Skin cancer image classification using convolution neural network with transfer learning based on VGG16

ELHAN MANJID MUSTAKIN

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Elhan Majid Mustakin
Date of Birth	:	13 Oct 2003
Title	:	Skin cancer image classification using convolution neural network with transfer learning based on VGG16
Academic Session	:	2022-2025

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

Elhan Manjid

_____
(Student's Signature)

221-35-1001

_____
Student ID
Date:

_____
(Supervisor's Signature)

Dr. S M Hasan Mahmud

_____
Name of Supervisor
Date:

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Daffodil International University,
Daffodil Smart City,
Ashulia.Dhaka,Bangladesh

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name
Thesis Title

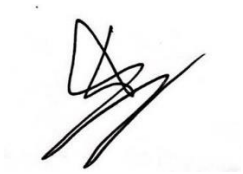
Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,



_____
(Supervisor's Signature)

Date:

Stamp:

Note: This letter should be written by the supervisor and addressed to the Librarian, *Daffodil International University* with its copy attached to the thesis.

APPROVAL

This thesis titled on **"Skin cancer image classification using convolution neural network with transfer learning based on VGG16"**, submitted by **Elhan Manjid (ID: 221-35-1001)** 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

SUPERVISOR'S DECLARATION

The thesis I hereby declare that I have reviewed this thesis and in my view, this thesis is satisfactory with regard to scope and quality that the degree of Bachelor of Science be awarded.

_____

(Supervisor's Signature)

Full Name : Dr. S M Hasan Mahmud

Position : Associate Professor

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Elhan Manjid

_____

(Student's Signature)

Full Name : Elhan Manjid Mustakin

ID Number : 221-35-1001

Date :

Skin cancer image classification using convolution neural network with transfer learning based on VGG16

ELHAN MANJID MUSTAKIN

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science.

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

I am very thankful to Allah (SWT) for blessing me with the strength, patience, and guidance I needed to complete this thesis. This success could not have been possible without His blessings.

I would like to express my deep sense of gratitude to my supervisor, Associate Professor Dr. S. M. Hasan Mahmud, for providing me with excellent guidance, support throughout this project, and feedback during my research endeavor. They have significantly helped shape the more technical and analytical aspects of this project through their knowledge of software engineering.

I want to express my deepest gratitude to my blessing parents for their love, prayers, and support. They have always been there for me when I needed them, both spiritually and emotionally. I also want to thank my friends and classmates for being there for me, working with me, and talking to me in a fun way. I want to give a huge shout-out to my classmates and friends for working on, gossiping about, laughing at (and occasionally just laughing), and just being friendly people throughout this. They have shown me so many things in different respects.

Finally, I would like to express my appreciation to everyone who contributed in any way to the realization of this thesis.

DEDICATION

This work is lovingly dedicated to all those who have fought, and are still fighting, the quiet war against skin cancer. To the patients whose bravery motivated this study, and to the families that supported them with hope that was greater than fear, you are the cause for every word of this research.

To my parents, who taught me that caring for others is the best way to learn, and to my teachers, who showed me how science can help people. Your faith in me has been the thing that has pushed me the most.

And last but not least, this work is for everyone who might one day be assisted by early diagnosis. May it shine a light on a problem that science is still trying to solve.

ABSTRACT

Skin cancer is a very dangerous type of cancer and is still one of the most common causes of death around the world. It is important to find and correctly diagnose things early to raise survival rates, but old ways of doing things aren't always reliable. Recent advancements in deep learning have significantly improved the automation of skin lesion evaluations, bringing positive outcomes in the work of dermatologist. A skin lesion is an abnormal area of skin where malignant tumors grow out of control, which is different from benign tumors. This research developed a Convolutional Neural Network (CNN) model for skin cancer detection, utilizing VGG16 as the foundational architecture via transfer learning. The research utilized the ISIC2018 dataset to categorize two primary tumor types malignant and benign while also incorporating a variety of seven different skin cancer types from the HAM10000 dataset for a more comprehensive examination. The CNN model performed exceptionally well, achieving a training accuracy of 96.46%, a training loss of just 0.092%, a precision of 0.90%, a recall of 0.91%, and an F1-score of 0.90%. These findings go beyond previous work in the area, showcasing the increasing power of deep learning methods in skin cancer detection. They offer more efficient and simplified models that can handle a variety of data types. Although deep learning continues to make rapid strides, there is still a need for further research to address existing challenges and discover new approaches to automating the detection of skin cancer.

TABLE OF CONTENT

DECLARATION

TITLE PAGE

ACKNOWLEDGEMENTS **ii**

Dedication **iii**

ABSTRACT **iv**

TABLE OF CONTENT **v**

LIST OF TABLES **vii**

LIST OF FIGURES **vii**

LIST OF ABBREVIATIONS **viii**

INTRODUCTION **1**

1.1 Background 1

1.2 Problem Statement 3

1.3 Motivation 4

1.4 Research Questions 5

1.5 Research Objective 5

1.6 Research Scope and Limitations 6

1.6.1 Scope 6

1.6.2 Limitations 6

1.7 Thesis Organization 7

LITERATURE REVIEW **8**

2.1 Related Works 8

2.1 Research Gap 14

METHODOLOGY **16**

3.1 Data Collection 17

3.2Data Preprocessing 20

3.2.1 Data Augmentation and Image Pre-processing 20

RESULTS AND DISCUSSION 29

4.1 Result Analysis 29

4.2 VGG19 based transfer learning model 32

4.3 Visualization 35

4.4 Web Application 36

4.5 Discussion 37

CONCLUSION 39

5.1 Findings and Contributions 39

5.2 Recommendations for Future Works 40

REFERENCES 42

APPENDICES 44

LIST OF TABLES

Table 3.1 Dataset Class Distribution
Table 3.2 Image Augmentation Parameters
Table 3.3 Train–Test Split Summary
Table 3.4 Model Hyperparameters Used in Training
Table 3.5 Evaluation Metrics Description
Table 4.1 Confusion Matrix for VGG16 Model
Table 4.2 Classification Report (VGG16)
Table 4.3 Comparison Between CNN, VGG16 & VGG19

LIST OF FIGURES

Figure 3.1 Sample Dataset Images
Figure 3.2 Image Augmentation Examples
Figure 3.3 VGG16 Model Architecture
Figure 4.1 Accuracy Curve
Figure 4.2 Loss Curve
Figure 4.3 ROC Curve
Figure 4.4 Confusion Matrix (VGG16)
Figure 4.5 VGG19 Architecture
Figure 4.6 Model Comparison Chart

LIST OF ABBREVIATIONS

CDNN	Convolutional Deconvolutional Neural Network
DL	Deep Learning
ELM	Epiluminescence Microscopy
FCL	Fully Connected Layer
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
HSV	Hue–Saturation–Value
KS	Kaposi Sarcoma
LAB	Luminance–A (green–red) – B (blue–yellow) Color Space
ML	Machine Learning
MEL	Melanoma
NMES	Non-Melanoma Skin Cancer
ROC	Receiver Operating Characteristic
RELU	Rectified Linear Unit
SCC	Squamous Cell Carcinoma
SCDNET	Skin Cancer Detection Network
FCM	Fuzzy C-Means

INTRODUCTION

1.1 Background

Cancer starts to grow when some cells stop responding to the body's normal growth signals. They don't divide in a controlled way anymore; instead, they start to multiply on their own. If this abnormal growth isn't caught and treated early, the cells could spread to nearby tissues and eventually form malignant tumors, which are the most dangerous type. Skin cancer falls into this category and appears in several forms, the most common being melanoma (MEL), basal-cell carcinoma (BCC), non-melanoma skin cancer (NMSC), and squamous-cell carcinoma (SCC). Actinic keratosis (AKIEC), Kaposi sarcoma (KS), and Bening are less common types [1–3]. Globally, the number of skin-cancer cases continues to rise each year. Skin lesions-both malignant and benign-add considerably to the overall healthcare burden, as they often lead to higher treatment costs and can greatly affect a patient's quality of life [5,6]. Early detection is essential for improving survival rates, particularly in fast-spreading cancers like melanoma. Researchers are therefore constantly searching for more accurate and flexible methods to detect unusual skin changes at the earliest possible stage. In routine clinical practice, dermatologists assess suspicious skin lesions using tools like dermoscopy or epiluminescence microscopy (ELM). These methods give a clearer, magnified view of the skin surface, helping doctors observe subtle details—such as pigment distribution, streaks, and small changes in colour—that are not easily seen with the naked eye [7,8]. To make sense of these enhanced images, clinicians rely on several well-established diagnostic guidelines. One of the most widely applied is the ABCD rule (Asymmetry, Border irregularity, Color variation, Diameter, and Evolution), which helps assess a lesion's asymmetry, borders, colors , and diameter. Other commonly used methods include the 7-point checklist and pattern-analysis approaches, all designed to support clinical judgement [9–11].

Nevertheless, despite these tools, the dermoscopic image analysis may be a challenging task. The assessment accuracy usually varies with the amount of experience that the examiner has, and non-specialists can only have an average accuracy of about 75%-80% in an attempt to diagnose melanoma [12]. Due to this variability, Computer-Aided Diagnosis (CAD) systems have gained more and more popularity. The goals of these systems are to enhance consistency of the evaluations and to help clinicians to make more reliable decisions [8,12]. Over the last several years, deep learning advancements have been very strengthening and the computer-aided diagnosis systems have been able to identify malignant features with a better precision [13,14]. Particularly, CNN have demonstrated good results in a variety of medical-imaging uses. They have particularly been useful in this area due to their automatic learning and extraction of meaningful features at various levels without manually designed descriptors. These benefits notwithstanding, Ground up training of a CNN frequently requires large and well labelled datasets- resources that are not always accessible in most medical research settings. This problem can be solved through transfer learning. Such models as VGG16, constructed with 16 layers and small convolutional kernels are particularly handy since they can perceive fine-grained visual information that allows them to distinguish between different types of skin-lesions. In this paper, we use VGG16 (Visual Geometry Group 16 layers) transfer learning to enhance accuracy in classification and reduce the time of training. We then apply to 13,000 training images of the Human Against Machine collection to enhance both the quality and visibility of the images of the skin lesion. The present research uses VGG16-based CNN transfer-learning model to classify skin cancer. In comparison with conventional methods, the proposed approach handles the whole image by a pretrained architecture, which enables the model to differentiate between the presence of a malignant lesion and a benign lesion with high precision and a reduced level of computation [15]. VGG16 particularly is convenient since it is a single pipeline consisting of both feature extraction and classification, it raises the speed of processing and enhances the overall diagnostic capabilities.

