

An Interpretable Enhanced Transfer Learning Approach for Improved Skin Cancer Classification

by

**Md. Rifat Aknda (212014003)
Md. Tanzim Ahmed Tomal (211014027)
Boishakhi Akther (212014040)
Shabnaz Akter Tisha (211014016)**

*Capstone project report (CSE 4098C) submitted in partial fulfillment of the
requirements for the degree of*

Bachelor of Science in Computer Science and Engineering

Under the supervision of

Prof. Muhammad Golam Kibria, PhD, SMIEEE



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY OF LIBERAL ARTS BANGLADESH**

SPRING 2025

© *Md. Rifat Aknda, Md. Tanzim Ahmed Tomal, Boishakhi Akther, and Shabnaz Akter*
Tisha

All rights reserved.

DECLARATION

Project Title An Interpretable Enhanced Transfer Learning Approach for Improved Skin Cancer Classification

Authors *Md. Rifat Aknda, Md. Tanzim Ahmed Tomal, Boishakhi Akther, and Shabnaz Akter Tisha*

Student IDs 212014003, 211014027, 212014040, and 211014016

Supervisor Prof. Muhammad Golam Kibria, PhD, SMIEEE

We declare that this capstone project report entitled *An Interpretable Enhanced Transfer Learning Approach for Improved Skin Cancer Classification* is the result of our own work except as cited in the references. The capstone project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Md. Rifat Aknda
212014003

Md. Tanzim Ahmed Tomal
211014027

Boishakhi Akther
212014040

Shabnaz Akter Tisha
211014027

Date: May 12, 2025



Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
688 Beribadh Road, Mohammadpur
Dhaka - 1207

CERTIFICATE

This is to certify that the capstone project report entitled **An Interpretable Enhanced Transfer Learning Approach for Improved Skin Cancer Classification**, submitted by **Md. Rifat Aknda** (Student ID: 212014003), **Md. Tanzim Ahmed Tomal** (Student ID: 211014027), **Boishakhi Akther** (Student ID: 212014040), and **Shabnaz Akter Tisha** (Student ID: 211014016) are undergraduate students of the **Department of Computer Science and Engineering** has been examined. Upon recommendation by the examination committee, we hereby accord our approval of it as the presented work and submitted report fulfill the requirements for its acceptance in partial fulfillment for the degree of *Bachelor of Science in Computer Science and Engineering*.

Prof. Muhammad Golam Kibria, PhD, SMIEEE
Professor & Head (Acting)

Md. Aktaruzzaman Pramanik
Lecturer

Md Kaviul Hossain
Lecturer

Place: Dhaka

Date: May 12, 2025

ACKNOWLEDGEMENTS

We would like to express our deep and sincere gratitude to our research supervisor, *Prof. Muhammad Golam Kibria, PhD, SMIEEE*, for giving us the opportunity to conduct research and providing invaluable guidance throughout this work. His dynamism, vision, sincerity and motivation have deeply inspired us. He has taught us the methodology to carry out the work and to present the works as clearly as possible. It was a great privilege and honor to work and study under his guidance.

We are greatly indebted to our honorable teachers of the Department of Computer Science and Engineering at the University of Liberal Arts Bangladesh who taught us during the course of our study. Without any doubt, their teaching and guidance have completely transformed us to the persons that we are today.

We are extremely thankful to our parents for their unconditional love, endless prayers and caring, and immense sacrifices for educating and preparing us for our future. We would like to say thanks to our friends and relatives for their kind support and care.

Finally, we would like to thank all the people who have supported us to complete the project work directly or indirectly.

Md. Rifat Aknda, Md. Tanzim Ahmed Tomal, Boishakhi Akther, and Shabnaz Akter Tisha
University of Liberal Arts Bangladesh

Date: May 12, 2025

To the one I've seen,
who walked through fire and smiled.

– Md. Rifat Aknda

Dedicated to
Myself.

– Md. Tanzim Ahmed Tomal

Dedicated to
My beloved parents.

– Boishakhi Akther

Dedicated to
My loving parents.

– Shabnaz Akter Tisha

ABSTRACT

Skin cancer is the most prevalent cancer worldwide, with early detection critical for improving survival rates. Currently used diagnostic methods, such as visual inspection and biopsies, have significant limitations, including human error and high costs. Although several existing deep learning models have been developed, they demonstrate notable limitations. An interpretable and improved transfer learning approach for binary skin cancer classification using VGG16 as the base model has been proposed in this research. The methodology focuses on addressing existing limitations in skin cancer classification technologies, with the ultimate goal of supporting healthcare professionals and potentially saving lives through advanced, accessible AI-driven diagnostic tools. Multiple optimization techniques have been applied to enhance the classification accuracy of dermoscopic images into benign and malignant skin cancer categories. The implemented transfer learning model has shown 90.63% classification accuracy, which is better than VGG16. A user-friendly web application has been developed that integrates the proposed classification model, allowing preliminary screening in real time.

Keywords: skin cancer, classification, skin cancer classification, binary classification, benign, malignant, Kaggle, convolutional neural network, deep learning, vgg16, transfer learning, explainable AI, grad-cam, medical diagnostics, dermoscopic image analysis, artificial intelligence, web application.

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Problem statement	3
1.3	Aims, objectives and Motivation	3
1.3.1	Aims	3
1.3.2	Objective	4
1.3.3	Motivation	4
1.4	Project Specification	5
1.5	Project Requirements	5
1.5.1	Hardware Requirements:	5
1.5.2	Software Requirements:	6
1.6	Project Budget	6
1.7	Project Scope	6
1.8	Project Timeline	8
2	Literature Review	10
3	Methodology	17
3.1	Concept Generation	17
3.1.1	Model Architecture Alternatives	17
3.1.2	Training Alternatives	18
3.2	Concept Reduction	18
3.2.1	Decision-Making Process	18
3.3	Final Concept Justification	19
3.4	Stages of the Methodology	19
3.5	Justification and Novelty	20

4	Design and Development	21
4.1	Proposed Transfer Learning Model	21
4.2	Application of Pretrained VGG16 Architecture	22
4.3	Fine Tuning of Layers	23
4.4	Subsystems and Workflow	24
4.4.1	Data Preprocessing Subsystem	24
4.4.2	Feature Extraction Subsystem	25
4.4.3	Classification Subsystem	25
4.4.4	Training and Optimization Subsystem	26
4.5	Engineering Analysis	26
4.5.1	Formulations and Parameters	26
4.5.2	Assumptions	27
4.5.3	Design Viability	27
4.6	Proposed System Architecture	27
4.7	Graphical User Interface (GUI) of the System	28
5	Result analysis	31
5.1	Hyperparameter Tuning	31
5.2	Model Accuracy and Model Loss Analysis	32
5.2.1	Training and Validation Accuracy	33
5.2.2	Training and Validation Loss	34
5.2.3	Test Accuracy and Test Loss	35
5.3	Comparison of Confusion Matrix and Performance Metrics with VGG16	36
5.3.1	Comparison of the Confusion Matrices	36
5.3.2	Comparison of Performance Metrics	37
5.4	Heatmap Analysis	39
5.5	Comparison with the State-of-the-Art Approaches	40
6	Conclusions	43
6.1	Social, Legal, Ethical, and Environmental Issues	43
6.1.1	Social Issues	43
6.1.2	Legal Issues	44
6.1.3	Ethical Issues	44
6.1.4	Environmental Issues	45

6.2	Project Outcomes	45
6.3	Future Work	46
6.4	Conclusion	46
References		48

List of Figures

1.1	Timeline of Project Phases	8
3.1	Stages of the Methodology	19
4.1	Proposed Transfer Learning Model Architecture	22
4.2	VGG16 Architecture	23
4.3	Fine Tuning of Proposed Transfer Learning Model	24
4.4	Data Preprocessing Subsystem	25
4.5	Feature Extraction Subsystem	25
4.6	Classification Subsystem	25
4.7	Training and Optimization Subsystem	26
4.8	Proposed System Architecture	27
4.9	Graphical User Interface (GUI) of the System	29
5.1	Training and Validation Accuracy	33
5.2	Training and Validation Loss	34
5.3	Test Accuracy and Test Loss	35
5.4	Comparison of Confusion Matrices between VGG16 and Proposed Model	36
5.5	Heatmap Analysis	40

List of Tables

1.1	Project Budget	6
2.1	Comparative Studies Analysis	16
5.1	Comparison of performance metrics between VGG16 and the Implemented Model.	38
5.2	Comparative result analysis of relative works	42

Chapter 1

Introduction

This chapter discusses the growing global incidence of skin cancer and the challenges associated with traditional diagnostic approaches. It highlights the need for an automated system capable of accurately classifying skin cancer as benign or malignant. The research focuses on developing a model based on transfer learning with VGG16, with key objectives including enhancing classification accuracy, integrating explainable artificial intelligence (XAI), and enabling real-time deployment. The scope of the project extends to clinical, educational, and public health applications. In addition, this chapter also outlines the project timeline, detailed project requirements, and the associated budget table, providing a comprehensive overview of the project's planning and resource allocation.

1.1 Introduction

Cells form the basic building blocks that make up our skin tissue. As the outer layer of our body, the skin protects us from infections and harmful radiation. Skin cancer is the most common type of cancer and is becoming more and more widespread [Anand et al. \(2022a\)](#); [International Agency for Research on Cancer \(IARC\) \(2022\)](#); [Skin Cancer Foundation \(2023\)](#). It occurs when skin cells grow abnormally, often due to DNA damage caused by ultraviolet radiation [Cancer Research UK \(2024a\)](#); [Skin Cancer Foundation \(2024\)](#). The two types of skin cancer are malignant and benign [Gupta et al. \(2023\)](#). The incidence of malignant skin cancers has increased significantly in recent decades [Cancer Research UK \(2024b\)](#). According to the World

Health Organization (WHO), between 2 and 3 million non-melanoma skin cancers and 132,000 melanoma skin cancers are reported globally each year [World Health Organization \(2022\)](#); [American Academy of Dermatology \(2023\)](#). Malignant skin cancer causes almost 75% of skin cancer-related deaths [Davis et al. \(2019\)](#). Early detection is crucial, as survival rates drop from 98% in the early stages to 17% in the advanced stages, highlighting the importance of accurate and timely diagnosis [National Cancer Institute \(2023\)](#).

Traditional methods like visual inspection and biopsies have limitations, including human error, high costs, and delays in results [Skin Cancer Foundation \(2025\)](#). Automated approaches that use artificial intelligence (AI) and deep learning have shown great potential to address these issues [Lee et al. \(2023\)](#). Deep learning, particularly convolutional neural networks (CNNs), has been shown to be effective in the analysis of medical images, the identification of patterns, and the distinguishing between image types, which is crucial to detect skin cancer [Esteva et al. \(2017\)](#). Among them, VGG16 stands as the leading CNN models for the ability to detect complex features in the dermoscopic images, and it has performed well in the classification of skin cancer types [Patel & Wong \(2023\)](#). Transfer learning, by using the pre-trained models such as VGG16, ResNet, and DenseNet, improves the ability of the systems to work with small-sized data [Chen et al. \(2023\)](#). Even while such technologies may make incredible leaps, however, they remain hampered with serious limitations, including challenges in classification accuracy, model overfit, reliance on large annotated data, complexities in model decision interpretability, and challenges in real-time implementation [Hara et al. \(2022\)](#); [Yadav & Tiwari \(2020\)](#); [Singh & Patel \(2021\)](#); [Zhang & Li \(2019\)](#).

A new and explainable transfer learning approach for binary skin cancer classification using VGG16 (last 4 conv layers) as the backbone model has been proposed in this research for enhanced skin cancer classification. It is built on the backbone VGG16, with new layers, optimization methods, and state-of-the-art activation units fine-tuned in such a manner as to enhance the benign and malignant classification accuracy.

