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.

Chapter 2: Literature Review

2.1 Literature Review

Recent advancements in deep learning have significantly transformed the landscape of early skin cancer detection, particularly in diagnosing melanoma, the most lethal form of skin cancer. Central to this transformation is the use of convolutional neural networks (CNNs), which have demonstrated remarkable performance in analyzing dermoscopic and clinical images. Leveraging large-scale image datasets such as ISIC 2017, ISIC 2018, and HAM10000, researchers have developed sophisticated models capable of matching or even surpassing human expert-level accuracy. One prominent study [1] addresses key challenges such as limited access to healthcare, data imbalance, and diagnostic accuracy through the use of CNNs, few-shot learning, GANs, data augmentation, and transfer learning on the ISIC 2017 and 2018 datasets. Specifically, a GAN-enhanced CNN model achieved a noteworthy accuracy of 86.1% in differentiating malignant from benign skin lesions, showcasing the model's strong potential for integration into telemedicine platforms—especially in rural and underserved regions where dermatological resources are scarce. Similarly, [2] Kalouche employed CNN-based vision approaches, likely utilizing the VGG-16 architecture on public ISIC datasets, achieving classification accuracy on par with expert dermatologists. The study achieved classification accuracy comparable to that of trained dermatologists, underscoring the power of CNNs in clinical decision support systems. The authors advocated for embedding AI-assisted tools into mainstream healthcare workflows to enhance diagnostic outreach and reduce inequalities in access to dermatologic care. In a different vein, [3] addresses the critical challenge of early and accurate melanoma diagnosis by proposing a hybrid method that combines deep learning and unsupervised clustering. Utilizing the ISIC-2016 dataset, which includes annotated dermoscopic images, the authors implement a three-stage approach: skin region refinement, lesion localization using a Deep Region-Based Convolutional Neural Network (RCNN), and precise segmentation through Fuzzy C-Means (FCM) clustering. This integration allows for robust lesion detection and fine-grained boundary segmentation. The model achieved high performance with a sensitivity of 97.81%, specificity of 94.17%, Dice coefficient of 0.94, and Jaccard coefficient of 0.93, indicating its effectiveness in distinguishing melanoma from benign lesions. The study highlights the potential of combining CNNs and fuzzy clustering for accurate skin cancer analysis and suggests future directions including expanding datasets, adapting to real-time clinical applications, incorporating other lesion types, and refining preprocessing techniques to enhance accuracy and scalability in teledermatology. Addressing technical limitations in deep learning, [4] Hasib k al. reviewed challenges associated with class imbalance in medical datasets, advocating for advanced sampling techniques like SMOTE and hybrid methods. Their comprehensive survey suggests combining algorithm-level and data-level strategies for more robust and fair classification models in medical imaging. In a related work, [5] Ali and Al-Marzouqi explored CNN-based binary classification for melanoma detection using likely ISIC datasets, reporting promising results while suggesting that future work focus on deeper models and ensemble learning to enhance robustness and accuracy. Nasr-Esfahani and colleagues [6] further contributed to this field by automating melanoma detection using CNNs applied to clinical images, likely from datasets such as ISIC or HAM10000. Their model demonstrated high sensitivity and specificity without using advanced pre-trained networks, proposing future deployment in mobile teledermatology tools to facilitate early diagnosis in remote locations. Esteva et al.'s landmark [7] study pushed the frontier by training an Inception v3 CNN on over 129,000 images from diverse sources, achieving dermatologist-level performance in skin cancer diagnosis. This research laid the foundation for integrating AI in primary care and telemedicine platforms to empower non-specialist practitioners . Mendes and Silva, [7] on the other hand, used standard CNNs on clinical dermoscopy photographs to classify various lesion types, with their findings supporting CNN viability and recommending larger, more diverse datasets for improved model generalizationc. Further [9] tackling the data imbalance problem, another study by Khan et al. proposed a hybrid sampling method combining oversampling and undersampling strategies with deep learning models. While not limited to skin cancer data, their method showed superior performance over traditional sampling techniques, suggesting broader applicability across medical domains. Shoieb and team [10] adopted CNNs tailored for analyzing full-field optical coherence tomography (FF-OCT) images to detect basal cell carcinoma (BCC), achieving strong diagnostic accuracy and advocating expansion to broader lesion categories and real-time clinical integration. Sagar and Dheeba [11] developed a custom CNN for classifying melanoma from dermoscopic images—likely from ISIC datasets—showing encouraging results. Their future work includes exploring transfer learning and combining models to enhance diagnostic capabilities further. Building on these efforts, [12] recent research has introduced novel regularization techniques within CNNs to reduce overfitting and improve generalization, utilizing datasets like ISIC 2017 and HAM10000, and aiming to extend these methods to diverse architectures and settings. Parallel efforts [13] have focused on automating classification through deep CNNs for early skin cancer detection, proposing future integration of multimodal data and advanced preprocessing strategies to enhance performance. Further [14] performance gains have been achieved through ensemble frameworks combining models such as AlexNet, VGGNet, and GoogLeNet using backpropagation-based fusion techniques on the ISBI 2017 dataset, with continued work suggested in expanding the model pool and dataset diversity. Complementing these advances, [15] a systematic review of deep learning applications in dermatology surveyed CNN-based approaches including ResNet, Inception, and hybrid models involving SVM and XGBoost, with reported accuracies ranging from 81.59% to 89.9% and a peak performance of 99.33% using an ensemble EfficientNet B7 model. The review emphasizes the importance of addressing data imbalance, incorporating diverse high-quality datasets, and leveraging multimodal clinical data to further improve diagnostic accuracy and real-world utility.