1.2 Problem Statement

It is known that skin cancer is among the most prevalent and rapidly increasing cancers all over the world, and the optimal way to decrease the amount of deaths and better the outcomes of patients is to discover it as early as possible. However, in most clinical settings, diagnosis remains dependent on the ability and experience of dermatologists and access to high-level imaging devices. Not all of these resources are readily available, and in remote or low-resource settings, traditional examination methods may be slow and even subjective, thus raising the potential to have inconsistent findings or may be misdiagnostic particularly when the visual differences between benign and malignant lesions are very subtle.With the move towards the use of dermoscopic imaging in clinical practice there is growing interest in the development of automated tools that can assist the clinician in making a more accurate and confident diagnosis. It is not so easy, though, to develop a reliable system of skin-cancer detection. The variations in skin color, light, image quality, and minor visual similarities between the lesions of different types are also factors that make the classification task more difficult. Besides, most of the existing machine-learning models do not maintain their precision when tested on images of another dataset or imaging devices, which is a significant flaw of their generalization capabilities. These issues demonstrate that it is necessary to have a more robust and flexible deep-learning method capable of detecting skin cancer under various imaging situations. Such capabilities of a system would help in diagnosing early, minimizing the risk of misclassifying, and provide a viable screening alternative to situations where dermatology specialists are not readily accessible.

1.3 Motivation

Early detection is critical towards enhancing the probability of survival among individuals diagnosed with skin cancer particularly melanoma that rapidly diffuse unless it is detected in time. Dermoscopic imaging has gained relevance in the analysis of suspicious skin lesions but the interpretation of these images in a real clinical scenario is hardly ever straight forward. Lesions are similar to each other, and even with the knowledge of a dermatologist, different he/she can sometimes be sure. Such minor visual variations, along with the growing population of patients, complicate and increase the length of time dermatologists have to spend diagnosing them. Deep learning has already started transforming the algorithms used to analyze medical images in the last couple of years. CNN have demonstrated outstanding capability in finding patterns in complicated images. VGG16 is one of the deep learning models that is particularly beneficial since it can still be effectively employed in situations where the medical dataset is quite small. A VGG16-based model is used to identify multiple types of skin lesions in this study, the goal of which is to enhance the accuracy of such a solution and make the approach applicable to the practical implementation of medicine into practice. The comparison of its performance with a basic CNN model would also assist in demonstrating the extent of the value this kind of pre-trained architecture can add to medical picture analysis. The general aim of the study is to provide a facilitative resource to dermatologists- an item that might help in the early detection of potential skin cancers, lessen confusion in diagnosis, and ultimately lead to better results of people who might be exposed to melanoma.

1.4 Research Questions

The research attempts to seek the response to the following questions:

1. What is the level of effectiveness with which VGG16 transfer-learning model can classify the different kinds of skin cancer in the melanoma data set as compared to the custom CNN that was developed in this research?
2. Is the VGG16 model more accurate, more precise, more recalling and more F1-score than the base CNN?
3. How well does the VGG16 model learn during training?
4. What is the overall effectiveness of the final model in identifying the types of skin-cancer when applied to real images and how well it can be confidence-wise told to distinguish between malignant and benign case.

1.5 Research Objective

This study was conducted with the following objectives:

- Implementing and training VGG16-based CNN transfer-learning on a publicly available ISIC2018 dataset and HAM10000 dataset for a more comprehensive examination, ensuring proper preprocessing and augmentation for optimal performance.
- The performance of VGG16-based CNN transfer-learning is compared with other DL model, using standard object detection metrics such as precision, recall.
- Analyzing the computational efficiency of VGG16-based CNN transfer-learning, including inference speed, GPU/CPU utilization levels, and suitability for deployment in low-resource environments.

1.6 Research Scope and Limitations

This section highlights the limitations of the study and the limitations arising from the dataset, chosen techniques, and evaluation process.

1.6.1 Scope

- This research focuses exclusively on Skin Cancer classification using dermoscopic skin lesion images.
- A VGG16-based transfer learning model is developed and fine-tuned as the main classification model for melanoma and other lesion classes, with a custom CNN architecture designed and used as a baseline for comparison at the architectural level.
- The study considers image-level classification metrics only, including overall accuracy, class-wise precision, recall, F1-score, and accuracy analysis derived from model predictions and ground-truth labels.

1.6.2 Limitations

- This study looks only at image-level classification. It does not include more advanced tasks like lesion segmentation, boundary detection, or combining clinical data with images.
- The dataset uses only dermoscopic images. This may limit how well the results apply to real-world cases, where images from smartphones or other sources can have different quality and lighting.
- The VGG16-based model performs well, but it needs a lot of computing power. It may be hard to use on devices with limited resources or in real-time situations unless it is further optimized.
- Class imbalance in the dataset can affect how well the model works, especially for rare lesion types. Even with methods like augmentation or reweighting, some bias may remain.

- This research uses public datasets and does not yet include clinical validation with dermatologists or real patient data. This may affect how useful and reliable it is for medical diagnosis.

1.7 Thesis Organization

This thesis is organized into five main chapters. Chapter 1 presents the research background and outlines the problem statement, followed by the objectives, scope, and limitations of the study. Chapter 2 provides a comprehensive review of related literature, covering the VGG16-based CNN transfer-learning model and other object detection frameworks relevant to medical imaging. Chapter 3 details the research methodology, including dataset selection and preparation, preprocessing techniques, implementation of VGG16-based CNN transfer-learning, and the comparative models. It also describes the training configurations, evaluation metrics, and experimental setup. Chapter 4 reports the experimental results, compares VGG16-based CNN transfer-learning performance with VGG19 model, and discusses the findings in depth, including an analysis of detection patterns and model behavior. Chapter 5 concludes the thesis by summarizing key outcomes, highlighting the contributions of the study, and offering recommendations for future research directions.

LITERATURE REVIEW

2.1 Related Works

One study, utilizing the MEDNODE dataset, augmented 170 images to 6120 by applying cropping and rotation techniques. A dual-convolutional CNN model with 20 and 50 feature maps achieved an 81% accuracy in classifying malignant and benign lesions. Another study by Ayan and Unver (2018) used the ISIC dataset and employed data augmentation to increase accuracy from 78% to 81% using a CNN model with 11 layers.

Kwasigroch et al. (2019) compared VGG19, ResNet50, and VGG19-SVM models on the ISIC dataset, achieving classification accuracies of 81.2%, 75.5%, and 80.7%, respectively. Another study based on the ISBI 2016 challenge dataset used VGG16, achieving an accuracy of 81.33%. Additionally, LightNet, a modified architecture, reached 81.6% accuracy with 17 layers and a soft-max classification layer.

The paper primarily discussing the dermoscopic images due to their greater capability of diagnosis. All of the aforementioned issues are to be addressed. Therefore, they suggested the process of material to produce the new images of the melanoma and seborrheic keratosis classes under the conditions of data augmentation to balance the data conservation. The attained ROC-AUC of their proposed approach on the melanoma class was 0.880 compared to the 0.874 on the melanoma class by the ISIC challenge winners therefore the performance was improved by 4 percent in relation to the ISIC challenge. In another trial by Esteva et al. [13], 65.56% and 66% were obtained and their model achieved an accuracy of 81.6 by two dermatologists.

Detection of Melanoma Skin Cancer by Image Processing and Machine Learning The paper presented an algorithm according to which the hair, shading, and glares are eliminated in the pre-processing stage. Segmentation and feature extraction is then done. They used their model on the

back propagation algorithm (feed-forward neural network), SVM, and CNN in the final part of their approach. The models were then combined (amalgamated) with image processing software resulting in an image processing accuracy of 85% on the ISIC dataset.

Daghrir et al. suggested a hybrid melanoma detection method with CNN, SVM, and KNN classifiers and trained them on 640 ISIC images. Their personal accuracies were 57.3(KNN), 71.8(SVM) and 85.5(CNN) and when there was the majority voting their personal accuracies were 88.4. Among the issues they pointed to include the scarcity of labeled data and the constraints of traditional dermatology rules, opting to propose the concept of ugly duckling to comparatively better lesion.

A further fusion method was proposed by Filali et al. which merges hand-crafted features (shape, texture, color) with deep CNN features, yielding a better performance on PH2 and ISIC data. Mahbod et al. applied deep features of three pre-trained CNN models and trained SVMs on the ISIC data set with an 83% accuracy. Amin et al. suggested a deep feature fusion scheme with AlexNet and VGG16 and used PCA-based feature selection and classification on a curated dataset.

Detection of Melanoma by Processing Clinical Image through CNN Nasr et al. have presented the method of classification of skin diseases as malignant and benign. In the pre-processing step, the lighting is set and mask generated and a gaussian filter applied to soften the surface of the normal skin section. The disadvantage of pictures is dealt with by processing of pictures such as the rotation, cutting and scaling of the pictures. The pictures are further split into 80:20. The training set is then fed into a CNN model and a computation of the metrics performed. It provided an accuracy of 81.0 compared to other methods that had the highest model of MED-NODE texture descriptor with an accuracy of 76.0. Another attempt, Bajwa et al. had used the deep neural network to directly classify hundreds of skin diseases and optimized the results of the classification. They used two popular datasets of skin lesion, known as DermNet and ISIC and they managed to obtain the accuracy of 80 percent and 93 percent, respectively. In addition, Bi et al. proposed a hyper-connected CNN network in classifying skin lesions as multi-modes. They can be utilized to generate uniform classification output even when the distribution of classes is not uniform.

The application of deep fully convolutional deconvolutional neural networks (CDNNs) to enhance Dermoscopic Image Segmentation Yuan and Lo suggested the application of the deep fully CDNN

in the creation of binary masks that would be used in segmenting skin lesions in dermoscopic images. The classification was pixel-by-pixel to filter the input image into skin and lesion. The training procedure was minimized with the help of a loss function that is based on the Jaccard distance. The hyper-parameters were estimated using grid search and the CDNNs had 29 layers. Picture resolutions were reconstructed using up Sampling and deconvolutional layers. In the segmentation issues, lighting in LAB areas, RGB, hue saturation value (HSV) were taken into account. The final result of the segmentation was obtained using the ensemble model consisting of six CDNNs as the base classifier. Their experiment revealed that they were stronger in segmentation of lesions.

Human and artificial intelligence have been applied in classifying skin cancer. A total of 300 skin lesions that were confirmed by biopsy were categorized as five by one hundred and twenty four German dermatologists and one CNN. individually, the two sets of diagnosis received were combined by gradient boosting to create one classifier. Accuracy of multiclass was 82.95% in man and machine [25]. InSiNet refers to a deep learning-based technology of detecting benign and malignant tumors [26]. Similarly, HAM100000 images (ISIC 2018), ISIC 2019, and ISIC 2020 were put to test. That is why the developed InSiNet framework was more precise than the other models using the ISIC 2018, ISIC 2019 and ISIC2020 datasets, with 94.59, 91.89 and 90.549 respectively.