1.2 Problem statement

Skin cancer, one of the most devastating forms of cancer with rapidly increasing global incidence rates, requires an automated system to identify skin cancer in benign and malignant stages to save lives [Anand et al. \(2022a\)](#). There are multiple deep learning methods to classify skin cancer as benign or malignant [Wu et al. \(2022\)](#). However, existing approaches demonstrate limitations in classification accuracy, lack of result interpretation, and absence of real-time deployment [Soumyya & et al. \(2024\)](#); [Tjoa & Guan \(2022\)](#); [Melarkode et al. \(2023\)](#).

The primary challenge is the need for an improved and deployable skin cancer classification system that addresses these limitations. In this research, an improved transfer learning approach has been proposed using VGG16 as the base model that improved the classification accuracy beyond existing benchmarks, used explainable AI for the interpretation of the results, and implemented web application integration for practical and real-time deployment. Optimized the proposed approach with multiple optimization techniques and advanced activation functions to achieve higher performance while maintaining computational efficiency.

1.3 Aims, objectives and Motivation

This section outlines the goals, objectives, and motivation behind the development of the skin cancer classification system.

1.3.1 Aims

The primary aim of this research is to develop an enhanced transfer learning approach for improved classification of skin cancer as benign or malignant, using VGG16 as the base model. The aim is also to integrate the implemented model into a practical mobile application to support real-world medical use.

1.3.2 Objective

- Design and develop the proposed transfer learning approach using VGG16 as the base model.
- Achieve higher classification accuracy compared to existing approaches.
- Increase both true positives (TP) and true negatives (TN) simultaneously.
- Implement explainable artificial intelligence (XAI) techniques to interpret model predictions.
- Evaluate system performance in terms of accuracy, recall, precision, F1 score, and support.
- Design and develop a user-friendly web application and integrate the transfer learning model for real-time usage.

1.3.3 Motivation

Skin cancer is one of the most devastating forms of cancer, with rapidly increasing incidence rates worldwide [Hasan et al. \(2023\)](#); [Siegel et al. \(2023\)](#). To reduce the death rate from skin cancer, it is essential to diagnose skin cancer in its benign stage as soon as possible [Naqvi et al. \(2023\)](#). Current diagnostic methods are based on biopsies, which are invasive, time-consuming, prone to human error, and can lead to the risk of misdiagnosis and improper treatment [Blundo et al. \(2021\)](#); [Brinker et al. \(2019\)](#). To save lives, an automated system that can detect skin cancer in its earliest stages is necessary [Mazhar et al. \(2023\)](#). Various researchers have conducted research using deep learning and transfer learning models [Akter et al. \(2023\)](#). However, the existing literature is limited in terms of classification accuracy and is often time consuming [Anand et al. \(2022b\)](#); [Dildar et al. \(2021\)](#). So, it is critical to design an automatic system that can deliver a fast judgment and considerably reduce errors in diagnosis. In this research, a deep learning-based model has been designed for the identification of skin cancer in benign and malignant stages using the concept of the transfer learning approach. Our approach aims to improve the accuracy of skin cancer classification by integrating advanced optimization techniques with transfer learning. The system incorporates explainable AI features to help doctors

understand the reasons behind each diagnosis and make more informed clinical decisions.

1.4 Project Specification

- Dataset: 3297 labeled images in 2 classes (benign, malignant).
- Model: Transfer Learning-Based Architecture (Base Model: VGG16).
- Software: PyCharm Professional, Jupyter Notebook.
- Programming Languages and Framework: Python, JavaScript, Flask.
- Markup and Style Sheet Languages: HTML, CSS.
- Libraries and Dependencies: Numpy, Pandas, scikit-learn, Seaborn, Matplotlib, Tensorflow, Keras, OpenCV (cv2).

1.5 Project Requirements

This section details the necessary hardware and software components required to implement and run the skin cancer classification system effectively.

1.5.1 Hardware Requirements:

- Dermoscope
- GPU
- Processor
- RAM
- Storage
- Smartphone or Computer

1.5.2 Software Requirements:

- PyCharm Professional
- Python (Version 3.10.11 or higher)
- Python Libraries: TensorFlow, Keras, NumPy, OpenCV, Flask, Matplotlib, Scikit-learn
- MySQL or SQLite (Optional)
- Web Browser

1.6 Project Budget

The project budget is detailed in Table 1.1, showing the allocation of funds across hardware, software, and Miscellaneous categories.

Category	Description	Estimated Cost (BDT)
Hardware	Dermoscope, GPU, Processor, Smartphone/Computer	150,000-200,000
Software	Pro version of software	8,000-12,000
Miscellaneous	Contingency	12,000
Total		170,000-224,000

Table 1.1: Project Budget

1.7 Project Scope

The scope of this research involves the development, validation, and deployment of an interpretable deep learning-based system for the binary classification of skin cancer as benign or malignant using dermoscopic images. The proposed solution addresses the critical limitations in current diagnostic methodologies by using transfer learning with the VGG16 architecture and incorporating explainable artificial

intelligence (XAI) techniques.

This work has significant implications across various domains within the healthcare and medical technology landscape, notably in the areas of:

- **Clinical Diagnostics Support**

- Assists dermatologists and general practitioners in early-stage skin cancer detection.
- Acts as a second opinion tool to reduce human error in clinical evaluations.
- Provides visual explanations of model decisions via Grad-CAM to support transparent diagnosis.

- **Telemedicine and Remote Healthcare**

- Allows skin lesion analysis in remote or underserved areas via a web application.
- Facilitates remote consultations by allowing healthcare professionals to upload dermoscopic images for instant analysis.

- **Educational and Training Tool**

- Serves as a learning resource for medical students and professionals by illustrating decision making of deep learning model.
- Incorporated Grad-CAM to outputs highlight key image features influencing predictions.

- **Early Screening and Mass Health Campaigns**

- Can be integrated into public health initiatives for large-scale skin cancer screenings.
- Reduce diagnostic constraints by providing quick and automated assessments in community health camps.

- **Research and Development**

- Provides a foundation for further research in medical AI, particularly in model interpretability and transfer learning enhancements.
 - Provides a benchmarking framework for future skin cancer classification models.
- **Digital Health Platforms**
 - Can be integrated into existing digital health ecosystems or hospital management systems.
 - Enhances electronic health records with automated, image-based diagnostic support.

1.8 Project Timeline

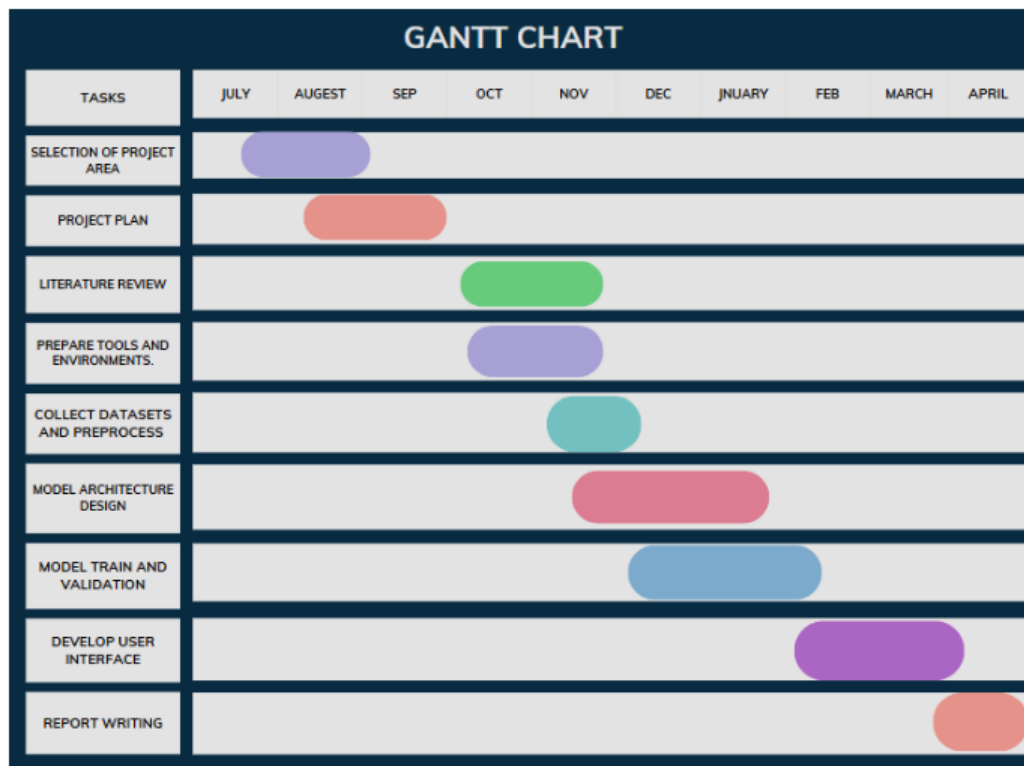


Figure 1.1: Timeline of Project Phases

The project timeline, including the research phases, is illustrated in Figure 1.1. This Gantt chart provides a clear and concise overview of the project timeline, covering from July to April. The project begins in July and August with the selection of the project area and planning. In September and October, the focus shifts to conducting a literature review and preparing the necessary tools and environments. During November and December, datasets are collected and pre-processed, ensuring that they are ready for the subsequent phase. In January and February, the model architecture is designed, followed by model training and validation in March. The project ends in April with the development of the user interface and the final report. Overlapping tasks are strategically mapped to optimize time and ensure steady progress. Key milestones are highlighted to ensure that deadlines are met, providing a structured approach to achieving the project's objectives.

Chapter 2

Literature Review

This chapter reviews recent advances in skin cancer classification using deep learning and computer vision, with a focus on transfer learning approaches. Various studies have used models such as VGG16, DenseNet-121, InceptionNet, and ResNet50, often enhanced through fine-tuning and hyperparameter optimization. Most used the Skin Cancer – Malignant vs. Benign Kaggle dataset (3,297 images), incorporating data augmentation techniques to boost performance. Enhanced deep learning models, including adaptations of VGG16, DenseNet, and ResNet, have demonstrated great potential in skin cancer classification tasks. These studies highlight the promise of AI in facilitating early diagnosis and supporting clinical decision making. However, more research is necessary to address the issues of robustness, scalability, and practical deployment in real-world medical settings. A table of comparative studies is presented at the end of this chapter.

[Anand et al. \(2022a\)](#) proposed an enhanced transfer learning approach using a modified VGG16 architecture for skin cancer diagnosis from dermoscopic images. The modifications included adding a flatten layer, two dense layers with LeakyReLU activation, and a dense layer with sigmoid activation to improve classification accuracy. Using Skin Cancer - Malignant vs. Benign dataset, containing 3,297 images sourced from the Kaggle (1,800 benign, 1,497 malignant) and data augmentation to expand the training set to 5,636 images, the model achieved its best performance with a batch size of 128 and 10 epochs, reaching 89.09% accuracy. This surpassed the original VGG16 and other approaches in accuracy, precision, sensitivity, and F1 score. However, the authors note there is still room for improvement in overall accuracy by enhancing both true positives and negatives simultaneously. The study

demonstrates the potential of transfer learning and architectural modifications in improving automated skin cancer diagnosis, though it may benefit from further optimization and real-world implementation and validation.

[Agarwal & Singh \(2022\)](#) explore the use of convolutional neural networks (CNNs) for classifying skin cancer images. Skin cancer, the most common human malignancy, is primarily diagnosed visually through clinical screening, dermoscopic analysis, and biopsy. This study aims to classify skin lesion images using CNNs, which show significant potential for image classification while accounting for environmental variability. The authors used images based on pixel values and disease labels, with the dataset Skin Cancer - Malignant vs. Benign, containing 3,297 images sourced from the Kaggle. The training involved multiple models with transfer learning, achieving a highest accuracy of over 86.65%. The dataset is publicly available to ensure credibility and reproducibility of results.

