**SpotCancerAI**

|  |
| --- |
|  |

**By:**

**Muhammad Usman**

**24761**

**Ali Sher Khan**

**39917**

**Hassan Dastagir**

**40124**

**Supervised by:**

**Mr. Hafiz Haseeb Tasleem**

**Faculty of Computing**

**Riphah International University, Islamabad**

**Spring/Fall 2025**

**A Dissertation Submitted To**

**Faculty of Computing,**

**Riphah International University, Islamabad**

**As a Partial Fulfillment of the Requirement for the Award of the Degree of**

**Bachelors of Science in Computer Science**

**Faculty of Computing**

**Riphah International University, Islamabad**

Date: [date of final presentation]

**Final Approval**

This is to certify that we have read the report submitted by ***Muhammad Usman(24761), Ali Sher Khan (39917), Hassan Dastagir (40124)***for the partial fulfillment of the requirements for the degree of the Bachelors of Science in Computer Science (BSCS). It is our judgment that this report is of sufficient standard to warrant its acceptance by Riphah International University, Islamabad for the degree of Bachelors of Science in Computer Science (BSCS).

**Committee:**

|  |  |
| --- | --- |
| **1** | Mr. Hafiz Haseeb Tasleem  (Supervisor) |
|  |  |
| **2** | Dr. Muhammad Musharraf  (Head of Department) |

**Declaration**

We hereby declare that this document “**SpotCancerAI**” neither as a whole nor as a part has been copied out from any source. It is further declared that we have done this project with the accompanied report entirely on the basis of our personal efforts, under the proficient guidance of our teachers especially our supervisor **Mr. Hafiz Haseeb Tasleem**. If any part of the system is proved to be copied out from any source or found to be reproduction of any project from anywhere else, we shall stand by the consequences.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Muhammad Usman**

**24761**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Ali Sher Khan**

**39917**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Hassan Dastagir**

**40124**

**Dedication**

Insert dedication Our final year project is dedicated to our parents, friends and teachers, whose love and support have been our pillars of strength. To our professors and especially supervisor "**Mr. Hafiz Haseeb Tasleem**", your guidance has shaped our academic journey.

**Acknowledgement**

First of all we are obliged to Allah Almighty the Merciful, the Beneficent and the source of all Knowledge, for granting us the courage and knowledge to complete this Project.

We extend our heartfelt gratitude to our project supervisor “**Mr. Hafiz Haseeb Tasleem”**, whose unwavering support, invaluable guidance, and continuous mentorship were indispensable to the successful completion of this project. Their dedication and commitment have been a driving force behind our work.

Furthermore, we want to say a big thank you to our family and friends. They have been our constant source of support and motivation, always encouraging us to do our best and be honest and hardworking.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Muhammad Usman**

**24761**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Ali Sher Khan**

**39917**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Hassan Dastagir**

**40124**

Table of Contents

[List of Tables vii](#_Toc195995175)

[Abstract 1](#_Toc195995176)

[Chapter 1: Introduction 2](#_Toc195995177)

[1.1 Introduction 3](#_Toc195995178)

1. [1.1.2 Opportunities 3](#_Toc195995179)
2. [1.1.3 Motivation 3](#_Toc195995180)
3. [1.1.4 Challenges 4](#_Toc195995181)

[1.2 Goals and Objectives 4](#_Toc195995182)

1. [1.2.1 Goals 4](#_Toc195995183)
2. [1.2.2 Objectives 4](#_Toc195995184)

[1.3 Scope of the Project 5](#_Toc195995185)

1. [1.3.1 Project Objectives 5](#_Toc195995186)
2. [1.3.2 Technological Components 6](#_Toc195995187)
3. [1.3.3 Implementation Phases 6](#_Toc195995188)
4. [1.3.4 Data Management 7](#_Toc195995189)
5. [1.3.5 Stakeholder Engagement 8](#_Toc195995190)
6. [1.3.6 Deliverable 8](#_Toc195995191)

[Chapter 2: Literature Review 9](#_Toc195995192)

[2.1 Literature Review 10](#_Toc195995193)

# 

List of Tables

[Table 1: Literature Review **Error! Bookmark not defined.**](#_Toc165128403)

[Table 2: Market Survey **Error! Bookmark not defined.**](#_Toc165128404)

[Table 3: This is Sample table caption **Error! Bookmark not defined.**](#_Toc165128405)

[Table 4: This is Sample table caption **Error! Bookmark not defined.**](#_Toc165128406)

**List of figures**

[Figure 1: Architecture Diagram **Error! Bookmark not defined.**](#_Toc165132973)