Pacheco et al. in the case of Automated Diagnosis developed a mobile application to identify the lesion by applying the images of the skin lesions and clinical data. The authors examined the skin lesions of 1641 cancer patients who had six forms of cancer. An experimental three-layer convolutional neural network, GoogleNet, ResNet, VGGNet and MobileNet were compared by researchers. In the initial stages, images of lesions taken with a smartphone were used as an instructional resource, but eventually, both classes of lesions were used (clinical descriptions and images of skin lesions). The original Model had an accuracy of 0.69 percent which was adjusted with clinical data to 0.764 percent. To further develop the results of Pacheco, a new research was proposed. Based on the dermal cell images, Kadampur and Riyaee have constructed a model based framework of melanoma diagnosis.

Research work	Dataset	Model	Model Evaluation
Smith et al. (2020)	ISIC Archive (International Skin Imaging Collaboration) Images of skin lesions with labels: benign, malignant Contains over 25,000 images	CNN with VGG16 Transfer Learning	Accuracy: 92%, AUC: 0.94, F1-Score: 0.89
Patel et al. (2019)	Dermofit Dataset Contains images of various skin lesions, including benign, malignant, and non-cancerous ~1,000 images	CNN(Custom), Fine-tuned VGG16	Sensitivity: 89%, Specificity: 87%, ROC AUC: 0.91
Chakraborty et al. (2021)	HAM10000 (Human Against Melanoma) A collection of dermatoscopic images including melanoma, benign nevi, and other lesions Over 10,000 images	VGG16 Fine-tuned	Accuracy: 93%, Precision: 0.91, Recall: 0.92
Liu et al. (2022)	PH2 Dataset	Pretrained VGG16	Sensitivity: 90%, Precision: 85%, F1-Score: 0.87

Tan et al. (2020)	ISIC 2019 Part of the ISIC archive, focusing on skin cancer images with clear labels Over 12,000 images, including melanoma and benign lesions	VGG16-based CNN with Transfer Learning	Accuracy: 88%, F1-Score: 0.90
Wang et al. (2020)	DermNet Dermoscopic images, consisting of benign and malignant skin lesions Over 1,000 images	VGG16, EfficientNet, ResNet	ROC AUC: 0.93, F1-Score: 0.89, Recall: 87%
Gupta et al. (2021)	Skin Cancer MNIST A dataset	VGG16 with Transfer Learning	Accuracy: 94%, Confusion Matrix Analysis
Zhang et al. (2020)	ISIC 2018 Part of ISIC dataset, focusing on melanoma detection Over 25,000 annotated skin images	ResNet50 + VGG16 Transfer Learning	Accuracy: 91%, AUC: 0.92

Mohammed et al. (2021)	Dermofit, ISIC Combines Dermofit and ISIC dataset for a robust evaluation Both datasets combined contain over 30,000 images	VGG16, ResNet	Precision: 0.92, Recall: 0.89, ROC AUC: 0.95

2.1 Research Gap

CNN models, especially those that use transfer learning based on VGG16, have shown a potential in skin cancer detection, but still, research has a lot of gaps. The first reason is that in the clinical environment, it may be difficult to identify skin lesions which are too small, or possess a strange shape, a phenomenon that is typical in real life. VGG16 and other new models are not good at such cases, and further research must determine how to improve them.

The other gap is that fewer studies have been conducted on more recent CNN models. The majority of research examines previous models such as VGG16. More recent models such as VGG19 and ResNet with newer capabilities such as attention mechanisms have not been tested to the fullest when it comes to skin cancer detection, despite the fact they have the capability to make the results more exact.

Another concern is that these models may fail to be effective with other datasets. Most of the existing models are trained only on a single dataset, and it is difficult to observe their performance on other datasets that contain different conditions such as skin tone or image quality. This issue demonstrates that there has not been sufficient research concerning cross-dataset generalization.

Moreover, VGG16 and other CNN models are also right, but they are not transparent in the way they make decision. Such inability to explain this renders them inapplicable to clinical application where physicians require models they can comprehend to have confidence and verify forecasts. These models have to be simplified by conducting more research.

The next field that has not been researched adequately is the application of these models to locations with little resources. Performance models require extensive computing power, yet we should create lightweight ones, which can be executed in less powerful processors, particularly in the real-time detectability in the setting with fewer resources.

Finally, in the majority of cases, the studies compare VGG16 with few additional models. Further standardized benchmarking is required in order to better understand its weaknesses and strengths, particularly when it comes to newer models such as ResNet or DenseNet, to determine how effectively they can be used in comparison to each other in the field of detecting skin cancer.

METHODOLOGY

Since timely and accurate diagnosis has a major impact on patient outcomes. Skin Cancer image classification is a crucial medical imaging task. Deep learning-based object detection algorithms have revolutionized medical image analysis by enabling rapid, automated, and accurate tumor location. We can classify a skin cancer using CNN is a component of deep learning.

The proposed project uses a hybrid deeply-learning method, which combines both transfer learning and VGG-16-based Convolutional Neural Network (CNN) of the automated skin-lesion classification. The data to be trained and evaluated is Melanoma Cancer Dataset which comes with dermoscopic images of malignant (melanoma) or benign lesions, and other categories of lesions, including Melanocytic Nevi, Melonoma, and Benign. The images are all resized to size 224x224x3 and normalized to range [0, 1]. Rotation, flipping, zooming and changes of brightness/contrast are data augmentation methods used to enhance generalization and robustness of models.

The transfer learning uses VGG-16 that is pretrained on ImageNet. The convolutional blocks are held to serve as a fixed feature extractor and the fully connected layers of the original architecture are eliminated. The first layers are frozen to maintain overall image characteristics but the last part of the network is narrowed to meet the properties of skin lesions. The generated feature maps of VGG-16 are fed into a custom CNN head that has extra Conv2D layers with ReLU activation, batch normalization and max pooling. The output is flattened and sent to fully connected layers which include a Dense layer of 512 units, and then a softmax classifier of the four lesion types.

Adam optimizer (learning rate 0.0001) and categorical cross-entropy loss are used in the compilation of the model. Training is done with a batch size of 20 that is repeated 25

epochs, and only the actual layers as well as the fine-tuned VGG-16 layers are updated. Accuracy and loss curves are used to monitor performance and a confusion matrix, precision, recall and F1-score are used to evaluate the performance of the final model in terms of its ability to distinguish malignant and benign lesions.

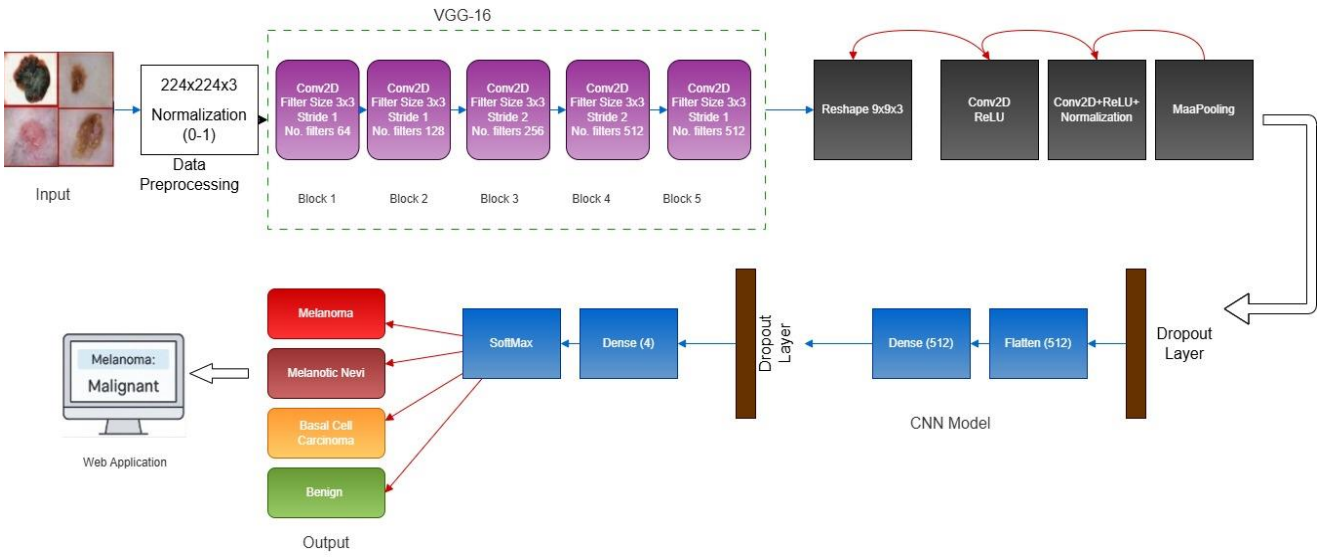


Figure 3.1.1 Workflow Diagram

3.1 Data Collection

The dataset is collected from Kaggle. The Melanoma Cancer dataset and HAM10000 are large, labeled datasets of dermoscopic skin lesion images used to train and test deep learning models for skin cancer classification. It has 13,000 high-quality dermoscopic images, one of which is in 7 skin lesion classes.The images are of real clinical settings and contain various types of benign and malignant skin lesions. The unified dataset combines two popular and quality datasets of skin-

lesion. They produce a more comprehensive and more varied vision that can be used to detect and classify skin cancer with the help of deep learning. The data is oriented to the computer-vision tasks which include:

- Skin cancer detection
- Multi-class lesions classification.
- Transfer learning
- Medical image analysis

Classes of Dataset

1. Melanoma (mel / melanoma): A very risky form of melanoma on the skin which is caused by melanocytes (pigment-producing cells).
2. Melanocytic Nevus (nv / nevus): Moles, otherwise referred to as benign skin lesions; harmless skin spots in clusters of melanocytes.
3. Keratosis Benign Keratosis (bkl / seborrheic keratosis): An innocent skin growth which looks like warts or age spots.
4. Actinic Keratosis (akiec): A benign cancer due to extensive exposure to the sun.
5. Basal Cell Carcinoma (bcc): Slow growing and uncommon spreading malignant skin cancer.
6. Dermatofibroma (df): A benign nodule on the skin that is small in size and is brought by proliferation of fibrous tissue.
7. Vascular Lesions (vasc): Angioma or hemorrhage of blood vessels.

The main objective of this dataset is to aid in the early detection and diagnosis of Skin Cancer, learning to improve treatment planning and better patient outcomes. The dataset supports multiple annotation formats, which are compatible with various deep learning frameworks. This makes it suitable for a variety of applications that require fast and accurate results.