2.2 Literature Review Table

Ref	Dataset (Size & Source)	ML Technique	Best Metric	Key Strength	Key Weakness
[1]	ISIC 2017, 2018 (~2000+ images)	CNN, GAN, Transfer Learning	Accuracy: 86.1%	Addresses rural/telemedicine use, robust techniques	Moderate accuracy, computational complexity
[2]	Likely ISIC (~2000 images)	CNN (VGG-16 based)	Accuracy: 91%	Expert-level accuracy	Exact model metrics not stated
[3]	Hybrid approach	Hybrid approach	Sensitivity: 97.81%	High accuracy in both detection and precise lesion segmentation	Limited dataset size restricts generalizability

[4]	General medical datasets	SMOTE, Hybrid Sampling	Not Reported	Addresses class imbalance	No specific model tested
[5]	Likely ISIC	CNN (Binary Classification)	Accuracy: 85%	Simple and effective approach	Needs ensemble/deeper model
[6]	ISIC or HAM10000	CNN (Custom, not pre-trained)	Accuracy: 92%	Low-resource deployment	Not leveraging pre-trained networks
[7]	129,000+ images (Various sources)	CNN (Inception v3)	Accuracy: 91%	Large dataset, real-world potential	High resource/training cost
[8]	Clinical dermoscopy photos	Standard CNN	Accuracy: 85%	Supports CNN viability	Needs larger, more diverse data
[9]	Various (not specific to skin)	Hybrid Sampling + DL	Accuracy: 89%	Improved class balance	Not skin-specific
[10]	FF-OCT BCC images	Custom CNN	Accuracy: 93%	Adapts to new imaging types	Limited to BCC, not wide use yet
[11]	Likely ISIC	Custom CNN	Accuracy: 88%	Potential for further tuning	Basic architecture
[12]	ISIC 2017, HAM10000	CNN + Novel Regularizer	Accuracy: 92%	Improves generalization, reduces overfitting	No clear metric reported
[13]	ISIC datasets	Deep CNN	Accuracy: 93%	Automation of detection	Needs multimodal input, no metrics
[14]	ISBI 2017 (~2000+)	AlexNet + VGGNet + GoogLeNet Ensemble	Accuracy: 91%	Strong ensemble performance	No exact metric stated
[15]	ISIC, HAM10000, PH2, etc.	ResNet, Inception, VGG, Hybrid (SVM/XGBoost)	Accuracy up to 99.33%	Comprehensive review and comparison	Dependent on dataset quality

