. Jusman et al., 2021) presented the effectiveness of Multi-Layer Perceptron and Deep Neural Networks in skin cancer classification. The paper was presented at the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), Nara, Japan. The research evaluates the performance of the machine learning models in identifying malignant and benign skin lesions. Comparing the classification accuracies by MLP and DNN, the authors show that with deeper architectures, complex image data are better handled. (K. W. Lee et al., 2020) presented the research on how data augmentation affects melanoma skin cancer prediction using CNNs for their paper presented at the 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology held in Kota Kinabalu, Malaysia**.** The study underlines the challenges brought about by imbalanced datasets in medical image analysis and shows how data augmentation techniques improve the performance of CNNs through increased diversity in training data. Their findings emphasize the importance of pre-processing strategies in enhancing prediction accuracy and provide valuable insights for developing robust AI models in dermatological diagnostics.

## 2.2 Image Enhancement Techniques for Medical Imaging

(Islam et al., 2019) presented advancements in medical image analysis through image enhancement techniques in their paper, "Image Enhancement Based Medical Image Analysis," at the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) in Kanpur, India. The study emphasizes the importance of enhancing image quality for accurate diagnostics and better visualization of medical images. It discusses contrast adjustment, noise reduction, sharpening filters, and the like that enable the better detection and classification of anomalies. (Pizer et al., 2003) reflect on the work and philosophies of the Medical Image Display and Analysis Group at the University of North Carolina in their article published in the IEEE Transactions on Medical Imaging. The paper chronicles some of the key innovations within medical imaging technologies, placing an emphasis on image display systems and their accompanying frameworks of analysis. Pizer shares insights into the group’s philosophies regarding the integration of user-friendly systems into clinical environments and their impact on medical practice. This reflective work provides a historical and philosophical perspective on the evolution of medical imaging. (Rubin et al., 2006) presented a new methodology for medical image contrast enhancement by means of histogram-adaptive segmentation during the 2006 IEEE International Conference on Information Reuse & Integration in Waikoloa, Hawaii. In this study, the image histogram is made to dynamically adapt so that regions of interest may enhance their contrast, while simultaneously suppressing noise. This method is particularly effective for medical applications where distinguishing subtle differences in tissue is critical.

(Jannin et al., 2002) explore the critical process of validating medical image processing techniques for image-guided therapy in their paper published in the IEEE Transactions on Medical Imaging. The paper reviews methodologies for the validation of image processing tools, emphasizing the need for accuracy, reproducibility, and reliability in clinical applications. Their findings contribute to the establishment of guidelines and frameworks for assessing medical image processing systems with regard to whether they meet clinical safety and efficacy requirements. (Li Jupeng et al., 2006) explore the methods of correction and calibration in biomedical image processing in their paper at the First International Conference on Innovative Computing, Information and Control, ICICIC'06, Beijing. The study researches geometric corrections and the methods of camera calibration that can be used to enhance the accuracy of biomedical imaging systems. Their work underlines the importance of precise calibration in developing high-quality medical imaging systems.

## 2.3 Comparison Analysis of CNN Architectures in Image Classification

(Kohsasih et al., 2022) present performance comparisons of several architectures of CNN for white blood cell classification in medical images. This paper was presented during the 2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM) held in Laguboti, North Sumatra, Indonesia. The authors discuss the performance of several CNN models, such as ResNet and VGGNet, with regard to their accuracy, precision, and computational efficiency. The authors have also identified that an optimal architecture of CNN is very important in medical diagnosis, where accuracy is the key. Their results have provided a way forward for the successful implementation of CNNs in healthcare contexts. (Lee et al., 2021) present an analysis of CNN architectures for emotion classification based on thermal face imaging during the 2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) held in Gangwon, Korea. By comparing AlexNet, VGGNet, and EfficientNet models, the authors identify the best architecture for thermal image processing. The study highlights the ever-increasing importance of CNNs in emotion recognition, finding applications in security, healthcare, and human-computer interaction. Their work shows the ability of CNNs to adapt to non-visual data, underlining recent developments in facial emotion classification by thermal imaging. The works of (Luthfi et al., 2023) are presented in a study on CNN for coconut ripeness level classification at the Eighth International Conference-ICIC, Manado, Indonesia, in 2023. The paper reviews several state-of-the-art CNN architectures applied to agricultural image analysis, such as MobileNet and ResNet. It thus looks at the performance of various models to identify efficient solutions to automate ripeness detection. The research highlights how CNN-based systems can assist farmers and industries by improving productivity and decision-making in agriculture through accurate image-based classification.

## 2.4 Limitations and Challenges: CLAHE and CNN Architecture

(Srikanth et al., 2024) discussed the enhancement of night-time vehicle image quality using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method. The paper outlines that poor light conditions affect vehicle imaging in terms of poor visibility and diminished detail recognition. The CLAHE method, as used in this work, effectively enhances contrast and brings out finer details necessary for applications such as license plate recognition and traffic monitoring. This paper makes a great contribution to furthering image processing techniques in low-light conditions and opens up new real-world applications for traffic management and surveillance. (Umrani et al., 2024) presented a hybrid approach using CLAHE in combination with transfer learning using VGG19 and DenseNet201 for ocular disease classification. The authors have applied CLAHE for pre-processing to enhance the contrast of the image, which will help in better feature extraction for deep learning models. Their findings have pointed out the potentiality of this hybrid approach to achieve high accuracy and reliability in diagnosing ocular diseases, emphasizing its value in medical image processing and diagnostic systems. (A. C. S. et al., 2023), he had given an extensive review about the latest developments in computer vision about CNN at Annual International Conference on Emerging Research. Evolution of CNN Architectures in the paper, presented the key evolution of state-of-the-art CNN architectures such as ResNet, DenseNet, and EfficientNet in object detection, segmentation, and classification of images. It puts weight on innovations such as the attention mechanism and lightweight models, which have improved performance in a resource-constrained environment.

(Venkatasunilsrikanth et al., 2023) present an enhanced deep CNN architecture for breast cancer detection at the Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) in Trichy, India. The study addresses the challenges of early and accurate diagnosis in breast cancer imaging, proposing a novel CNN design optimized for feature extraction and classification. Their approach has underlined the potential of CNN advancements for improving diagnostic systems to make them more reliable and efficient for clinical applications.

# 3 Research Methodology

The Methodology section provides a detailed blueprint of the approach taken to develop and assess the skin cancer detection system. This framework/system is designed to ensure that each step is methodically planned and executed to address the project's research objectives effectively. It begins with the dataset, which involves choosing an appropriate and diverse dataset that can represent real world scenarios. This is followed by data pre-processing, two approaches – first with raw dataset (Unprocessed Images) and second with CLAHE (Image Enhancement Technique). For first, the raw data will work on data cleaning, EDA, loading data, mapping images ID with meta file, data splitting and data generator for model training, building and evaluating. Finally, the section delves into the implementation of methodologies, highlighting the use of advanced CNN architectures - ResNet, DenseNet, and EfficientNet, alongside techniques like CLAHE image enhancement. By implementing this CNN methodology and CLAHE technique, this project will perform model building and model evaluation. And the project goal is to represent that – which approach gets the high score in-terms of model evaluation (Recall, Precision, Accuracy & Recall).