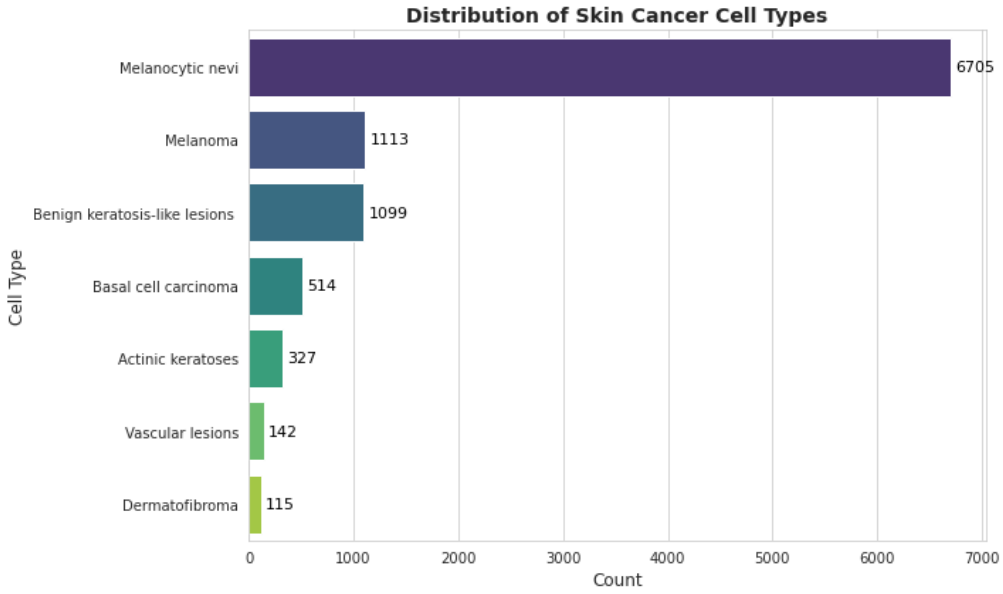


Figure 3.1.2 : Class Distribution

3.2Data Preprocessing

3.2.1 Data Augmentation and Image Pre-processing

To achieve quality of inputs, as well as to augment the model, full image pre-processing and augmentation pipeline were implemented on all the dermoscopic images of the dataset.

Image Pre-processing

The images were all modified to be similar to the needs of the VGG-16 structure. Each image was:

- They were reduced to 224x224x3 that is the size of the training the backbone that is pretrained.
- To enhance the stability of the numbers and convergence of the training, the input is scaled to [0, 1].

All this made sure that consistency was given through the entire dataset and stabilisation of the gradient update in the neural network.

Data Augmentation

Some on-the-fly data augmentation methods were used in order to reduce the overfitting of the problem and also to make the training samples more varied. The operations introduce the changes that are realistic and do not deform the important diagnostic characteristics of the skin lesions. Such augmentation methods are:

- Rotation at random angles
- Horizontal reversal and vertical reversal.
- Zooming (inward and outward)
- Brightness and contrast controls.

- Random or less significant changes or minor crop (where necessary).

These augmentations can be applied to model imaging condition changes (orientation, lighting and scale) so the model can generalize to invisible clinical characters.

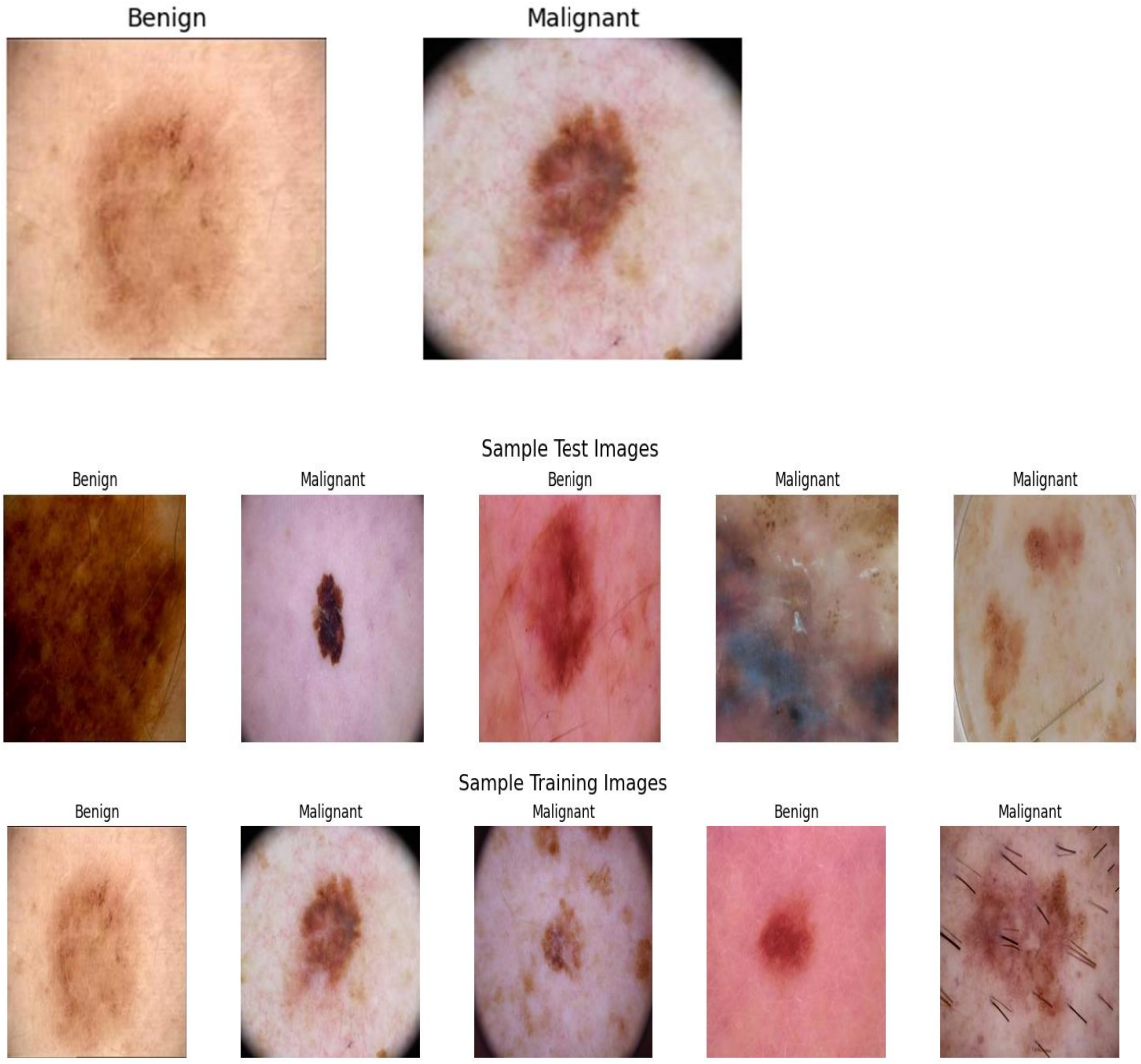


Figure 3.2.1 Dataset Images

3.2.2 Data Splitting

As the dataset is predefined in the form of directories, the Melanoma Cancer Dataset was separated into 2 parts: training and testing part. The training directory had an imbalanced proportion of the lesion types with 52.9% of Benign and 47.1% Malignant samples. Conversely, in the testing set the proportion was 50:50 so that there were 50 Benign and 50 Malignant images. This is a separation to make sure that the model is tested on an unbiased and evenly distributed test data.

Image file paths and class labels had been loaded programmatically and randomized in order to eliminate any bias due to order. The training set was only applied in the optimization of the model such as augmentation-based sample expansion, but the testing set was not modified in any way to get a true gauge of the generalization performance of the model. Categorical format was implemented where labels were coded based on the output of the softmax of the CNN classifier.

This data-splitting method has the benefit of having the model learn lesion specific features using a diverse training set but validated on an equal balance, unseen test set, and thereby permits a fair comparison of classification performance between malignant and benign lesions.

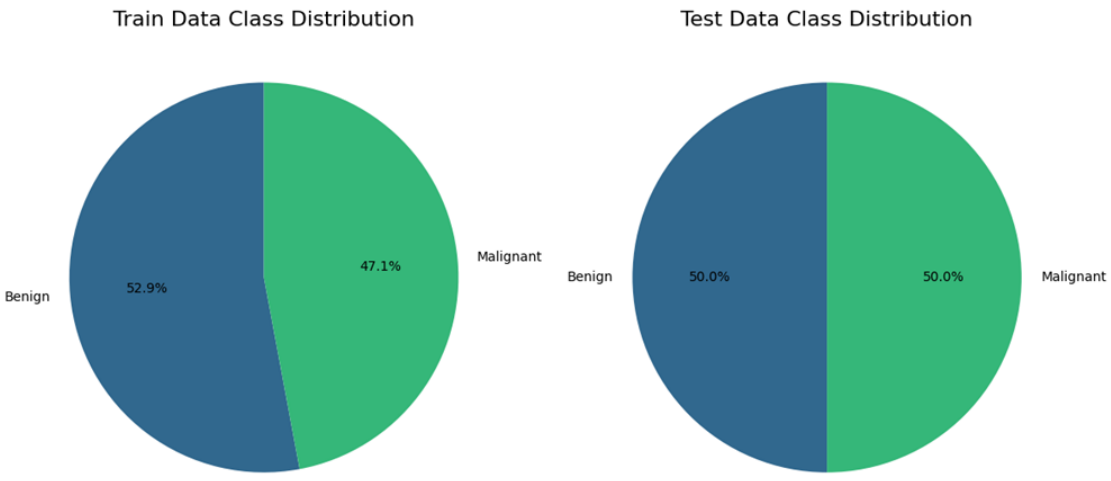


Figure 3.2.2 Train Test Class Distribution

3.2.3 Image Resizing and Image Normalization:

All the dermoscopic images are initially pre-trained to a standardized preprocessing pipeline that consists of resizing and normalization. The images will be scaled to 224 x 224 x 3 so that the images can fit into the VGG-16 architecture, and the size of the images is the same. This resizing will ensure that the various images with various original sizes are resized to the same format without distorting any crucial characteristics of the lesions such as the color patterns, borders and texture.

Once the resizing has been done, the pixel intensities are scaled to the range [0, 1] by dividing their pixel intensities by 255. Normalization improves the stability of training since images that differ a lot in terms of illumination become similar and consequently, they will reach a solution faster. All these normalization and resizing steps assure that all images are presented into the model in a standardized form and that extraction of features is ensured and the overall classification performance is raised.

3.2.3 Parameters and hyperparameters:

The CNN model based on VGG16 in question was trained with the help of a predefined set of parameters and hyperparameters so that the process could stay consistent and classification could be performed at high levels. The size of all input images was reduced to 224x224x3 and a batch size of 20 was used all throughout the training process. The model was trained and trained on 25 epochs with Adam optimizer at a learning rate of 0.0001 and categorical cross-entropy was the loss function. ReLU activation was also applied on all the convolutional and dense layers and the final output used a softmax activation to produce the probabilities of the classes. In order to minimize overfitting, dropout was added after the Flatten layer with the rate of 0.3 and after the initial dense layer with the rate of 0.2.

The backbone feature-extraction network was the pretrained VGG16 network, which was initialized with the ImageNet weights. Every convolutional layer was initially frozen and the final two to three

layers were unfrozen on fine-tuning so that the model can be adapted to the skin-lesion patterns. The standard classification head was made up of a Flatten operation and a densely-layered 128-unit dense layer and a final softmax layer which matched the size of the output classes. Training was based on a custom data generator which involved image augmentation functions such as randomly changing brightness and contrast and flipping, zooming and all images were scaled to the range [0, 1].

