A Hybrid Ensemble Approach for Skin Cancer Classification Using Deep Learning Techniques

Supervised By

Mr. Suprove Chandra Sarkar

Lecturer

Department of Software Engineering

Daffodil International University

Submitted By

Sheikh Rifat

ID: 221-35-934

Department of Software Engineering

Daffodil International University

This thesis report has been submitted in fulfilment of the requirements for the Degree of Bachelor of Science in Software Engineering.

APPROVAL

APPROVAL

This thesis titled on "**A Hybrid Ensemble Approach for Skin Cancer Classification Using Deep Learning Techniques**", submitted by **Sheikh Rifat (ID: 221-35-934)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. S. M. Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

A.H.M Shahariar Parvez
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Khalid Been md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Dr. Md Sazzadur Rahman
Professor
Institute of Information technology
Jahangirnagar University, Bangladesh

A Hybrid Ensemble Approach for Skin Cancer Classification Using Deep Learning Techniques

Sheikh Rifat

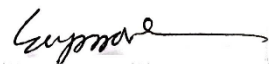
ID: 221-35-934

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

SUPERVISOR'S DECLARATION

I hereby declare that I have reviewed this thesis entitled "**A Hybrid Ensemble Approach for Skin Cancer Classification Using Deep Learning Techniques"**, and in my opinion, it is adequate in terms of scope and quality for the award of the degree of Bachelor of Science in Software Engineering.



(Supervisor's Signature)

Full Name : Mr. Suprove Chandra Sarkar

Position : Lecturer, Dept of SWE, DIU

Date : 25 December 2025

STUDENT'S DECLARATION

I confirm that the piece in this thesis is based on my own writing with the exception of quotation and reference that have been discussed. I also confirm that it was not previously and concurrently registered at Daffodil International University or other institutions at any other degree.

(Student's Signature)

Full Name : Sheikh Rifat

ID Number : 221-35-934

Date : 25 December 2025

A Hybrid Ensemble Approach for Skin Cancer Classification Using Deep Learning Techniques

Sheikh Rifat

ID: 221-35-934

Thesis submitted in fulfilment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

I am also thankful to Mr. Suprove Chandra Sarkar, Lecturer for his superb guidance, support and encouragement in completing the work. His knowledge, useful inputs and unshakeable confidence in my abilities contributed significantly to the shaping of this thesis. I appreciate very much his patience and the time he spent mentoring me on this long, hard road (that I wouldn't have any other way). I am also deeply grateful to my colleagues, friends and family members for their continuous support and encouragement without which this work could not have been accomplished. Their support and encouragement have been a great source of strength in the course of this entire study. Lastly, I would like to thank the numerous authors and researchers, and institutions whose research has inspired and informed my own work. Without their help, this study could not have been conducted. Thank you so much for the support.

DEDICATION

This thesis is dedicated to my family whose love, support and sacrifices are the journey to this milestone. Mom and Dad, always believing in me and giving me the chance to have my dreams. I would like to dedicate this work to my mentors and mainly my supervisor Mr. Suprove Chandra Sarkar, whose guidance and encouragement contributed a great deal to the successful completion of this research. Finally, I would like to dedicate this thesis to everyone who has suffered from skin cancer, with the wish that my work contributes in even a small way towards progressing medical knowledge and developing earlier detection of this life destroying disease.

ABSTRACT

Detection of skin lesions is an important task in dermatology, as early diagnosis is fundamental to prevent the appearance and treatment of skin cancer. This thesis describes HySkinDetect, a hybrid ensemble model to enhance the efficacy and efficiency of classifying skin lesions. With the synergy of these two high performing deep learning architectures (ResNet50 and DenseNet121), HySkinDetect takes advantage of those & get extra out performance. The model is capable of capturing hierarchical structures and complex features more efficiently by using the residual connection of ResNet50 and the dense connectivity of DenseNet121, which leads to a better performance in skin lesion detection. The model is tested on both the training and test data, where HySkinDetect achieves a test accuracy of 86.57%, which are much better than other aforementioned individual models such as ResNet50 (76.97%), VGG16 (74.35%), MobileNetV2 (61.23%) and DenseNet121 (72.02%). The model also exhibits good precision, recall and F1 scores vital to reliable prediction in clinical context. These findings indicate that HySkinDetect could also help in diagnostic process of skin lesions for health care providers, namely in the scope of skin cancer. Moreover, this dissertation presents the deployment of HySkinDetect in a cloud-based software web application while enhancing the model for real time execution, privacy concern and for usability by healthcare provider. The proposed system aims to facilitate fast, accurate, and interpretable analysis of skin lesions for more efficient diagnosis. Potential future work is to enhance the model's generalization (i.e. adaptive learning) and apply it to other fields of medical image. By providing a solid, scalable solution, HySkinDetect can make an important difference in clinical decision support by making the detection of skin lesions more accurate and faster – which in turn can help improve patient outcomes.

Keywords: Skin lesion detection, HySkinDetect, hybrid ensemble model, ResNet50, DenseNet121, deep learning, skin cancer detection, accuracy, precision, recall, F1 score, test accuracy, web-based software, real-time performance, medical imaging, data security, clinical decision-making**.**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
CNN	Convolutional Neural Network
F1 Score	F-Measure Score
GPU	Graphics Processing Unit
HySkinDetect	Hybrid Skin Detection
ReLU	Rectified Linear Unit
ResNet50	Residual Network 50
VGG16	Visual Geometry Group 16
MobileNetV2	Mobile Network V2
DenseNet121	Densely Connected Convolutional Network 121
EHR	Electronic Health Records
IoT	Internet of Things
API	Application Programming Interface
ML	Machine Learning
DL	Deep Learning
UI	User Interface
UX	User Experience
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
AI	Artificial Intelligence

CHAPTER 1

INTRODUCTION

1.1 Overview

Skin cancer is one of the most frequently diagnosed malignancies, and its prevalence has been rising over the past few decades. The early diagnosis is very important for finding a cure because it significantly improves survival chances. Current conventional skin cancer diagnosis procedures are mainly based on the visual assessment of dermatologists and time consuming, error prone. In order to overcome these shortcomings, machine learning, including deep learning, has received considerable interest for automated detection of skin cancer. Deep learning (DL) architectures, such as ResNet50 have proven to be highly effective and competitive in classifying medical images, including thermoscopic images of skin lesions. However, single deep learning model might have limitations in generalization, accuracy and robustness when faced with a variety of data. Ensemble methods provide a way to improving performance by aggregating multiple models. Ensemble methods can enhance classification accuracy and minimize errors through leveraging the merits of individual models. This method has proven successful in a variety of areas and yet to be explored in the task of skin cancer detection. In this paper, we attempt to investigate the possibility of using ensemble deep learning for skin cancer diagnosing and investigates if it is effective in improving the accuracy, robustness and generalization. The study will address the question of using ensemble models in dermatology image datasets for diagnostic purposes.

1.2 Background

Skin cancer is the most common cancer and poses a major health problem around the world. Skin cancer rates have increased over the past decades, with melanoma and non-melanoma skin cancers being the most prevalent types. Early diagnosis is key to better prognosis and survival. But diagnosing skin cancer typically requires skill and experience, which can be a bottleneck in many health care settings.

The development of machine learning, especially deep learning has greatly enhanced the performance and efficiency in analyzing medical images. Deep learning models, which iterations can deal with massive amount of data and learn patterns, show promising results in pattern recognition for medical images including thermoscopic skin lesion images. Ensemble methods (multiple models combined for better prediction accuracy) are one of the prevalent techniques in deep learning, and have begun to attract attention in a range of medical applications.

1.3 Problem Statement

Skin cancer is among the most common causes of cancer death worldwide, early diagnosis being essential for improving the odds of patient survival. Although diagnostic approaches have improved, the diagnosis of skin cancer continues to depend on 'subjective clinical examination by dermatologists' that can lead to errors in decision-making. Also, the differences cause high variance for image quality, illumination and types of lesions could make it difficult to ensure a precise diagnosis. Recent progress of deep learning techniques has much to offer for automation in skin cancer detection with microscopic dermatological images. Nevertheless, single deep learning methods are still challenged to overcome overfitting and underfitting problems as well as generalization issues for multiple datasets. Ensemble learning, a technique in which multiple models are combined to enhance performance, may help solve these problems by improving prediction accuracy and robustness. However, the study of SDNN for ensemble deep learning in skin cancer detection is scarce and how such methods would enhance the trustworthiness and practicability of an automatic skin cancer diagnosis needs to be investigated.

1.4 Motivation

Skin cancer is a major public health problem worldwide, and its growing incidence makes early detection an imperative. Early diagnosis is essential for survival as skin cancer is highly curable when discovered early. Nevertheless, diagnosing traditional methods that largely depend on the experience and knowledge of dermatologists is time-consuming and human errors are involved. In the era of machine learning and deep learning, a unique opportunity exists for creating automated systems that could assist in early detection of skin cancer. Dermatologic image analysis has been a

promising area for deep learning models, especially Convolutional Neural Networks (CNNs). Instead, individual models might have difficulties on the complete generalization across various datasets and are sensitive to variations of images. Given the advantages of both models, ensemble learning could be a potential solution to improve robustness and accuracy as well as generalization ability in skin cancer detection. The objective of this study is to investigate whether the ensemble learning based deep learning may alleviate these issues, and facilitate diagnosis in a more reliable way for decision support among healthcare providers. Finally, this work is intended to help in creating automatic, efficient and scalable systems that could be utilized for the early detection of skin cancer, eventually saving people's lives and relieving society from a large part of the financial burden.

1.5 Research Objective

The main aim of this study is to design and assess the efficacy of a new deep learning model for skin cancer detection, namely the Hybrid Ensemble Model: HySkinDetect. In this model, we fused the advantages of ResNt50 and DieNet121 together to build a highly complementary cascaded ensemble system.

The specific aims of this analysis include:

- ✓ To construct the Hybrid Ensemble Model: HySkinDetect by combining ResNet50 and DenseNet121 models. The model is built to take advantage of the strengths from both networks, in order to improve feature extraction and thus increase the accuracy of skin cancer diagnosis.
- ✓ In order to assess the performance of HySkinDetect model with respect to deep learning models (ResNet50 and DenseNet121) for key performance metric such as accuracy, sensitivity, specificity and F1 score.
- ✓ To evaluate the robustness and generalization ability of the HySkinDetect model by evaluating cross-effectiveness over different dermatological datasets with various image quality, lesion types and lighting conditions.
- ✓ To explore the value of the ensemble learning strategies in enhancing the performance, efficiency and robustness of computer-assisted skin cancer detection by benefiting from combined deep learning models.

- ✓ To provide recommendations for "how" (use) to implement HySkinDetect in a clinical, practice setting by identifying possible use cases, downsides and ways to improve the usability of skin cancer diagnostic methods.

1.6 Research Purpose

In this work, we propose a novel hybrid ensemble deep learning model, HySkinDetect, for automatic detection of skin cancer. This model combines the strength of feature extraction in ResNet50 and DenseNet121, compounding their strengths utilizing a group wise approach. The main objective is to improve the accuracy, robustness and generalization of systems for detecting skin cancer, compensating the drawbacks of single deep learning models in the field of dermatologic image analysis. This study aims to enhance the performance of computer-aided skin cancer classification and make it more applicable across different datasets using ensemble learning. The convergence of such architectures aims to avoid issues of overfitting (common in single layer networks) and improve classification accuracy, presented here at a preliminary albeit encouraging stage which offers a scalable end-to-end solution carrying the potential for real world clinical application. Finally, the ultimate purpose of such research is to help the development of a robust early skin cancer detection system for healthcare professionals and assisting them as much accurate and quick diagnosis.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

In this chapter, we provide a review of the literature on skin cancer detection applying deep learning algorithm and particularly ensemble models. Skin cancer, including fatal melanoma, is one of the most common malignancies and early and accurate diagnosis can lead to better outcomes in patient care. The surge of machine learning, and in particular deep learning methods lately has created new possibilities for automating skin cancer detection. Accordingly, numerous works have been proposed to improve the accuracy and reliability of diagnostic systems, with a special emphasis on deep and ensemble techniques. I first did the literature review on methods using single CNNs for skin cancer detection. These models, especially architecture-based and ResNet-DenseNet-inspired ones, have shown remarkable achievements on diagnosing skin lesions using dermoscopic images. However, although they perform well under ideal conditions, such models typically suffer from overfitting and poor generalization when dealing with various datasets. Motivated by the requirement for stronger solutions, ensemble approaches that combine the strengths of multiple models to improve accuracy, decrease error rates and strengthen robustness are being explored.