[Figure 2: Farmer Use Case **Error! Bookmark not defined.**](#_Toc165132974)

[Figure 3: Admin Use Case **Error! Bookmark not defined.**](#_Toc165132975)

[Figure 4: Database Design **Error! Bookmark not defined.**](#_Toc165132976)

[Figure 5: Methodology Diagram **Error! Bookmark not defined.**](#_Toc165132977)

[Figure 6: List of Styles **Error! Bookmark not defined.**](file:///C:\Users\DELL\Desktop\Tomato%20Care_CS_Mr.Muhammad%20Usman%20Karim.docx#_Toc165132978)

[Figure 7: IEEE Reference style **Error! Bookmark not defined.**](#_Toc165132979)

## Abstract

Skin cancer is one of the most common and dangerous cancer in worldwide, but early detection can improve treatment outcomes. SpotCancerAI is a deep learning-based project designed to help identify skin cancer from dermoscopic images using the HAM10000 dataset. This project focuses on building an application that preprocesses medical images, segments lesions, and classifies them into different types of skin cancers. By combining image processing techniques like grayscale conversion, Gausian Blur, and inpainting with modern machine learning models, SpotCancerAI aims to provide an accurate and efficient tool for early diagnosis. The system is intended to support dermatologists and increase accessibility to skin cancer screening, especially in areas with limited medical resources.

# 

# 

# Chapter 1: Introduction

## 1.1 Introduction

SpotCancerAI is an inventive project that uses machine learning techniques to detect skin cancer from images of skin lesions. The goal is to improve early diagnosis and provide a reliable tool for healthcare professionals. By analyzing a large dataset (HAM10000) of dermatological images, SpotCancerAI focuses on accurately classifying and segmenting lesions to determine whether they are benign (non-cancerous) or malignant (cancerous). The project grips on advanced image processing methods, including grayscale conversion, gausian blur, and inpainting, to increase the quality of the images before applying machine learning algorithms. Finally, SpotCancerAI aims to assist in the early detection of skin cancer, potentially saving lives by enabling quicker and more correct diseases.

### 1.1.2 Opportunities

* **Early Detection of Skin Cancer**  
  SpotCancerAI can identify skin cancer at an early stage, which is essential for increasing survival rates. Early detection often leads to simpler and more successful medical care.
* **Support for Healthcare Professionals**  
  The system can act as a determination-support tool for dermatologists and experts by highlighting doubtful lesions, reducing human error, and improving diagnostic correctness.
* **Improved Access in Underserved Areas**  
  In regions with limited approach to skin doctors or specialized care, SpotCancerAI could be integrated into mobile or telemedicine platforms, helping people receive initial evaluations without needing to travel.
* **Scalability and Speed**  
  Unlike standard diagnosis methods, machine learning systems like SpotCancerAI can process thousands of images quickly, making them highly flexible for hospitals and clinics handling large number of patients.
* **Educational Tool**  
  SpotCancerAI can also have a work as an educational support for medical students and trainees, offering a practical understanding of how skin wound are classified and identified using AI.
* **Cost-Effective Screening**  
  Computer screening with SpotCancerAI could lower medical care costs by reducing the need for unnecessary biopsies and in-person consultations when wounds are found to be benign.
* **Continuous Improvement with Data**  
  The model can be continually improved and retrained with more diverse and updated datasets, leading to better performance over time, especially across different skin tones and lesion types.

### 1.1.3 Motivation

The motivation at the back of SpotCancerAI project lies in the serious need for early and correct detection of skin cancer, particularly melanoma, which can be life-threatening if not diagnosed in time. Traditional diagnostic methods often depend on expert dermatological evaluation, which can be subjective and limited by availability, especially in neglected regions. SpotCancerAI aims to make use the power of artificial intelligence and computer screening to create an accessible, reliable, and efficient tool for skin lesion examination. By computerized screening the detection process using advanced image processing and deep learning techniques, the project seeks to support medical professionals, reduce diagnostic errors, and ultimately improve patient outcomes through faster and more compatible identification of possibly cancerous skin lesions.

### 1.1.4 Challenges

The SpotCancerAI project faces some challenges that impact its development and successfulness. One major challenge is **data quality and diversity**—skin wound datasets may lack presentation across different skin tones, age groups, and rare cancer types, which can lead to biased or less accurate models. Another difficulty is the **complexity of medical image processing**, as skin wounds can vary greatly in appearance due to lighting, image resolution, and surrounding skin features. **Segmentation of wound** is particularly difficult, requiring precise isolation of the region of interest, which is critical for accurate classification. Additionally, **model understandability and clinical validation** are essential, as medical professionals need to trust and understand AI-driven decisions before adopting them in practice. Finally, **regulatory and ethical concerns** around patient data privacy and the deployment of AI in healthcare must be carefully managed to make sure safe and responsible use of the system.