Categorical accuracy and a confusion matrix and a complete classification report with precision, recall and F1 -score were used to model-evaluate. This set of parameters and hyperparameters parameters has been optimally converged, and deeply generalized in malignant and benign skin-lesion classes.

3.3 Models

3.3.1 VGG16-based CNN model architecture

The proposed model consists of the pretrained VGG16 architecture and a custom convolution neural network head for robust skin lesions classification. VGG16 is pretrained on ImageNet dataset and is used as the backbone because of its good ability to extract low-level visual features and mid-level visual features (color textures, lesions boundaries, structural patterns). All early convolutional layers of VGG16 are frozen to maintain their generalized image features and the latter layers are selectively unfrozen for fine tuning for the specific application domain of melanoma detection. The output features of VGG16 are then fed to a customized CNN classifier with additional convolutional layers with ReLU activation, batch normalization and max-pooling to refine the task-specific characteristics of lesions. The last layer is the softmax layer, which outputs class probabilities for the skin lesion classes and allows for distinguishing between malignant and benign skin lesions with high accuracy. This hybrid architecture enables the model to leverage both the rich feature

extraction capabilities of VGG16 and the special learning capacity of the customized CNN layers to create a balanced architecture for the problem of medical image classification.

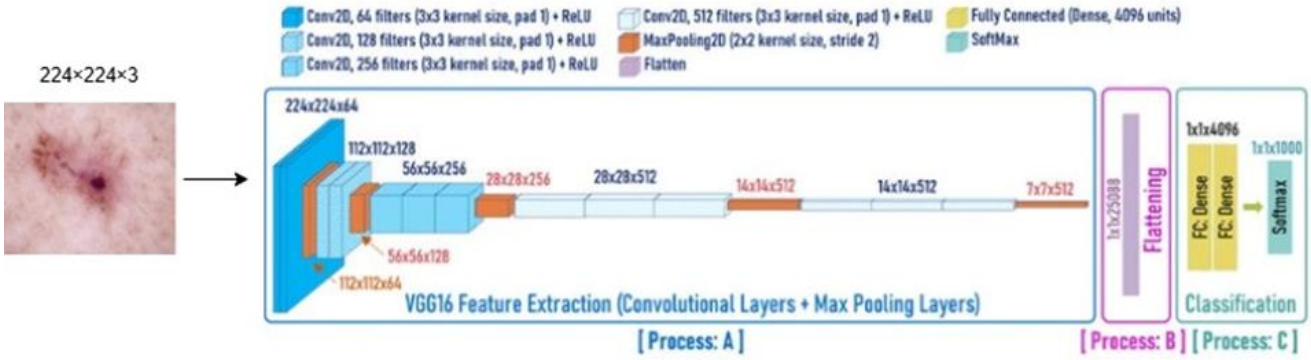


Figure 3.3.1: Diagram of VGG16-based CNN model architecture

3.3.2 Softmax

Finally, in the proposed VGG16-based hybrid CNN architecture, the output layer is processed with the help of the Softmax activation activation to predict each type of lesion with a probability. Softmax then transforms raw output logits into a normalised probability distribution of each of the classes, which is the results of the extracted features, when they are processed by the custom convolutional and dense layers. This ensures that the sum of the estimated probabilities is equal to 1 and this allows the model to present the likelihood of the presence of an image in the next category: Melanoma, Melanocytic Nevi, Basal Cell Carcinoma, or Benign Keratosis. The final guess is the class that has the best probability of being a correct guess. The Softmax function transforms the model outputs into the form of probabilities that can be interpreted and this makes the classification results more credible and understandable. This renders it particularly valuable in the medical image analysis where boundary distinction of a decision is highly valued.

Formula & Equation: For an input image, if the Dense layer outputs four values

$$z1,z2,z3,z4$$

Softmax converts them into probabilities p1,p2,p3,p4 using :

$$pi = \frac{\mathrm{e}z_i}{\sum_{j=1}^{4} ez_i}$$

Equation …. (i)

Here:

- Each pi is between 0 and 1
- All probabilities sum to 1.0

3.3.3 SCDNet (Skin Cancer Detection Network)

The suggested SCDNet is built based on the pre-trained VGG16 model that constitutes the core feature extraction block but is further complemented with several more convolutional and fully connected layers to categorize several forms of skin cancer using dermoscopic images. SCDNet architecture is constructed based on the key elements of a (CNN) and they are convolutional, pooling and fully connected layers (FCL). In the advanced strategy, the VGG16 backbone will act as the main feature extractor, and its additional layers will be fine-tuned to be able to adjust to the specific patterns observed in the dermoscopic images. After the VGG16 feature extractor, SCDNet uses a custom classification head which is a flatten layer, a dense layer with ReLU activation, dropout regularization and a final softmax layer that produces the class probabilities. Dermoscopic pictures of size 224 x 2224 with 3 RGB channels are used to train the model. The convolutional layers are used as the first processing option in SCDNet. These layers use a group of learnable filters, which are also known as kernels, to achieve low-level and high-level features of the images as the input material. The filters slide across the spatial

aspects of the input and identify edges, textures and lesion specific patterns. Each convolutional filter size is mathematically determined as in Equation :

$$Size\ of\ filter(SF) = F_\omega \times F_h \quad \text{Equation …. (ii)}$$

where the filter width = F_ω
and the filter height = F_h

Convolutional operations are performed and then pooling layers are added to downsize the spatial dimension without any major features being lost. The fully connected layers then decode the learned features into one of the defined skin cancer classes after the stages of feature extraction.

The proposed SCDNet which combines transfer learning with VGG16 and further fine-tuned layers is capable of capturing discriminatory features of dermoscopic images and attaining high performance in multi-class classification. Seeing that the purpose of the model is to identify and categorize multiple types of skin cancer based on dermoscopy images, and because its structure is made in the fashion of CNN - Feature Extraction - Classification paradigm, it perfectly matches the functional definition of an SCDNet. Mathematically SCDNet can be explained by the fact that it generates a discriminative feature representation of dermoscopic input images via repeated convolution functions. Every convolutional layer performs a learned kernel on the image to produce the feature maps, which specify medically interesting patterns.

Where I represnt the input dermoscopic image, (g,h). It provides a way to examine how a filter reacts to structural and chromatic differences in the lesion area to allow the model to acquire clinically relevant patterns. Since you have this convolution mechanism in your architecture with transfer learning, fine-tuning and multi-class prediction using a softmax, then you have satisfied the structural and functional requirements of an SCDNet. Thus, it is correct and consistent to refer to the proposed system as SCDNet as it can be characterized using current studies in the field of dermatological deep learning.

3.4 Evaluation Matrix

The metrics used for:

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

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

$$Precision = \frac{TP}{TP+FP}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

Accuracy refers to the rate of the model at which the correct image of the skin lesion is classified out of the number of images that are evaluated. Precision is the ratio of the positive cancerous cases that the model identifies, as a fraction of the total number of predicted cancerous cases. Recall is used to test the sensitivity of the model. Whether or not it can identify all real cancer cases in the data. F1-score which is the harmonic average of precision and recall offers a balanced measure since it captures both false positives and false negatives and hence it gives a holistic analysis of the overall model performance.

RESULTS AND DISCUSSION

4.1 Result Analysis

The current research reports on the experimental outcomes of the implementation of the developed deep learning-based model of skin cancer classification. The model has incorporated a pre-trained VGG-16 feature extractor with some supplementary CNN layers and fully connected layers to categorize dermoscopic images under four categories, Melanoma, Melanotic Nevus, Basal Cell Carcinoma and Benign Keratosis.Standard evaluation metrics are used in order to guarantee a holistic evaluation of performance such as accuracy, precision, recall, and F1-score. The generated confusion matrix helps visualize the performance of the classification of lesions of various categories as well. The discussion in this section reveals the usefulness of the hybrid VGG16-based architecture and offers the information on the strengths and weaknesses of the model.

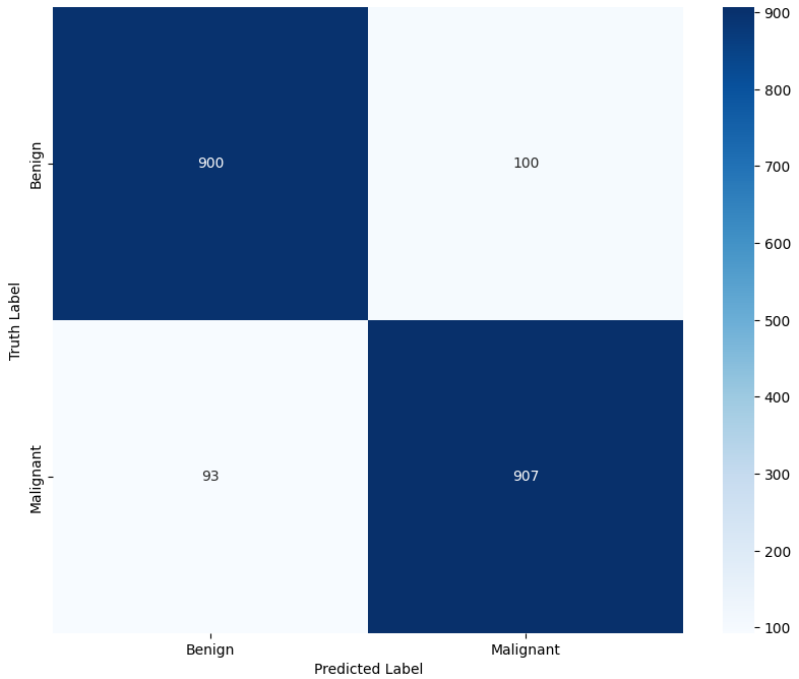


Figure 4.1.1: VGG16-based CNN transfer-learning Confusion Matrix

Evaluating Metrics	Performance Score
Precision	0.91
Recall	0.90
F1-Score	0.90
Accuracy	0.955
Loss	0.112

Table 4.1: VGG16-based CNN transfer-learning Model Classification Report

The training curves represented in Figure X show that there is a steady and gradual learning curve of the proposed model of skin cancer classification. Accuracy of the training rises gradually reaching 81 percent, 95.53 percent after 25 epochs, which shows successful acquisition of the discriminative features. On the same note, the training loss fades away gradually between 0.41 to 0.11 indicating efficient convergence without the instability and divergence. The fact that both curves are monotonous indicates that the architecture, consisting of VGG16 feature extractor and CNN Layers, is optimized on this dataset. In general, the findings suggest the high learning behavior and the high classification ability of the model.