[Kalouche \(2020\)](#) proposed a deep learning technique to classify melanoma cancer from skin images utilizing machine learning algorithms for early detection. In this study, logistic regression, a fine-tuned VGG-16 CNN model and deep neural networks were trained with the IS IC dataset. The VGG-16 model achieved 78% balanced accuracy for classifying melanoma skin lesions while suffering from dataset imbalance and image variability. This study conveys the possibility of effectively applying deep learning to improve the performance of melanoma detection, and to continue improving the dataset for increased size diversity to achieve impactful reductions in melanoma misclassification.

[Ibrahim et al. \(2023\)](#) used the Kaggle skin cancer dataset to classify skin cancer into benign and malignant categories using transfer learning approach with the VGG16 architecture. In 3,297 photos in Malignant vs. Benign dataset were taken from Kaggle and 360 benign and 300 malignant testing images. In addition to 1,440 benign and 1,197 malignant training images. Modify the VGG16 model for this particular classification task, the researchers added new layers and changed hyperparameters to the pre-trained model which was initially trained on ImageNet. The model classified skin lesions with comparatively little data as evidenced by its 84.242% accuracy on the testing dataset. By using VGG16 model, the researchers

were able to take advantage of learnt properties like texture, color, and shape which decreased risk of overfitting. This research did have certain drawbacks such as rather small dataset size which might have an impact model applies to a wider range of clinical situations. The computing needs for training and testing were not addressed and the model's performance. Then This was not contrasted with alternative architectures which may have affected its suitability in low-resource environments.

[Bello et al. \(2024\)](#) propose a skin cancer classification using fine-tuned transfer learning with DenseNet-121. The authors highlight the base model by adding a flatten layer and three dense layer connected layers employing LeakyReLU activation functions. Also it improves training stability. The final output layer uses a sigmoid activation function for binary classification. The authors used Malignant vs. Benign dataset, containing 3,297 images from the Kaggle. Their approach was evaluated against other models including CNN, SVM, Random Forest, VGG16, ResNet34, Inceptionv3, and EfficientNetB0. The fine-tuned DenseNet-121 achieved an accuracy of 87%, class-wise accuracies of 88% benign lesions, 85% for malignant lesions and an AUC of 0.87 which indicating superior overall classification capability. The study gives the effectiveness of integrating additional layers and fine-tuning in deep learning models for skin cancer diagnosis. Its offering improved accuracy and reliability in clinical settings.

[Bazgir et al. \(2024\)](#) uses the InceptionNet architecture to propose a skin cancer classification system. The authors gathered and preprocessed dermatoscopic pictures of benign and malignant skin lesions from the 3,297 photos in the publically accessible Skin Cancer-Malignant vs. Benign dataset that they obtained from Kaggle. After applying image augmentation techniques, they implemented and fine-tuned the Inception Network for skin cancer detection. The optimized InceptionNet model achieved an accuracy of 85.94%, a sensitivity of 86.89%, and a specificity of 81.98% when utilizing the Nadam optimizer for classifying skin lesions as melanoma or non-melanoma. The Inception Network-based system demonstrated superior classification performance compared to traditional CNN architectures, particularly in distinguishing melanoma from benign lesions with high precision. Compared to

previous work on skin cancer classification, the method showed improved diagnostic accuracy while maintaining computational efficiency. The authors suggest this approach could be integrated into clinical decision support systems and potentially developed into mobile applications for preliminary skin cancer screening. However, limitations include dependence on high-quality imaging data and the need for validation across diverse patient populations.

[Acosta et al. \(2021\)](#) proposed a two-stage deep learning framework for melanoma diagnosis using dermatoscopic images. The first stage used Mask R-CNN to segment the lesion region, while the second applied a ResNet152 model to classify the cropped image as benign or malignant. Trained on Skin Cancer - Malignant vs. Benign dataset, containing 3,297 images sourced from the Kaggle, their eVida M6 model proved to be a reliable predictor, achieving an excellent balance between overall accuracy (0.904), sensitivity (0.820), and specificity (0.925), outperforming previous models. However, the authors note that merging additional datasets like PH2 did not improve performance due to preprocessing differences. They also highlight that ensure generalizability in clinical settings, aware augmentation and additional validation are required.

[Hasan et al. \(2019\)](#) proposed machine learning and image processing approaches to make a skin cancer diagnosis system. The convolutional neural networks was used to classify the features of the impacted skin cells that were collected from dermoscopic pictures using segmentation. With an accuracy of 89.5% using the Skin Cancer Malignant vs. Benign dataset which included 3,297 images from Kaggle. The model showed promise for early skin cancer detection. The researchers also recognize certain drawbacks including the dependence on a small dataset and the requirement for additional testing to improve the system's resistance and generalizability clinical contexts.

[Unnisa et al. \(2025\)](#) examined the effect of the Good Tuning Convolutional Neural Community (CNN) parameters to detect skin cancer. Their experiments have shown that CNN increases the amount of filters in the Convnet 2D teams, improves the CNN architecture - by incorporating layers and getting rid of dropout - ever increased category accuracy, which increases it from sixty% to eighty% to eighty% - soluble

vs. Take snap shots and snap. The probe outlines the importance of hyperparameter setting in adaptation to CNN performance for the clinical photo category. However, the author is well-known limitations, including the need for more verification in different data sets and scientific surroundings to ensure the strength and generality of the version.

[Khan et al. \(2025\)](#) investigated the use of transfer learning for skin cancer detection from low-resolution images. Using the HAM10000 dataset (10,015 dermoscopic images), they evaluated VGG16, DenseNet169, DenseNet121, and ResNet50 on grayscale images downsampled to 80x80, 64x64, and 32x32 pixels. Converted from 800x600 RGB to grayscale for memory efficiency, the dataset was split into 80% training, 10% validation, and 10% testing. VGG16, with ImageNet pre-trained weights, achieved the highest accuracy (80.46%, 78.56%, 74.15% for the respective resolutions), followed by DenseNet169. The study highlights transfer learning's effectiveness in classifying skin cancer from low-resolution images, supporting portable diagnostic applications. However, testing was limited to HAM10000, and resolutions below 32x32 were not explored. The authors recommend further studies on additional datasets and lower resolutions to enhance portability.

[Djaroudib et al. \(2024\)](#) applied VGG-16 with transfer learning to diagnose basal cell carcinoma (BCC), melanoma (MEL), and nevi (NV), investigating the impact of data quality versus quantity. Three models were trained on datasets of varying sizes, with the model using 13,232 images achieving the highest validation accuracy (94%) and testing accuracy (84.5%), along with a precision of 86%, recall of 85%, and F1 score of 84.5%. The study emphasized the importance of image quality and preprocessing for improving deep learning model performance in medical imaging. However, it was limited by dataset bias, as the Kaggle-sourced images lacked representation of diverse skin types and imaging conditions, potentially affecting generalizability. Future research should explore more diverse datasets, data augmentation, and alternative architectures to enhance robustness and applicability.

[Mahmud et al. \(2023\)](#) propose an interpretable deep learning approach for skin cancer classification, integrating explainable AI (XAI) to enhance model transparency. Using four pre-trained models—XceptionNet, EfficientNetV2S, InceptionResNetV2,

and EfficientNetV2M—they apply image augmentation to address class imbalance and improve generalization. XceptionNet achieved the highest accuracy (88.72%). The study underscores deep learning and XAI's role in early detection, improving patient care and reducing healthcare costs. However, challenges include dataset bias, high computational demands, and the need for clinical validation. Future work should refine generalizability and address class imbalance more effectively.

Kanchana et al. (2024) developed a skin cancer classification system using EfficientNet B0–B7 models combined with patient-specific transfer learning and a custom preprocessing pipeline. The models were fine-tuned on a multiclass dermatological dataset and evaluated using multiple metrics. EfficientNet-B7 achieved the highest performance with 84.4% top-1 accuracy, offering improved diagnostic accuracy and reduced model size. While effective for common conditions like melanoma and benign keratosis, the model struggled with diseases like eczema and psoriasis due to class imbalance. The study suggests strong potential for clinical use, despite the need for more balanced data.

Comparative Studies Analysis

We performed a comparative studies analysis to evaluate existing approaches, as summarized in Table 2.1.

Features / Papers	"Skin Cancer - Malignant vs. Benign" dataset	Accuracy > 90.40%	VGG16	Transfer Learning	Explainable AI	Real-Time Deployment
Anand et al. (2022a)	✓	×	✓	✓	×	×
Agarwal & Singh (2022)	✓	×	×	✓	×	×
Kalouche (2020)	✓	×	✓	×	×	×
Ibrahim et al. (2023)	✓	×	✓	✓	×	×
Bello et al. (2024)	✓	×	×	✓	×	×
Alrabai et al. (2025)	✓	×	✓	×	✓	×
Bazgir et al. (2024)	✓	×	×	×	×	×
Acosta et al. (2021)	✓	×	×	×	×	×
Hasan et al. (2019)	✓	×	×	×	×	×
Aydin & Yildiz (2023)	✓	×	×	×	×	×
Unnisa et al. (2025)	✓	×	×	×	×	×
Khan et al. (2025)	×	×	✓	×	×	×
Djaroudib et al. (2024)	×	×	✓	✓	×	×
Mahmud et al. (2023)	×	×	×	×	✓	×
Kanchana et al. (2024)	×	×	×	✓	×	×
An Interpretable Enhanced Transfer Learning Approach for Improved Skin Cancer Classification (This research)	✓	✓	✓	✓	✓	✓

Table 2.1: Comparative Studies Analysis

Chapter 3

Methodology

This chapter outlines the methodology used in the development of the skin cancer classification model. It includes the process of concept generation and reduction, with an emphasis on selecting the optimal design and training strategies. The chapter also discusses the stages of model development, from data preprocessing to model evaluation, and justifies the novelty and effectiveness of the final approach in improving skin cancer classification accuracy and real-time deployment in clinical settings.

3.1 Concept Generation

In the initial phase, we explored several potential design solutions for the classification of skin cancer using deep learning models. These design alternatives focused on optimizing the performance of the model while ensuring that it could run efficiently in real time. The key design aspects considered were the following:

3.1.1 Model Architecture Alternatives

- Base Model Choice: VGG16 vs. ResNet vs. InceptionNet.
- Layer Freezing Strategy: Freeze only the initial layers vs. freeze all layers except the last four.
- Activation Functions: LeakyReLU vs. ReLU.

- Regularization Techniques: L2 regularization vs. Dropout vs. BatchNormalization.

3.1.2 Training Alternatives

- Optimizer Choice: Adam vs. SGD.
- Learning Rate Schedules: Adjustment of Constant vs. Decay-based learning rate.
- Batch Size and Epochs: Evaluating combinations that balanced model performance with computational efficiency.

We used several techniques to analyze and evaluate these alternatives, ensuring that the final design solution addressed the existing limitations that we identified.

3.2 Concept Reduction

After exploring various conceptual solutions, we applied a structured decision-making process to narrow down the options. The selection was based on several criteria, such as performance metrics, computational efficiency, robustness, scalability, and upgradability.

3.2.1 Decision-Making Process

- Base Model Selection: We chose VGG16 as the base model due to its proven effectiveness in image classification and the availability of pre-trained weights.
- Layer Freezing: We froze all layers except for the last four to retain pre-trained knowledge while fine-tuning for skin cancer classification.
- Activation Function: LeakyReLU was selected to prevent the dying ReLU problem, the vanishing gradient problem, and improve the training stability.
- Regularization: We implemented BatchNormalization, Dropout, and L2 regularization to stabilize training and prevent overfitting.

- **Optimizer and Training Strategy:** Adam was chosen for its adaptive learning rate and the learning rate was fine-tuned to ensure smooth convergence.

3.3 Final Concept Justification

The final model was chosen based on the combinations of optimal accuracy, efficiency, and the ability to perform well for demographics. The decision to focus on those key design elements provided the best balance between performance and computational feasibility for real-time deployment.

3.4 Stages of the Methodology

The stages of the proposed methodology are illustrated below in Figure 3.1.