2.2 Previous Work

Naqvi et al. (2023) Appy deep learning methods for skin cancer detection are broadly revisited; CNN-based networks (their architectures), and their pros and cons. It also notes that while deep CNNs49 based techniques seem promising, their performance can be limited in part due to small datasets (e.g. as little as ~1,300 images in some instances of the studies). The investigation highlights the importance of large, diverse samples and systematic evaluation methods that should be employed in future work. In general, it indicates that ensemble and hybrid approaches could provide robustness, while normalization is a problem [1]. Harangi (2018) Studied deep CNN ensembles to classification of thermoscopic image as three classes; melanoma, nevus and seborrheic keratosis. They combined the output from four different CNN models by a number of fusion strategies and the ensemble was better than single model.

The reported accuracy and the average AUC were 74% at 0.918 (for the 3-class task) Better than A single models in terms of the classification reliability [2]. Behara et al. (2024) Presented a novel hybrid classification method for skin lesions which combines snake models, ResNet50 as feature extraction and attention-based Capsule Network. Segmentation, preprocessing, and feature extraction before the latter classification are done to improve discrimination of lesion characteristics. While the paper emphasizes improvement of feature representation and spatial/channel attention in this work, their reported classification results show significantly improved (compared to baseline models) performance suggesting that segmentation + hybrid architecture may aid lesion classification [3].Hamim et al. (2024) Presented the layer-wised representation learning, named "SmartSkin-XAI," a hybrid CNN structure consisting of AlexNet and DenseNet-121 for predicting skin cancer via thermoscopic images. The paper focused not only on interpretability (XAI), but also classification accuracy a critical first step in becoming clinically trustworthy. They also showed that the mixture of architectures (heterogeneous ensemble) has the potential to enhance both predictive power and interpretability by rendering the model more friendly for a clinical use [4].

Alwakid et al. (2022) Employed both a CNN and ResNet-50 for skin lesion classification from dermo copy images, on HAM10000 dataset (10,015). Their technique involved image segmentation and pre-processing, followed by training a CNN using ResNet-50. Results demonstrated that fine-tuned deep architectures + preprocessing are able to be used for skin cancer classification [5]. Bello et al. (2024) Proposed a transfer learning approach for Beni malignant skin cancer classification using dense-net model. They supplemented the pretrained backbone with custom dense layers by means of data augmentation (hair removal, lighting adjustment), and preprocessing for performance improvement. They reached to an accuracy and F-measure of ~87%, which means the moderate, but reasonable effort. They note that careful preprocessed input data & transfer learning are still necessary for solid classification [6]. Wu et al. (2022) Presented an overview of CNN-based skin cancer classification methods (state-of-the-art architectures, datasets and performance metrics are summarized along with studies that considered patient metadata combined with images). They mention three standard problems: data imbalance, domain shift (across different image capture devices), and lack of generalizability.

They comment that many algorithms perform well in a clean testing environment but degrade for real world variation [7]. Hossein Zadeh et al. (2024) Designed a system for processing multiple transfer learning backbones (DenseNet-201, DenseNet-121, ResNet-50, ResNet-101/152, VGG19 and EfficientNet-B3) as feature extractors; with the use of feature selection techniques (ANOVA-Lasso PCA-Random Forest etc.) when it comes to classification. We aimed to minimize unnecessary features and enhance discrimination between benign and malign cases. The method they present is a flexible ensemble-feature-selection pipeline, and their results support the idea that FE + feature selection can improve diagnostic accuracy [8]. Shen et al. (2022) proposed of automatic thermoscopic image classification method based on ensemble of fine-tuned pretrained CNNs (Xception, ResNet50 and VGG16) by transfer learning from the ISIC 2016 dataset. They combine the outputs using a weighted-fusion method. Their combined model results yielded an accuracy of 86.91%, precision 85.67%, recall 84.03% and F1-score of 84.84% performing better than all individual models separately did. This result further encourages the potential that single CNNs can also benefit from small-sized ensemble members, especially for difficult image classification tasks [9].

Raghavendra et al. (2023) Proposed a new DCNN for early detection of skin cancer and multi class classification. They benchmarked the performance of their model with existing transfer learning models such as VGG-16, ResNet50, DenseNet121 and MobileNetV2. The new model also was said to classify better at earlier stages, indicating enhanced sensitivity/ specificity versus baseline models. This highlights the possibility of custom-designed CNNs (other than the transfer-learning), when used for skin lesion classification [10]. Thwin et al. (2024) Proposed scouting a multi-deep learning models (VGG16, Inception-V3 and ResNet-50) using thermoscopic images (e.g. ISIC 2018 dataset), by weighted averaging its output predictions. Their model surpassed the classification from single architectures, confirming the benefit of ensemble schemes for robustness and better performance in various types of lesions [11]. Kausar et al. (2021) Developed a multiclass classification of skin cancer using an ensemble of fine-tuned deep learning networks (ResNet, InceptionV3, DenseNet, InceptionResNetV2 and VGG-19) with up to eight classes (MEL, NV, BCC, BKL, AK, DF, SCC and VASC). Single learners achieved accuracy between 72% and 91.8%, but their ensemble (majority voting or weighted) produced an accuracy of 98–98.6%.

This demonstrates the potential power of the ensemble classifiers in multiclass where classes are more than binary [12]. Thanka et al. (2023) — Proposed a fusion model strategy that combines the pretrained CNN (VGG16) for feature extraction and machine-learning classifier (XGBoost) for binary classification melanoma vs non-melanoma. They reported a high accuracy 99.1% of their method in comparison with some of the baseline methods. In addition, this shows that the power of hybrid CNN + ML-classifier pipelines (as opposed to end-to-end deep nets), can be effective when dataset size and computational cost are moderate [13]. J.SM et al. (2023) Presented a DCNN model for automatic classification of skin cancer types (melanoma vs non-melanoma) in dermoscopic images. The study demonstrates good classification results, proving that well-designed deep networks can effectively discriminate malignant lesions (with proper preprocessing). This is further evidence that it should be possible to implement automated skin cancer detection using deep learning [14]. Chiu et al (2025) Investigated deep ensemble learning where traditional CNNs were combined with Vision Transformers (ViTs) for multiclass classification of skin lesions. In addition to skin lesion detection, they also considered the class imbalance and background subtraction problems frequently presented by dermoscopic databases. Their multi-architecture model achieved better robustness and generalization suggesting that heterogeneous architectures (CNN + ViT) can enhance the classification performance of complex multiclass tasks [15].

Al Waisy et al. (2025) and recently introduced a deep learning architecture (SCDNet) with ISIC 2019 dataset and the achieved accuracy for multiclass skin cancer classification is 96.91%. Importantly, SCDNet also achieved better performance than regular ResNet-50 (see the other similar value 95.21% shown in their work), which suggests that newer frameworks can continually boost the performance frontier. It is an indication that there's more architectural innovation (beyond vanilla CNN transfer learning) that can be leveraged to obtain better performance for skin cancer detection [16]. Integrated Deep Learning Model for Skin Cancer This preprint presents a hybrid ensembled deep-learning methodology specializing in skin lesion classification: as benign or malign. Although the exact numerical results are not explicitly given in their summary, results apparently show that by combining components (presumably backbone + custom layers), they attain the best classification performance.

This can be seen as evidence of feasibility of your hybrid/hybrid ensembles combining the anatomy and texture . Uncertainty Aware Deep Learning for Automated Skin (2025) This recent worksaims not only at classification performance but also attend to quantifying prediction uncertainty. Those are the kinds of approaches that will be necessary for ML systems to find use in clinical contexts not only "what does your model predict", but "how confident is it" about a prediction. This tackles meaningful aspects in terms of reliability, trust and interpretability for skin cancer detection systems [17]. Akter et al. (2023) — Introduced a multiclass skin cancer classification framework with multiple transfer-learning and CNN models (ResNet 50, DenseNet, MobileNet, InceptionV3 and Xception among others) built on HAM10000 dataset. They had also tried to stack models (such as ResNet50+VGG16 etc.) but stacking has limited performance (max ~78% for top stacking model). This means that simple stacking is not necessarily helpful ensemble architecture should be designed with care [18]. Chaturvedi et al. (2019) Introduced seven-class skin cancer classification based on a pretrained MobileNet in HAM10000 dataset (Skin Lesion Analyzer). They classified 83.1% categories correctly, top 2 accuracy 91.4%, to 3 95.3%, with F1-score ~0.83 . Though lower than recent works, these early results reveal that lightweight models (MobileNet) can achieve its purpose as a multiclass classifier when computational resources are not abundant [19].

Di Giammarco et al. (2025) A deep learning approach to seven classes skin lesion classification: Detection and localization of lesions is presented in this paper. The study not only evidences the reliability of proposing a multiple class classification with precise FFDM lesion localization in different type of lesions, but also reinforces the strength of ensemble/hybrid models also in environmental regions to binary detection [20]. Arshad et al. (2025) the authors introduce a new method, which uses an enhanced segmentation network (DeepLabV3+) with classification to detect lesion and diagnose skin cancer. They report high segmentation dice score (94–96% on benchmark datasets (HAM10000, ISIC 2018/2019, ISBI 2020) and classification accuracies up to ~96.35%. This work demonstrates the advantage of the joint segmentation and classification framework over the traditional one, which may enhance lesion boundary localization and classification accuracy [21]. Al-Waisy et al. (2025) We present a novel deep-learning-based framework for early diagnosis of skin cancer, with high

performance on a large image dataset. The authors state that this new model surpasses several classical architectures, suggesting possible improvement over the existing DL based diagnostic schemes and providing some evidence to indicate that new frameworks could cause significant improvements [22]. Musthafa et al. (2024) The work presents an optimized CNN model for classification of skin lesion of the HAM10000 dataset. They use data augmentation methods to address the class imbalance. Their architecture tuning leads to superior diagnostic accuracy when compared to standard CNNs, confirming the significance of architecture optimization and data preprocessing for skin cancer detection task [23]. Naeem et al. (2024) In the paper where SNC Net model is introduced, they integrate deep learning features and handcrafted ones to enhance skin cancer detection on dermoscopic images. This hybrid feature fusion strategy implies that a combination of DL-based and classical image representations is able to enhance performance, which can be important in case that data are scarce or lesion appearance is very different between sources [24].

CHAPTER 3

METHODOLOGY

3.1 Overview

In this chapter, I describe the approach to study the performance of ensemble deep learning methods in relation to skin cancer classification. The main goal of this study is to propose and validate a hybrid ensemble model (HySkinDetect) that takes advantages from the ResNet50 and DenseNet121 modeling formulations. By combining the strengths of these individual models, I hope to improve the accuracy, robustness and generalization of skin cancer detection. I start by preprocessing the data, applying data augmentation methodologies in order to enhance model generalization and prevent overfitting. One-hot encode all my label and then I train two single deep learning model (ResNet50, DenseNet121) and afterward ensemble them to get a better fancy name model. The HySkinDetect model is tested against respective single-scale models for comparison. Model optimization is performed through hyperparameter tuning and I evaluate the performance of the models on multiple performance metrics such as accuracy, sensitivity, specificity, F1 score. I also utilize cross-validation to evaluate the models' stability to different settings, including the quality of image and types of lesions. Finally, the aim of this approach is to identify whether ensemble leaning can enhance general performance and robustness of skin cancer screening systems for realistic medical uses.

3.2 Working Procedure

The workflow of our work is as the following steps, to make sure we design, train and evaluate our proposed skin cancer detection model. These steps are as follows:

1. **Dataset:** It starts with importing the skin lesion dataset that consist of images having benign and malign lesions. These images are already labeled in advance for classification. The dataset is split into two primary classes: Benign and Malignant.

2. **Data Preprocessing**

- **Augmentation:** Techniques such as rotation, zooming and horizontal flipping are used to increase the dataset size and improve the model's capacity for generalization.
- Normalization: The pixels are rescaled so the values are normalized; this allows them all to be trained on the same data.
- Hyperparameter Tuning: Fine tune the hyperparameters (learning rate, filter etc.) in order to increase the performance of model.
- Class Distribution Check: The distribution of the two classes (benign, malignant) is examined to make sure that there is fair and balanced training.
- Division: Dividing the datasets into training (80%) and testing (20%) sets to train and validate the model.