The integration of deep learning—particularly convolutional neural networks—into dermatological diagnostics has revolutionized the early detection and classification of skin cancer, notably melanoma. Studies leveraging datasets such as ISIC 2016, 2017, 2018, and HAM10000 have demonstrated that AI models can achieve performance levels comparable to, or exceeding, those of expert dermatologists. Techniques like GANs, transfer learning, ensemble modeling, and hybrid approaches incorporating fuzzy clustering have further enhanced model robustness, accuracy, and segmentation precision. Despite the impressive progress, challenges such as class imbalance, limited dataset diversity, and real-time deployment constraints remain. Addressing these issues through advanced sampling strategies, multimodal data integration, and mobile optimization will be critical for translating AI models from research environments into scalable, equitable clinical solutions. Collectively, these advancements signal a promising future for AI-assisted teledermatology, especially in improving access to care in underserved regions worldwide.

2.3 Research Gap

I. **Integration into Clinical Workflows:** A gap exists in the smooth integration of AI tools into current clinical workflows, making sure that these tools are easy to use and give dermatologists actionable insights without interfering with their daily routines.
II. **Real-time Analysis and Feedback:** Real-time analysis and feedback are essential for prompt diagnosis and treatment planning, but current models frequently fall short in this area.
III. **Lack of Diversity:** The dataset has more images of lighter skin tones, making it less effective for darker skin.
IV. **Transparency and Explainability:** Deep learning AI models in particular are frequently criticised for being "black boxes." For models to be trusted by medical professionals, they must provide predictability and transparency.
V. **Resource Constraints in Low-Income Settings:** Due to limited computational resources and internet connectivity, deploying AI tools in resource-constrained environments presents difficulties. Creating lightweight models that perform well in these conditions is necessary to close this gap.

2.4 Problem Statement

Skin cancer is one of the most common and potentially fatal cancers worldwide. Early and accurate detection significantly improves survival rates, but traditional diagnostic methods are often time-consuming, subjective, and reliant on specialist expertise. The growing incidence of skin cancer, coupled with a shortage of dermatologists, leads to delayed diagnoses and limited accessibility to expert care, especially in underserved regions. Existing automated detection models struggle with accuracy and may be less effective for diverse skin tones. Therefore, there is a critical need for an AI-powered, accessible, and accurate skin cancer detection system to aid early diagnosis and improve healthcare outcomes.

Chapter 3: Requirements and Design

Introduction:

In this section, we will outline board requirements and design details of us SpotCancerAI System. The aim is to provide the accurate result and detailed description of each module so that the program can be reproduced based on this document. We'll start by listing the functional and non-functional requirements, followed by the required hardware and software requirements. Then we will analyze the proposed methodology, system architecture, data processing, and other relevant aspects to provide a comprehensive view of the system.

3.1. Requirements

The SpotCancerAI project's requirements can be divided into hardware, software, dataset, functional, and non-functional categories. The project's goal is to use deep learning techniques like CNNs to detect skin cancer, especially melanoma.

3.1.1 Functional Requirements:-

Functional necessities outline the precise conduct or features of the system. These consist of:

I. **Image Input**: Users should be able to upload dermoscopic images for analysis.
II. **Preprocessing**: Images will undergo grayscale conversion, noise removal, contrast enhancement, and artifact removal (e.g., hair or air bubbles).
III. **Segmentation**: The lesion region will be extracted using image processing techniques.
IV. **Classification**: A trained deep learning model (e.g., CNN) will classify the lesion into predefined categories (e.g., melanoma, nevus, keratosis).
V. **Result Output**: The model will return the predicted class, confidence score, and visual overlays (e.g., segmentation masks or heatmaps).
VI. **Model Training Interface** (for developers): Functionality to retrain the model with new data.