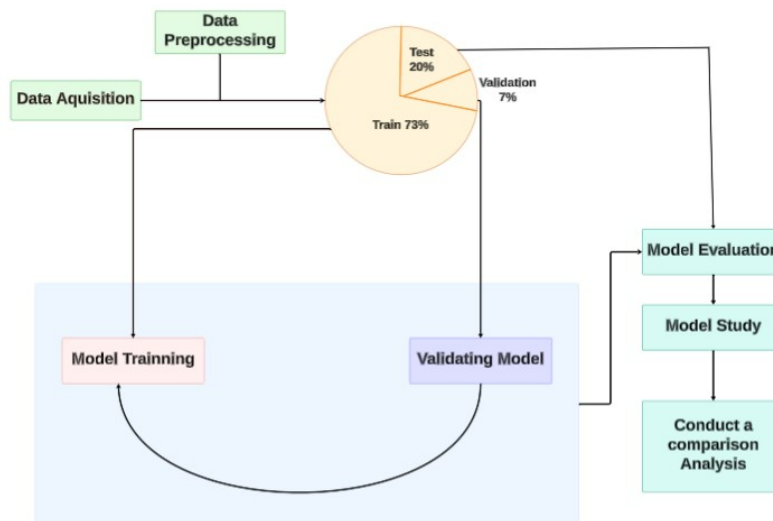


Figure 3.1: Stages of the Methodology

The methodology follows a structured process to ensure optimal model performance. First, labeled data acquired from Kaggle. Next, we did all the preprocessing steps needed. The processed data was divided into training sets (73%), validation

sets (7%), and test sets (20%). The model was trained using training data, learning patterns, and relationships, followed by validation to fine-tune the hyperparameters and prevent overfitting. After validation, the model was evaluated on test data using multiple performance metrics. Finally, we performed a comparison analysis with other state-of-the-art models.

3.5 Justification and Novelty

This research presents an improved transfer learning approach for the classification of skin cancer, using VGG16 as the base model. By incorporating advanced optimization Technology and Activation Work improves the Model Classification. Accuracy increases both real positivity and real negative. To Make sure lecturer, explain AI (character comb) is integrated, provides trans-Insight into the process of decision -making of parents' visual model. Other than this, The model is distributed in a real -time online application, which ensures convenient. Propactism in the clinical environment. This approach addresses the most important limits of the previous-Instead methods, improved both clinical accuracy and purpose Practice.

Chapter 4

Design and Development

This chapter presents the design and development of the proposed system for the classification of skin cancer. It begins with a description of the proposed transfer learning model and the integration of the pre-trained VGG16 architecture for feature extraction. The fine-tuning approach applied to specific layers is discussed to highlight performance improvements. The structure of the subsystem and the overall workflow are illustrated, followed by an engineering analysis to assess the technical feasibility and viability of the design. The proposed system architecture is then outlined, followed by the presentation of the graphical user interface (GUI) designed to support efficient user interaction and real-time deployment.

4.1 Proposed Transfer Learning Model

An interpretable and improved transfer learning-based model was developed for the binary classification of skin cancer as benign or malignant. The model was trained and evaluated using a Kaggle dataset comprising 3,297 dermoscopic images of skin lesions [Fanconic \(2019\)](#). These images were pre-processed and divided into training, validation, and test sets, ensuring a balanced distribution of both classes. The general architecture of the proposed model is illustrated in Figure 4.1.

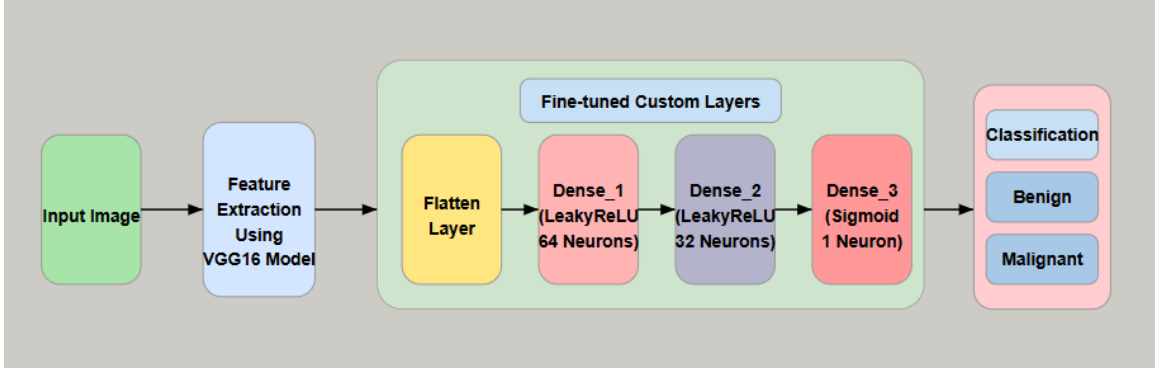


Figure 4.1: Proposed Transfer Learning Model Architecture

The feature extraction phase uses only the final four layers of the VGG16 architecture for feature extraction due to their ability to capture high-level abstract features relevant to skin lesion classification. Following feature extraction, a custom classifier head was implemented to interpret the extracted features for binary classification. The classifier head begins with a flatten layer that transforms the multi-dimensional feature maps into a one-dimensional vector. This is followed by two fully connected dense layers with 64 and 32 neurons, respectively, both using the LeakyReLU activation function to enhance training stability and mitigate the problem of vanishing gradient. The final layer is a dense output layer with a single neuron, activated by sigmoid function, which classifies skin cancer as benign or malignant based on previous learned characteristics. This structure provides a balance between model complexity and performance, making it suitable for real-time and clinical deployment scenarios.

4.2 Application of Pretrained VGG16 Architecture

To enhance the classification of skin lesions into benign and malignant categories, this study uses a transfer learning approach with the VGG16 model. The architecture of the VGG16 model, illustrated in Figure 4.2, consists of five convolutional blocks with 3×3 filters and ReLU activations, each followed by 2×2 max-pooling operations to progressively reduce spatial dimensions while increasing feature depth.

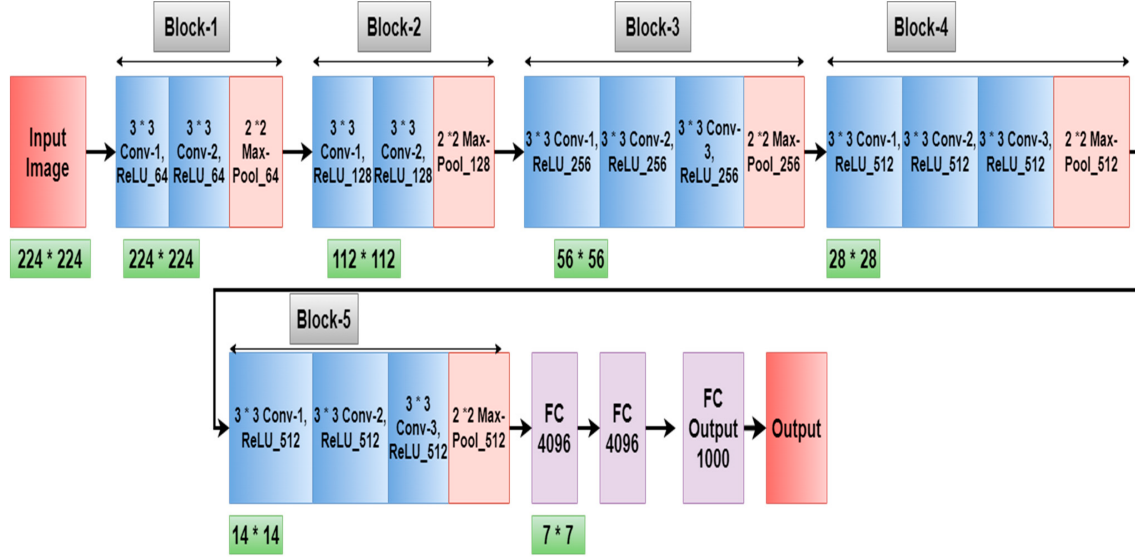


Figure 4.2: VGG16 Architecture

The final four layers of the pre-trained VGG16 model were utilized for feature extraction in this research. These layers were selected because of their ability to capture high-level abstract features relevant to the classification of skin lesions. By freezing the remaining layers, the model retains the general feature representations learned from the ImageNet dataset while ensuring efficient and stable learning with limited medical data. The extracted features were passed through the custom classifier head tailored for binary classification. This transfer learning strategy provides a balance between the use of deep representational power and the assurance of efficiency, enabling effective classification even with limited domain-specific data.

4.3 Fine Tuning of Layers

Figure 4.3 illustrates the fine-tuning of the layers (including the last 4 layers of VGG16) of the proposed model.

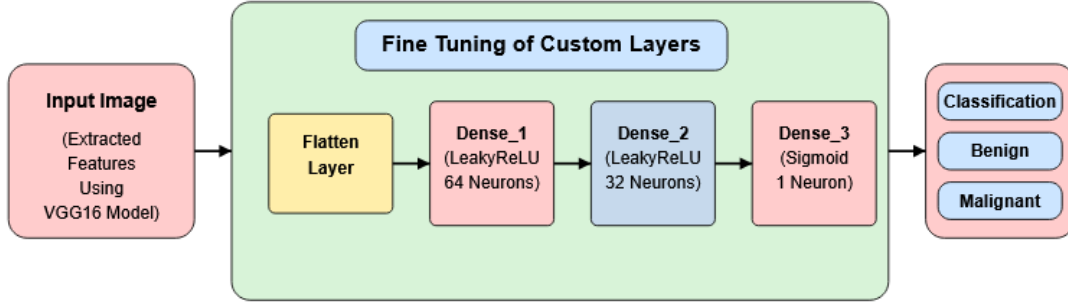


Figure 4.3: Fine Tuning of Proposed Transfer Learning Model

The extracted features from the VGG16 model are passed to a flatten layer, followed by two dense layers with 64 and 32 neurons, respectively, using the LeakyReLU activation function. Implemented batch normalization, dropout, and L2 regularization to stabilize training and prevent overfitting. The final dense layer consists of a single neuron with a sigmoid activation function. Based on these operations, the input image is classified as benign or malignant.

4.4 Subsystems and Workflow

The proposed system is composed of several subsystems, shown in Figures 4.4 to 4.7, each essential for the efficient operation and general performance of the model.

4.4.1 Data Preprocessing Subsystem

- **Input Size:** The images are resized to (224, 224, 3), compatible with VGG16.
- **TensorFlow Pipeline:** Auto-tuning is applied to optimize data loading for efficient processing.

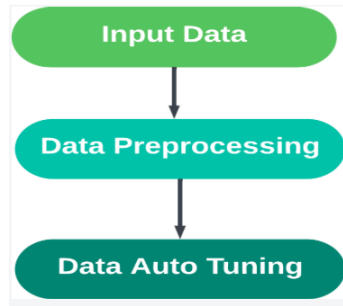


Figure 4.4: Data Preprocessing Subsystem

4.4.2 Feature Extraction Subsystem

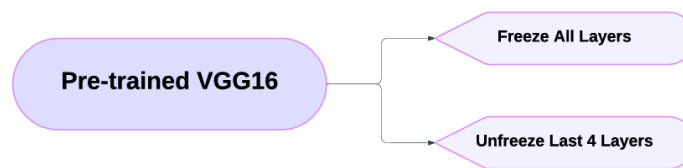


Figure 4.5: Feature Extraction Subsystem

- VGG16 Base Model: All layers are frozen except for the last four, which are fine-tuned to capture domain-specific features related to skin cancer.

4.4.3 Classification Subsystem

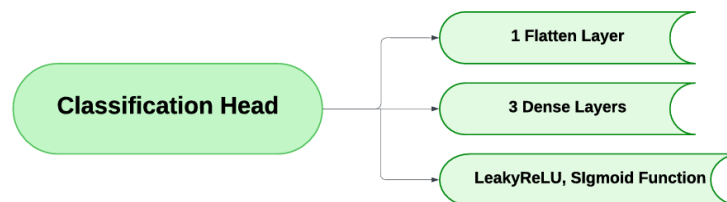


Figure 4.6: Classification Subsystem

- Flatten Layer: Converts 2D feature maps into a 1D vector for dense layers.