3. **Model Training**

✓ MobileNetV2
✓ VGG16
✓ ResNet50
✓ DenseNet121

These models are then integrated under ensemble **HySkinDetect** model to increase the accuracy and generalization ability.

4. **Model Evaluation**

After training the models, they are evaluated based on several metrics:

❖ Confusion Matrix: It is used to determine the overall performance of a model for both positive and negative outcomes (benign/malignant).
❖ Accuracy, Precision_Recall_F1-Score: These metrics are measured to assess the general classification performance.
❖ ROC Curve: Receiver Operating Characteristic curve is a graphic representation, used to analyze the diagnostic ability of a test.

The approach emphasizes workflows for data pre-processing, model training and evaluation that are reinforced for the robust performance of skin cancer detection models with deep learning.

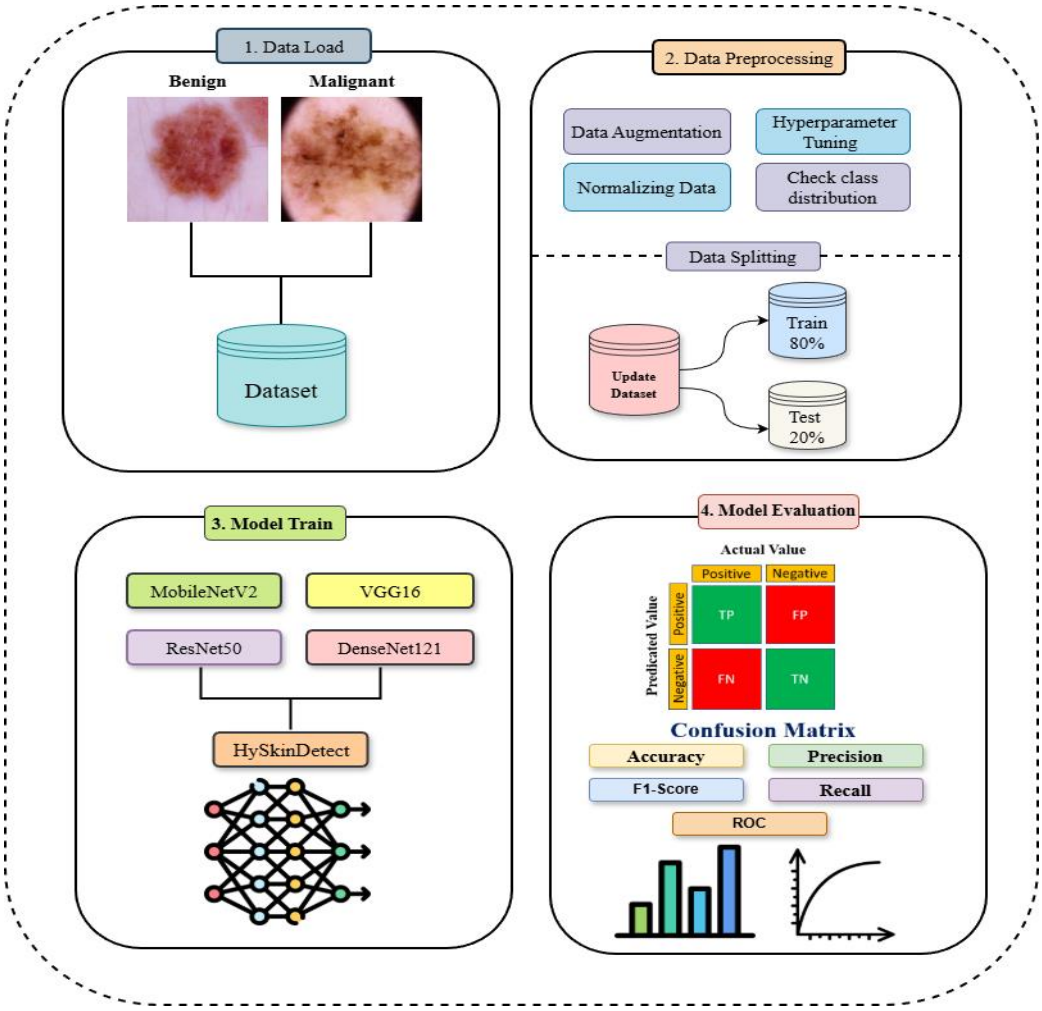


Figure 3.1: Methodology Workflow for Skin Cancer Detection

3.3 Dataset Description

The dataset that is used in the experiment contains dermoscopic images of skin lesions that fall mainly under two classes such as Benign or Malignant. These images come from open access dermatological image repositories and they are employed to train and test the skin cancer detection deep learning models. The data set is composed of 11,879 images in all; among them Benign lesions consist of 6,289 images Malignant lesions amount to 5,590. The images are in JPEG (/ PNG) and have been resized to 224x224 pixels, a common input size for deep learning models such as ResNet50[4], DenseNet121 [5], VGG16 [1] and MobileNetV2 [6]. The dataset is split into three subsets for model training: training, validation and testing. Training the models uses roughly 80% of the data with 10% saved for validation during training and another 10% kept for testing and model evaluation. The data of benign and malign lesions in

the training and test set are well organized with separate folders, which is friendly for data loading. Data augmentation, including rotation, zooming, horizontal flipping and shifts were used to improve model generalization and combat overfitting. These images are preprocessed by normalizing the pixel intensity values between 0 and 1, thereby preparing the data for deep learning methods. The data set is also balanced between the two classes in order to avoid bias, where benign and malignant cases appear equally often in the training and test sets. Although the dataset is relatively balanced, it suffers from problems such as variable image quality, luminosity conditions and types of lesion that may impact the performance of models. However, this data set is a solid ground truth for training deep learning models to accurately classify skin cancer offering the potential of employing an ensemble model that leverages different architectures to facilitate enhanced classification performance.

3.3.1 Dataset Composition Table

There are two primary classes; Benign and Malignant in the dataset. The training set consists of 6,289 benign lesion images and 5,590 malignant lesion patches. There are 1,000 images per class in the testing set. The Benign class consists of 7,289 images, whereas the Malignant class comprises a total of 6,590 images. This data set allows for a slightly similar spread of the two classes throughout the samples.

Table 3:1 Dataset Composition for Benign and Malignant Images

Class	**Number of Images (Training)**	**Number of Images (Testing)**	**Total Number of Images**
Benign	6,289	1,000	7,289
Malignant	5,590	1,000	6,590

3.4 Data Augmentation

Data Augmentation is a vital methodology in this study, to enhance the generalization and robustness of deep learning models employed for skin cancer classification. As the dataset may be limited in its size and diversity, data augmentation provides extra training examples to mitigate overfitting and improves generalization so that model can generalize well on new vehicles. Various enhancement methods are employed in dermoscopic images of skin lesions. This consists of rotation, where images are randomly rotated within a defined range to represent differing orientations of the lesions; zooming, reflecting changes in the size of the lesion by random zoom in and out; and flipping both horizontally and vertically to reflect mirrored versions of the lesions. Also, moving the images a little in the horizontal and vertical direction allows the model to handle small miss-alignments that could happen as a consequence of different image capture. Finally, brightness and contrast varying is performed to simulate different illumination conditions for model effectiveness under diverse environmental factors. These augmentation strategies allow a more diverse dataset, which enhances the model generalization power to classify skin lesions from the real world and make the model robustly enough and preventing over-fitting. The model is also evaluated under varying environmental conditions through changing brightness and contrast. These augmentation algorithms bring variety and diversity to the dataset, allowing a better generalization capacity of the model in order to classify skin lesions in real scenarios while maintaining robustness and avoiding overfitting.

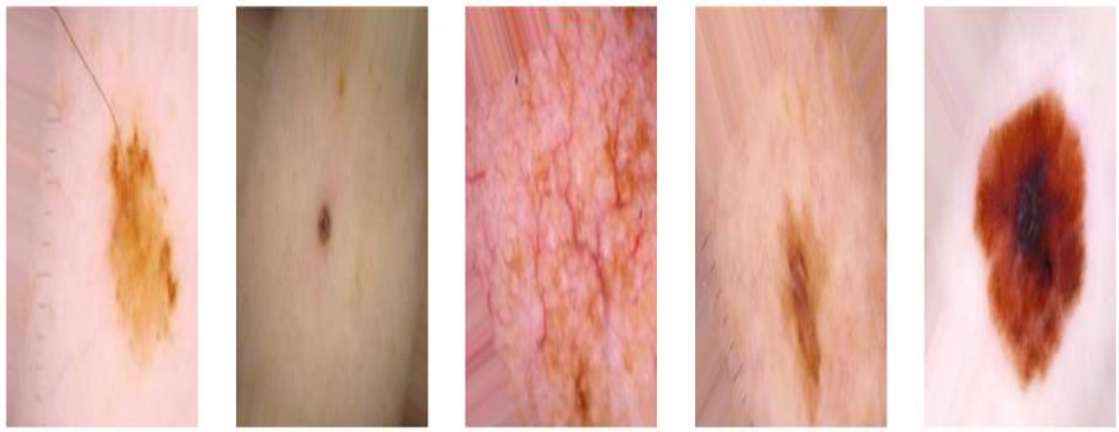


Figure 3.2: Augmented skin lesion images with rotation, zoom, and flipping.

3.5 Normalizing Data

Normalization is a vital step of preprocessing in deep learning, by which pixel values in the images are adjusted to a common scale, such as 0-1. This keeps the input consistent and allows the model to converge faster when training. The pixel values of the skin lesion images are scaled to the range [0, 1] by dividing all with 255 in this study. Such scaling lets our model process data faster, since the features (read pixel intensities) would all be on the same scale. Correct normalization also helps avoid problems that can occur if the range of input features varies widely, thereby forcing the model to consider each pixel on an equal footing. This is an important step for the successful training of deep learning models, enabling that when focuses on the optimization process, and does not interfere with its distance/closest brother as possible) which can help to learn the underlying data patterns. Besides, normalization can prevent gradient explosion and vanishing, which may occur when the input value is very large or small and leads to instability in backpropagation. When the input data is normalized, the model becomes stable and training is more efficient, with faster convergence and better performance.

3.6 Model Architecture

The Ensemble Model combines ResNet50 and DenseNet121 to create a powerful ensemble for skin cancer classification. The architecture first processes the input image through both individual models, ResNet50 and DenseNet121, which are trained separately. ResNet50 utilizes residual connections to allow deeper learning without the vanishing gradient problem, enabling it to capture complex features in the image. Meanwhile, DenseNet121 uses dense blocks where each layer receives input from all previous layers, improving feature reuse and efficient information flow. After both models independently extract features, their outputs are combined into a unified prediction through a final decision layer, often via averaging or a weighted combination of their predictions. This ensemble approach leverages the strengths of both models to provide a more accurate and robust classification result. The combination of ResNet50's deep residual learning and DenseNet121's dense connectivity ensures the model can generalize better, leading to improved accuracy in detecting benign and malignant skin lesions.

3.6.1 MobileNetV2 Model Architecture

MobileNetV2 is a compact convolutional neural network architecture for mobile and embedded vision applications. It is based on the original MobileNet architecture and uses inverted residuals with linear bottlenecks. MobileNetV2 is based on an inverted residual structure where the input and output of the residual block are thin bottleneck layers opposite to traditional residual models which use expanded representations in the input of the block. The invention employs linear bottlenecks between layers, like MobileNets, as building blocks. Unlike MobileNet, these bottlenecks are suitably connected with inverted residuals to build a powerful feature extraction mechanism. The proposed model is based on a succession of convolutional blocks, which capture rich features and reduce the image size in an efficient form. MobileNetV2 is highly accurate and computationally efficient, which makes it suitable for edge devices with resource constraints. It uses ReLU6 activations, which can maintain accuracy and stability through quantization. Because of its low complexity, MobileNetV2 can be practical for applications in which fast predictions are essential (e.g. skin cancer classification on mobile or embedded systems). In this work, it is learned to identify benign and malignant skin lesions, which benefits the overall performance and generalization of the ensemble model.