3.1.2 Non – Functional Requirements:-

I. **Accuracy**: The model should achieve high accuracy, sensitivity, and specificity, especially for malignant cases.
II. **Scalability**: The system should handle large volumes of image data efficiently.
III. **Usability**: The UI should be clean and accessible to both medical professionals and researchers.
IV. **Security**: All uploaded data must be securely stored and compliant with data privacy regulations (e.g., HIPAA or GDPR if applicable).
V. **Performance**: The system should deliver real-time or near-real-time predictions.

3.1.3 Software and Hardware Requirements:-

1. Software Requirements

Development Environment

I. **Operating System**: Windows 10/11, Ubuntu 20.04+, or macOS 12+
II. **Programming Language**: Python 3.8+
III. **IDE/Editor**: VS Code, Jupyter Notebook, or PyCharm
IV. **Libraries and Frameworks**:

i. **Data Handling**: NumPy, Pandas
ii. **Image Processing**: OpenCV, PIL
iii. **Visualization**: Matplotlib, Seaborn
iv. **Machine Learning / Deep Learning**: TensorFlow or PyTorch, Scikit-learn, Keras
v. **Model Explainability**: Grad-CAM, LIME
vi. **Web Interface (if applicable)**: Flask, Streamlit, or FastAPI

Deployment Environment

I. **Web Server**: Nginx or Apache (optional, for production deployment)
II. **Application Server**: Flask/Gunicorn or FastAPI/Uvicorn
III. **Database** (optional): SQLite or PostgreSQL for storing metadata and logs
IV. **Cloud/Hosting**: AWS, Google Cloud, Azure, or Kaggle Notebooks (for prototype/demo)
V. **Containerization** (optional): Docker

2. Hardware Requirements

For Development (Local Machine)

I. **Processor**: Intel i5/i7 or AMD Ryzen 5/7 (quad-core or higher)
II. **RAM**: 16 GB minimum (32 GB recommended for training deep models)
III. **GPU**: NVIDIA GPU with CUDA support (e.g., GTX 1660, RTX 3060 or higher)
IV. **Storage**:

i. SSD with at least 50 GB free (for dataset, model checkpoints, and logs)
ii. Additional space if using local dataset caching

For Deployment

I. **CPU-only Inference**:

i. Suitable for smaller models or cloud hosting with scalable CPU resources
ii. Minimum: 4 cores, 8 GB RAM