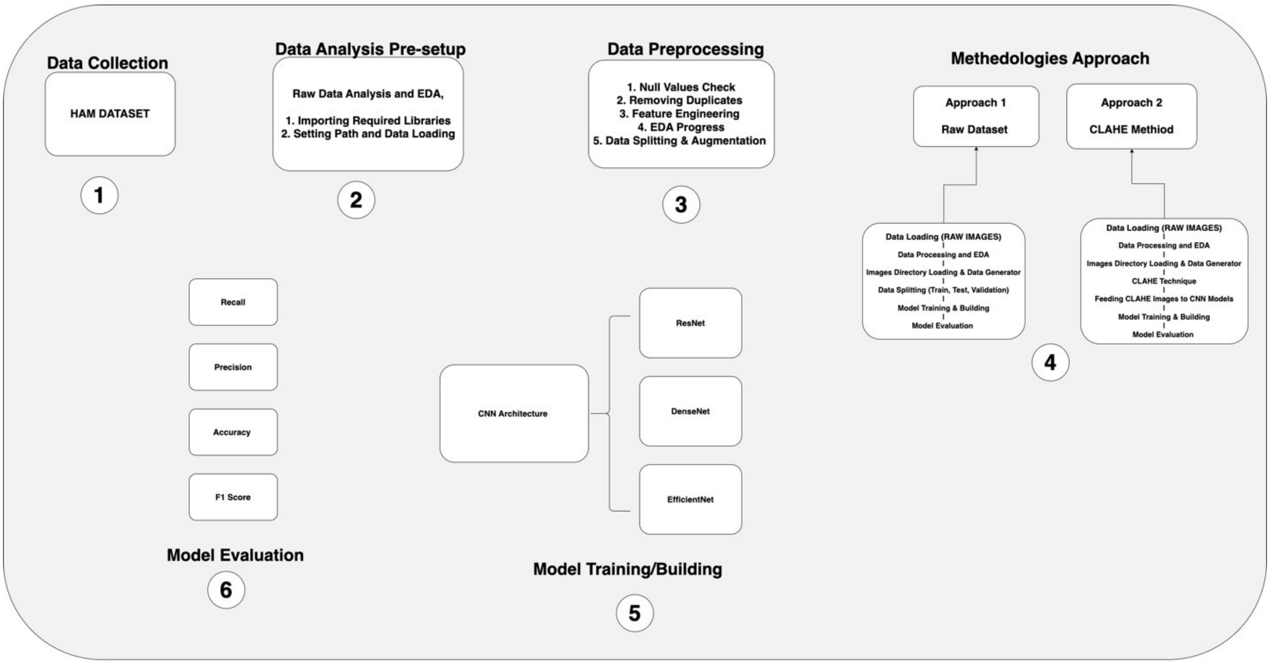


Figure 1: Research Methodology Flow

## 3.1 Dataset Selection

In choosing an appropriate dataset for this project, the HAM10000 dataset has been approved as a suitable dataset for dermatological research and medical image classification. This dataset contains about 10,015 dermoscopic images to illustrate all the various types of skin lesions that include melanoma, basal cell carcinomas, and many others. These lesions are frequently seen and clinically relevant to the diagnosis of skin cancers, making the database relevant for the research. Each image is attached to metadata containing the image ID, the type of the lesion, patient's age, and the location in the human body where it was taken, which can further be analyzed. For the purpose of processing, this dataset is divided into two sections metadata and image files (two directories). Where all metadata are kept in a CSV file, the images are properly organized into two directories.[2]

## 3.2 Data Pre-processing

Data preprocessing prepares the data before building a model in constructing a model. For the case of skin cancer detection, the main aim is to ensure uniformly high-quality consistency of images before feeding them into the CNN models for classification. The model will then have learnt patterns useful and generalizes well on unseen data.

### 3.2.1 Data Cleaning

Data cleaning is a major step in any machine learning project, particularly in medical imaging, as the quality and integrity of data can hugely affect the performance of the model. Data cleaning aims to majorly cover these areas in this project:

[3]https://www.kaggle.com/datasets/surajghuwalewala/ham1000-segmentation-and-classification

- Handling Errors: Images can sometimes be incomplete and affect the model's learning ability. If this is the case, these handling missing images can be identified and removed from the dataset.
- Validating Image Consistency: Images in a dataset should be consistent in their size and format. This ensures that each photo is adequately processed and standardized before being brought into the model.
- Duplicates Removal: Duplicates appear because of issues during data collection or pre-processing, are also removed.
- Creating Label: The dataset labels should also be consistent with the image they correspond to. Any images that are wrongly labelled would need to ensure the correct categorization of images into Melanoma, Nevus, Seborrheic Keratosis, etc., classes are important for the models' correctly trained models.

Cleaning the dataset can limit the conditions under which noise or inconsistencies cause the model to fail, thus making prediction more reliable and accurate.

```
<bound method NDFrame.head of          lesion_id      image_id     dx dx_type   age     sex localization
0      HAM_0000118  ISIC_0027419    bkl   histo  80.0    male        scalp
1      HAM_0000118  ISIC_0025030    bkl   histo  80.0    male        scalp
2      HAM_0002730  ISIC_0026769    bkl   histo  80.0    male        scalp
3      HAM_0002730  ISIC_0025661    bkl   histo  80.0    male        scalp
4      HAM_0001466  ISIC_0031633    bkl   histo  75.0    male          ear
...            ...           ...    ...     ...   ...     ...          ...
10010  HAM_0002867  ISIC_0033084  akiec   histo  40.0    male      abdomen
10011  HAM_0002867  ISIC_0033550  akiec   histo  40.0    male      abdomen
10012  HAM_0002867  ISIC_0033536  akiec   histo  40.0    male      abdomen
10013  HAM_0000239  ISIC_0032854  akiec   histo  80.0    male         face
10014  HAM_0003521  ISIC_0032258    mel   histo  70.0  female         back

[10015 rows x 7 columns]>
```

Figure 2: Head of Dataset

```
Missing Values:
lesion_id        0
image_id         0
dx               0
dx_type          0
age             57
sex              0
localization     0
dtype: int64

Number of duplicate rows: 0

Missing Values After Cleaning:
lesion_id       0
image_id        0
dx              0
dx_type         0
age             0
sex             0
localization    0
dtype: int64
```

Figure 3: Data Cleaning

### 3.2.2 CLAHE Technique

The CLAHE technique, or Contrast Limited Adaptive Histogram Equalization, was applied to enhance the dermoscopic image's quality. CLAHE works well in medical imaging; it increases sharpness in confined regions, thereby increasing contrast of complicated structures. Unlike global contrast enhancement methods, CLAHE does not cause an over-amplification of noise and clarity of the image. Using OpenCV, every grayscale image was processed with CLAHE. Clip limit and grid size parameters were optimized to achieve a best result with added artifacts. Further these images in enhanced state gave more details of lesions with respect to edges and structures, and all these contribute to an accurate classification. Separate saving of processed images resulted in a parallel dataset creation for comparative analysis to raw images. This was the step that enhanced the sensitivity of the model towards key features and enhanced accuracy in classification.