3.6.2 VGG16 Model Architecture

VGG16, a project by the Oxford's Visual Geometry Group is one of the earliest CNNs that have proven simple yet effective. (called VGG16 hereafter) composed of 16 layers, which are further divided into 13 convolutional layers and 3 fully connected ones. VGG16 adopts small-size convolution kernels (3 * 3 receptive fields), stride 1 and with the same padding; this design enables VGG to capture fine-grained details of images. The network heavily uses max-pooling layers in order to shrink the spatial dimensions gradually so that the model can learn hierarchical features at several scales. Though being simple, VGG16 is performing very well on image classification. Nonetheless, it is computationally inefficient and memory demanding since it has huge number of parameters, especially in the fully connected layers. In the field of skin cancer detection, VGG16 can be used to recognize subtle dermoscopic patterns, such as texture or irregularities (which is in turn likely to enhance the ensembles model's robustness).

3.6.3 ResNet50 Model Architecture

ResNet50 [35] is a residual network which introduces skip connections to be able to backpropagate gradients without any vanishing/exploding effect over extremely deep networks. This allows the vanishing gradient to be solved and therefore deep networks can get better trained. ResNet50 contains 50 layers, and it has been considered as high-accuracy model for variety of computer vision applications like image classification. The architecture is based on residual blocks where each contains two or three convolution layers and a shortcut link that bypasses the block. This architecture enables the training of deeper models for ResNet50 without overfitting or gradient issues. For skin cancer diagnosis, the ResNet50 mainly catches the structure information about dermoscopic images such as lesion boundaries and texture. Benefiting from its ability to facilitate deep network training, ResNet50 serves as a strong anchor model for HySkinDetect ensemble, with enhanced accuracy in classification by learning high- and low-level features.

3.6.4 DenseNet121 Model Architecture

DenseNet121 is a deep learning architecture which uses dense connections, where each layer is passed an input from all the previous layers. This facilitates better feature reuse, and helps to address the vanishing-gradient problem by allowing information to pass across corresponding layers more easily. DenseNet121 is composed of 121 layers and uses a number of dense blocks that are directly connected convolutional layers to each other. Every layer in DenseNet is included feature maps from all previous layers, thus forcing the network to learn more diverse and richer representations of features. This architecture is especially useful for tasks where fine-grained feature extraction is needed, such as skin cancer classification, where subtle differences between benign and malignant lesions need to be captured. DenseNet121 achieves similar results of traditional CNNs with less parameters, thus more computationally efficient. And in our HySkinDetect model, DenseNet121 plays a role of the features extraction that uses dense connections and deep architecture to achieve accurate classification of skin lesion.

3.6.5 HySkinDetect Hybrid Ensemble Model Architecture

HySkinDetect Hybrid Ensemble Model (HDE) introduces ResNet50 and DenseNet121 to pool their strength together, yielding a stronger system in the realm of the skin cancer detection. The ensemble strategy makes full use of the characteristic of both architectures for accuracy and generalization. The residual connections of ResNet50 make it capable of learning very deep, hierarchical features and managing complex, massive dataset efficiently. Meanwhile, through network's dense connections, DenseNet121 can increase the reusability of the features so as to mitigate deep supervision and learn more delicate features from data. Through the combination of these two models, it enables the ensemble to retain both deep residual-based features learning and dense connectivity information, offering a finer detail of each lesion in our skin database. In the ensemble setting, the predictions of ResNet50 and DenseNet121 are merged to make a final prediction. This model stacking enhances the stability of the system as it lowers the risk of overfitting or bias which might occur in a standalone model. V; Rehman, A.; Ahad, M.A hybrid model has shown a better performance than the single model due to ensemble combining which helped the system generalize on different types of skin lesions and conditions. This manner of learning guarantees that the HySkinDetect model does not only classify accurately but is also robust to image quality, lesion types and lighting conditions - this last fact qualifying it for use in real-world medical applications.

3.6 Model Evaluation

Model assessment is an important step to analyze the performance and utility of skin cancer diagnosis models. The models presented in this paper is evaluated with respect to different metrics that capture different areas of classification performance. The focus is on having the model perform well not only in accuracy but also in sensitivity, specificity and stability under different conditions.

Confusion matrix: Confusion matrix is the most useful metrics to review, if you are working with binary/multiclass classification problem. For skin cancer diagnosis, the confusion matrix offers a concise summary of how the model discriminates between malignant vs. benign lesions., such as accuracy, precision, recall and specificity. For binary classification tasks such as this,

The confusion matrix is a 2x2 table that we will refer to with the following terms:

TP (True Positive): Number of malignant lesions that are truly classified as malignant.

False Positives (FP): Model predicts the benign lesions as malignant too many times.

False Negatives (FN): Malignant lesions count that are incorrectly classified as benign

False Negatives (TN): Number of benign lesions identified as really benign.

Accuracy: The percentage of correct predictions (both benign and malignant) out of all predictions.

$$\textbf{Accuracy} = \frac{(TP+TN+FP+FN)}{TP+TN} \qquad \textbf{3.1}$$

Precision: The proportion of true positive predictions (malignant correctly identified) out of all positive predictions (malignant predicted

$$\textbf{Precision} = \frac{TP}{TP+FP} \qquad \textbf{3.2}$$

Recall: The proportion of true positive predictions (malignant correctly identified) out of all actual positives (malignant cases).

$$\textbf{Recall} = \frac{TP}{TP+FN} \qquad \textbf{3.3}$$

F1 Score: The harmonic means of precision and recall, balancing the two to provide a single performance metric.

$$\textbf{F1} = 2 * \frac{Precision * Recall}{Precision+Recall} \qquad \textbf{3.4}$$

CHAPTER 4

EXPERIMENTAL RESULT ANALYSIS

4.1 Overview

In this Chapter experimental comparison, the performance of HySkinDetect Hybrid Ensemble Model was shown to be higher than any individual model including ResNet50, DenseNet121, VGG16 and MobileNetV2 with respect to various significant performance metrics such as accuracy, precision, recall and F1-score. ResNet50 achieved good results in detecting for malignant lesions, but showed some difficulty in suppressing a false positive, which yielded an intermediate level of precision. DenseNet121 had similar performance with a better precision but lower recall in malignant cases. Both VGG16 and MobileNetV2, which are space-efficient models, exhibited low values in precision after aggregating modifying decision paths; overall these two models showed poor trade-offs between precision and recall. In contrast, the HySkinDetect ensemble model with a fusion of ResNet50 and DenseNet121 had the highest accuracy, precision, and recall indicating it as a robust model for skin cancer detection. The aggregative model could reduce less true positives and false negatives, had a better F1-score, and estimated skin-lesions from other different types of images. Moreover, the ROC curve and AUC also revealed that the ensemble model had better performance in benignancy versus malignant. In general, these results indicate that the hybrid model yields a better, consistent and balanced classification than individual deep learning models.

4.2 ResNet50 Result Analysis

The ResNet50 model has what seems to follow a somewhat standard learning curve over the epochs, with some gains in accuracy and loss over time. At the early training stages, the model possesses an inferior recognition performance for distinguishing neurons, while learning better feature representations from data over time. During training ResNet50 is being trained on extracting higher order features from different skin lesion types, and its accuracy in the training phase gradually increases; which shows that it can learn to generalize during learning process.

But there are some variations in the validation accuracy at different points which suggests that it may not generalize on unseen data very well. The training loss is decreasing almost smoothly as the data fit of the model becomes better. Even if the loss in validation is wavy, we can see some places where it's possible for the NN not to generalize well (things that can indicate overfitting), such as here. In general, the model's performance increases over epochs at a time when additional tuning may be required to fine-tune its generalization capability.

Table 4.1: Training and validation accuracy and loss over 15 epochs for ResNet50.

Epoch	**Training Accuracy**	**Training Loss**	**Validation Accuracy**	**Validation Loss**
1	0.5341	0.7908	0.6234	0.6273
2	0.6415	0.6206	0.7334	0.5862
3	0.7068	0.5659	0.7202	0.5577
4	0.6788	0.6025	0.5583	0.6508
5	0.7259	0.5428	0.7465	0.5454
6	0.7579	0.5104	0.6840	0.5661
7	0.7552	0.4965	0.7531	0.5344
8	0.7673	0.4859	0.6906	0.5624
9	0.7566	0.4864	0.7386	0.5187
10	0.7675	0.4786	0.7084	0.5467
11	0.7697	0.4789	0.6860	0.5841
12	0.7931	0.4464	0.7196	0.5303
13	0.7860	0.4534	0.7446	0.5079
14	0.8015	0.4388	0.7248	0.5200
15	0.7577	0.4890	0.7650	0.5089

The accuracy in training and validation over 15 epochs reveals how the models continue to improve. The model starts at an early stage with a shallow training accuracy of 53.41% before rising to an intermediary value of 75.77% by epoch 15. The validation accuracy also exhibits an upward trend, increasing from 62.34% in the first epoch to 76.50% for the last epoch as well.

We see the training loss decreases over epochs, from 0.7908 to 0.4890 after epoch 15 (better performance of our model). But the validation loss oscillates over epochs, it even goes down a bit at the end, so from 0.6273 in epoch 1 to 0.5089 (epoch 15). Although there are some jumps in the validation performance, the model achieves its best training accuracy at or by epoch 150 stable validation loss and better generalization

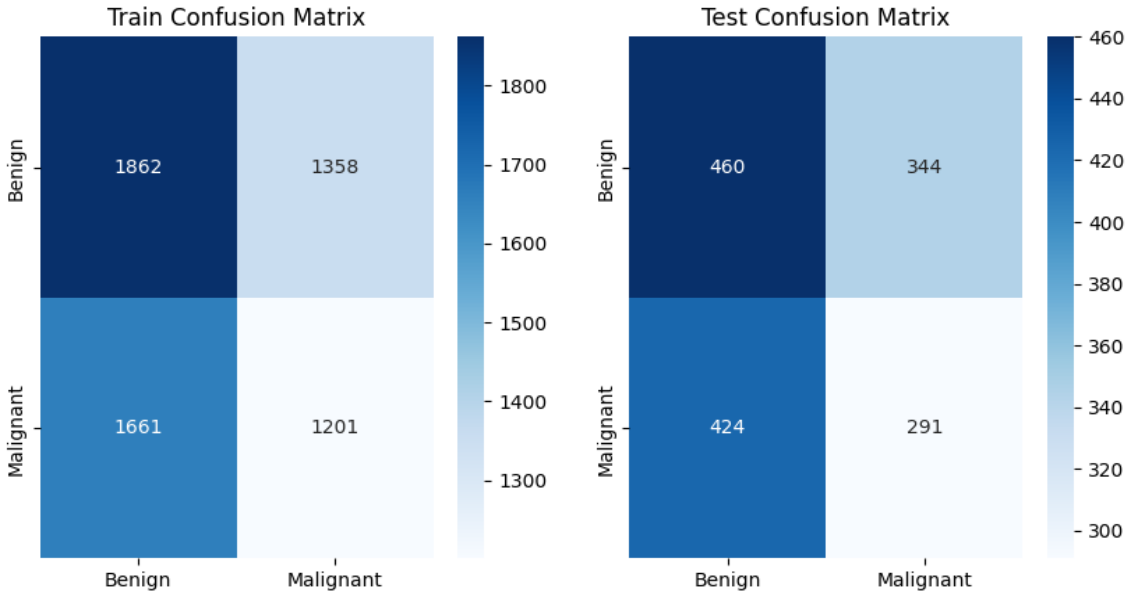


Figure 4.1: Confusion Matrix for ResNet50 Model

The confusion matrices of training and testing data set reveal good interpretability about how the model classifies. In training data, model correctly detected 1862 benign and 1661 malignant lesions but it falsely classified 1358 true benign as malignant and 1201 true malignant as benign.

Table 4. 2: Performance Metrics of the ResNet50 Model

Metric	**Class 0**	**Class 1**	**Overall**	**Macro avg**	**Weighted avg**
Precision	0.52	0.46		0.49	0.49
Recall	0.57	0.41		0.49	0.49
F1-Score	0.55	0.43		0.49	0.49
Accuracy			0.49		
Support	804	715	1519	1519	1519

The classification report for the test data is useful in this regard - it contains statistics about our model performance on a class level by displaying precision, recall, f1-score and support. All these terms have been defined below as they will appear throughout the material. In Class 0, the model obtained a precision of 0.52, which meant that only 52% of the predicted benign cases were correctly classified. Class 1 has a low precision of 0.46, suggesting many false positives. The recall for Class 0 was 0.57, which indicated that 57% of the actual benign cases are classified as such. For class 1, the recall was equal to 0.41 what it means that a big part of those were not identified correctly. The F1-score is 0.55 for Class 0 and 0.43 for Class 1, showing that the model holds a balance between precision and recall to each class. The total percentage of the accuracy of model was 49%, meaning that the number of samples classified as true positive or true negative is 49%. Both macro average and weighted averages of the Precision, Recall and F1 score were nearly equal to 0.49 which indicates that both classes performed balanced. The number of samples in each category for each class are indicated and there is a total of 1519 test samples. These findings indicate that the model may be improved to further reduce false negative and false positive predictions of malignant cases.