## 1.2 Goals and Objectives

### 1.2.1 Goals

The Goals of SpotCancerAI are as following :-

* Detect skin cancer using Machine learning and deep learning models.
* Classify different types of skin wounds from images.
* Preprocess images (grayscale, gausian blur, inpainting) for clarity and accuracy.
* Segment lesion areas to isolate them from background skin.
* Support early and correct diagnosis for dermatologists.
* Improve and contribute to Computerized screening or Application in healthcare.
* Share intelligence and tools with the research and developer community.

### 1.2.2 Objectives

* To use the **HAM10000** dataset for training and testing skin diagnosis detection models.
* To clean and enhance the images using preprocessing methods like grayscale conversion, gausian blur, and inpainting.
* To correctly separate (segment) the skin wounds from the rest of the image.
* To train deep learning models that can categorized different types of skin lesions.
* To estimate the model’s performance using accuracy, precision, recall, and F1-score.
* To improve the model results by tuning its hyperparameters.
* To build a complete system that goes from image input to final result.
* To support early detection of skin cancer and help in use of medical field.

## 1.3 Scope of the Project

The Scope of the Project SpotCancerAI are as following :-

* **AI-Based Skin Cancer Detection:** Uses deep learning to classify skin lesions.
* **Fast & Accurate Results:** Provides quick analysis to support medical decisions.
* **User-Friendly Interface:** Simple and easy-to-use system for both doctors and patients.
* **Data Security & Privacy:** Ensures patient information is kept safe.
* **Mobile & Web Compatibility:** Can be used on smartphones and computers.

### 1.3.1 Project Objectives

* To develop an AI-based system for the early detection of skin wound using dermoscopic images.
* To apply preprocessing techniques such as grayscale conversion, gausian blur, and inpainting for improving image quality.
* To perform correct segmentation of skin wounds from background skin to focus on relevant areas.
* To classify skin wounds into different categories using deep learning models.
* To estimate the performance of the model using standard metrics like accuracy, precision, recall, and F1-score.
* To optimize model performance through setting a hyperparameters.
* To create a complete, end-to-end pipeline from image input to final categorical output.
* To provide a knowledge in medical AI research and support early and efficient disease of skin cancer.

### 1.3.2 Technological Components

* Dataset:

1. HAM10000 – A large collection of dermoscopic images used for training and training the model.

* Programming Language:

1. Python – Used for data processing, model development, and evaluation.

* Libraries and Frameworks:

1. NumPy, Pandas – For data manipulation and analysis.
2. OpenCV – For image preprocessing tasks like grayscale conversion, gausian blur, and inpainting.
3. Matplotlib, Seaborn – For data visualization.
4. Scikit-learn – For preprocessing, model evaluation, and metrics.
5. TensorFlow / Keras or PyTorch – For building and training deep learning models.

* Image Preprocessing Tools:

1. Grayscale conversion
2. Gausian blur (for hair and noise removal)
3. Inpainting (to restore cleaned image regions)

* Deep Learning Models:

1. Convolutional Neural Networks (CNNs) – Used for image classification and lesion detection.
2. (Optional) U-Net or similar architectures – For image segmentation.
3. Model Evaluation Metrics:
4. Accuracy, Precision, Recall, F1-score – To assess the performance of the classification model.

* Development Environment:

1. Jupyter Notebook
2. Google Colab
3. Kaggle Kernels – For interactive development and experimentation.

* Hardware:

1. GPU (if available) – To accelerate model training and improve performance.

### 1.3.3 Implementation Phases

1. **Problem Understanding & Dataset Selection**

* Study the problem of skin cancer detection.
* Select a dataset (**HAM10000**) for testing and training the model.

1. **Data Preprocessing**

* Load and run the dataset.
* Apply preprocessing techniques such as:
  + Grayscale conversion
  + Gausian Blur
  + Inpainting

1. **Lesion Segmentation**

* Implement segmentation techniques to extract the wound from the skin image.

1. **Model Development**

* Design and train a **Convolutional Neural Network (CNN)** for wound categorization.

1. **Model Evaluation**

* Test the trained model using estimated metrics such as:
  + Accuracy
  + Precision
  + Recall
  + F1-Score
* Analyze results to identify perfection and imperfection.

1. **Model Optimization**

* Tune hyperparameters to improve model performance.
* Apply regularization or data augmentation if needed.

1. **Integration & Final Pipeline**

* Combine all steps into one streamlined process.
* Ensure the pipeline works efficiently from input image to diagnosis.