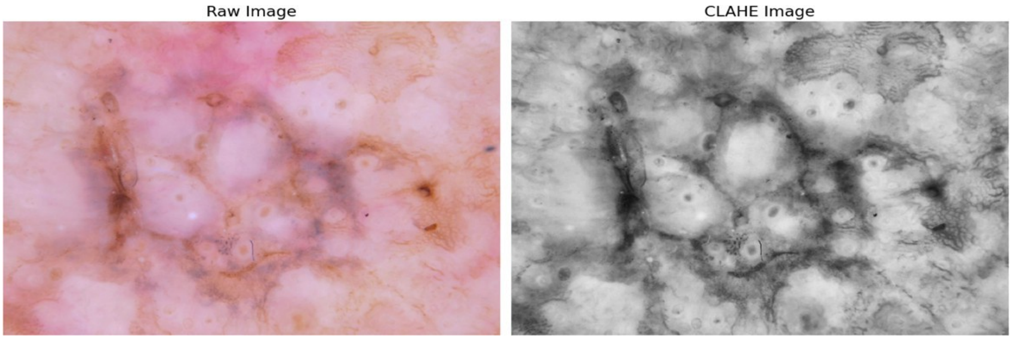


Figure 4: Comparison Raw Image VS CLAHE Image

### 3.2.3 Data Generator and Data Loading

Data loading and generators have been designed to assist in working with huge data. Images were read into directories through OpenCV, resized to dimensions that complemented the CNN architectures, and normalized to scale pixel values between 0 and 1 before splitting up the dataset into training, validation, and testing. Stratified sampling based on lesion classes was used to ensure the samples were homogenous in characteristics to cater for balance in representation. The data augmentation was done using TensorFlow with transformations such as flipping, rotation, zooming, and shearing applied to the training dataset to create more image variations. Custom data loaders accessed large image files well while maintaining fast I/O operation during the training. In total these methods guaranteed the consistent and effective presentation of the dataset to the model so as to bolster training and evaluation.

```
Training dataset size: 7010 images
Validation dataset size: 1002 images
Test dataset size: 2003 images
```

Figure 5: Loaded Images

### 3.2.4 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was applied here to get the better understanding the dataset's structure and key characteristics. A number of visualizations were developed from libraries focusing on demographic features and lesion distributions.

The patients' age distribution was visualized using a histogram with a corresponding kernel density estimation (KDE). It showed the age ranges in which epidermal lesions tend to occur most prevalently as well as the trends across different cohorts. Knowledge about the age distribution is critical for identifying particular types of lesions associated with specific ages. This sex distribution was compared using a count plot, signifying whether the enrolled patients were male or female. It gave hints to gender-specific prevalence, a vital point in the development of unbiased models. The spatial lesion localization was also captured in the dataset; this is the anatomical site where the lesion occurs. This is further illustrated with a count plot that depicted the amount of lesion occurrence across different parts of the body. This will show the most common site of lesions and could indicate some features that will be critical for an understanding of the developed model.

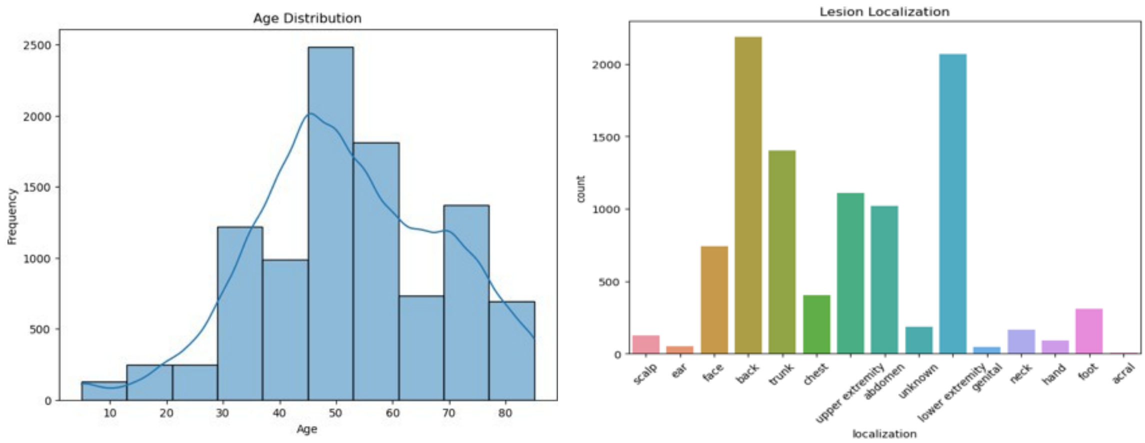


Figure 6: Age EDA

Inference boxplots showing age versus sex and age versus site of lesion were created to analyze visible relationships even further. This revealed agglomeration of lesions in different age groups and genders as well as possible age-related patterns in lesion characteristics. A grouped count plot by sex and localization also indicated how the lesion distributions differed between males and females across different body parts.