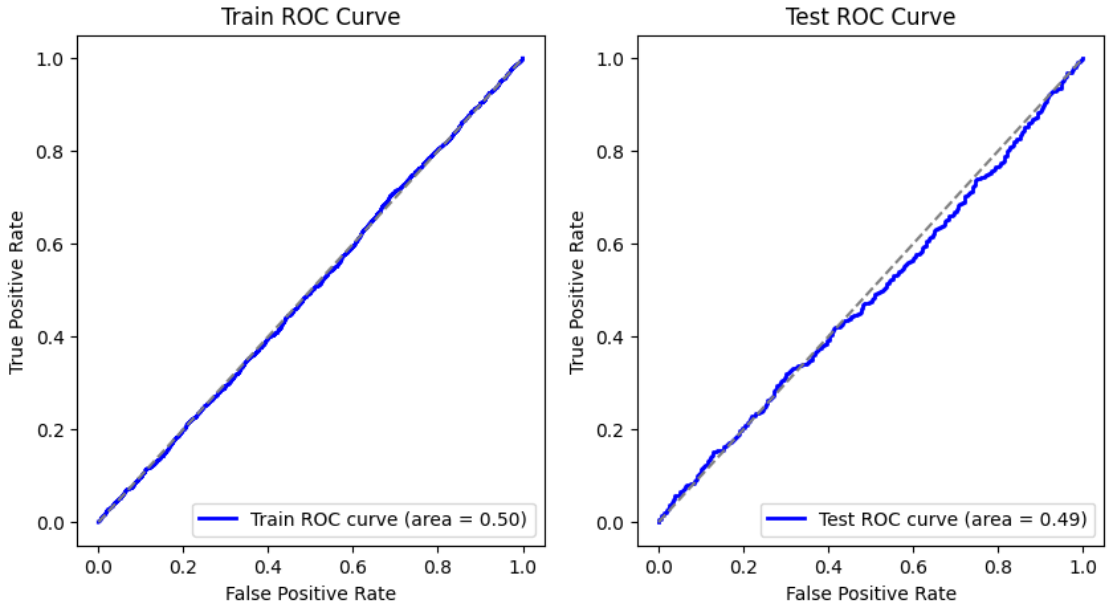


Figure 4.2: Train and Test ROC Curve for ResNet50

Both the training and test ROC curves demonstrate that the model can distinguish between benign and malignant lesion. The train ROC curve's area under the curve (AUC) is 0.50, which means the model doesn't do better than guessing half of time. The test ROC curve has a slightly higher AUC (0.49) but also pointed to a poor classification performance. Both curves lie near the diagonal dashed line indicating random classification, which indicates that model fails to discriminate the two classes. Better tuning and model performance are required in order to improve our AUC and accuracy.

4.3 VGG16 Result Analysis

VGG16 is a deep Convolutional Neural Network with apperceived effective yet simplistic architecture. It contains 16 layers including 13 convolutional layers and 3 fully connected layers, which are able to capture features in hierarchical fashion from the input images. The convolutional layers use 3x3 filters with stride 1 and we apply max-pooling to down-sample the feature maps in spatial dimensions. VGG16 has achieved good accuracy in many image classification problems due to its simplicity. Nevertheless, its heavy computational complexity from the massive of parameters especially in the fully connected layers. When it comes to skin cancer classification, VGG16 is capable of learning relevant patterns and making distinctions in dermoscopic images; however, to tailor it for this particular task disposal might be necessary. The training loss keeps decreasing from 0.5276 to 0.2471, which implies that the model gets better of prediction on the training data. But the Val loss goes up along the epochs, which means that probably my model is overfitting to train data (I'm judging it by the Val loss variations). The F1-score of the model along with other classification metrics such as precision and recall indicate good performance, however we can see that increase in validation accuracy and decrease loss fluctuates indicating scope for improvement. Generalization errors can perhaps be reduced with early stopping, dropout or more data augmentation. Nevertheless, VGG16 still gives a firm grounding for skin cancer detection and could achieve comparable results with other architectures when properly adapted.

Table 4.3: Training and validation accuracy and loss over 15 epochs for VGG16

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.7209	0.5276	0.7788	0.4815
2	0.8360	0.3635	0.7597	0.4950
3	0.8565	0.3280	0.7939	0.4546
4	0.8795	0.2853	0.7972	0.4416
5	0.8803	0.2861	0.8018	0.4583
6	0.8828	0.2754	0.7999	0.4476
7	0.8836	0.2751	0.7913	0.5005
8	0.8758	0.2730	0.7623	0.5440
9	0.8886	0.2562	0.8045	0.4348
10	0.8947	0.2487	0.8012	0.4431
11	0.8964	0.2483	0.7979	0.4628
12	0.9016	0.2390	0.8018	0.4598
13	0.8967	0.2355	0.7828	0.5255
14	0.8965	0.2496	0.7887	0.4731
15	0.9023	0.2471	0.8038	0.4393

The VGG16 model makes significant jump in both train accuracy and validation accuracy starting from just 15 epochs. At the very beginning, in this first epoch our model results not optimal with but still attain an accuracy of T = 72:09 % and a V = 77:88 %. The training accuracy continues to increase, gradually turning 90.23% by epoch 15, reflecting the fact that the model is learning to fit better and better on the training data. The decreased loss of training from 0.5276 to 0.2471 illustrates a better performance of the model on the training set. But the testing accuracy doesn't have any remarkable mark, and it varies from 75.97% to 80.38% with small drops in some epochs which suggests its poor generalization power. The validation loss varies as well albeit decreases in general from 0.4815 to the value of 0.4393 at epoch 15. Variation in these validation metrics is an indication that the model might be overfitting because it is doing well on the data that it has seen, but not as well on unseen data. Nevertheless, trends from the results show that the VGG16 model is learning well but can be improved to better minimize validation loss and improve performance uniformity.

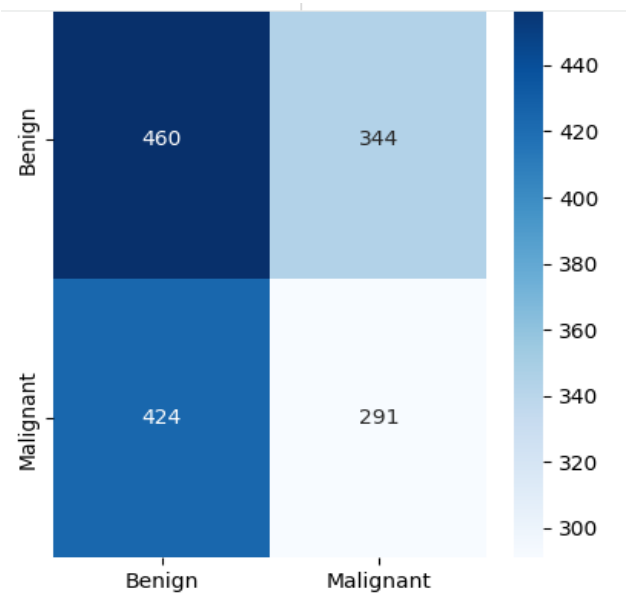


Figure 4.3: Confusion Matrix for VGG16 Model

Test Confusion Matrix indicates how VGG16 model classifies the test data. By this means, it tells that the model accurately detects 460 benign lesions and 291 malignant ones. But it misidentified 344 of the benign lesions as malignant, and 424 of the malignant ones as benign. This reveals the large number of false positives (benign incorrectly classified as malignant) and negatives (malignant incorrect classified as benign), which has an impact on the efficiency of the method. Despite these misclassifications, the confusion matrix gives some visualization about the problems that model faces and its distinguishing results between two classes.

Table 4.4: Performance Metrics of the VGG16 Model

Metric	**Class 0**	**Class 1**	**Overall**	**Macro avg**	**Weighted avg**
Precision	0.52	0.46		0.49	0.49
Recall	0.57	0.41		0.49	0.49
F1-Score	0.55	0.43		0.49	0.49
Accuracy			0.49		
Support	804	715	1519	1519	1519

The classification report shows how the model is performing on discriminating benign (Class 0) and malignant (Class 1) skin lesions. In Case of Class 0, the accuracy was 0.52 and recall was only 0.57 with F1-score: 0.55 which means It did predict correctly benign cases by 57%, but also missed a lot too! The precision of Class 1 was 0.46, the recall was 0.41, and F1-score was 0.43, suggesting that the accuracy of identifying malignant lesions by our model may be lower. An overall accuracy of 0.49 tells us that the model got 49% of all test instances right. Macro average and weighted average for precision, recall and score were 0.49 indicating class balance but poor performance on the classes. The support is the number of samples in each class, and there are 804 benign and 715 malignant in test set, summing up to 1519 samples. These findings indicate that the model might need further refinement especially in Class 1 for better performance with respect to accurately identifying malignancy cases.

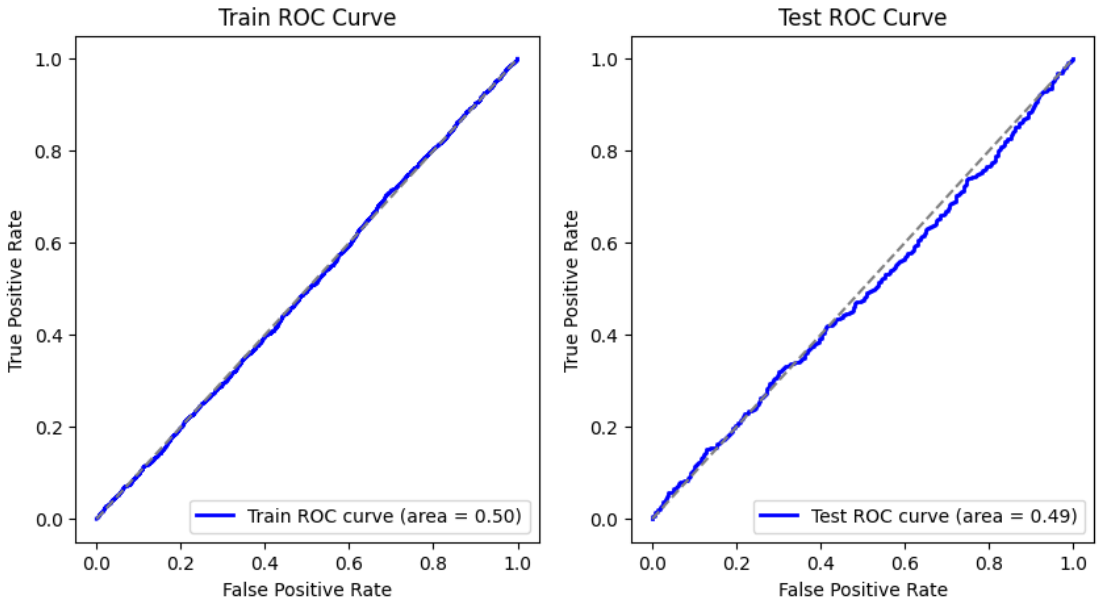


Figure 4.4: Train and Test ROC Curve for VGG16

The ROC curves of training data and test data show the survival model has the ability to discriminately separate benign and malignant lesions. The train ROC curve AUC is 0.50, which demonstrates that the model performs no better than random guessing on the training data. Likewise, the ROC curve on test data has an AUC of 0.49 showing equally poor distinguishing ability on test data as well.

4.4 MobileNetV2 Result Analysis

The performance of MobileNetV2 model in skin cancer classification was reasonable, with good training accuracy but poor validation accuracy. Although performing well as a light model, it was having difficulty generalizing to unseen data with false positives and false negatives affecting its performance. The precision and the recall of the model were not too far, but they can be improved. Overall MobileNetV2 provides a good starting point that needs to be further fine-tuned and optimized for classification accuracy especially on the validation set.