- Dense Layers: Three layers (64, 32, 1 unit) with LeakyReLU, BatchNormalization, and L2 regularization.
- Output Layer: Sigmoid activation for binary classification (benign or malignant).

4.4.4 Training and Optimization Subsystem

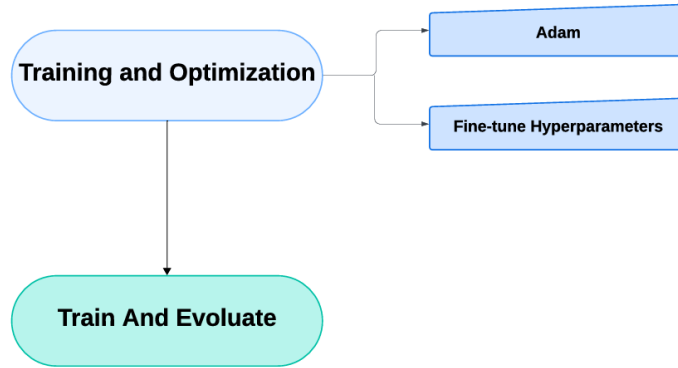


Figure 4.7: Training and Optimization Subsystem

- Optimizer: Adam optimizer with a fine-tuned learning rate.
- Loss function: Binary Cross Entropy to address the binary classification problem.
- Epochs and Batch Size: Set to 10 epochs (optimal) with a batch size of 32 (optimal) for efficient training.

4.5 Engineering Analysis

4.5.1 Formulations and Parameters

- Loss function: Binary Cross-Entropy

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N (y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))) \quad (4.1)$$

- Regularization: L2 regularization

$$R = \lambda \sum_{j=1}^n w_j^2 \quad (4.2)$$

- Learning Rate: Fine-tuned to ensure stable convergence.

4.5.2 Assumptions

- Balanced the data set across benign and malignant categories.
- Input images are resized to (224, 224, 3) without losing relevant features.

4.5.3 Design Viability

- Model Stability: Ensured stability by batch normalization and LeakyReLU activation function.
- Overfitting Prevention: Overfitting mitigated using Dropout and L2 Regularization.

4.6 Proposed System Architecture

The overall architecture of the proposed Skin Cancer Classification System is illustrated in Figure 4.8.

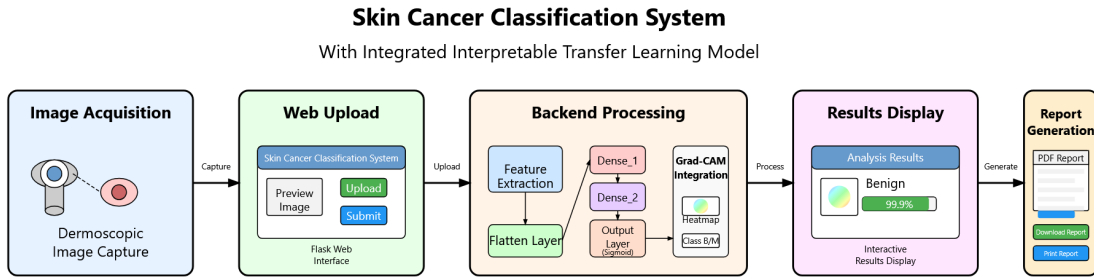


Figure 4.8: Proposed System Architecture

The system is structured into five major modules: Image Acquisition, Web Upload, Backend Processing, Results Display, and Report Generation. The pipeline

initiates with the Image Acquisition phase, where dermoscopic images are captured using a dermoscope or specialized imaging devices to ensure high-quality inputs. These captured images need to be uploaded to the Web Upload module, which provides a Flask-based web interface allowing users to submit images for analysis.

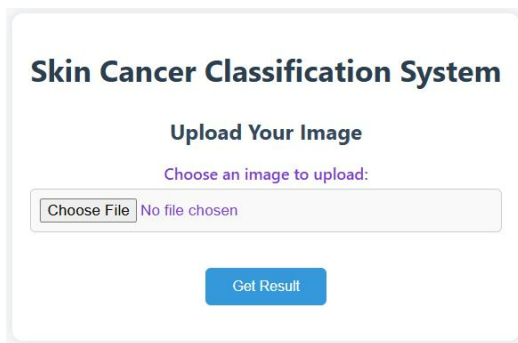
Upon submission, the images are processed in the Backend Processing module. Here, feature extraction is conducted, and the extracted features are passed through a customized classifier composed of a flatten layer, two dense layers with LeakyReLU activation function, and a final dense output layer utilizing a sigmoid activation function for binary classification (benign or malignant). In parallel, Grad-CAM (Gradient-weighted Class Activation Mapping) is integrated to generate interpretability heatmaps, enabling visual explanation of the model's decision.

Following back-end processing, the results, including the classification outcome, confidence score, uploaded image preview, and Grad-CAM heatmap for the input image, are presented in the Results Display module through an interactive interface made with Flask. Users can observe both the classification results and the associated visual interpretations, improving the transparency and reliability of the diagnostic decision.

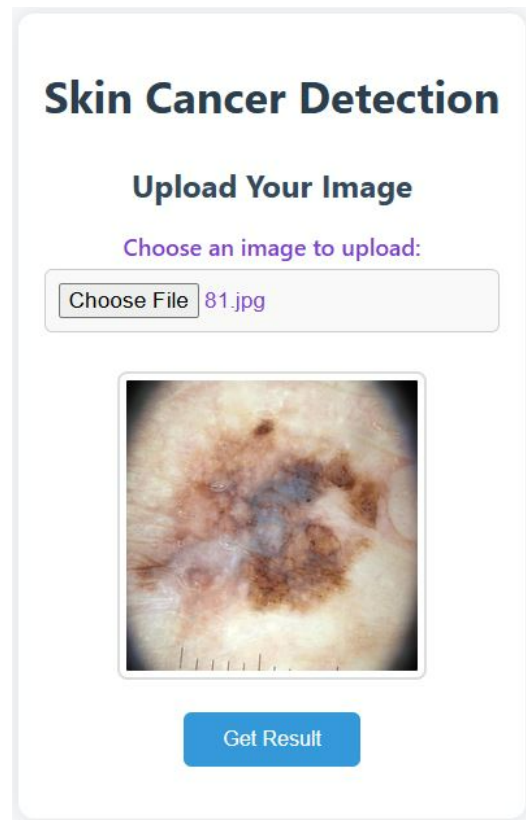
Finally, the Report Generation module facilitates the creation of comprehensive reports. Users are given options to download or print a PDF report summarizing the analysis results, supporting documentation for clinical decision-making or further reference. This systematic integration of the implemented interpretable and enhanced transfer learning model into an accessible web-based framework ensures that the proposed system is both user-friendly and clinically valuable, offering reliable diagnostic support for the preliminary screening of skin cancer.

4.7 Graphical User Interface (GUI) of the System

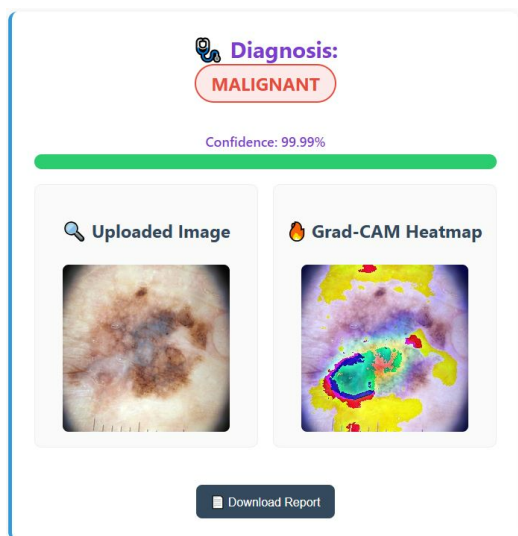
Figure 4.9 illustrates the graphical user interface (GUI) developed for the Skin Cancer Classification System, demonstrating the key stages of user interaction.



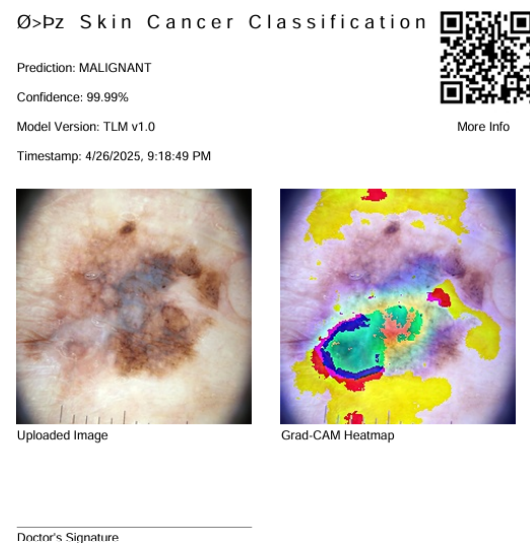
(a) Upload Image



(b) Preview Uploaded Image



(c) Analysis Results



(d) Downloaded Report

Figure 4.9: Graphical User Interface (GUI) of the System

The first subfigure, 4.9a, presents a clean and minimalistic interface where users can select and upload a skin image for analysis, featuring a simple "Choose File" button and a "Get Result" action button. The second subfigure, 4.9b, shows the selected image displayed within the interface, allowing users to visually confirm the upload before proceeding with the classification. The third subfigure, 4.9c, presents the diagnosis results, including the predicted class (e.g. benign or malignant), a confidence score visualized by a progress bar, and Grad-CAM heatmap visualizations highlighting the regions of interest within the lesion that influenced the model's decision. Finally, the fourth subfigure, 4.9d, displays the generated classification report, summarizing the diagnosis, confidence level, model version, timestamp, and including side-by-side images of the original upload and the Grad-CAM heatmap, along with a placeholder for a doctor's signature, supporting clinical verification. Together, these interface components ensure a seamless workflow from image upload to result interpretation and report generation, focusing on user-friendliness, transparency, and clinical usability.

Chapter 5

Result analysis

This chapter presents a comprehensive analysis of the experimental results obtained during the development of the proposed skin cancer classification system. It begins with hyperparameter tuning, detailing the optimization strategies employed to enhance model performance. Following this, the behavior of the model was examined by accuracy and loss analysis, covering the training, validation, and testing phases to assess the learning efficiency and generalization capability. A detailed comparison of confusion matrices and performance metrics with the baseline VGG16 model is provided to highlight the improvements achieved through the proposed methodological enhancements. Heatmap analysis using Grad-CAM is also conducted to visualize the regions that influence the model's predictions, contributing to the interpretability of results. Finally, the performance of the proposed model compared with existing state-of-the-art approaches to demonstrate its competitiveness and effectiveness in the classification of skin cancer.

5.1 Hyperparameter Tuning

Different parameters, such as optimizer, batch size, and number of epochs, were adjusted to optimize the performance of our transfer learning model. The Adam optimizer was chosen over the traditional Stochastic Gradient Descent (SGD) because of its superior performance in training transfer learning algorithms. Adam combines the characteristics of RMSProp and AdaGrad, dynamically adapting

the learning rate dynamically to improve convergence. The Adam optimizer is mathematically expressed as follows:

$$p_t = \alpha_1 p_{t-1} + (1 - \alpha_1) \left(\frac{\delta L}{\delta w_t} \right) \quad (5.1)$$

$$q_t = \alpha_2 q_{t-1} + (1 - \alpha_2) \left(\frac{\delta L}{\delta w_t} \right)^2 \quad (5.2)$$

In the Equations 5.1 & 5.2, α_1 and α_2 are the decay rates, δL is the loss function derivative, δw_t is the weights derivative at time t , δw_t signifies the weights, p_t is the collection of gradients, and q_t is the sum of squares of past gradients.