1. 8. **Documentation & Reporting**

* Document all phases, methods, and results.
* Prepare reports or presentations to share findings and show the system.

### 1.3.4 Data Management

The data management plan for the **SpotCancerAI** project revolves around the HAM10000 dataset, which provides dermoscopic images and associated metadata such as wound types and lesion location. The dataset is organized into folders for raw images, processed outputs, segmentation masks, training and testing splits, and metadata. Preprocessing includes mapping lesion codes to readable labels, converting images to grayscale, applying gausian blur, and using inpainting to remove artifacts like hair. All images are resized to a consistent shape (e.g., 224x224) to standardize model input. The data is split into training (70%), validation (15%), and testing (15%) sets using stratification to preserve class balance. Label mapping converts shorthand codes like nv and mel into meaningful classes such as “benign” and “Melanoma.” For model robustness, data augmentation techniques such as flipping, rotation, scaling, color jitter, and noise are applied. Versioning tools like DVC or Github are recommended to track data changes, with cloud or external backups maintained. Since the HAM10000 dataset is publicly available and anonymized, it meets more principles.

### 1.3.5 Stakeholder Engagement

We heard about a patient who ignored a small skin spot, thinking it was harmless, but later it was diagnosed as late-stage skin cancer. Many people delay checkups due to lack of awareness, high costs, or limited access to doctors. Existing AI models are also hard to use and inaccurate for darker skin. This inspired us to create a fast, simple, and accessible AI tool for early skin cancer detection, helping people get diagnosed quickly and accurately. Some Key Features are as following:

* **AI-Based Skin Cancer Detection:** Uses deep learning to classify skin lesions.
* **Fast & Accurate Results:** Provides quick analysis to support medical decisions.
* **User-Friendly Interface:** Simple and easy-to-use system for both doctors and patients.
* **Data Security & Privacy:** Ensures patient information is kept safe.
* **Mobile & Web Compatibility:** Can be used on smartphones and computers.

### 1.3.6 Deliverable

* **System Architecture Documentation:** Detailed design documents outlining the system’s architecture, components, and integration points.
* **Training Materials:** Comprehensive training manuals and resources for law enforcement personnel.
* **Pilot Test Reports:** Evaluation reports from pilot testing phases, including performance data and identified issues.
* **Deployment Plan:** A detailed plan for full system deployment, including timelines, resources, and responsibilities.
* Compliance Reports: Documentation of compliance with legal and more principles, including privacy impact assessments and bias evaluations.

# Chapter 2: Literature Review

## 2.1 Literature Review

Paper 1:

The paper titled *"Skin Cancer Detection Using Deep Machine Learning Techniques"* addresses the critical problem of early detection of skin cancer, particularly melanoma, by leveraging the power of deep learning. The authors aim to overcome challenges such as limited access to healthcare in rural areas, dataset imbalance, and the need for accurate and accessible diagnostic tools. To achieve this, the study utilizes the ISIC (International Skin Imaging Collaboration) dataset, specifically including ISIC 2017 and ISIC 2018, which are widely recognized for skin lesion classification research. The techniques employed in the paper include Convolutional Neural Networks (CNNs), few-shot learning, and Generative Adversarial Networks (GANs), along with methods such as data augmentation and transfer learning to improve model performance even with limited data. The models primarily used in the study are CNN-based architectures, with comparisons to other advanced models like ResNet and DenseNet. Notably, a GAN-enhanced approach achieved a classification accuracy of 86.1%, demonstrating its effectiveness in distinguishing between benign and malignant skin lesions. As for future directions, the authors propose further advancements by refining GAN architectures, enhancing model robustness, and addressing data imbalance issues, aiming to improve the reliability and integration of AI models in real-world healthcare systems, especially through telemedicine platforms.

Paper 2:

Kalouche’s paper focuses on solving the problem of early skin cancer detection using deep learning, specifically vision-based approaches. The work leverages public dermoscopy datasets, likely the ISIC datasets, to train models. The main technique used is deep learning through convolutional neural networks (CNNs), with the VGG-16 architecture likely adopted to perform classification tasks. The final results show that the system achieved high classification accuracy, comparable to that of expert dermatologists. The future direction proposed involves deploying such AI models into real-world clinical decision support systems, to assist dermatologists and extend diagnostic capabilities into remote and underserved areas​.

Paper 3:

This document by the World Health Organization (WHO) does not directly solve a computational problem but addresses the global public health issue of UV radiation exposure leading to skin cancer. There is no dataset or model used, as the paper is purely informative. The technique applied involves awareness campaigns and public education. The final result emphasizes the prevention of UV-related skin cancers through behavioral changes like minimizing sun exposure and using protective clothing. Future directions focus on continuing global education efforts to reduce the incidence of UV-induced skin cancer​.