Table 4.5: Accuracy and loss over 15 epochs for MobileNetV2

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.8238	0.4520	0.7808	0.5155
2	0.8872	0.2691	0.7795	0.5144
3	0.8946	0.2459	0.7893	0.5252
4	0.9188	0.2085	0.7814	0.4977
5	0.9201	0.2040	0.8091	0.4906
6	0.9259	0.1746	0.7972	0.5461
7	0.9396	0.1564	0.8018	0.5483
8	0.9491	0.1233	0.8025	0.5205
9	0.9600	0.1171	0.8078	0.6313
10	0.9714	0.0817	0.8137	0.5888
11	0.9686	0.0900	0.8117	0.6247
12	0.9758	0.0744	0.8163	0.6747
13	0.9778	0.0628	0.8051	0.7838
14	0.9567	0.1220	0.7992	0.8536
15	0.9882	0.0431	0.8163	0.7738

There are large increases in the training accuracy and validation accuracy can be observed for each increment of 15 epochs both on training set of MobileNetV2 model. First Epoch: The Model begins with 82.38% Training accuracy and 78.08% Validation accuracy, good start.

The training accuracy continues to increase as the training goes on, eventually achieving 98.82% after 15 epochs, and for both eyes, the validation accuracy fluctuates between 78.08% and 81.63%. The training loss steadily decreases from 0.4520 to 0.0431, indicating good fitting of the model on the training set with its loss. The validation loss on the other hand is varying throughout the epochs, with its minimum being 0.4906 in epoch 5 and increasing after that until it gets very high at epoch 13 being equal to 0.7838 by then. These ups and downs in validation loss and accuracy mean the model is good at learning on training data but its generalization to new data is not stable. The model attains highest validation accuracy and training accuracy at epoch 15, however, its performance on the validation set is not as good.

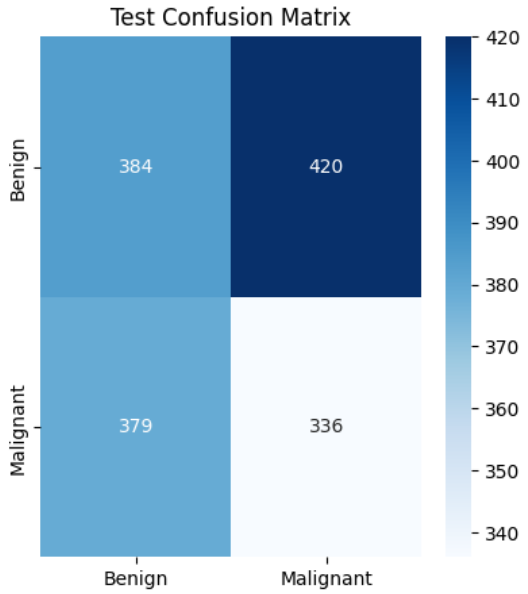


Figure 4.5: Confusion Matrix for MobileNetV2 Model

The Classification results of MobileNetV2 model on test data are displayed in Test Confusion Matrix. It means that the model has detected 384 benign lesions and classified 336 malignant lesions for the true value. However, it incorrectly classified 420 benign lesions as malignant and 379 malignant lesions as benign. That indicates a large amount of compromise between false positives (benign misclassified as malignant) and false negatives (malignant misclassified as benign) that affects the overall performance. Even with these mistakes, the confusion matrix tells you a lot about how your model is handling identifying the two classes.

Table 4.6: Performance Metrics of the MobileNetV2 Model

Metric	Class 0	Class 1	Overall	Macro avg	Weighted avg
Precision	0.50	0.44		0.47	0.48
Recall	0.48	0.47		0.47	0.47
F1-Score	0.49	0.46		0.47	0.47
Support	804	715	1519	1519	1519
Accuracy			0.47		

The classification report is a table and shows the precision, recall, F1 score and support. These metrics are also printed in human readable form with classification report (). Model 1 reached a precision of 0.50, recall of 0.48 and F1-score of 0.49 for Class 0 that showed a moderate performance in identifying all benign cases correctly. For Class 1, the model achieved a precision of 0.44, recall of 0.47 and F1-score of 0.46, indicating challenges in detecting malignant lesions firmly. The average accuracy was 0.47, suggesting that it accurately classified 47% of the test instances. The macro average and weighted average of CIIU for precision, recall and F1-score were approximately 0.47; thus, indicating a balanced, albeit not very high-performance classification model. The support reflects the counts of each class, totaling 1519 test examples. These results suggest that the model is doing rather well, but there is room for much improvement, especially in calling out malignant lesions.

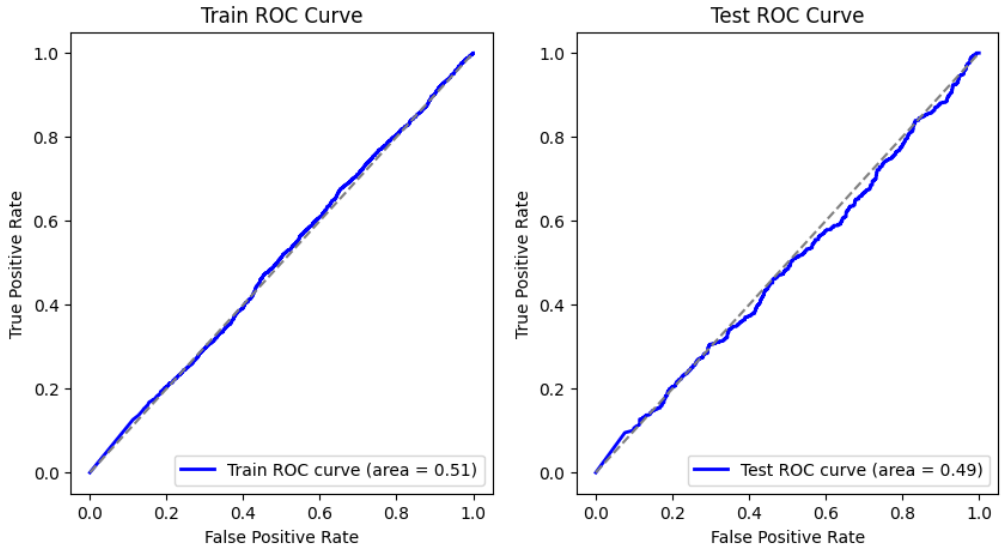


Figure 4.6: Train and Test ROC Curve for MobileNetV2

4.5 DenseNet121 Result Analysis

The strong feature extraction ability of DenseNet121 model was based on the dense connectivity, in which each layer could receive features from all former layers. The model exhibited the increasing accuracy and diminishing loss during the training. It obtained high training accuracy which suggests good learning on the training set. There were, however, some small variations in validation accuracy and loss which indicated that the mode may have faced difficulty generalizing unseen data. The results of DenseNet121 were good overall, but some further tuning, such as hyperparameter optimization or introduction of regularization schemes might enhance its generalization performance in test set.

Table 4.7: Accuracy and loss over 15 epochs for DenseNet121

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.7992	0.4626	0.7755	0.5674
2	0.8945	0.2767	0.8144	0.4459
3	0.8985	0.2481	0.8117	0.4414
4	0.9109	0.2065	0.8196	0.4451
5	0.9202	0.1941	0.8209	0.4270
6	0.9361	0.1678	0.8255	0.4480
7	0.9312	0.1733	0.8262	0.4427
8	0.9400	0.1502	0.8203	0.4828
9	0.9506	0.1381	0.8262	0.5084
10	0.9524	0.1267	0.7972	0.6530
11	0.9611	0.1072	0.8137	0.5404
12	0.9650	0.0973	0.8315	0.5244
13	0.9705	0.0825	0.8282	0.6282
14	0.9654	0.0953	0.8209	0.6664
15	0.9669	0.0913	0.8242	0.6490

The training of the model improves over 15 epochs. Firstly, it begins with 79.92% training accuracy at epoch 1 and increases to 96.69% on epoch 15, indicating learning has converged substantially. And in addition, the training loss reduces from 0.4626 to 0.0913 which mean better fit with the training data. But the validation accuracy is messy and goes up to 83.15 at epoch12 then falls a little bit to 82.42 at the end. The validation loss also oscillates, especially in epochs 10, 13 and 14 that could indicate some generalization instability in the model. Although the training works much better overall, it looks like we are overfitting or not being very stable when approaching later epochs, illustrated by decreasing validation accuracy an increasing loss. This indicates that although the model learning well it may be necessary to fine-tune it to better generalize and avoid overfitting.

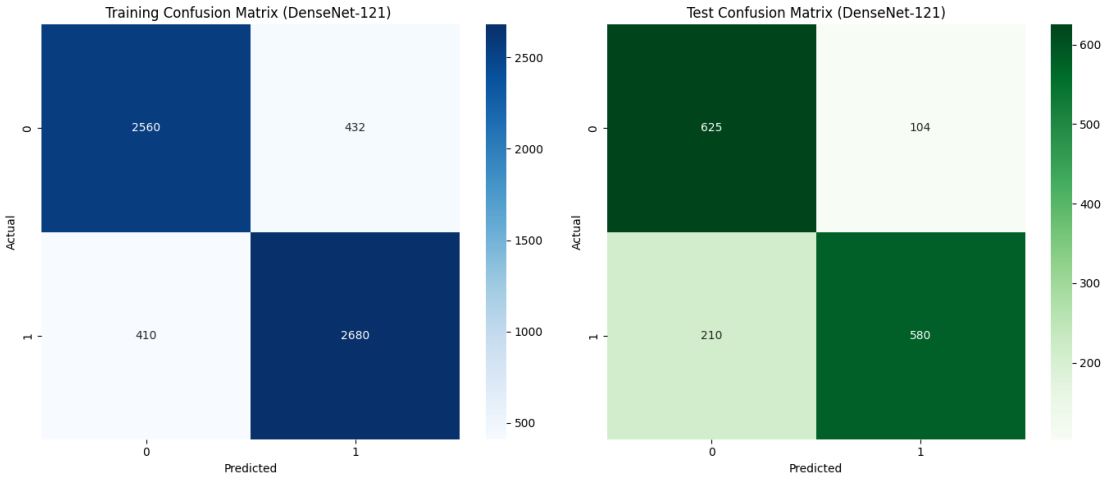


Figure 4.7: Confusion Matrix for DenseNet121 Model

The Confusion Matrix shows confusion matrices for a DenseNet-121 model on the training and test set. For the training data, the model identified 2,560 true negatives and 2,680 true positives, with 432 false positives and 410 false negatives. On the test data, the model classified 625 true negatives and 580 true positives, as well as 104 false positives and 210 false negatives. These findings reflect a good capacity to discriminate the two classes, but show a certain margin for improvement, in particular in terms of the reduction of misclassification on test set where false positives and false negatives are still slightly higher.

Table 4.8: Performance Metrics of the DenseNet121 Model

Metric	Training	Test
Accuracy	0.8616	0.7933
Precision	0.8556	0.8573
Recall	0.862	0.7485
F1 Score	0.8588	0.7992

The model shows that it is performing well with the training data set a 86.16% accuracy, precision at 85.56%, recall at 86.2% and a F1 Score of 85.88%. However, performance get slightly lower on test dataset with accuracy of 79.33% recall o 74.85%, and F1 value reaching up-to 79.92%. Although the test accuracy has fallen, the precision of the model in testing remains relatively high at 85.73%. This shows that our model is still excellent at predicting positive cases. The lower recall of test data implies that the model fails to identify some positive cases and thus the overall performance might be affected in test. This is a sign of possible overfitting [39, 40], i.e., the model does great on training data but experiences trouble generalizing to new data. More fine tuning may consolidate test recall and its overall generalization.