Batch size is the most critical hyperparameter that affects both convergence speed and generalization. A larger batch size improves computational speed due to parallelization, but may lead to suboptimal generalization. In contrast, smaller batch sizes improve convergence, but increase training time. To identify the optimal batch size, we have experimented with different batch sizes like 8, 16, 32, 64 and 128. The results indicated that a batch size of 32 provided the best balance between computational efficiency and accuracy.

The number of epochs also plays a crucial role in the performance of the model. Training for too few epochs can lead to underfitting, while excessive epochs can cause overfitting. We experimented with 8, 10, 16, and 20 epochs, ultimately selecting 10 epochs because it yielded the best accuracy while preventing overfitting.

5.2 Model Accuracy and Model Loss Analysis

This section analyzes the performance of the model by presenting graphs of training and validation accuracy, training and validation loss, as well as test accuracy and test loss, to evaluate the learning behavior and generalizability of the proposed model.

5.2.1 Training and Validation Accuracy

The graph 5.1 illustrates the training and validation accuracy of the implemented skin cancer classification model.



Figure 5.1: Training and Validation Accuracy

The training accuracy starts at approximately 0.80 and steadily increases with each epoch, reaching around 0.975 by the eighth epoch. Similarly, the validation accuracy begins near 0.80 and shows a consistent improvement, peaking at about 0.950. The close alignment between the two curves indicates effective learning. The results demonstrate the strong performance of the model in classifying skin cancer, with both the training and validation accuracy achieving high values at the end of the training process.

5.2.2 Training and Validation Loss

The graph 5.2 illustrates the loss of training and validation of the skin cancer classification model.



Figure 5.2: Training and Validation Loss

The training loss begins around 0.5 and decreases consistently with each epoch, dropping to approximately 0.2 by the eighth epoch. Similarly, the validation loss starts near 0.5 and follows a steady downward trend, reaching about 0.3. The parallel decrease in both curves indicates stable learning. These results validate the high accuracy of the figure 5.1, confirming the effectiveness of the model in minimizing errors during training while maintaining generalization of the validation data.

5.2.3 Test Accuracy and Test Loss

The graph 5.3 compares the test accuracy and test loss of the skin cancer classification model at a specific evaluation point.

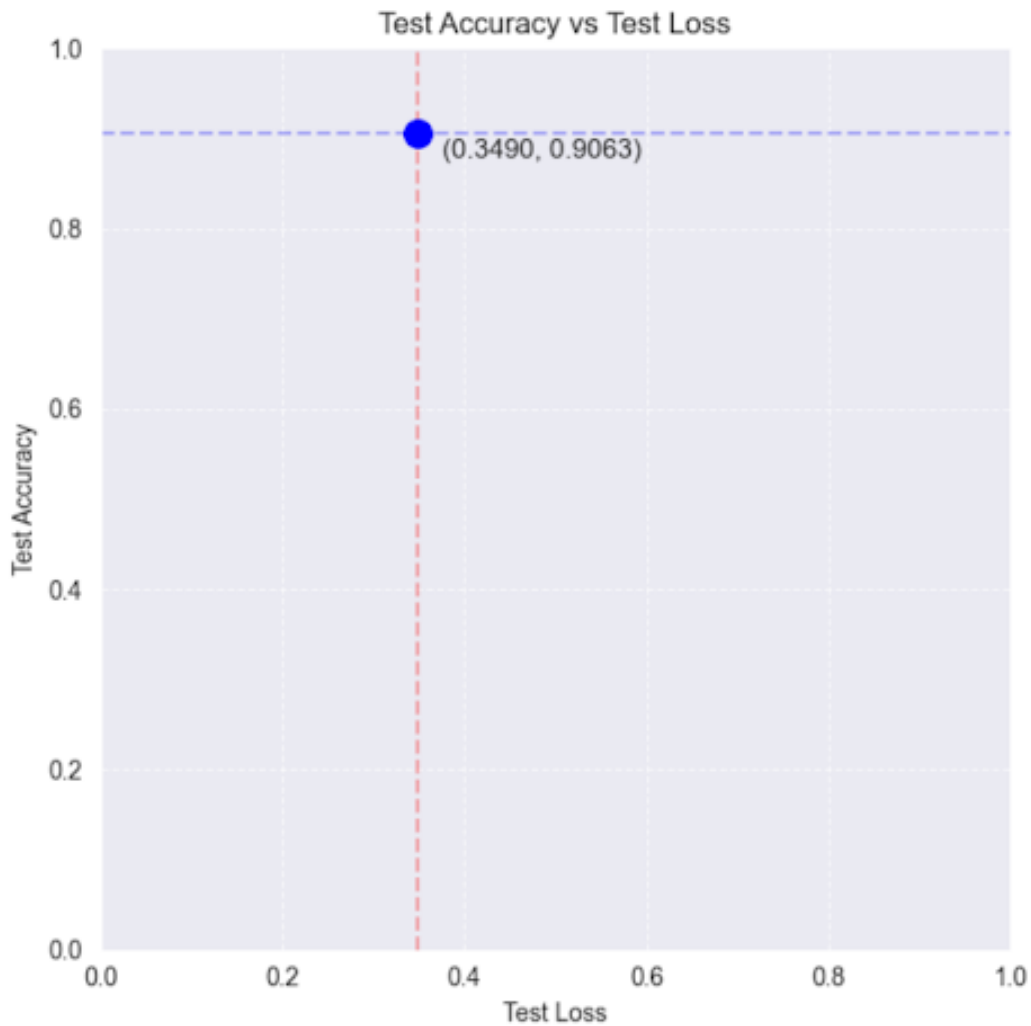


Figure 5.3: Test Accuracy and Test Loss

The test accuracy reaches 0.9063, indicating strong performance in classifying skin cancer images, while the test loss is 0.3490, reflecting relatively low prediction errors. The high accuracy paired with moderate loss suggests the model generalizes well to unseen data, maintaining robust classification capability without significant overfitting. These metrics align with the earlier training/validation results, confirming the model's reliability in real-world applications.

5.3 Comparison of Confusion Matrix and Performance Metrics with VGG16

A comparative analysis of confusion matrices and performance metrics between the implemented transfer learning model and the baseline standard VGG16 model was conducted in this research. Both models were trained and evaluated on the same dataset, 'Skin Cancer : Malignant vs. Benign', from Kaggle, which comprises 3,297 dermatoscopic images (1,680 malignant and 1,617 benign) sourced from the ISIC Archive [Fanconic \(2019\)](#). The dataset was divided into 2,637 training images and 660 test images, and both models were evaluated using the same test set to ensure stable and reliable performance evaluations. The hyperparameters for both models were fine-tuned in a consistent manner to ensure a fair and unbiased comparison.

5.3.1 Comparison of the Confusion Matrices

Figure 5.4 illustrates the confusion matrices for the baseline standard VGG16 model and the implemented transfer learning model, showing the distribution of predictions across the two classes.

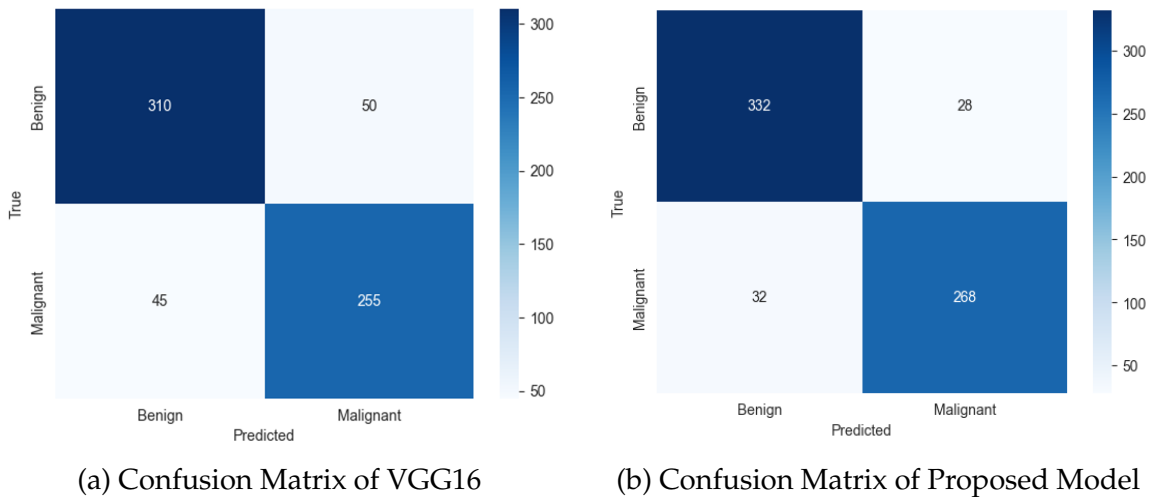


Figure 5.4: Comparison of Confusion Matrices between VGG16 and Proposed Model

The confusion matrix shown in Figure 5.4a represents the results for the baseline VGG16 model, where 310 benign samples and 255 malignant samples were correctly classified. However, 50 benign cases were misclassified as malignant, and 45 malignant cases were incorrectly predicted as benign. In contrast, the confusion matrix shown in Figure 5.4b corresponds to the implemented transfer learning model. It successfully classified 332 benign and 268 malignant samples, while only 28 benign and 32 malignant samples were misclassified.

Compared to the baseline VGG16 model, the implemented transfer learning model achieved a significant reduction in both types of misclassification. Most importantly, there was a substantial decrease in malignant cases incorrectly labeled benign (32 compared to 45), which is particularly important for the early detection and diagnosis of malignant skin cancer. Similarly, benign misclassifications decreased significantly (28 compared to 50), reducing the potential for unnecessary procedures and patient anxiety. These findings further validate the ability of the implemented model to generalize well on unseen data, offering a notable advancement over the standard VGG16 architecture.

5.3.2 Comparison of Performance Metrics

To quantitatively assess the performance differences between the two models, we calculated several key metrics as shown in Table 5.1. These metrics were derived from the confusion matrix parameters using Equations 5.3 to 5.10 [Anand et al. \(2022a\)](#), respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (5.3)$$

$$\text{Misclassification Rate} = 1 - \text{Accuracy} \quad (5.4)$$

$$\text{Sensitivity (Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5.5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5.6)$$

$$\text{False Negative Rate} = 1 - \text{Sensitivity} \quad (5.7)$$

$$\text{False Positive Rate} = 1 - \text{Specificity} \quad (5.8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5.9)$$

$$\text{F1 Score} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \quad (5.10)$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Comparison table of performance metrics

Metric	VGG16	Implemented Model	Difference
Accuracy	85.61%	90.63%	+5.02%
Misclassification Rate	14.39%	9.37%	-5.02%
Sensitivity (Recall)	85.00%	89.33%	+4.33%
Specificity	86.11%	92.23%	+6.12%
False Negative Rate	15.00%	10.67%	-4.33%
False Positive Rate	13.89%	7.77%	-6.12%
Precision	83.61%	90.54%	+6.93%
F1 Score	84.25%	89.93%	+5.68%

Table 5.1: Comparison of performance metrics between VGG16 and the Implemented Model.

The implemented model demonstrates a notable improvement in overall accuracy compared to the standard VGG16, with an increase of 5.02%. This represents a significant enhancement in binary classification performance, effectively reducing the misclassification rate from 14.39% to 9.37%. Meaningful improvements are observed in all key diagnostic metrics. The sensitivity (recall) increased from 85.00% to 89.33%, representing a 4.33% point improvement in the model's ability to correctly identify positive cases. Similarly impressive improvement in specificity

from 86.11% to 92.23%, a 6.12% point gain in correctly identifying negative cases.

These improvements result in notable reductions in error rates. The false negative rate decreased from 15.00% to 10.67%, an important enhancement for medical diagnostic settings where missing malignant cases can have serious consequences. Similarly, the false positive rate demonstrated a decrease from 13.89% to 7.77%, representing a reduction of 6.12% points in incorrectly identifying benign cases as malignant. The precision metric shows a substantial improvement from 83.61% to 90.54%, a 6.93% point increase in the proportion of correctly identified positive cases among all predicted positives. This is complemented by the F1 score, the harmonic mean of precision and recall, which rose from 84.25% to 89.93%, a 5.68% point gain, indicating a superior balance between precision and recall.

