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

List of figures

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:
  I. HAM10000 – A large collection of dermoscopic images used for training and training the model.
- Programming Language:
  I. Python – Used for data processing, model development, and evaluation.
- Libraries and Frameworks:
  I. NumPy, Pandas – For data manipulation and analysis.
  II. OpenCV – For image preprocessing tasks like grayscale conversion, gausian blur, and inpainting.
  III. Matplotlib, Seaborn – For data visualization.
  IV. Scikit-learn – For preprocessing, model evaluation, and metrics.
  V. TensorFlow / Keras or PyTorch – For building and training deep learning models.
- Image Preprocessing Tools:
  I. Grayscale conversion
  II. Gausian blur (for hair and noise removal)
  III. Inpainting (to restore cleaned image regions)
- Deep Learning Models:
  I. Convolutional Neural Networks (CNNs) – Used for image classification and lesion detection.
  II. (Optional) U-Net or similar architectures – For image segmentation.
  III. Model Evaluation Metrics:
  IV. Accuracy, Precision, Recall, F1-score – To assess the performance of the classification model.
- Development Environment:
  I. Jupyter Notebook
  II. Google Colab
  III. Kaggle Kernels – For interactive development and experimentation.
- Hardware:
  I. GPU (if available) – To accelerate model training and improve performance.

1.3.3 Implementation Phases

I. **Problem Understanding & Dataset Selection**

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

II. **Data Preprocessing**

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

III. **Lesion Segmentation**

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

IV. **Model Development**

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

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

VI. **Model Optimization**
- Tune hyperparameters to improve model performance.
- Apply regularization or data augmentation if needed.

VII. **Integration & Final Pipeline**
- Combine all steps into one streamlined process.
- Ensure the pipeline works efficiently from input image to diagnosis.

VIII. 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.