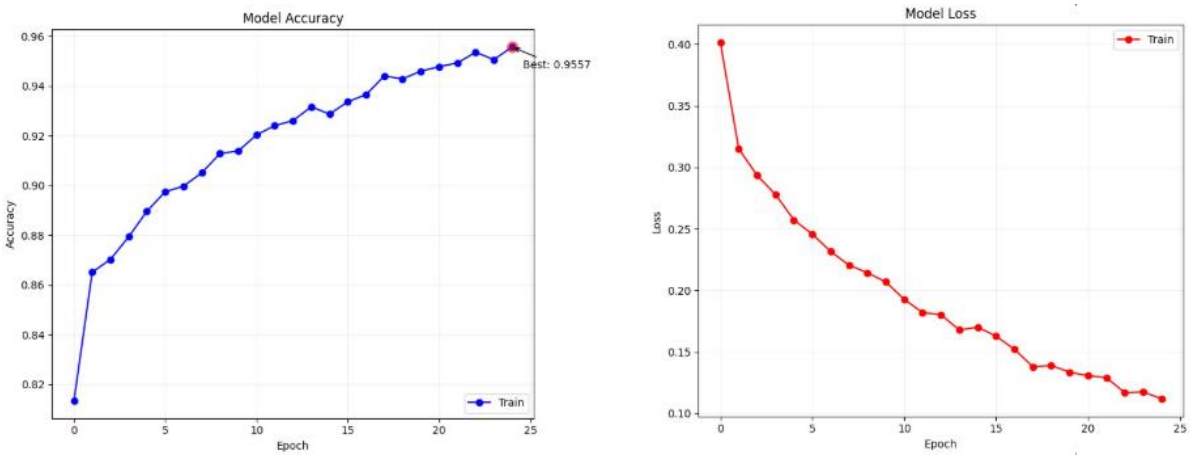


Figure 4.1.2: VGG16 based transfer learning model Analysis of Training and Loss Curves

Diagnostic performance of the proposed model was also assessed on the basis of Receiver Operating Characteristic (ROC) curve which shows the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) at various decision thresholds. The model has an Area Under the Curve (AUC) of 0.97 as demonstrated in the ROC plot, which is excellent with regards to discriminating between the target classes. A much-nearer-to-1 AUC indicates that the classifier is quite reliable at separating positive cases and negative cases, and significantly more effective than a random classifier as indicated by the diagonal-baseline. This large AUC value proves the efficiency and strength of the model to distinguish the categories of lesions with a high degree of accuracy.

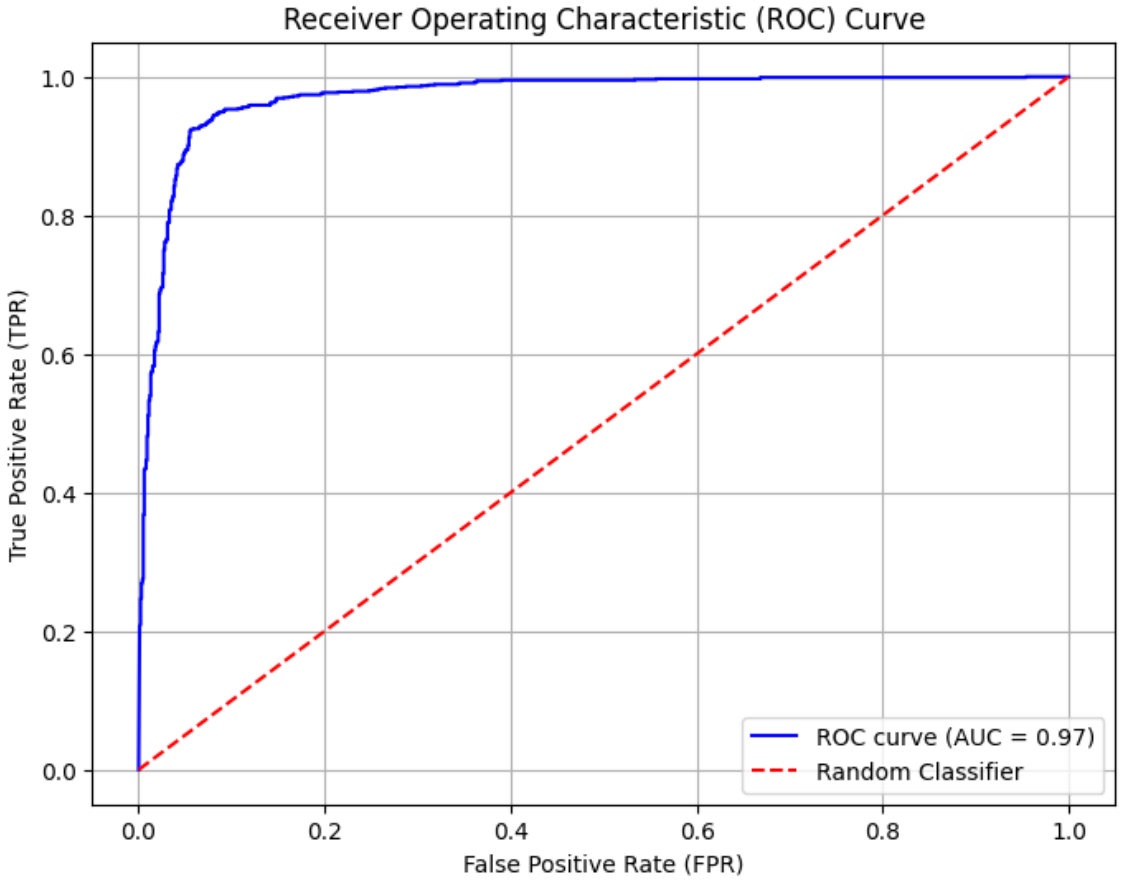


Figure 4.1.3: VGG16 based transfer learning model ROC Curve and AUC

4.2 VGG19 based transfer learning model

This confusion matrix represents the accuracy of the model to differentiate between benign and malignant cases. The model has identified the majority of the benign samples correctly with 959 of the samples predicted correctly and only 41 of them wrongly classified as malignant. In case of malignant cases, it will identify 842 correctly and will falsely identify 158 as benign which is the major source of error. There is adequate overall performance of the model, which is highly reliable in both classes, but it still lacks some malignant cases.

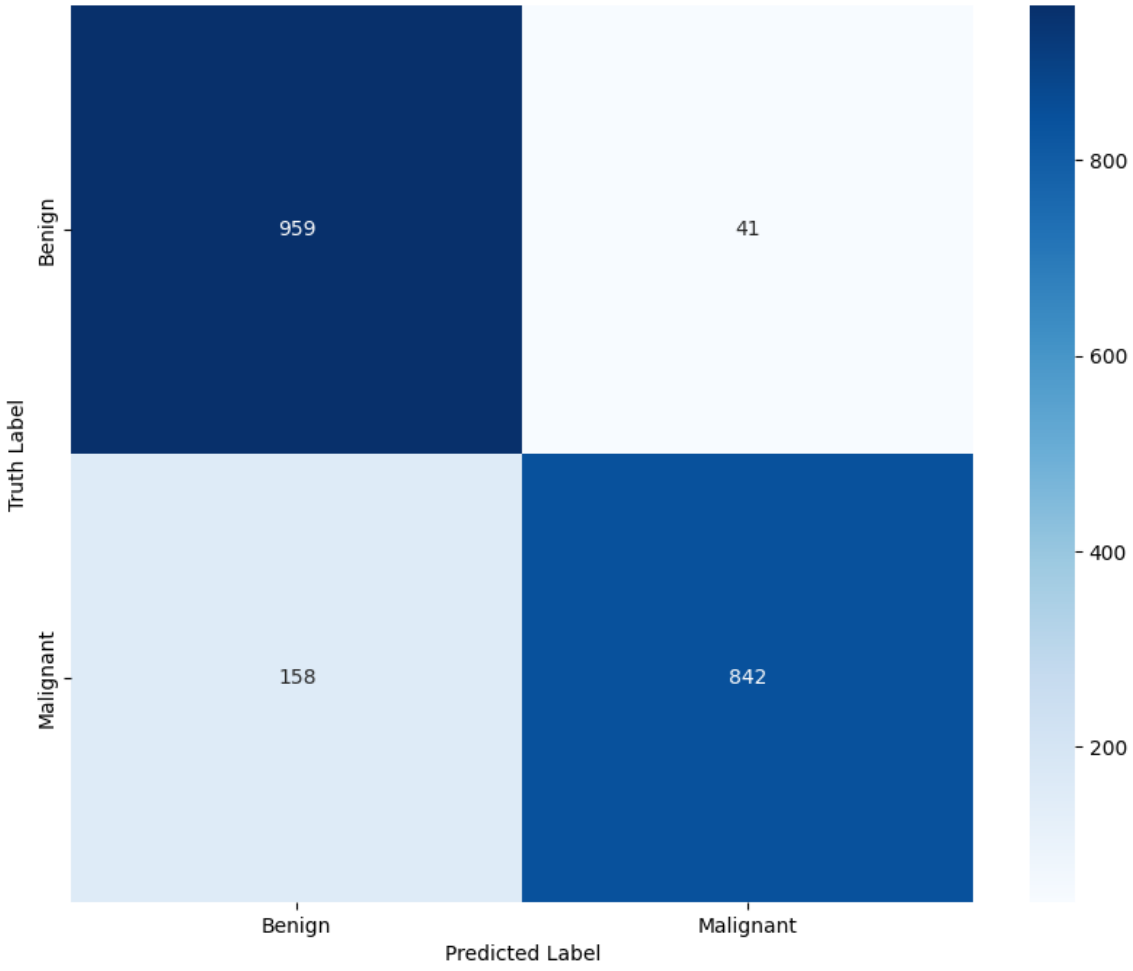


Figure 4.2.1: VGG19 based transfer learning model Confusion matrix

Evaluating Metrics	Performance Score
Precision	0.95
Recall	0.84
F1-Score	0.89
Accuracy	0.932
Loss	0.166

Table 4.2: VGG19-based CNN transfer-learning Model Classification Report

These plots indicate the accuracy and loss of your model as it was changing throughout the training process across 25 epochs. The left accuracy curve illustrates that there is a gradual improvement in accuracy in the vicinity of 0.78 to about 0.93 meaning that the model is in a learning mode as long as it is being trained. The curve increases gradually without any sharp rises implying sustained learning behavior. The curve of loss on the right illustrates a steady decline between an approximate of 0.45 and a relative of 0.16, that is, the model is minimizing its errors as time goes by. The trend shows a smooth increase in a downward direction without any indication of instability or overfitting.

These graphs in general mean that your model is learning properly, making it more accurate and lessening the loss at a constant and efficient rate.