With an accuracy of 90.63% compared to VGG16's 85.61%, the implemented model demonstrates significantly improved overall performance. This 5.02% point increase in accuracy represents a substantial advancement in classification capability, demonstrating comprehensive improvements in all performance metrics that are critical for medical diagnostic applications.

5.4 Heatmap Analysis

To better understand the decision-making process of the model for the classification of malignant skin cancer, a heatmap analysis was performed using Grad-CAM (Gradient-weighted Class Activation Mapping) illustrated in Figure 5.5.

The first image 5.5a shows the original dermoscopic input image of skin cancer in malignant stage, which the implemented transfer learning model also classified as malignant. The second image 5.5b is the heat map of Grad-CAM, and indicates the region the model is focused on for its prediction. The red, blue and yellow hues indicate the region with an increasing level of importance, respectively, and infer that the model prioritized the region to varying extent in predicting the lesion as malignant.

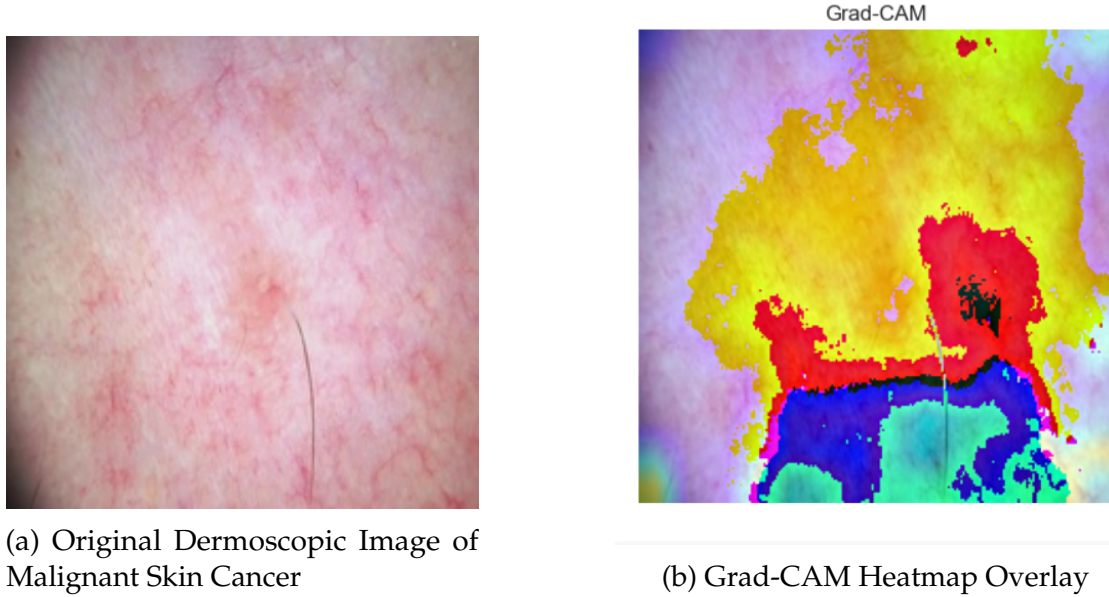


Figure 5.5: Heatmap Analysis

From the heatmap overlay, it can be observed that the model was concentrated on the unusual borders and pigmentation of the lesion, which are crucial determinants of malignancy in the clinic [Garbe et al. \(2022\)](#). The regions in the heatmap, marked in red, indicate a high consistency with dermatologic wisdom, solidifying the interpretability and validity of model classification. Beyond validating model attention to clinically important features, analysis improved the interpretability of the AI system, an important factor in the release of models for high-stakes use, such as the early diagnosis of skin cancer.

5.5 Comparison with the State-of-the-Art Approaches

Comparative analysis on the accuracy was performed with other state-of-the-art deep learning-based solutions for binary skin cancer classification in order to find the validity of the approach suggested here. All works discussed here employed the "Skin Cancer: Malignant versus Benign" dataset, comprising 3,297 dermoscopic images, from the Kaggle database, originally downloaded from the International Skin Imaging Collaboration (ISIC) archive [Fanconic \(2019\)](#). This consistency in the choice of the data set makes a proper comparison between the model solutions

feasible. The comparative data are shown in Table 5.2, the evaluation represented in graphical format.

For example, [Anand et al. \(2022a\)](#) used a modified VGG16 architecture enhanced with additional dense layers and LeakyReLU activation, reporting an accuracy of 89.09%. Similarly, [Alrabai et al. \(2025\)](#) evaluated multiple models, including ResNet50, VGG16, and Vision Transformers (ViT-b16, ViT-b32), achieving the highest accuracy of 89.09% with ResNet50. [Agarwal & Singh \(2022\)](#) utilized multiple CNN architectures with transfer learning techniques and achieved an accuracy of 86.65%, while [Bello et al. \(2024\)](#) fine-tuned DenseNet-121 and achieved 87%. Other approaches, such as [Kalouche \(2020\)](#), who explored logistic regression, deep neural networks and a fine-tuned VGG16 model, and [Ibrahim et al. \(2023\)](#), who applied transfer learning with a fine-tuned VGG16 architecture, reported lower accuracies of 78% and 84.24%, respectively. In contrast, [Hasan et al. \(2019\)](#) implemented a traditional CNN-based classification system and achieved 89.5%, while [Aydin & Yildiz \(2023\)](#) combined local histogram-based descriptors with an XGBoost classifier to reach 90%. [Acosta et al. \(2021\)](#) used a two-stage framework that involved lesion segmentation with Mask R-CNN followed by classification using ResNet152, resulting in an accuracy of 90.4%.

The proposed model used the last four layers in the VGG16 for features, along with its dense classifier, and achieved an accuracy level of 90.63%, improving the performance of many leading state-of-the-art methods. The findings justify the consistency and strength of the model used, providing evidence for the model's ability in delivering an accurate and interpretative skin cancer classification.

Papers	Dataset	Neural Network Models	Descriptions	Accuracy (%)
Anand et al. (2022a)	Kaggle (ISIC)	Modified VGG16	Enhanced transfer learning with additional dense layers and LeakyReLU activation	89.09
Agarwal & Singh (2022)	Kaggle (ISIC)	CNN with Transfer Learning	Utilized multiple CNN architectures with transfer learning techniques	86.65
Kalouche (2020)	Kaggle (ISIC)	Logistic Regression, Deep Neural Networks, Fine-tuned VGG16	Melanoma classification using various models, with fine-tuned VGG16, achieved the highest accuracy	78.00
Ibrahim et al. (2023)	Kaggle (ISIC)	Fine-tuned VGG16	Applied transfer learning with fine-tuned VGG16 architecture	84.24
Bello et al. (2024)	Kaggle (ISIC)	DenseNet-121	Fine-tuned DenseNet-121 with additional dense layers and LeakyReLU activation	87.00
Arabai et al. (2025)	Kaggle (ISIC)	ResNet50, VGG16, ViT-b16, ViT-b32	Evaluated multiple pre-trained models with XAI techniques, ResNet50 achieved the highest accuracy	89.09
Bazgir et al. (2024)	Kaggle (ISIC)	InceptionNet	Implemented and fine-tuned InceptionNet with image augmentation	85.94
Acosta et al. (2021)	Kaggle (ISIC)	Mask R-CNN, ResNet152	Two-stage framework with lesion segmentation and classification	90.40
Hasan et al. (2019)	Kaggle (ISIC)	CNN	Image processing and CNN-based classification system	89.50
Aydin & Yildiz (2023)	Kaggle (ISIC)	LDN with XGBoost	Histogram-based local descriptors with XGBoost classifier	90.00
Implemented Model in this research	Kaggle (ISIC)	VGG16 (last 4 layers) + Custom Dense Classifier	Interpretable model using high-level VGG16 features and custom classifier head with LeakyReLU and sigmoid activation	90.63

Table 5.2: Comparative result analysis of relative works

Chapter 6

Conclusions

This chapter provides the overview of the issues that can arise because of the real time implementation of the system. In addition, this chapter also includes the final outcomes, future works and a comprehensive conclusion of the research.

6.1 Social, Legal, Ethical, and Environmental Issues

In this research, we have addressed the issues that can arise including social, legal, ethical, and environmental issues. In the subsections below, these issues and solution for them is discussed in details.

6.1.1 Social Issues

The suggested system design and implementation would greatly improve the initial screening for skin cancer. Application in real-time, however, is flawed in the sense that the system might be inexact and prone to misdiagnosis. Misclassifications may occur if the model is not tuned optimally. Either false positives may result in unnecessary treatment, while false negatives may result in failure to diagnose, with adverse impact on patient health. In this research, both the model and the system have been optimized to minimize the risk of biased outcomes, and explainable artificial intelligence techniques have been incorporated to support informed decision-making. Furthermore, while technology has the potential to improve health outcomes, there is also the risk of deepening the digital divide. To mitigate

this, we have ensured that the application is designed to be accessible to all, maximizing its social impact.

6.1.2 Legal Issues

The deployment of the system as a mobile application in Bangladesh raises several legal and regulatory considerations. Healthcare technologies, especially those used for medical diagnostics, must meet stringent safety and efficacy standards. Since this research developed a web-based application for the medical domain, the system is under the supervision of the Directorate General of Drug Administration (DGDA) and the Bangladesh Medical and Dental Council (BMDC). For the system to be considered reliable for medical use, the model must undergo rigorous testing, validation, and potentially approval from relevant authorities prior to clinical deployment. In addition, privacy and data security are paramount when handling personal health information (PHI) through the web application. Research ensures compliance with existing frameworks, such as the Digital Security Act (DSA) and the upcoming Data Protection Act, by implementing robust data anonymization, encryption, and secure storage measures, while obtaining explicit user consent for data collection and usage. The consequences of failure to follow legal and ethical standards can be dangerous, so data security and safety ensured in this research

6.1.3 Ethical Issues

The ethical concerns of the skin cancer classification system in the health care system in Bangladesh involves several ethical issues. One of the most crucial among them is transparency, healthcare professionals and the patients must understand how the AI model generates its predictions. For interpretability, Grad-CAM have been implemented in the model, providing the user with insight into why the model labeled the tumor as benign or malignant. Transparency is necessary in order for trust in the technology to be built, and it will allow healthcare professionals to integrate it into clinical decision-making with confidence. Informed consent is an important ethical consideration. Users of the web application fully understand how their data will be collected, stored, and used. If the data is being used for model training for further

model building or for the purposes of research, clear consent will be obtained. Upward preservation of user privacy and data and user autonomy is consistent with ethical AI practice as well as compliance needs. In addition, the training data set used in the work here is heterogeneous, and it will continue to expand in size and diversity in order to perform well in all demographics. In curating the data set with care and constantly monitoring the model's performance, the work has constructed an AI tool that is equitable, inclusive, and beneficial for all consumers in Bangladesh.

6.1.4 Environmental Issues

The environmental impact of the proposed system is an important factor to address. Deep learning algorithms for image classification, are computationally intensive, power consuming tasks. This research has attempted to make the system efficient in order for it to be able to function on ease of use and use little power. Such optimization improves accessibility and minimizes the environmental impact of handheld devices' usage for prolonged durations. In addition, high smartphone and electronic devices usage generates electronic waste (e-waste) since the user always purchases new devices. Therefore, the system has been optimized for low usage, and green technology habits have been promoted. It incorporates proper electronic waste recycling and ensuring the application runs in low-processing, older devices, thereby avoiding the need for frequent replacement of the hardware. With energy-saving AI as its major objective and proper usage of technology, the work aims at limiting its environmental impact while providing the application for AI-based skin cancer detection for everyone in Bangladesh.