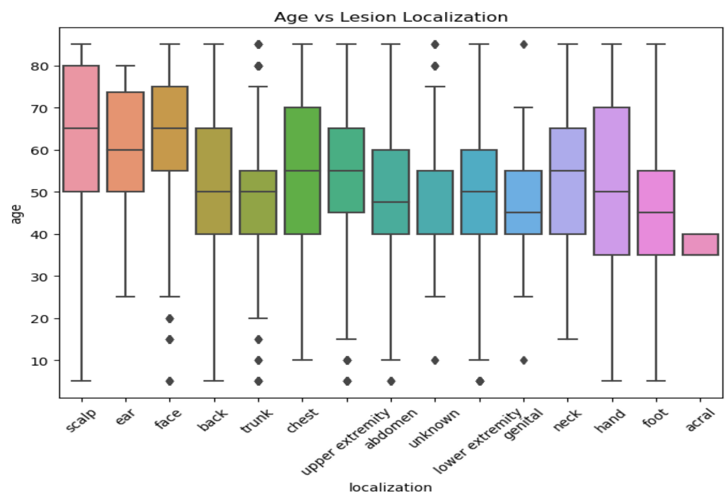


Figure 7: Age vs Lesion Localization

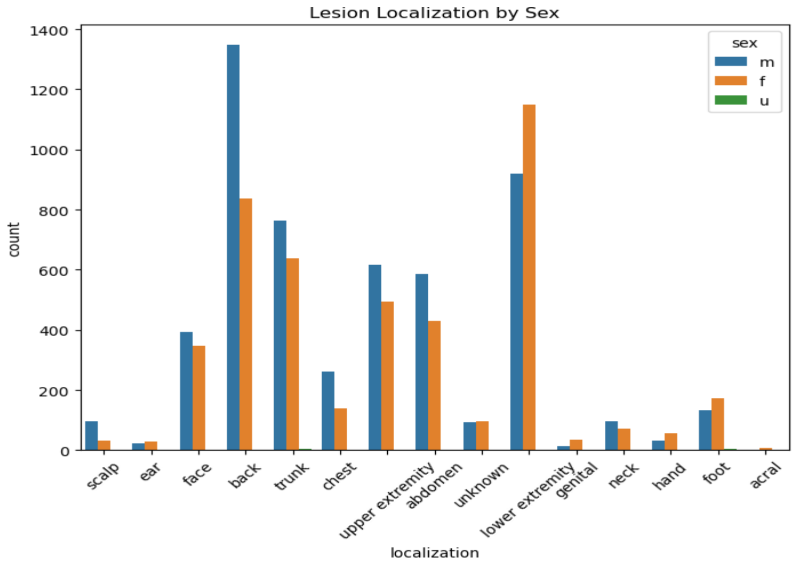

Figure 8: Sex vs Localization

## 3.3 Methodology Implementations

This project's methodology aims to integrate advanced image enhancement techniques with robust CNN architectures in the skin cancer detection system. This section focuses on the implementation of chosen models in conjunction with image enhancement strategies.

### 3.3.1 CNN Architecture (ResNet, DenseNet, EfficientNet)

Convolutional Neural Networks (CNNs) are among the most widely employed methods for classification tasks. This project has chosen three prominent CNN architectures–ResNet, DenseNet, and EfficientNe due to their demonstrated effectiveness in image classification, particularly in the realm of medical imaging applications.

- ResNet: It skips connections allow to a bypass of certain parts of the network, thus feeding the inputs from previous layers directly into deeper layers. It helps prevent deep networks from becoming absent-minded, and hence making them deep architectures. Using these residual blocks, models based on ResNet can prove very accurate in the very deep architectures.
- DenseNet: One more effect of such kind in deep neural networks is known as DenseNet which introduced a direct network. Thus, all the layers take input from all the previous layers and refer to feature reuse efficiency and parameter saving, that is often termed as the best parameter-wise efficiency-of-use. Also, it shows a good performance when handsome data in the domain of usage is few.
- EfficientNet: This amalgamation and balancing of the depth, breadth, and shape of a network is termed EfficientNet. It optimizes using a composite-based measure and makes it smaller in terms of its dimensions compared to any other architecture.

Each model is optimized using transfer learning, where the model is previously trained on large datasets such as ImageNet. Transfer learning helps models initialize with known features from multiple images, speeding up the training process and improving performance on smaller data sets

### 3.3.2 CNN with CLAHE

The primary goal of applying CLAHE is to enhance the contrast of the skin lesions, making it easier for the CNN models to identify and classify them. The images acquired from the CLAHE process will train better models, as clearer patterns are supposed to elicit meaningful learning by the models. Images undergo CLAHE processing, followed by training and evaluation of CNN models trained and evaluated based on CLAHE enhanced data.

# 4 Design Specification

It highlights the two radically distinct methods of using convolutional neural networks (CNNs) for skin lesion classification. The first one relies on raw images whereas the second method integrates pre-processing through Contrast Limited Adaptive Histogram Equalization (CLAHE). In both cases, pre-trained CNN architectures such as ResNet50, DenseNet121, or EfficientNetB0 were employed to see how pre-processing affected the accuracy of classification.

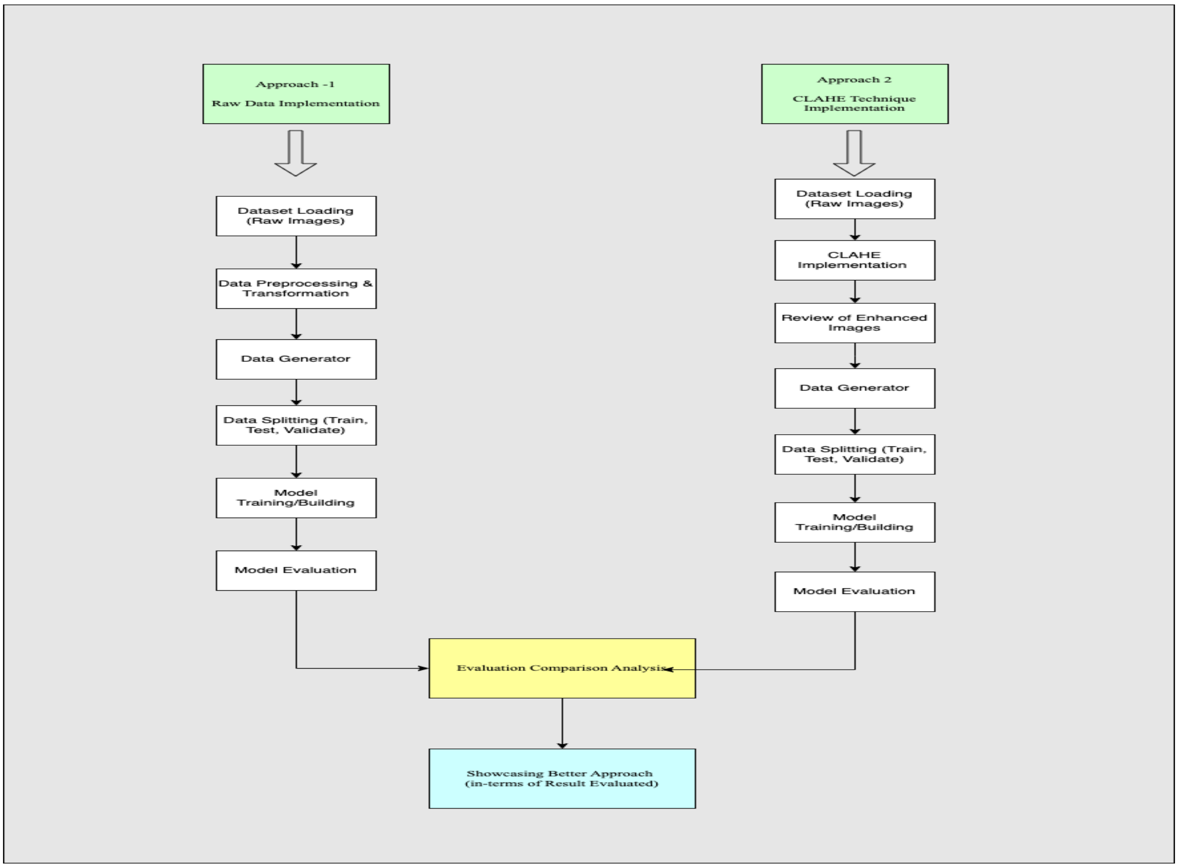


Figure 9: Approaches Flow

## 4.1 Model Approach 1 - Raw Dataset with CNN

The unprocessed raw dataset approaches utilize the raw images for training convolutional neural networks without pre-processing, then classify lesions according to the existing features inside the lesions from such training. This is a measure for comparison of results achieved by using a method enhanced by CLAHE. To this end, HAM10000 dataset was used which is an extensive repository for skin lesion images. Now reduce it to 224 by 224 pixels in dimensions now resize the image. Add the data generator to train the model since the dataset is really quite huge in batched mode. To prepare the dataset, the paths to the metadata file and the image directories are specified. In this way, when entering this data into the dataframe, also an image paths column is added by extracting it from the respective image id against both image directories, thus maintaining a sound correlation between metadata and proper image files. After that, a dictionary type mapping is created such that diagnosis labels of dx are converted into numeric classes assigning to each integer to its unique diagnosis form.