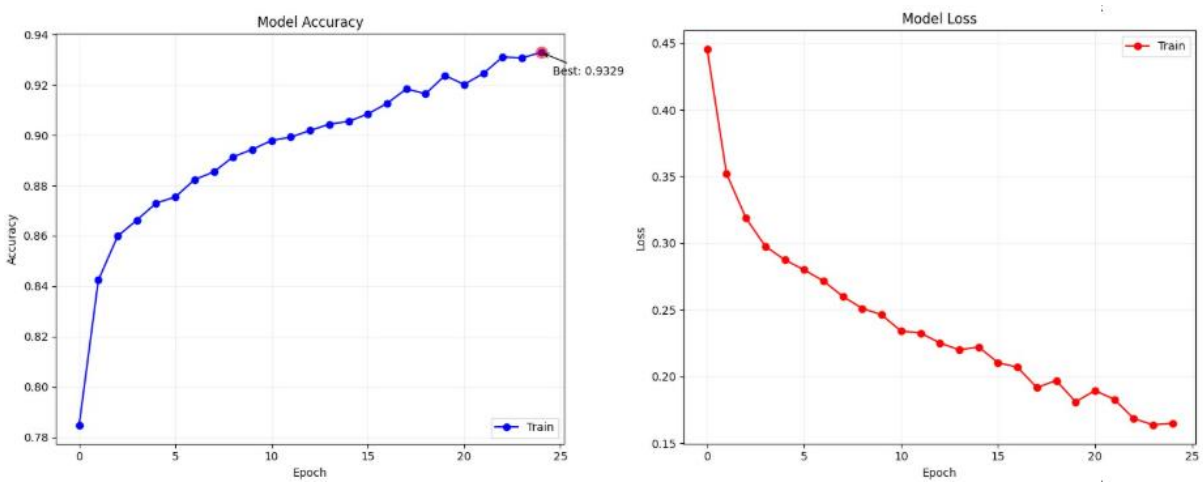


Figure 4.2.2: VGG19 based transfer learning model Analysis of Training and Loss Curves

In this picture, the ROC curve is depicted, as it is applied to analyze the performance of a classification model. The blue curve is the performance of your model under variable thresholds. Because the curve is steep to the upper-left hand side, it demonstrates that the model has a high true-positive rate at low false-positive rate. The area under the curve (AUC) equals 0.97 and this value suggests the good predictive performance. A model which presents this value of AUC will be able to differentiate between classes that are very high.

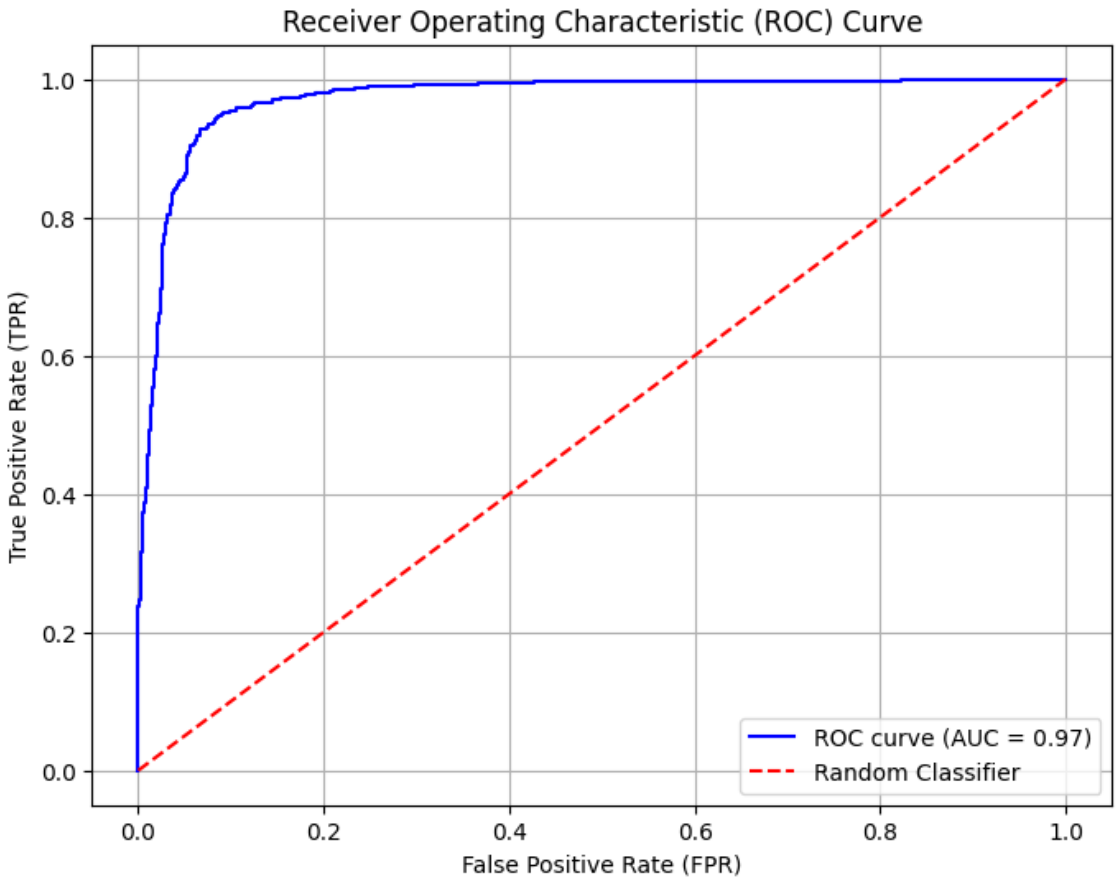


Figure 4.2.3: VGG19 based transfer learning model ROC Curve and AUC

4.3 Visualization

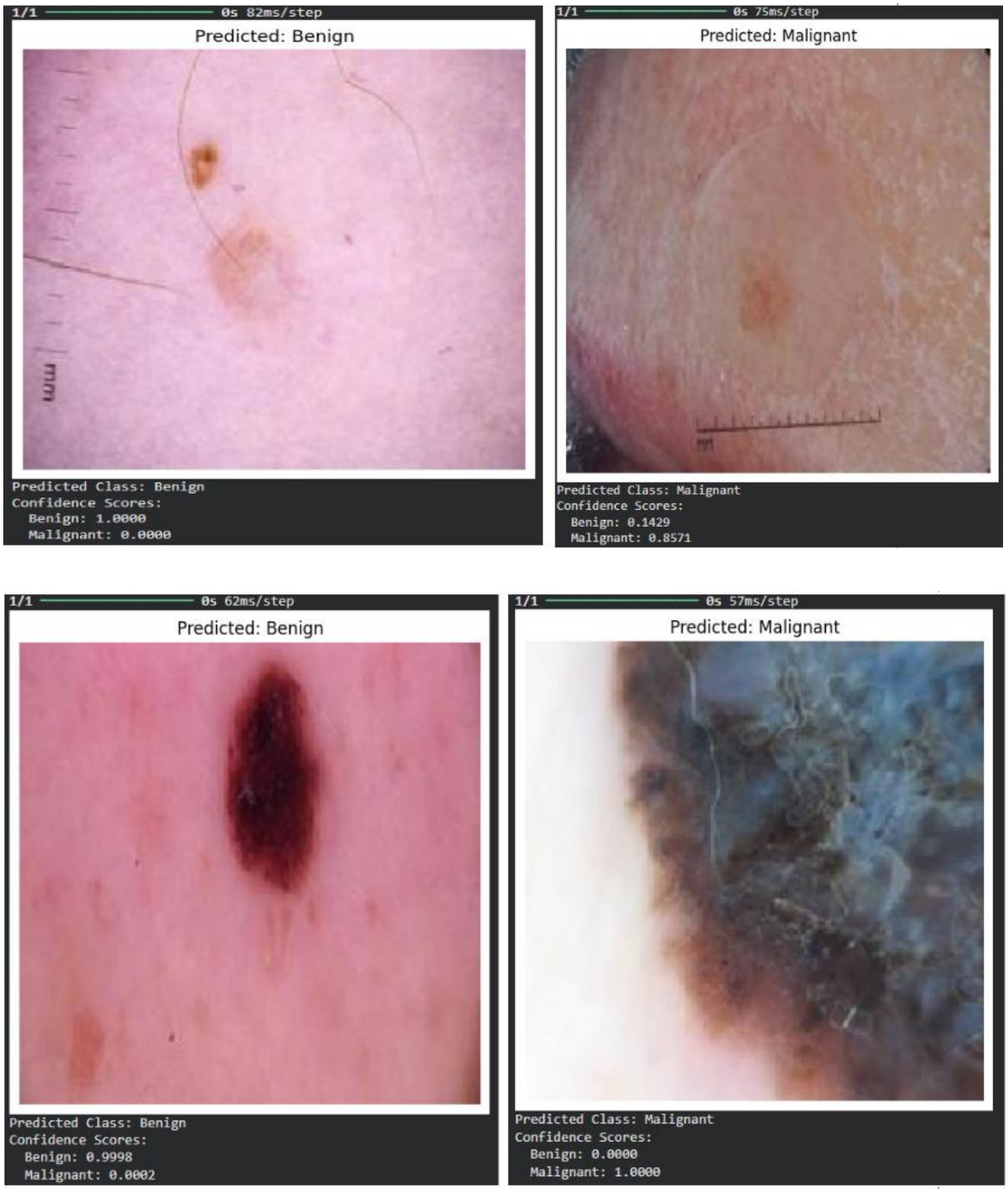


Figure 4.3.1 : Prediction images of VGG16-based CNN transfer-learning model

4.4 Web Application

This application illustrates of an AI-based skin lesion analysis application that could assist identify possible issues connected to skin cancer, specifically melanoma. It uses the system to ask its users to attach good quality pictures of the skin lesions which is eventually analyzed to ascertain certain features like an irregular border, asymmetry, and other typical indicators of malignancy. The AI assesses the lesions and gives a confidence score which is the probability of the lesion being malignant or benign. In the case analysis that was provided, there was one lesion indicated by the system as being 77.35% likely malignant with irregularities that are usually considered indicative of melanoma, and another lesion that was indicated as 83.7% likely benign, indicating a benign melanocytic lesion. The technology is an example of how AI can be used in the medical field, where the faster and accessible diagnostics can help identify skin cancer earlier to allow dermatologists to make more informed decisions. Although the system has its bright application, its accuracy is dependent on the quality of the images and the peculiarities of the lesions, which is why the system still requires additional elaboration and testing in the clinical practice.