6.2 Project Outcomes

A binary classification model for skin cancer as benign or malignant is successfully developed based on an optimized VGG16-based transfer learning. The model is very accurate in classification while being light in computation and is equipped with Grad-CAM for interpretability for clinical decision support. The deliverable

of this research includes an optimized VGG16 based transfer learning model integrated with Grad-CAM. User-friendly web application with a simple graphical user interface was implemented, incorporating the optimized model for medical uses. The outcomes indicate the direction towards the potential for skin cancer classification in preliminary stages, supporting dermatologists in diagnosis, and contributing towards AI-based medical diagnostics.

6.3 Future Work

An enhanced skin cancer classification system has been designed and implemented in this research, although there are still scope for improvement. Two main areas for improvement are identified. First, incorporating a dermoscope based IoT system directly into the skin cancer classification system pipeline to streamline the process by enabling real-time scanning and analysis of skin cancer stages. This will terminate the need for users to manually capture and upload images that improves the efficiency and convenience. And second, although the current model demonstrates higher classification accuracy compared to the state-of-the-art, there is still scope for improvement.

6.4 Conclusion

This research presents an interpretable and enhanced VGG16-based transfer learning model for binary classification of malignant and benign skin lesions, the aim being accuracy and interpretability. Leveraging the transfer learning process with VGG16 and fine-tuning the new and last four layers, the model presented here provides encouraging output in the binary classification of malignant and benign skin lesions. Utilizing the explainability in AI (XAI) methods, such as, Grad-CAM, the presented model has been able to provide meaningful explanation to the model decision, providing it with enhanced interpretability and confidence in its prediction. The performance of the model has been validated on several metrics, ensuring

significant improvement compared to state-of-the-art. The approach shows efficient in utilizing the pre-trained weights of VGG16, ensuring successful skin image feature extraction while improving the model performance. In addition, the integration of the model through an easily usable web interface made with Flux, provides an executable solution for the practitioners, making the output not only accurate, but comprehensible too. Overall, the paper makes an effort towards the design of automated binary classification models for skin cancer, with the potential to greatly help in the prognosis and early detection of skin cancer.

References

- Acosta, M. F., Caballero Tovar, L. Y., Garcia-Zapirain, M. B., & Percybrooks, W. S. (2021). Melanoma diagnosis using deep learning techniques on dermatoscopic images. *BMC Medical Imaging*, 21, 6.
- Agarwal, R., & Singh, N. (2022). Convolutional neural networks for skin cancer classification: A comparative study. *Journal of Biomedical Informatics*, 130, 104045.
- Akter, M. S., Shahriar, H., Sneha, S., & Cuzzocrea, A. (2023). *Multi-class skin cancer classification architecture based on deep convolutional neural network*. <https://doi.org/10.48550/arXiv.2303.07520>. (arXiv preprint arXiv:2303.07520)
- Alrabai, A., Ehtioui, A., & Kallel, F. (2025). Exploring pre-trained models for skin cancer classification. *Applied Sciences and Innovations*, 8(2), 35. doi: 10.3390/asi8020035
- American Academy of Dermatology. (2023). *Skin cancer: Facts and statistics*. <https://www.aad.org/media/stats-skin-cancer>.
- Anand, Bhatnagar, R., & Arora, A. (2022a). An enhanced transfer learning based classification for diagnosis of skin cancer. *IEEE Access*, 10, 13579–13592. doi: 10.1109/ACCESS.2022.3147852
- Anand, Bhatnagar, R., & Arora, A. (2022b). Skin cancer detection using deep learning and image processing. *Journal of King Saud University - Computer and Information Sciences*. doi: 10.1016/j.jksuci.2022.05.002
- Aydin, & Yildiz. (2023). A comparative analysis of skin cancer detection applications using histogram-based local descriptors. *Diagnostics*, 13(19), 3142. doi: 10.3390/diagnostics13193142
- Bazgir, E., Haque, E., Maniruzzaman, M., & Hoque, R. (2024). Skin cancer classification using inception network. *World Journal of Advanced Research and Reviews*, 21(02), 839–849. doi: 10.30574/wjarr.2024.21.2.0500
- Bello, A., Ng, S.-C., & Leung, M.-F. (2024). Skin cancer classification using fine-tuned transfer learning of densenet-121. *Applied Sciences*, 14(17), 7707. doi: 10.3390/app14177707

- Blundo, A., Cignoni, A., Banfi, T., & Ciuti, G. (2021). Comparative analysis of diagnostic techniques for melanoma detection: A systematic review of diagnostic test accuracy studies and meta-analysis. *Frontiers in Medicine*, 8, 637069.
- Brinker, T. J., Hekler, A., Enk, A. H., & Berking, C. (2019). Skin cancer classification using convolutional neural networks: Systematic review. *Journal of Medical Internet Research*, 21(10), e11936. doi: 10.2196/11936
- Cancer Research UK. (2024a). *How does the sun and uv cause cancer?* <https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/sun-uv-and-cancer/how-does-the-sun-and-uv-cause-cancer>.
- Cancer Research UK. (2024b). *Melanoma skin cancer statistics*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer>.
- Chen, Y., Roberts, T., Lin, M., & Xu, J. (2023). Transfer learning for medical imaging: Challenges and solutions. *Pattern Recognition Letters*, 155, 45–53.
- Davis, C., Shalin, S. C., & Tackett, A. J. (2019). Current state of melanoma diagnosis and treatment. *Cancer Biology & Therapy*, 20(11), 1366–1379. doi: 10.1080/15384047.2019.1640032
- Dildar, M., Akbar, S., Mehmood, F., Razzak, I., Gilani, S. O., & Yousaf, M. H. (2021). Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health*, 18(10), 5479. doi: 10.3390/ijerph18105479
- Djaroudib, A., et al. (2024). Transfer learning with vgg-16 for basal cell carcinoma and melanoma diagnosis. *Neural Networks*, 164, 117–128.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. doi: 10.1038/nature21056
- Fanconic. (2019). *Skin cancer: Malignant vs. benign*. <https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign>. Kaggle. (Accessed: 2025-05-12)
- Garbe, C., Amaral, T., Peris, K., Hauschild, A., Arenberger, P., Basset-Seguín, N., ... others (2022). European consensus-based interdisciplinary guideline for melanoma. part 1: Diagnostics: Update 2022. *European journal of cancer*, 170, 236–255.
- Gupta, S., Chaturvedi, A. Y., & Sharma, P. (2023). Skin cancer: Causes, prevention, and treatment. *International Journal of Dermatology Research*, 56(4), 201–212.

- Hara, K., Fukui, K., & Takahashi, M. (2022). A review of overfitting issues in deep learning for medical imaging. *Computational Biology and Medicine*, 146, 105655. doi: 10.1016/j.compbimed.2022.105655
- Hasan, Kabir, A., Rahman, M. M., & Hossain, M. S. (2023). Skin cancer detection using hybrid deep learning model. *Computers in Biology and Medicine*, 161, 106867. doi: 10.1016/j.compbimed.2023.106867
- Hasan, Mahamudul, S. D., Barman, Islam, S., & Reza, A. W. (2019). Skin cancer detection using convolutional neural network. In *Proceedings of the 2019 5th international conference on computing and artificial intelligence (iccai)* (pp. 254–258). doi: 10.1145/3330482.3330525
- Ibrahim, A., et al. (2023). Transfer learning with vgg16 for malignant and benign skin lesion classification. *Computers in Biology and Medicine*, 155, 106589.
- International Agency for Research on Cancer (IARC). (2022). *Skin cancer*. <https://www.iarc.who.int/cancer-type/skin-cancer/>.
- Kalouche, R. (2020). Deep learning for melanoma classification: A study on the isic dataset. *Artificial Intelligence in Medicine*, 109, 101962.
- Kanchana, K., Kavitha, S., Anoop, K. J., & Chinthamani, B. (2024). Optimizing skin cancer diagnosis with efficientnet b0-b7 and patient-specific transfer learning. *Asian Pacific Journal of Cancer Prevention*, 25(5), 1795–1802. doi: 10.31557/APJCP.2024.25.5.1795
- Khan, et al. (2025). Skin cancer detection using transfer learning on low-resolution images. *Biomedical Signal Processing and Control*, 80, 104072.
- Lee, M. K., Kim, J., Patel, R., & Wang, C. (2023). Artificial intelligence in dermatology: Recent advances and future directions. *Nature Medicine*, 29(5), 678–690.
- Mahmud, F., Mahfiz, M. M., Kabir, M. Z. I., & Abdullah, Y. (2023). Interpretable deep learning approach for skin cancer classification using explainable ai. *IEEE*, 26, 1–6.
- Mahmud, F., et al. (2023). Interpretable deep learning with xai for skin cancer classification. *Pattern Recognition*, 135, 108932.
- Mazhar, T., Haq, I., Ditta, A., Mohsan, S. A. H., Rehman, F., Zafar, I., ... Goh, L. P. W. (2023). The role of machine learning and deep learning approaches for the detection of skin cancer. *Healthcare*, 11(3), 415. doi: 10.3390/healthcare11030415
- Melarkode, N., Srinivasan, K., Qaisar, S. M., & Plawiak, P. (2023). Ai-powered diagnosis of skin cancer: A contemporary review, open challenges, and future research directions. *Cancers*, 15(4), 1183. doi: 10.3390/cancers15041183

- Naqvi, S. M., Rehman, A., Khan, M. A., & Saba, T. (2023). A deep learning approach for skin cancer classification using transfer learning. *IEEE Access*, 11, 11234–11245. doi: 10.1109/ACCESS.2023.3244956
- National Cancer Institute. (2023). *Melanoma survival rates*. <https://seer.cancer.gov/statistics/>. National Cancer Institute.
- Patel, K., & Wong, T. J. (2023). Performance of cnn architectures in medical image classification. *Journal of Machine Learning Research*, 21(4), 134–150.
- Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), 17–48. doi: 10.3322/caac.21763
- Singh, P., & Patel, H. (2021). Understanding the interpretability of deep learning models in medical diagnostics. *Sustainability*, 15(7), 5930. doi: 10.3390/su15075930
- Skin Cancer Foundation. (2023). *Skin cancer facts & statistics*. <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>.
- Skin Cancer Foundation. (2024). *Uv radiation & your skin*. <https://www.skincancer.org/risk-factors/uv-radiation/>.
- Skin Cancer Foundation. (2025). *Skin cancer: An overview — melanoma, basal cell, and squamous cell carcinoma*. <https://www.skincancer.org/skin-cancer-information/>.
- Soumyya, S., & et al. (2024). Skinnet-14: A deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time. *Neural Computing and Applications*, 36, 18935–18959. doi: 10.1007/s00521-024-10225-y
- Tjoa, E., & Guan, C. (2022). Explainable artificial intelligence in skin cancer recognition: A systematic review. *Artificial Intelligence in Medicine*, 127, 102153. doi: 10.1016/j.artmed.2022.102153
- Unnisa, Z., Tariq, A., Sarwar, N., Din, I., Serhani, M. A., & Trabelsi, Z. (2025). Impact of fine-tuning parameters of convolutional neural network for skin cancer detection. *Scientific Reports*, 15, 14779. doi: 10.1038/s41598-025-14779-3
- World Health Organization. (2022). *Skin cancers: Key facts*. <https://www.who.int/news-room/fact-sheets/detail/skin-cancers>. World Health Organization.
- Wu, Y., Chen, Y., Zeng, Y., Pan, Y., Wang, Y., & Zhao, Y. (2022). Skin cancer classification with deep learning: A systematic review. *Frontiers in Oncology*, 12, 932773. doi: 10.3389/fonc.2022.932773

- Yadav, P., & Tiwari, M. (2020). Challenges in machine learning for medical image analysis: Overcoming data scarcity. *International Journal of Health Sciences*, 14(6), 502–507. doi: 10.2147/IJHS.S284016
- Zhang, Z., & Li, J. (2019). Challenges and solutions in real-time deployment of deep learning in medical diagnosis. In *Springer handbook of medical imaging* (Vol. 2, pp. 1035–1045). Springer. doi: 10.1007/978-3-030-37188-3_2