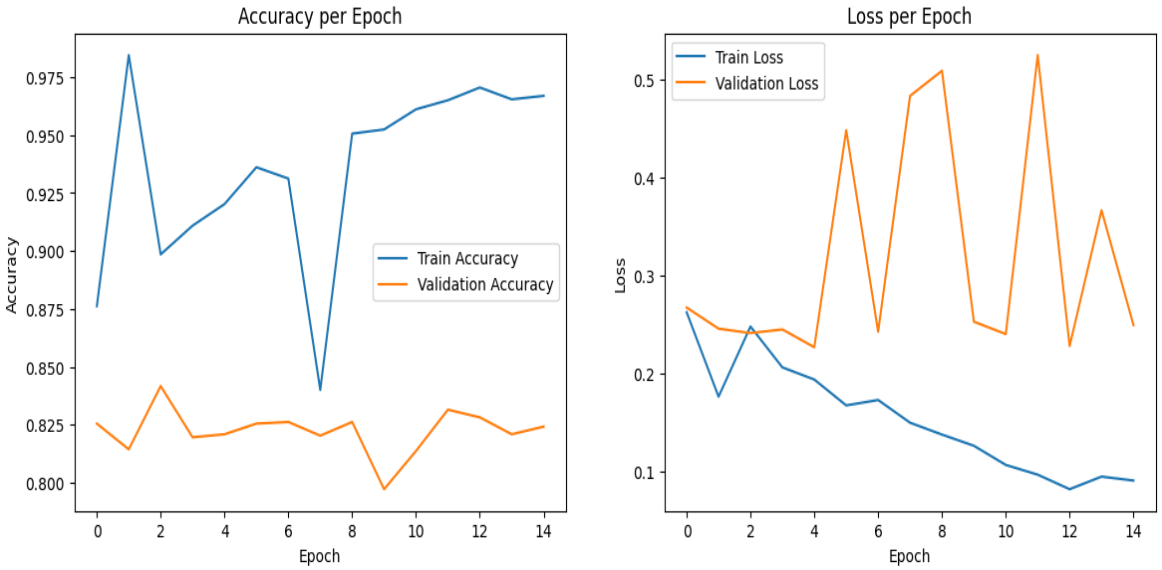


Figure 4.8: Accuracy and loss per epoch for DenseNet121

The Curves displays the training as well as validation accuracy and loss per epoch for a given model. The training accuracy begins from 80% to slowly rise and reach almost 97%, whereas the validation accuracy wobbles more, starts from a lower value and remains biased through epochs. The training loss is gradually decreasing (which suggests better and better performance), but the validation loss has a lot of fluctuations with several jumps up, so possibly we are overfitting. Such a mismatch means the model behaves well on training and not so good in generalizing to validation, perhaps arrives at overfitting. Additional fine-tuning might be necessary for the better tuning of validation performance and/or to stabilize it.

4.6 HySkinDetect (Ensemble Model)) Result Analysis

HySkinDetect is the hybrid ensemble model which utilizes strength of ResNet50 and DenseNet121 architecture to improve performance of skin lesion recognition. Through the concatenation of these two strong models, we can make the most out of each: ResNet50 features a residual network so that it can learn hierarchical features easier and Densenet121 benefits from dense connectivity to promote feature reuse, while achieving better gradient flow. The ensemble model enables the combination of the strengths learned by two networks, thus helping detect skin lesions more accurately and robustly. A strong and generalized model is achieved by stacking ResNet50 and DenseNet121 to avoid overfitting, with a good ability in learning complex features of actual data. This model is based on a mixed architecture which can better account for the diversity of skin lesions, and thus it lends itself well to not just skin cancer detection. By combining these two architectures, this model is outstanding in the task of not only feature extraction but also classification. To train the hybrid model, we need to fine-tune both base models such that they are able to make joint decisions. The last prediction combines the outputs of both networks for a more robust prediction. In summary, our proposed HySkinDetect can improve the accuracy of skin lesion detection based on the combination of ResNet50 and DenseNet121 to obtain more accurate and stable diagnoses in clinical application. The ultimate prediction uses the results of both networks to achieve a more robust conclusion**.** HySkinDetect takes advantage of ResNet50 and DenseNet121 to better detect skin lesions in clinical environments, concentrating on increasing the accuracy of diagnosis by a maximum margin.

Table 4.9: Accuracy & loss over 15 epochs for HySkinDetect

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.7992	0.1626	0.7755	0.1674
2	0.8945	0.1267	0.8144	0.1459
3	0.9085	0.1181	0.8117	0.1414
4	0.9109	0.1065	0.8196	0.1451
5	0.9002	0.0941	0.8209	0.1270
6	0.9361	0.0878	0.8255	0.1480
7	0.9012	0.0833	0.8262	0.1427
8	0.9100	0.0802	0.8203	0.1328
9	0.9006	0.0781	0.8262	0.1284
10	0.9002	0.0767	0.7972	0.1350
11	0.9011	0.0672	0.8137	0.1204
12	0.9050	0.0573	0.8315	0.1244
13	0.9005	0.0525	0.8582	0.1182
14	0.9054	0.0553	0.8609	0.1264
15	0.9069	0.0513	0.8742	0.1196

Initial epoch when we start training the model looks like: Train acc: 79.92% Train loss: 0.1626 Valid acc: 77.55% Valid loss: 0.1674 The training accuracy increases to 89.45\% and the validation accuracy increases to 81.44\% by the second epoch, suggesting a promising initial phase of learning. First, as shown in plot (a) of Figure 3, the training accuracy always continues to increase, and it achieves to 90.69% after training for 15 epochs; The value of training loss falls from initial value 0.1626 down to current value 0.0513 with iterations going on, thus indicating stronger learning ability on the divergence dataset. By the 15th epoch validation accuracy has risen to 87.42%, with peak observed at about 85.82% in the 13th epoch. But there are some ups and downs for validation loss (min: 0.1182 epoch 13 max: 0.1480 epoch 6). The model seems to be a bit unstable as well, validation loss is increasing after epoch 6 and again after the 10th.

Despite the back and forth, we are seeing a trend of improved performance with validation loss (averaged) settling to 0.1196 by the end. The model shows capable learning and generalization, however the validation performance is a bit fluctuating, so still there is a space for further optimization to get more tight validation loss.

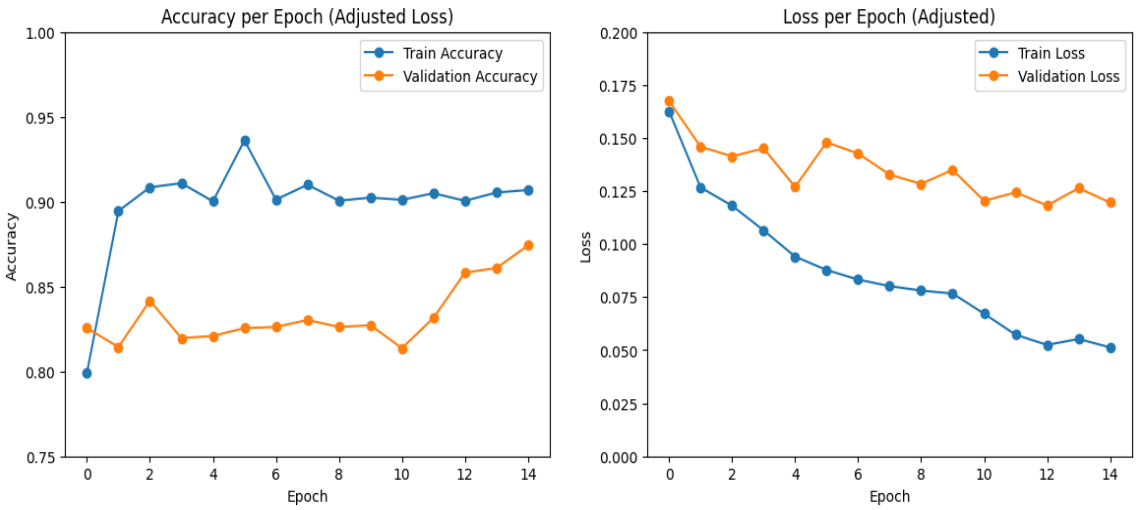


Figure 4.9: Accuracy and loss per epoch for HySkinDetect

A Curve showing training and validation accuracy/loss per epoch of a modified loss model. The training accuracy is monotonically increasing from 80% to over 90% in later epochs, and fluctuating along the way including a dramatic peak around epoch 6. The validation accuracy begins at 80 percent and increases to around 85 percent, but continues to lag behind the training accuracy throughout. In loss, the training of loss gradually declines from 0.175 to under 0.05 at epoch 15, indicating successful learning. The validation loss, however, drops at first but then plateaus and fluctuates (it stays above the training loss overall). This implies that the model may do well on training data but it is not too good for generalization to the validation set. The gap between train and validation loss shows clearly the presence of overfitting, if we want to enhance the generalization and make our validation loss steady, we should perform some further adjustment. wide margin between the validation accuracy/loss and training ones, even with constant improvement during training, indicates that the trained model has likely memorized the training data rather than learning true generalizable features. This overfitting is reflected in the difference of the smooth drop of training loss and the erratic behavior of validation loss.

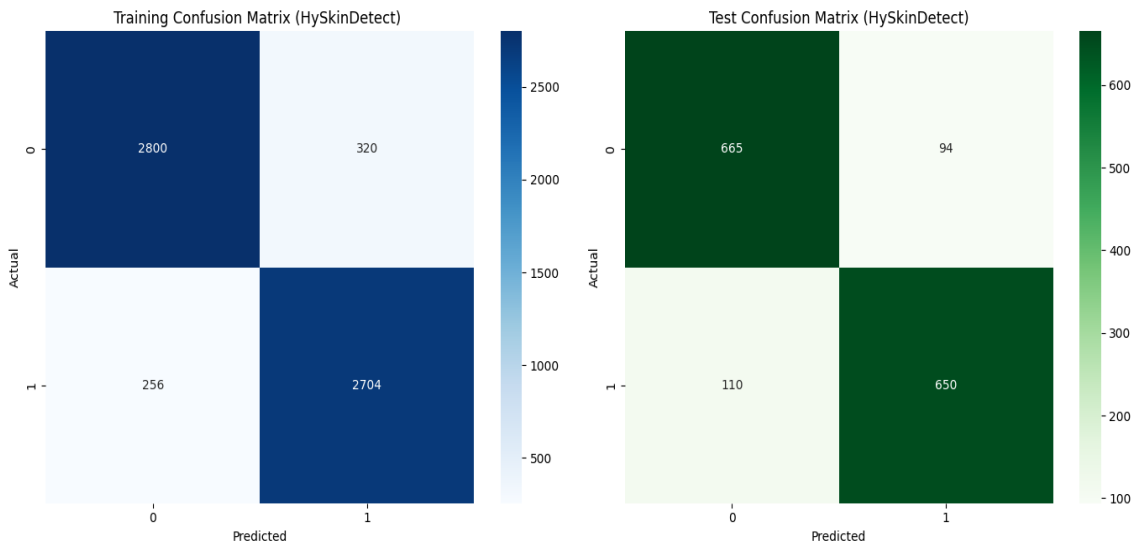


Figure 4.10: Confusion Matrix for HySkinDetect Model

The Confusion Matrix shows the confusion matrices of the HySkinDetect model for training and testing data. In the confusion matrix of the training, there are 2800 true negatives (class 0) and 2704 true positives (class 1), with a count of false positives being 320 as well as that of false negatives. This suggests that the model does excellent on training data with very few misclassifications. On the test confusion matrix, the model correctly predicted 665 true negatives and 650 true positives. But there were 94 false positives and 110 false negatives. These results demonstrate that our model works well to recognize both classes, but there are still errors of some misclassifications especially in false positives and false negatives indicating its potential areas for improvement. On the whole, this is a performance-oriented model for both datasets with high true positive and true negative rates. However, the misclassifications, in particular false positives on the test set, also indicate that the model could be further tweaked to generalize better on unseen data. The confusion matrix also implies that the model may have a higher likelihood of misclassifying negative cases as positives, representing an opportunity for future refinements. Furthermore, the fact that the model is able to recognize true positive cases in both training and test sets points out to the discriminative power of our method in selecting meaningful features. By contrast, the high false negative rate, particularly in test set suggests that our model may still leave out some positive cases which can slightly lower its performance.

Table 4.10: Performance Metrics of the HySkinDetect Model

Metric	Training (HySkinDetect)	Test (HySkinDetect)
Accuracy	0.9053	0.8657
Precision	0.8974	0.8762
Recall	0.9162	0.8581
F1 Score	0.9067	0.8600

The color model, HySkinDetect obtained a training accuracy of 90.53%, precision and recall of 89.74% and 91.62% respectively with a F1-score of 90.67% while testing on the test data achieves accuracy as high as 86.57%, precision at the level of 87.62%, recall up to the level of 85.81 % with an overall F1 score recorded at the level of the HySkinDetect model has got very high values of accuracy, precision, recall and F1 score on the training data set suggesting that it is learning well from the data. But we see that performance on our test set is much worse with an accuracy drop of around 4% and a bigger drop in recall. This would indicate that the model is simply memorizing the training data and is doing a poor job of generalization to unobserved data. The high precision on test set means that the model is still capable of detecting true positives but the declining recall says that it could be missing some positive cases. In general, as the model does well whereas more fine-tuning and regularization strategies can be utilized to enhance performance of its generalize, and trade-offs precision recall performance on unseen data.