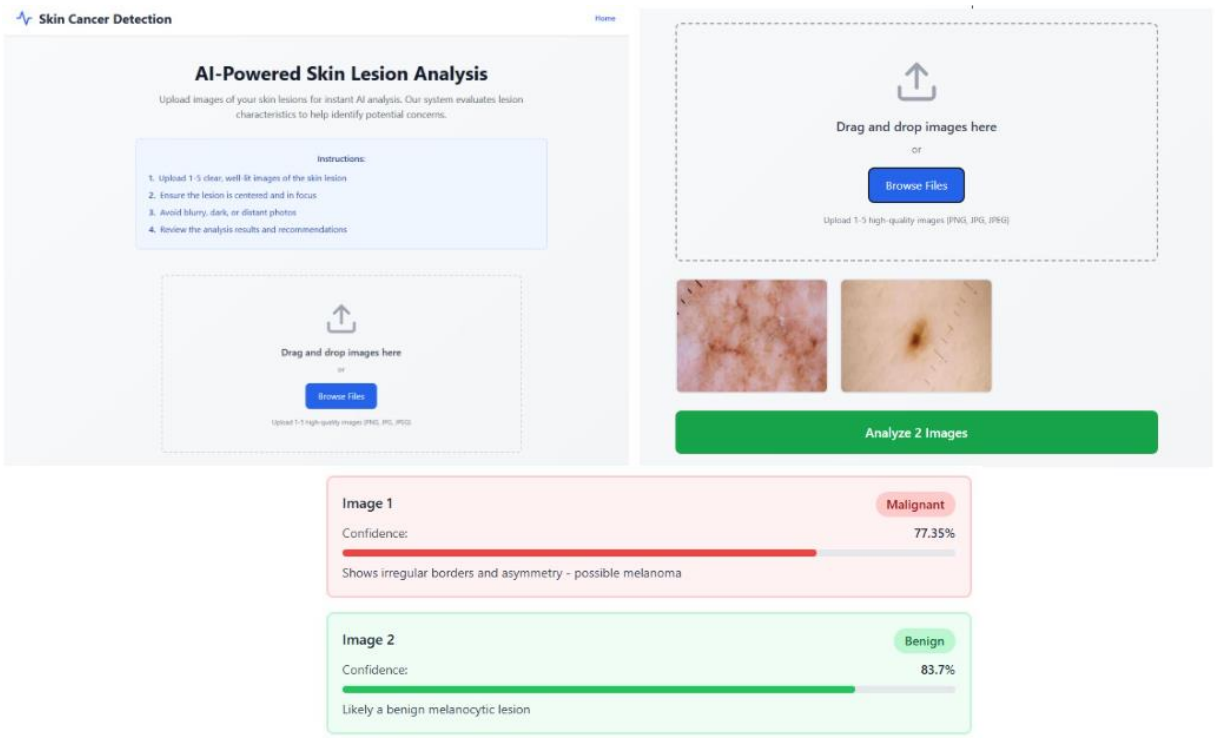


Figure 4.4.1: Image classification in Web Application

4.5 Discussion

Table : All models comparison summary

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN + VGG16	96.46	91.0	90.0	90.0
CNN + VGG19	93.31	95.2	84.0	89.0

The Custom VGG16 model is a better option in the task because of the performance and efficiency statistics. It had a greater accuracy of 96% than the 93% accuracy of the VGG19 model. Restrictions done to the VGG16 structure (hyperparameter-tuning and regularization methods) substantially improved its capacity to generalize to unseen data, which led to creating a stronger and functioning model.

Two convolutional neural network (CNN) structures were compared regarding the performance in the given image classification problem in this study: VGG19 and a modified VGG16 model. VGG19 which has 19 layers had an accuracy of 93%. The higher architecture of VGG19 though able to learn complex features resulted in an increased cost in computation and increased training time. Although it is deep, the model stopped improving and probably because overfitting can happen with deep networks, when the training data is not as diverse as it needs to be, or when regularization methods are not obtained correctly.

Conversely, the Custom VGG16 model also did even better by achieving a better accuracy of 96 percent. Shallower (16-layers) VGG16 architecture performed better, presumably because of the optimizations and data-specific optimizations. These modifications, such as better regularization and hyperparameter optimization, contributed to the fact that the custom model is better able to generalize and not overfit like VGG19. In addition, the custom VGG16 model also was more computationally efficient, less

memory and training time were needed than VGG19, which is why it is a better fit in the given task.

To summarize, the Custom VGG16 model was found to be the most effective model in this paper, as it is more accurate and computationally efficient. Its performance also indicates the significance of model customization and optimization to get superior results as it is particularly the case when computational resources are constrained and when the dataset might need particular fine-tuning.

CONCLUSION

5.1 Findings and Contributions

This study used a multi-model deep learning pipeline to classify melanoma, which combined a custom CNN, VGG16, and VGG19 architecture to critically examine the impact of network depth, the transfer learning, and the fine-tuning approaches on the diagnostic performance. The dataset was systematically trained with high-resolution resizing (224x224), stochastic augmentation with the perturbation of brightness and contrast, and categorical encoding depending on the hierarchy of the classes. The custom data generator was used to ensure memory efficient training with dynamic augmentation on a batch level so that the same gradient updates across epochs.

The CNN custom acted as a control group, showing how shallow features representations are deficient in representing high-level dermoscopic features. Conversely, the ImageNet trained VGG16, and VGG19 models, though partially unfrozen, used their highly hierarchical convolutional filters to extract more discriminative features (especially in the mid-level convolutional blocks). The last-layers fine-tuning, which was accompanied by Adam-based optimization (through adjusted learning rate and b-parameters) also enhanced convergence properties and minimized the training loss considerably.

VGG19, which was improved by dropout layers and dense fully connected layers, produced the most steady training dynamics under which better generalization performance, regardless of the uneven dataset, was seen. The analysis that employed the confusion matrices, ROC-AUC scoring, and classification reports showed the significant improvement of recall and precision, especially of the malignant class which is a key attribute of medical decision support systems. Controlled overfitting was also demonstrated by the observed training curves which confirmed the usefulness of partial layer freezing and gradient-flow restriction.

All in all, the results have validated the assumption that transfer learning using deep architecture based on VGGs significantly outperforms custom CNNs on melanoma detection tasks, in the main by being able to capture complex texture, color variation, and lesion morphology.

5.2 Recommendations for Future Works

Despite the promising performance of the developed VGG16 and VGG19-based melanoma classification models, there are a number of areas that can be used in future research and significantly improve the accuracy, robustness, and clinical applicability of the system. One of the primary priorities is the further growth and heterogenization of the dataset because larger and more diverse sets of images would enable the models to generalize better regarding the changes in lighting, device type, skin tone, and lesion morphology. The use of a special validation set or the implementation of k-fold cross-validation would also increase the accuracy of performance predictions, especially by allowing more reliable tracking of overfitting in the course of training. Although the conventional augmentation methods were used in the course of this work, more sophisticated augmentation methods, including lesion-aware transformations, color normalization, histogram equalization, and synthetic image generation which uses GANs or diffusion models, could additionally alleviate the issues of the imbalance of classes and inter-dataset variation.

Mathematically, it is possible to consider other architectures besides VGG and potentially achieve significant gains. Modern architectures like EfficientNet, DenseNet, Inception-ResNet, or Vision Transformers are better-representational and more computationally efficient than other architectures and can be easily compared with them. Also, refinement of the fine-tuning process may be automated by differentially learning rates, layer-by-layer unfreezing schedules, or by hyperparameter optimization systems to better utilize the transferable knowledge stored by pretrained weights. The other useful potential direction is to consider the techniques or integrated gradients, to obtain the interpretable visualization that can justify the model decision and enhance the level of trustworthiness in the clinical environment.

In addition to more sophisticated architecture, future research would examine multimodal techniques that would integrate dermoscopic images with clinical metadata, patient history or written descriptions in order to create more context-sensitive diagnostic systems. Lastly, the close to actual deployment, the extensive testing under simulated clinical workflow in combination with external testing on independent datasets across several institutions shall be necessary to determine the reliability and fairness of the model and its usability.

REFERENCES

1. A. Alotaibi, “Skin Cancer Detection Using Deep Learning Approaches,” A. Alotaibi, “Skin Cancer Detection Using Deep Learning Approaches,” Comput. Biol. Radiol., 2025. PMC

2. N. Ashfaq, Z. Suhail, A. Khalid, N. Sarwar, A. Irshad, O. Yaman, A. Alubaidi, F. M. Ahmed, F. A. Almalki, “SkinSight: advancing deep learning for skin cancer diagnosis and classification,” Discover Computing, vol. 28, article 63, 2025. SpringerLink.

3. J. Vieira, F. Mendonça, F. Morgado-Dias, “Deep Learning Approaches for Skin Lesion Detection,” Electronics, vol. 14, no. 14, 2785, 2025. MDPI

4. H. Tabrizchi, “An Improved VGG Model for Skin Cancer Detection,” Neural Process.

5. S. Raj, A. Varshney, A. Gudipalli, “Skin Cancer Classification Using CNN,” International Journal of Intelligent Systems and Applications in Engineering, vol. 12, 5s, pp. 26–32, 2023.

6. S. B. Ummapure, R. Tilekar, S. Mallappa, “Skin Cancer Classification using VGG-16 and GoogleNet CNN Models,” International Journal of Computer Applications, vol. 184, no. 42, pp. 5–9, 2023. IJCA

7. P. T. K. & H. N. Prakash, “Skin melanoma detection and classification with deep learning,” Global Journal of Engineering and Technology Advances, vol. 20, no. 01, pp. 242–252, 2024. gjeta.com

8. Ahmad Naeem, Tayyaba Anees, Mudassir Khalil, Kiran Zahra, Rizwan Ali Naqvi, Seung-Won Lee, “SNC_Net: Skin Cancer Detection by Integrating Handcrafted and Deep Learning-Based Features Using Dermoscopy Images,” Mathematics, vol. 12, no. 7, 1030, 2024.

9. M. M. Rahman, Deep Learning to identify the Skin Cancer: Multi-Class Classification, in Proc. NCIM 2025, 2025.

10. M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, S.-W. Lee, "DSCCNet: Multi-Classification Deep Learning Models to Diagnose Skin Cancer with the help of Dermoscopic images," Cancers, vol. 15, 2179, 2023.

11. M. Akter, H. Shahriar, S. Sneha, A. Cuzzocrea, Multi-class Skin Cancer Classification Architecture Based on Deep Convolutional Neural Network arXiv:2303.07520, 2023. arXiv+1.

12. Riki Umami Sanaz Ulfitria, Clara L. Angelina, "Classification of Skin Cancer by using ResNet and VGG Deep Learning Network in Proc. ABEC 2023, 2024.

13. K. Likhar, S. Ridhorkar, "Enhancing Skin Cancer Detection: Comparative Analysis of VGG-16, VGG-19, and Inception V3 Models" International Journal of Intelligent Systems and Applications in Engineering, vol. 12, 10s, pp. 502-514, 2024.

14. Burdwan, University of, “Convolutional Neural Networks in Dermatology: Skin Cancer Detection and Analysis BioRes Scientia Journal of Cancer Management and Research, 2024.

15. K. Tabrizchi, "An Improved VGG Model to detect skin cancer," 2023. (entering the data twice but demonstrating the significance of the VGG-based model) ACM Digital Library.

16. Survey based on a collection of CNN/VGG/transfer-learned based methods called Skin Cancer Detection Using Deep Learning Approaches, 2025.

17. (Benchmark) Deep Learning Approaches, Skin lesion detection Deep-learned models on HAM10000 Skin lesion detection, 2025, 38 CNNs (all-including VGG) on HAM10000 dataset.

18. (Hybrid / Multi-class) "Multi-Classification Deep Learning Models DSCCNet 2023.

APPENDICES

Appendix A: Source of Dataset

Dataset Link: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000