The dataset was split into 80% training and 20% test subsets, and augmentation techniques such as rotation, zoom, shear, and horizontal flips were applied dynamically using Keras's Image Data Generator class to enrich the training data and mitigate overfitting. The models were trained for 5 epochs. Among the models, DenseNet121 achieved the best test recall at 76%, while ResNet50 and EfficientNetB0 both achieved recall value as 67% respectively. Whereas Precision, accuracy, and F1-score metrics also where similar towards the recall.

## 4.2 Model Approach 2 - CLAHE with CNN

This approach performs skin cancer image classification using pre-trained Convolutional Neural Networks (CNNs), namely ResNet50, DenseNet121, and EfficientNetB0, along with CLAHE (Contrast Limited Adaptive Histogram Equalization) for image preprocessing. It starts by importing essential libraries for data handling, image processing, model building, and evaluation. Metadata, containing image IDs and diagnosis labels, is loaded from a CSV file, and the labels are encoded numerically for model compatibility. The dataset is split into training, validation, and test sets using stratified sampling to maintain class distribution. A function is defined to map image IDs to their respective file paths across two directories. Images are preprocessed using CLAHE to enhance contrast, converted to grayscale, and resized to 128x128 pixels for uniformity. A data generator is implemented to feed batches of preprocessed images and their one-hot encoded labels into the model during training, ensuring efficient memory usage.

The process of building the model utilizes transfer learning by using the pre-trained architectures ResNet50, DenseNet121, and EfficientNetB0 as base models, global average pooling, dense layers, and dropout layers as methods of preventing overfitting. They are compiled with the Adam optimizer and the categorical cross-entropy loss function, and early stopping and checkpointing callbacks are attached to save the best-performing models. Visualization functions display raw and CLAHE-enhanced images to highlight preprocessing effects. Confusion matrices can be visualized by enhanced plots for better understanding and insights into model performance. Metrics such as accuracy, precision, recall, and F1 score are computed for evaluation. The code executes model training and evaluation for each architecture, using the prepared dataset. It computes the test accuracy, creates the predictions, and evaluates all the models with robust metrics for a complete comparison of the CNN architectures in the classification of skin cancer images effectively. This modular implementation makes it easily reproducible and interpretable - thus applicable for tasks in medical imaging.

# 5 Implementation

The implementation phase is the backbone of the project, comprising the practical steps taken to preprocess data, train models, and evaluate their performance. This section provides a detailed explanation of the processes and tools used to execute the research objectives, divided into three sub-sections: developing a data pipeline, training and evaluating CNN models, and presenting outputs.

## 5.1 Development of Data Pipeline

The project collection has been cautiously catered for fulfilling the requirements needed to present data to deep learning models. The essential raw data incorporated metadata and image directories, which have been thoroughly combined into a structured dataset. The metadata contained data such as lesion IDs, image file paths, diagnosis types, and additional features like age and lesion localization of patients. The first stage in cleaning data included dropping duplicates, removing missing values, and correcting inconsistencies in the data. For instance, rows with missing image paths or negative age values were excluded from the dataset to keep data integrity. In order to simplify the classification task, a novel binary classification column was created denoting whether lesions malignant or benign. This conversion further assisted in adding directness and clinical relevance to the investigation. The dataset was then partitioned into training, validation, and test subsets, keeping an equal number of both classes across all splits. The stratification was paramount since it saved the models from being biased in the process of learning and thus endangering its generalizability.

Further modifications to the dataset were realized by including various methods of data augmentation, such as flipping, rotation, cropping, and zooming, among others. These modes of augmentation helped in improving the variability of data and, thus, robustness of the models. Particularly, it was very helpful in combating overfitting due to the small size of existing medical imaging datasets. There were two strategies for preprocessing images. The first assimilated raw images into Keras's ImageDataGenerator, which offered real-time augmentation overtraining. Such a pipeline relieved its users from manually resizing the images into desired sizes and normalizing pixel values. The second approach was implemented by Pipline Preprocessing of OpenCV, where the images were converted from color to grayscale and then enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) . This would prove beneficial by improving lesion boundary visibility and induced informativity of the data when put through the model. So such a dual-paradigm would indeed make a comparison between preprocessing techniques in their impact on model performance.

## 5.2 Model Training and Evaluation

The main implementation activity comprised training and testing three different state-of-the-art Convolutional Neural Network (CNN) architectures, namely ResNet50, DenseNet121, and EfficientNetB0, which demonstrate robustness in image classification tasks due to their inherent feature extraction capability from complex medical images. These models were pretrained on the ImageNet dataset that serves as the strong foundation for feature extraction. The training process was performed sensitively to both preprocessing methods. The first of these collected raw image data with live runtime augmentations; the other, however, reported results from models that had been trained on CLAHE-enhanced grayscale images alone. The training employed the Adam optimizer and a learning-rate scheduler, which dynamically changed the learning rate based on training status. Early stopping marked the validation performance and caused stopping if improvement was not forthcoming.

Examined was the model performance on several metric dimensions derived from the confusion matrix. Accuracy gives an overall assessment of the performance of a model, while precision, recall, and F1-score provide more fine-grained granularity distinctions in the classification capabilities. The precision shows the model gets malignant reality correct, recall in fines how sensitive model performance might be, and a composite measure given through the Recall, Precision, Accuracy and F1-score. These metrics were especially important when evaluating the ability of models in detecting malignant lesions because false negatives could lead to critical consequences in clinical practice. This impact was observed in the results from applying pre-processing techniques to the models. Models trained from CLAHE-enhanced images show better recall and precision, especially for malignant cases. This points out the potential of CLAHE concerning improving the diagnostic value of skin cancer detection systems.