Paper 4:

Khan and colleagues address the issue of class imbalance, a major barrier in machine learning models, particularly in fields like medical imaging where rare disease cases are few. No specific dataset was used; instead, a variety of imbalanced datasets were reviewed. Techniques discussed include sampling methods like SMOTE, hybrid sampling, and deep learning adaptations for imbalanced classification tasks. No single model is proposed, but the paper surveys CNNs, decision trees, and ensemble methods. There are no experimental final results as it is a review paper. The future direction suggests developing more integrated hybrid approaches combining data-level and algorithmic solutions to better tackle imbalance problems​.

Paper 5:

Ali and Al-Marzouqi solve the problem of detecting melanoma early through regular convolutional neural networks (CNNs). Although the exact dataset is not specified in the excerpt, it is likely one of the ISIC challenges (2016/2017 datasets). They applied a basic CNN model without using complex pre-trained architectures. The technique involved training CNNs directly on dermoscopy images for binary classification (melanoma vs non-melanoma). Their final results showed notable accuracy in melanoma detection. The authors suggest that future research should explore deeper neural networks and ensemble approaches to further enhance detection accuracy and robustness​.

Paper 6:

Nasr-Esfahani and team focused on automating melanoma detection using convolutional neural networks (CNNs) applied to clinical images. The dataset is not explicitly mentioned, but based on context it likely involves ISIC or HAM10000 collections. Their technique centered on CNN-based feature extraction and classification. A basic CNN model was utilized without complex transfer learning at the time. Their final results demonstrated superior sensitivity and specificity compared to traditional computer vision approaches. Future directions include integrating their models into mobile applications to aid in teledermatology and early melanoma detection in remote regions​.

Paper 7:

Esteva’s landmark study aimed to match dermatologist-level accuracy in diagnosing skin cancers through deep neural networks. The dataset used was one of the largest, containing more than 129,000 clinical images drawn from sources like ISIC and other curated databases. The technique employed was deep learning using convolutional neural networks combined with transfer learning. Specifically, an Inception v3 architecture, pre-trained on ImageNet and then fine-tuned on skin lesion images, was used. The final results showed that the AI model achieved performance on par with board-certified dermatologists. Future directions involve deploying AI systems in primary care or telemedicine platforms to assist non-specialist doctors​.

Paper 8:

Mendes and Silva tackled the classification of various skin lesions using CNNs on clinical dermoscopy images. The datasets used were clinical photographs, although specifics like ISIC or PH2 aren't stated. Their technique was straightforward convolutional neural networks trained to classify images into different lesion categories. The model architecture specifics were not deeply discussed, indicating a standard CNN was used. The final results confirmed the viability of CNNs in diagnosing skin conditions. For future work, the authors recommend increasing the size and diversity of training datasets to further improve model generalization​.

Paper 9:

This study by Khan and colleagues tackled the challenge of imbalanced data classification by proposing a hybrid sampling method combined with deep learning. Various datasets were discussed, not limited to skin cancer data. Their technique was a blend of over-sampling and under-sampling strategies with deep neural networks. Although they used generic deep learning classifiers, no specific CNN or RNN model was emphasized. The final results showed that hybrid methods outperformed traditional oversampling and undersampling techniques in handling imbalance. The paper recommends extending hybrid sampling combined with deep learning approaches across different domains where imbalance is critical.

Paper 10:

Shoieb et al. addressed the problem of automatically detecting basal cell carcinoma (BCC) using full-field optical coherence tomography (FF-OCT) images. They used datasets composed of FF-OCT images obtained from clinical samples. The primary technique applied was deep learning using CNNs, with a CNN architecture tailored to handle the special characteristics of OCT imagery. Their final results demonstrated significant diagnostic performance, making CNNs a valuable tool for interpreting OCT images. Future directions include expanding the system to detect a broader range of skin lesions beyond BCC and improving processing speed for clinical deployment​.

Paper 11:

Sagar and Dheeba’s research focused on classifying melanoma from dermoscopic images using convolutional neural networks. While the specific dataset isn’t cited, it is likely from the ISIC archives. Their technique was direct CNN-based classification without extensive pre-processing. The model was a customized CNN suited for the classification of melanoma and benign lesions. The results demonstrated promising accuracy, suggesting CNNs are effective for such medical image analysis. In future work, they plan to explore transfer learning, using deeper networks, and possibly combining CNNs with other machine learning models to further boost detection accuracy