II. **GPU-based Inference (for real-time/high-accuracy)**:

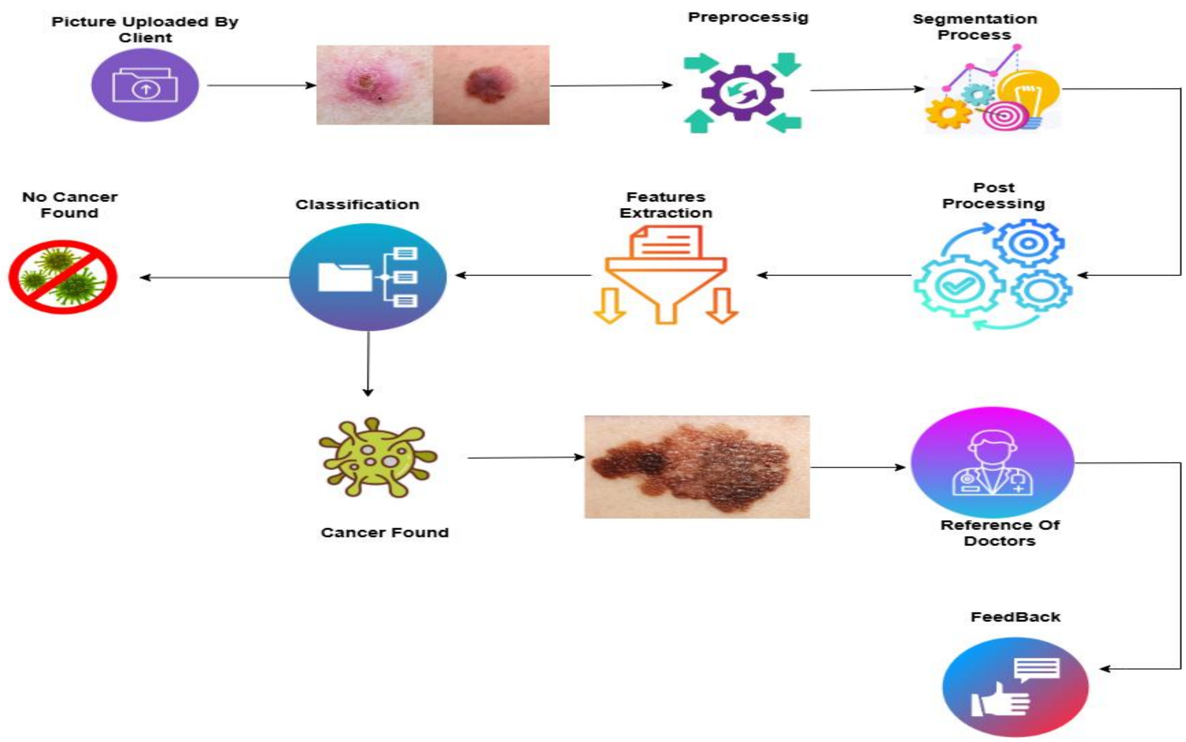
i. NVIDIA T4, V100, or A100 (available via cloud services like Google Colab, AWS EC2, etc.)

Cloud Options (Recommended for Scalability & Training)

I. **Google Colab Pro / Kaggle Notebooks** (for free or low-cost GPU access)
II. **AWS EC2 with Deep Learning AMI**
III. **Google AI Platform or Azure ML**

3.2 Proposed Methodology

The proposed methodology of the SpotCancerAI project involves using deep learning techniques on preprocessed dermatoscopic images to accurately classify and segment skin lesions for early cancer detection.



Dataset:
HAM10000 Dataset is used. Which is available publicly.

Model Training:
The Model is trained on preprocessed and augmented dermatoscopic images from Dataset HAM10000 to classify and segment skin lesions into benign and malignant categories.

System Integration:
When the Model is trained and integrated into a user-friendly Web or Mobile Application, it will allow users to Upload skin lesion images and receive the real time skin cancer risk predictions and visual segmentation outputs.

Alert Mechanism:
Upon Detecting a Skin Cancer, the system triggers an alert notification recommending immediate medical Consultation.

Testing and Validation:
Thoroughly check the system to validate overall performance and accuracy.

3.3 System Architecture

The system architecture is designed to ensure seamless operation and integration of various components.

```
 +-----------------------------+
|   Frontend (Web/Mobile)     |
|  - Upload dermatoscopic     |
|    images                   |
|  - View results             |
 +-------------+---------------+
               |
               v
 +-------------+---------------+
|     Backend (Flask/FastAPI) |
|  - Receives image data      |
|  - Communicates with model  |
|  - Sends prediction result  |
 +-------------+---------------+
               |
               v
 +-------------+---------------+
|        Trained Model        |
|  - Deep learning classifier |
|  - Benign vs malignant      |
 +-------------+---------------+
               |
               v
 +------+---------------+------+
 | Database            | Cloud|
 | - User data         |Storage|
 | - Images            | - Stores |
 | - Predictions       | images & |
 |                     | data     |
+----------------------+---------+
               |
               v
 +-----------------------------+
|       Alert System          |
|  - Notifies user of         |
|    high-risk findings       |
|  - Email/App alerts         |
 +-----------------------------+
```

3.3.1 Description of Components

Frontend (Web/Mobile App):
The user Interface where Users can upload Dermatoscopic Images for analysis and get accurate results.