## 5.3 Outputs and Tools Utilized

Models trained on performance metrics and visualizations as final outputs of implementation will help interpret results. The interference of training and validation accuracy with loss curves maps the understanding of a model. The confusion matrix presents a complete summary about true positives, false positives, and negatives to lay down the capability of classifying models. Classification reports give evaluation summaries in metrics for different classes, highlighting areas of strength and weak points. The implementation would be achieved with several tools and frameworks to simplify the workflow. Python offers the primary programming language whereby all model construction and training are done with TensorFlow and Keras. Image preprocessing using especially CLAHE is done in OpenCV. Pandas would be for data manipulation: all visualizations would be achieved through Matplotlib-Seaborn.

With these tools, implementation is efficient and reproducible and clear-cut in its components. The outputs amply served to demonstrate the potentiation of CLAHE preprocessing and CNN architectures. This further not only improved the accuracy of the model but also improved precision and recall for malignant lesions, which gives insights into future possibilities for advancing skin cancer classification systems. Structuring implementation in these phases has helped the project achieve its desirable endpoint of developing a robust and reliable skin cancer detection framework.

# 6 Evaluation

Evaluation metrics is crucial in assessing the performance of the model which provides insights in how well the model is able to predict. This section gives overview of the approaches of machine learning used and their evaluation metrics used in assessing them. Several evaluation metrics such as accuracy, precision, recall and f1 score were used to evaluate the model's performance. Confusion matrix is plotted to evaluate the actual and the predicted values of the model. This has two subsections discussing both ensemble and stacking approaches of machine learning.

- Accuracy: This metric is used to evaluate the performance of regression as well as classification model, i.e. it shows a ratio between correctly predicted values to total values.
- Precision and Recall: Precision gives the ratio of a positive value to the total number of identified positive values. It gives the accuracy of positive predictions made by the model. Recall is the ratio of true positive values against the actual positive values, thus called sensitivity.
- F1 Score: This Score helps in establishing a balance between precision and recall. It can also be referred to as the harmonic mean as it considers false positives and false negatives under the evaluation of model performance.
- Confusion Matrix: It is used to access performance of classification Model through plotting matrix between actual values and predicted value, which also given by true positives and true negatives as well as false positives and false negatives.

## 6.1 Case Study 1: Unprocessed Images with CNN

During this initial case study, CNNs were tested on raw, unfiltered skin lesion images. The images included in this study were never subject to any processing step: contrast enhancement or noise reduction, etc. Thus, one can obtain the baseline performance of the CNNs to classify skin lesions based on raw image features. Three models were tested: ResNet50, DenseNet121, and EfficientNetB0: all of which are distinguished in their architectures and strengths in regard to different approaches in image classification. Each of the models has undergone training with the raw images, which were split into training, validation, and test sets. Stratified splitting has been employed to uphold the distribution of lesion types-Melanoma, Vascular Lesions, Basal Cell Carcinoma, Benign Keratosis, Dermatofibroma, Melanocytic Nevi, and Seborrheic Keratosis across the sets. Evaluation of models was done according to several performance measures, such as Accuracy, Precision, Recall, and the F1 score.

### 6.1.1 Results Obtained

**ResNet50**: The ResNet50 model attained percentages of 67 in recall and precision as well as 67 in accuracy and F1 score. This indicates that ResNet50 has a good percentage in correctly classifying skin lesions but fails slightly in classifying them into seven classes. The model could successfully identify 1341 images but failed to classify 662, indicating that a majority of the images fail to identify lesions, especially concerning the more challenging ones like Melanoma and Vascular Lesions.

**DenseNet121**: In fact, DenseNet121 surpassed ResNet50 with a recall of 76%, precision of 76%, accuracy of 76%, and an F1 score of 76%. It correctly identified 1,515 images and misclassified 488. The improved performance again indicates dense connectivity across layers in DenseNet121, which would have resulted in better feature extraction and classified the images in a more robust manner.

**EfficientNetB0**: EfficientNetB0 showed performance similar to ResNet50, having 67% recall, 67% precision, 67% accuracy, and an F1 score of 67%. The model classified correctly 1,341 images as positive and misclassified 662. Thus, despite providing a more efficient architecture, this model gave results almost parallel to those obtained from ResNet50, stressing the point that for this particular dataset EfficientNetB0 would not yield any substantial merit on the ground of classification accuracy.

| Approach 1: Raw Images with CNN | | | | | | |
|---|---|---|---|---|---|---|
| **CNN Model** | **Recall** | **Precision** | **Accuracy** | **F1 Score** | **Correctly Identified Images** | **Misclassified Images** |
| ResNet | 67% | 67% | 67% | 67% | 1341 | 662 |
| DenseNet | 76% | 76% | 76% | 76% | 1515 | 488 |
| EfficientNet | 67% | 67% | 67% | 67% | 1341 | 662 |

Figure 10: Performance Metrics

### 6.1.2 Performance Evaluation

**Comparative Analysis of Classes Performance**: DenseNet121 outperformed all models under consideration (ResNet50 and EfficientNetB0) with respect to recall, precision, accuracy, and F1 score. However, all performed poorly in significant misclassification cases especially for complex lesions, characteristics which are impossible to make significant differential diagnosis with regard to raw image features.

**Misclassification Problem**: The misclassified images contain all seven categories, especially those categories having less number of samples, which indeed underscore the difficulty in accurate detection for such types of lesions.

**Comparison between Models**: Although ResNet50 and EfficientNetB0 reported the same performance metrics of 67 percent for recall, precision, accuracy, and F1 score, DenseNet121 showed a significant improvement over both models. And DenseNet121 enabled it to adequately capture the complex patterns and relationships within the image data during classification.

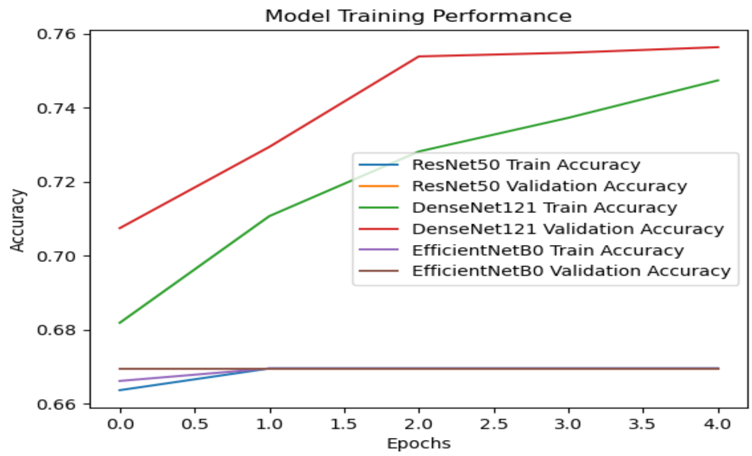


Figure 11: Model Training Plot

## 6.2 Case Study 2: CLAHE Images with CNN

CLAHE improved image contrast while limiting noise amplification to capture finer details that are quintessential for differentiating images of lesions between melanoma and basal cell carcinoma. By boosting local contrast, CLAHE intended to improve the ability of CNNs to detect lesions in skin and classify them. Images were processed by three CNN architectures namely ResNet, DenseNet, and EfficientNet. Performance of models was analyzed as Recall, Precision, Accuracy, and F1 Score, which were primary metrics to judge the models' efficiency in correctly identifying and classifying skin lesions. Enhancing such images with CLAHE was as well expected to improve the accuracy of differentiation attained by CNN models on similarly appearing skin lesions within medical purposes.