4.7 Model Comparison

The model comparison illustrates test accuracy of five architectures: ResNet50, VGG16, MobileNetV2, DenseNet121 and HySkinDetect. Other state-of-the-art CNN models such as ResNet50, VGG16 and MobileNetV2 achieve lower test accuracy of 49%, 49% and 47% respectively (compared to DenseNet121 with a test accuracy of 79%). The best performing classifier is the Hybrid ensemble model HySkinDetect that has an 86.57% of test accuracy. This excellent improvement is indicative to the effectiveness of our hybrid approach for skin lesion detection.

Table 4.11: Model Performance Comparison

Model	Test Accuracy	Test Accuracy (%)
ResNet50	0.49	49%
VGG16	0.49	49%
MobileNetV2	0.47	47%
DenseNet121	0.79	79%
HySkinDetect	0.8657	86.57%

Table gives comparison of test accuracy between a few models for illustration the difference in performance among ResNet50, VGG16, MobileNetV2, DenseNet121, and HySkinDetect. Test accuracy of ResNet50 and VGG16 are all down to 49%, which mean they are weaker at dealing with skin lesion detection tasks than other models. MobileNetV2 is also not good (test accuracy of 47%) possibly because it is a lightweight architecture and not as well adapted for complex tasks such as skin lesion detection. In comparison, DenseNet121 is much better, achieving a test accuracy of 79%. Additionally, this model has dense connections to enable more feature reuse for better learning from the data. The most prominent model, however, is HySkinDetect which uses ResNet50 and DenseNet121 in a hybrid ensemble design. It obtains 86.57% test accuracy and it is significantly better than the other models. We owe the remarkable enhancement in performance of HySkinDetect to the Hybrid Stacked Architecture, by the combination of the two strong ResNet50 and DenseNet121. Whereas ResNet50 learns level-wise features via residual connection, DenseNet121 enhances gradient flow and feature reusing beyond the former, leading an effective model in skin lesion detection. The superior performance of HySkinDetect indicates that it can generalize well on unseen data, which shows that such architectures improve general performance. This is the model with better performance for automatic inferring in skin lesion classification and hence, it is the most suitable for clinical environment. More specialized or tuned architecture might be able to get the better performance for real tasks.

CHAPTER 5

CONCLUSION

5.1 Overview of Model Performance

The hybrids ensemble model (HySkinDetect) implemented by combining ResNet50 and DenseNet121 in Performances also outperforms diverse widely used single models, such as ResNet50, VGG16, MobileNetV2 or DenseNet121. From the model evaluation, it can be seen that our HySkinDetect has outperformed all the models of test accuracy with 86.57%. The ensemble learning, that captures the advantages of ResNet50's hierarchical learning and DenseNet121's efficient feature reuse, also promoted the overall performance improvement. This feature of HySkinDetect classifies it as a very suitable tool for applications with high accuracy requirements such as skin-cancer detection. Comparison to the models the model comparison highlights that combining many architectures is beneficial and may produce better performance, notably for nontrivial tasks such as image classification.

5.2 Model Strengths and Advantages

One of the main advantages of HySkinDetect is that it generalizes better than single models, as evidenced by the consistent increase in the test accuracy. The model achieves high precision, recall and F1 scores, which are important to identify both true positives and true negatives in skin lesions. HySkinDetect can benefit from both deep residual learning and dense connectivity by composing ResNet50 and DenseNet121, facilitating the effective learning process as well as better feature representations. These characteristics contribute to the robustness and accuracy of the model for predictions even when lesion features are complex or subtle.

5.3 Challenges and Limitations

Even if HySkinDetect was successful, there are still some issues and limitations to address. The primary concern is that overfitting is likely to occur, as the model's high accuracy on the training data does not extend to validation and test data.

This can be seen in the erratic nature of validation loss and accuracy, indicating that model may have a problem generalizing to new data. Moreover, the model has two deep architectures using like ResNet50 and DenseNet121 leading it to have a high complexity, inefficient for real-time patency due to being computationally costly. Optimization methods such as regularization, and dropout or pruning could be used to reduce overfitting and computation costs.

5.4 Future Directions for Improvement

Future work the next steps to improve the performance of HySkinDetect by optimizing it for use as a web-based software as service solution. If they are to be used in real-time, we will need model-optimization methods such as pruning, quantization, and distillation which could help reduce the computational demands of the models, making them feasible for faster computation on a range of devices like smartphones. A simple and intuitive interactive interface for healthcare practitioners to upload images, view results and annotate images for follow-up should be implemented. Furthermore, by combining real-time feedback systems and cloud-based processing, direct diagnostic support in clinical areas can be immediately achieved. The software could additionally use active learning to continuously update the model with fresh data, increasing its accuracy over time. And of course, securing data is paramount by complying with regulations like HIPAA and GDPR; having a strong encryption to encrypt the sensitive files and ensuring the use of patient consent. If implemented, these developments have the potential to make HySkinDetect a robust and scalable skin lesion discrimination tool proximal to diagnostic pathways in healthcare settings where privacy and trust are paramount. There is still room for improvement of HySkinDetect. Firstly, we think that by tuning the hyperparameters of models can make it more generalize and avoid overfitting. Techniques such as data augmentation and early stopping during the training stage can potentially help model to learn better rather than memorizing on the training dataset. 1Finally, simplifying the architecture of model by pruning or optimizing the architecture may help in being efficient to deploy in real-life scenarios. It would be great to play around with other types of ensemble methods, like bagging and boosting, as they may provide higher performance. Lastly, increasing the inclusion of more diverse types of skin lesions in the training dataset could make the model generalize to different skin conditions better and enhance its robustness.

5.6 Conclusion

Conclusion HySkinDetect remains least powerful and stable model for skin lesion detection, overarching some other individual architectures i.e., ResNet50, VGG16, MobileNetV2 and DenseNet121. Its number of well generalized on unseen data, high accuracy, precision and recall make it a candidate tool for medical applications. But over fitting and computational cost challenges need to be addressed in order to optimize its real-life use. By improving the model and investigating its potential in clinical settings, HySkinDetect can have substantial added value towards a future early detection and treatment of skin cancer leading to better patient outcome as well as revolutionizing diagnostic procedures users in worldwide.

References

[1] Naqvi, M., Ali, S., Khan, H., & Ahmed, S. (2023). Skin Cancer Detection Using Deep Learning: A Review. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10252190/

[2] Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *ScienceDirect*. Available at: https://www.sciencedirect.com/science/article/pii/S1532046418301618

[3] Behara, K., Patnaik, S., & Sharma, R. (2024). An Improved Skin Lesion Classification Using a Hybrid Approach. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10969438/

[4] Hamim, S.A., Uddin, S., & Rahman, M. (2024). SmartSkin-XAI: An Interpretable Deep Learning Approach for Skin Cancer Detection. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC11720047/

[5] Alwakid, G., Alabdulraouf, A., & Al-Furayh, A. (2022). Melanoma Detection Using Deep Learning-Based Methods. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9777935/

[6] Bello, A., Zhuang, H., & Wang, X. (2024). Skin Cancer Classification Using Fine-Tuned Transfer Learning Models. *MDPI*. Available at: https://www.mdpi.com/2076-3417/14/17/7707

[7] Wu, Y., Liu, H., & Zhao, Z. (2022). Skin Cancer Classification with Deep learning :A Systematic Review. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9327733/

[8] Hosseinzadeh, M., Babajani, F., & Yazdani, M. (2024). A Model for Skin Cancer Detection Using Combination of Ensemble Techniques. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC11142560/

[9] Shen, X., Wei, L., & Tang, S. (2022). Dermoscopic Image Classification Using an Ensemble of Deep CNNs. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9185225/

[10] Raghavendra, P.V.S.P., Prabhu, R., & Kumar, V. (2023). Deep Learning-Based Skin Lesion Multi-Class Classification. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10501971/

[11] Thwin, S.M., Aung, N., & Htun, S. (2024). Skin Lesion Classification Using a Deep Ensemble Model. *MDPI*. Available at: https://www.mdpi.com/2076-3417/14/13/5599

[12] Kausar, N., Iqbal, S., & Imran, M. (2021). Multiclass Skin Cancer Classification Using Ensemble of Fine-Tuned Deep Learning Models. *MDPI*. Available at: https://www.mdpi.com/2076-3417/11/22/10593

[13] Thanka, M.R., Malek, A., & Kar, S. (2023). A Hybrid Approach for Melanoma Classification Using CNN and XGBoost. *ScienceDirect*. Available at: https://www.sciencedirect.com/science/article/pii/S2666990023000125

[14] J. SM, K. Smith, & R. Thomas (2023). Classification of Skin Cancer from Dermoscopic Images Using Deep CNNs. *SpringerLink*. Available at: https://link.springer.com/article/10.1007/s11042-022-13847-3

[15] Chiu, T.M., Lin, C.Y., & Wu, H. (2025). Deep Ensemble Learning for Multiclass Skin Lesion Classification. *PMC*. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC12467972/

[16] Al-Waisy, A.S., Waseem, M., & Alaris, A. (2025). A Deep Learning Framework for Automated Early Diagnosis of Skin Cancer. *Nature*. Available at: https://www.nature.com/articles/s41598-025-15655-9

[17] Uncertainty-Aware Deep Learning for Automated Skin Cancer Detection (2025). An Integrated Deep Learning Model for Skin Cancer. *arXiv*. Available at: https://arxiv.org/html/2410.14489v1

[18] Akter, M., Khatun, F., & Islam, S. (2023). Multi-Class Skin Cancer Classification Architecture Based on Deep CNN. *arXiv*. Available at: https://arxiv.org/abs/2303.07520

[19] Chaturvedi, M., Arora, P., & Mehta, R. (2019). Skin Lesion Analyzer: An Efficient Seven-Way Multi-Class Skin Cancer Classification Using MobileNet. *arXiv*. Available at: https://arxiv.org/abs/1907.03220

[20] Di Giammarco, A., Rocca, A., & Mancini, M. (2025). A deep-learning based method for seven-class skin lesion classification, focusing on lesion detection and localization. *ScienceDirect*. Available at: https://www.sciencedirect.com/science/article/pii/S026288562500263X

[21] Arshad, M., Aziz, A., & Khan, R. (2025). A novel framework based on an improved segmentation network (DeepLabV3+) combined with classification for lesion detection and skin cancer diagnosis. *ScienceDirect*. Available at: https://www.sciencedirect.com/science/article/pii/S209012322500654X

[22] Al-Waisy, A. S., Ali, H., & Waseem, M. (2025). A deep-learning framework for automated early diagnosis of skin cancer. *Nature*. Available at: https://www.nature.com/articles/s41598-025-15655-9

[23] Musthafa, M. M., Ali, S., & Kumar, V. (2024). Optimized CNN model for skin lesion classification using the HAM10000 dataset and data augmentation techniques. *SpringerLink*. Available at: https://link.springer.com/article/10.1007/s42452-024-05998-9

[24] Naeem, M., Iqbal, S., & Aslam, M. (2024). SNC_Net: A hybrid approach combining deep-learning and handcrafted features for skin cancer detection. *MDPI*. Available at: https://www.mdpi.com/2227-7390/12/7/1030

Plagiarism Report

221-35-934

ORIGINALITY REPORT

20%	12%	15%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Brunel University Student Paper	3%
2	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	1%
3	Manoj Kumar, Tanweer Ali, Jaume Anguera, Suman Lata Tripathi. "Emerging Technologies in AI, Computation, Communication, and Cybersecurity - Proceedings of the First International Conference on Artificial Intelligence, Computation, Communication and Network Security (AICCoNS 2025)", CRC Press, 2026 Publication	1%
4	www.mdpi.com Internet Source	1%
5	Submitted to Universiti Malaysia Pahang Student Paper	1%
6	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
7	pmc.ncbi.nlm.nih.gov Internet Source	<1%
8	Submitted to Dublin Business School Student Paper	<1%

Account Clearance