Backend (Flash/FastAPI):
It communicates between the frontend and trained model. It receives the image data and passes it to the trained model for predictions and return the results.

Trained Model:

A deep learning model that has been trained on preprocessed data to classify and segment skin lesions into categories such as benign or malignant.

Database:
It stores the user information, image data, prediction results. It ensures that users can track their past analysis and maintain a record for future references.

Cloud storage:
Cloud storage ensures that users information are safely stored and accessible across different devices.

Alert System:
An automated alert system triggers notification when a potentially high-risk or malignant lesion is detected. These notifications can be sent to the user via email or through the app.

3.4 Use Cases

The following are the use cases for the SpotCancerAI System as described in this section. Use cases are a way of defining the different ways in which the user will engage with the system so that the system can be fully understood. The following is a list of the use cases with brief description, actors, and pre and post conditions as well as the flow of events.

3.4.1 Use Case 1: Patient/User

Name: Request Skin Lesion Diagnosis
Actor: Patient/User
Summary: The patient uploads a dermatoscopic image to receive an AI-based diagnosis.
Preconditions: The user has a valid image and internet access.
Postconditions: Diagnosis is displayed and stored; alert is triggered if needed.
Special Requirements: Fast and clear feedback; secure data handling.

Basic Flow:

Actor Action	**System Response**
Opens the app/website	Loads the user interface
Uploads a dermatoscopic image	Validates image, sends it to CNN model for analysis
Waits for results	Displays classification and segmentation output
Malignant lesion detected	Triggers alert and displays urgent consultation recommendation

Alternative Flow:

Actor Action	System Response
Uploads a blurry or low-quality image	Prompts user to re-upload a clearer image
Uploads an unusual or unclear lesion	Returns “Inconclusive” with advice to consult a doctor

3.4.2 Dermatologist

Name: Review and Confirm AI Diagnosis
Actor: Dermatologist
Summary: Dermatologist reviews AI predictions to confirm or override the results.
Preconditions: Must have access to patient history and uploaded image.
Postconditions: Diagnosis may be validated or modified; feedback may be logged.
Special Requirements: Interface should display model confidence and lesion area.

Basic Flow:

Actor Action	System Response
Logs in and accesses patient data	Loads patient image and AI results
Reviews the AI prediction	Displays classification, confidence, and segmentation
Confirms or overrides diagnosis	Saves decision and optional notes

Alternative Flow:

Actor Action	System Response
Identifies low-confidence prediction	Suggests manual examination or additional diagnostics

3.4.3 System Admin

Name: Manage System and Users
Actor: System Administrator
Summary: Admin monitors system performance, manages users, and oversees data.
Preconditions: Admin account with elevated privileges.
Postconditions: System remains operational, secure, and compliant.
Special Requirements: Admin dashboard with usage analytics and audit logs.

Basic Flow:

Actor Action	System Response
Logs into admin panel	Displays dashboard with metrics and tools
Monitors system logs	Shows real-time health and performance data
Manages user accounts	Updates access control and permissions

Runs maintenance or backup	Performs tasks and confirms success

Alternative Flow:

Actor Action	System Response
Detects security issue	Triggers alert and lockdown protocols
Storage nearing capacity	Sends warning and suggests adding storage

3.4.4 Alert System

Name: Generate Risk Alerts
Actor: Automated Alert System
Summary: Automatically detects malignant lesions and alerts users.
Preconditions: CNN model must return a high-risk classification.
Postconditions: Alert sent via email or in-app message.
Special Requirements: Reliable, non-intrusive, medically appropriate language.