### 6.2.1 Results Obtained

The primary objective was to improve the models' ability to detect, and especially classify, skin lesions, especially subtle ones, by enhancing image contrast without adding noise. The metrics used to assess model performance included Recall, Precision, Accuracy, and F1 Score. Different CNN architectures with CLAHE-enhanced images produced the following results: follows:

ResNet gives a Recall of 75%, Precision of 73%, Accuracy of 75%, and an F1 Score of 73%. Thus, ResNet was successful in identifying more true positives but had slightly lesser precision than other models.

DenseNet performed slightly better, with a Recall of 77%, Precision of 75%, Accuracy of 77%, and an F1 Score of 75%. DenseNet demonstrated a better balance between identifying true positives and minimizing false positives, leading to higher classification accuracy.

EfficientNet achieved Recall of 67%, Precision of 45%, Accuracy of 67%, and F1 Score of 54%. Although EfficientNet showed decent recall, it had a significant drop in precision, indicating a higher number of false positives and affecting overall model performance.

| Approach 2: CLAHE with CNN | | | | | | |
|---|---|---|---|---|---|---|
| **CNN Model** | **Recall** | **Precision** | **Accuracy** | **F1 Score** | **Correctly Identified Images** | **Misclassified Images** |
| ResNet | 75% | 73% | 75% | 73% | 1509 | 494 |
| DenseNet | 77% | 75% | 77% | 75% | 1540 | 463 |
| EfficientNet | 67% | 45% | 67% | 54% | 1341 | 662 |

Figure 12: Performance Metrics

### 6.2.2 Performance Evaluation

**Model Performance Across Classes:** The models have shown variant performances across the seven classes of skin lesion. DenseNet has consistently outperformed all other models, particularly in terms of estimating melanoma (MEL) or basal cell carcinoma (BCC). The model has 77% Recall and 77% Accuracy, which suggests that it is able to find more true positives, especially for important lesions like melanoma. Some benign lesions, such as melanocytic nevi (NV) and seborrheic keratosis (BKL), were generally more challenging to classify, causing misclassifications.

**Challenges with Misclassification:** The difficulty of misclassification was highly experienced for lesions with very similar visual attributes, such as EfficientNet, which has 45% Precision, and therefore had problems differentiating between benign and malignant lesions leading to false positives. Misclassification occurred in ResNet and DenseNet, and this was due to higher Precision of 75% in DenseNet reducing the number of false positives giving it an additional advantage to be clinically more useful.

**Comparison of Models:** DenseNet has been proved to be the most reliable model in all aspects. It outperforms ResNet and EfficientNet in terms of Recall and Accuracy. ResNet was relatively lower than DenseNet, with a Recall of 75% and Accuracy of 75%. On the other hand, EfficientNet has a Recall of 67%, but has many false positives as evidenced by a Precision of 45%. So overall, DenseNet would be considered as the best model in detecting skin lesions for this research due to its balanced performance on all metrics.

In conclusion, while CLAHE improved model performance, challenges with misclassification remain, particularly for similar-looking lesions. DenseNet is recommended for optimal performance in skin lesion classification.

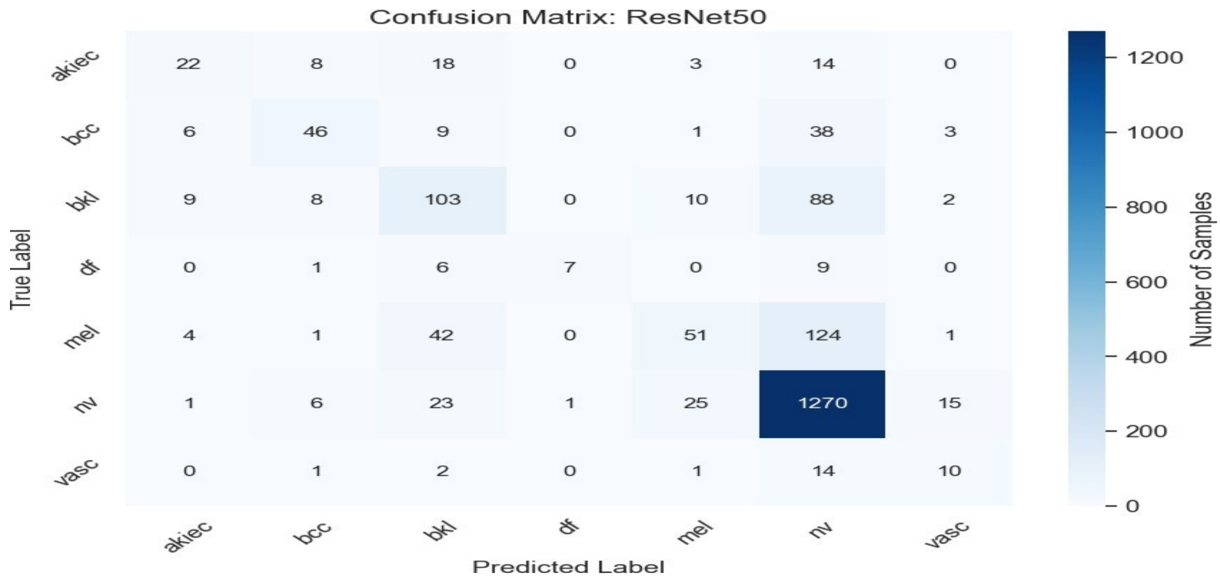

Figure 13:: ResNet Confusion

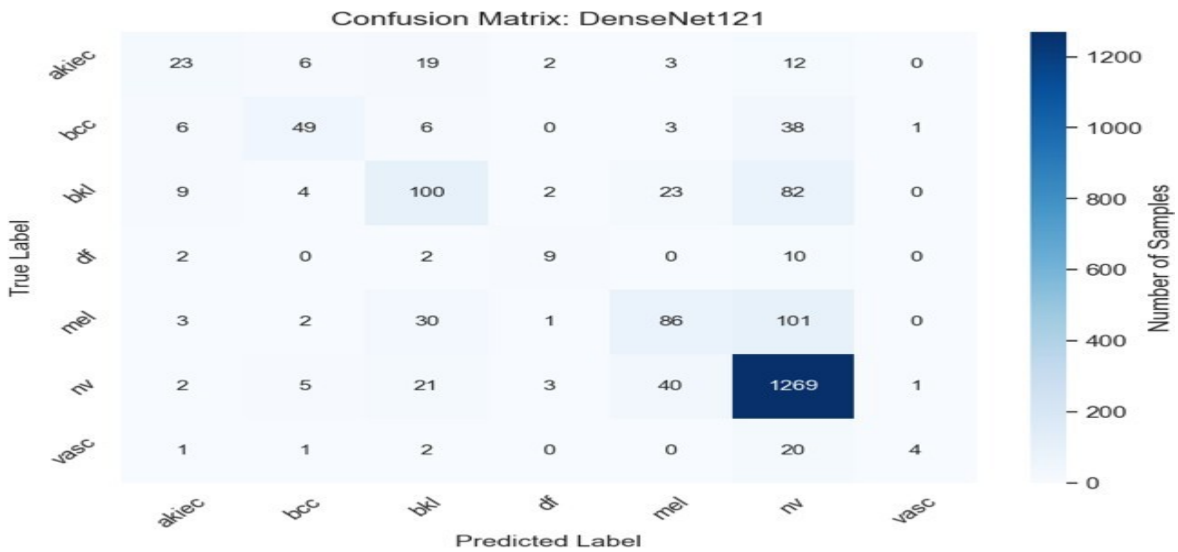

Figure 14: DenseNet Confusion Matric

## 6.3 Discussion

**Raw Images with CNN:** The CNN models performed quite reasonably on raw images. ResNet and EfficientNet achieved comparable levels of success, hovering around 67% for Recall, Precision, Accuracy, and F1 Score. DenseNet, on the other hand, achieved the highest performance with 76% scoring across all four metrics. Regardless of the above, these models struggled to understand classes that bore many similarities, especially benign lesions like seborrheic keratosis (BKL) and melanocytic nevi (NV), resulting in misclassification errors. All these show that raw images cannot capture small details that might be useful in distinguishing between very closely related lesions.

**CLAHE Images with CNN:** CNN Models showed a significant improvement with the application of CLAHE to the images. DenseNet again outperformed the rest of the models and achieved 77% in Recall, Precision and Accuracy, and 75% in F1 Score. The improvement helped the model in locating the key features, particularly in the case of melanoma (MEL) and basal cell carcinoma (BCC). However, EfficientNet continued to be on the lower side, dropping a lot in Precision and F1 Score which means that the benefits derived from CLAHE as it did to ResNet and DenseNet did not seem to apply to this model.

# 7 Conclusion and Future Work

This research investigated whether using Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance skin lesion images before classification with Convolutional Neural Networks (CNNs) would produce better outcomes. Between the two approaches-a model that took in raw images and a model which observed CLAHE-enhanced images-CLAHE showed significant model performance improvements. Interestingly, DenseNet outperformed all other models consistently, hence recording the highest recall, precision, accuracy, and F1 score across both approaches. Clahe indeed improved contrast and brought out finer details in the images making it easier for cnn models to differentiate between complex skin lesions, especially melanoma and basal cell carcinoma. However, the performance of EfficientNet was inconsistent, particularly in CLAHE enhanced images because it recorded lower precision and F1 scores compared to the other models.

Future studies on this topic could take many directions in further enhancing skin lesion classification accuracy. First, include advanced preprocessing techniques besides CLAHE-such as edge detection, adaptive filtering-to capture even finer details and increase model differentiability. Data set expansion with more diverse and balanced image classes could address problems related to class imbalance and misclassification. Further, model architecture fine-tuning and transfer learning could lead to improved robustness in models, especially when lesions that are too subtle need to be detected most certainly.

# References

Alzubaidi, L., Zhang, J., Humaidi, A.J., et al., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8, p.53. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8

Barbadekar, A., Ashtekar, V., and Chaudhari, A., 2023. Skin Cancer Classification and Detection Using VGG-19 and DenseNet. https://doi.org/10.1109/ICCINS58907.2023.10449996

Hoshyar, A.N., Al-Jumaily, A., and Sulaiman, R., 2011. Review on automatic early skin cancer detection. https://doi.org/10.1109/CSSS.2011.5974581

Islam, S.M., and Mondal, H.S., 2019. Image Enhancement Based Medical Image Analysis. https://doi.org/10.1109/ICCCNT45670.2019.8944910

Jannin, P., Fitzpatrick, J.M., Hawkes, D.J., et al., 2002. Validation of medical image processing in image-guided therapy. https://doi.org/10.1109/TMI.2002.806568

Jusman, Y., Firdiantika, I.M., Dharmawan, D.A., and Purwanto, K., 2021. Performance of Multi Layer Perceptron and Deep Neural Networks in Skin Cancer Classification. https://doi.org/10.1109/LifeTech52111.2021.9391876

Kohsasih, K.L., Zarlis, M., and Hayadi, B.H., 2022. Comparison of CNN Architecture for White Blood Cells Image Classification. https://doi.org/10.1109/ICOSNIKOM56551.2022.10034875

Kourounis, G., Elmahmudi, A.A., Thomson, B., et al., 2023. Computer image analysis with artificial intelligence: a practical introduction to convolutional neural networks for medical professionals. Postgraduate Medical Journal, 99(1178), pp.1287–1294. https://pmj.bmj.com/content/99/1178/1287

Kumar, R.S., Singh, A., Srinath, S., et al., 2022. Skin Cancer Detection using Deep Learning. https://doi.org/10.1109/ICEARS53579.2022.9751826

Kuruguntla, L., Yadala, M., N, H.N., et al., 2023. Efficient Classification Of Autism in Children based on Resnet-50 and Xception module. https://doi.org/10.1109/ICCCNT56998.2023.10306519

Lee, K.W., and Chin, R.K.Y., 2020. The Effectiveness of Data Augmentation for Melanoma Skin Cancer Prediction Using Convolutional Neural Networks. https://doi.org/10.1109/IICAIET49801.2020.9257859

Li, J., Chen, H., Wang, Y., and Wang, W., 2006. Rectification and Calibration for Biomedical Image Processing. https://doi.org/10.1109/ICICIC.2006.502

Luthfi, L., Imam Muslem, R., et al., 2023. Analysis of CNN Method for Image Classification of Coconut Ripeness Levels. https://doi.org/10.1109/ICIC60109.2023.10381964

Pizer, S.M., 2003. The Medical Image Display and Analysis Group at the University of North Carolina: Reminiscences and philosophy. https://doi.org/10.1109/TMI.2003.809707

Rubin, S.H., Kountchev, R., et al., 2006. Contrast Enhancement with Histogram-Adaptive Image Segmentation. https://doi.org/10.1109/IRI.2006.252482

Srikanth, R., Nagarjuna, B., et al., 2024. Enhancing of Night Time Vehicle Images using CLAHE Method. https://doi.org/10.1109/ICCCNT61001.2024.10725464

Umrani, S. and Pande, H.M., 2024. CLAHE - Enhanced Transfer Learning with VGG19 & DenseNet201 for Ocular Disease Classification. https://doi.org/10.1109/ICCCNT61001.2024.10726170

Venkatasunilsrikanth and Krithiga, S., 2023. Improved Deep CNN Architecture based Breast Cancer Detection for Accurate Diagnosis. https://doi.org/10.1109/ICAISS58487.2023.10250616