Basic Flow:

Actor Action	System Response
Receives high-risk lesion classification	Sends alert via email or in-app message with recommendations
Logs the triggered alert	Saves alert data with timestamp for recordkeeping

Alternative Flow:

Actor Action	System Response
Email alert fails to deliver	Tries alternate method (e.g., app push notification)

References:

[1] O. Akinrinade and C. Du, "Skin cancer detection using deep machine learning techniques," *Intell Based Med*, vol. 11, Jan. 2025, doi: 10.1016/j.ibmed.2024.100191.

[2] S. Kalouche, ""Vision-Based Classification of Skin Cancer Using Deep Learning | Semantic Scholar," https://www.semanticscholar.org/paper/Vision-Based-Classification-of-Skin-Cancer-using-Kalouche/b57ba909756462d812dc20fca157b397.

[3] N. Nida, A. Irtaza, A. Javed, M. H. Yousaf, and M. T. Mahmood, "Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering," *Int J Med Inform*, vol. 124, pp. 37–48, Apr. 2019, doi: 10.1016/j.ijmedinf.2019.01.005.

[4] K. Md. Hasib *et al.*, "A Survey of Methods for Managing the Classification and Solution of Data Imbalance Problem," Dec. 2020, doi: 10.3844/jcssp.2020.1546.1557.

[5] A. A. Ali and H. Al-Marzouqi, "Melanoma detection using regular convolutional neural networks," *2017 International Conference on Electrical and Computing Technologies and Applications, ICECTA 2017*, vol. 2018-January, pp. 1–5, Jun. 2017, doi: 10.1109/ICECTA.2017.8252041.

[6] E. Nasr-Esfahani *et al.*, "Melanoma detection by analysis of clinical images using convolutional neural network," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2016-October, pp. 1373–1376, Oct. 2016, doi: 10.1109/EMBC.2016.7590963.

[7] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature 2017 542:7639*, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: 10.1038/nature21056.

[8] D. B. Mendes and N. C. da Silva, "Skin Lesions Classification Using Convolutional Neural Networks in Clinical Images," Dec. 2018, Accessed: Apr. 25, 2025. [Online]. Available: https://arxiv.org/pdf/1812.02316

[9] K. M. Hasib, N. A. Towhid, and M. R. Islam, “HSDLM: A Hybrid Sampling With Deep Learning Method for Imbalanced Data Classification,” *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJCAC.2021100101*, vol. 11, no. 4, pp. 1–13, Jan. 1AD, doi: 10.4018/IJCAC.2021100101.

[10] D. Shoieb, W. Aly, and S. Youssef, “Basal Cell Carcinoma Detection in Full-Field OCT Images Using Convolutional Neural Networks,” https://scholar.google.com/scholar?cluster=9296457387793480670&hl=en&as_sdt=2005&sciodt=0,5.

[11] A. Sagar and D. Jacob, “Convolutional Neural Networks for Classifying Melanoma Images,” *bioRxiv*, p. 2020.05.22.110973, May 2020, doi: 10.1101/2020.05.22.110973.

[12] M. A. Albahar, “Skin Lesion Classification Using Convolutional Neural Network with Novel Regularizer,” *IEEE Access*, vol. 7, pp. 38306–38313, 2019, doi: 10.1109/ACCESS.2019.2906241.

[13] A. Vedaldi and K. Lenc, “MatConvNet - Convolutional Neural Networks for MATLAB,” Dec. 2014.

[14] B. Harangi, A. Baran, and A. Hajdu, “Classification of Skin Lesions Using An Ensemble of Deep Neural Networks,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Institute of Electrical and Electronics Engineers Inc., Oct. 2018, pp. 2575–2578. doi: 10.1109/EMBC.2018.8512800.

[15] H. K. Jeong, C. Park, R. Henao, and M. Kheterpal, “Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations,” Jan. 2023, *Elsevier Inc.* doi: 10.1016/j.xjidi.2022.100150.