**Egyptian E-Learning University**

Faculty of Computers & Information Technology

Skin Disease Detection Using CNN

**By**

|  |  |
| --- | --- |
| Name karim Mohsen Mahmoud hamed | ID 2102010 |
| Name Ziad AbdelWahab Abdelgayed | ID 2101899 |
| Name Sayed Mohamed Badawy | ID 2101935 |
| Name Abdalla Yahiya Atalaa | ID 2001514 |
| Name Mahmoud Mohamed Mahdy | ID 2102068 |
| Name Hosam Ramadan Abdalla | ID 2101891 |
| Name Mohamed Salah Darwish | ID 2102045 |

Supervised by

Dr. Alaa Zaki

Assistant

Eng. Rehab Galal

[Cairo]-2025

Abstract

This project focuses on enhancing early detection and classification of skin cancer using deep learning, specifically Convolutional Neural Networks (CNNs). Skin cancer, including types like melanoma, basal cell carcinoma, and squamous cell carcinoma, remains a major global health concern, and early diagnosis is critical for effective treatment.

The methodology involves several key steps. A large, diverse dataset of dermoscopic and clinical skin images was used. Preprocessing steps such as image resizing and class balancing through oversampling were applied. Three different models were trained: one basic CNN for validating input as skin images, and two others using fine-tuned ResNet101 architectures to classify skin diseases in dermoscopic and clinical images. These models demonstrated high accuracy and strong generalization across test data.

To translate this research into practical use, a web application was developed. It allows users to upload images of skin lesions for instant classification using the trained models. Additional features include the ability to locate the nearest medical laboratory based on the user's location and access to an educational section that provides comprehensive information on various skin cancer types.

Overall, this project showcases the effectiveness of CNNs in medical image analysis and demonstrates how AI-powered tools can improve diagnostic accuracy, support timely medical intervention, and raise public awareness, contributing to more accessible and efficient skin healthcare.

Acknowledgments

We extend our deepest gratitude to our esteemed supervisor, Prof. Dr. Alaa Zaki, whose invaluable guidance and support have been instrumental throughout this journey. We are equally grateful to ENG-

DR Rehab Galal, whose unwavering encouragement and belief in our potential have been a source of constant motivation. Their mentorship not only expanded our knowledge but also broadened our perspectives, for which we are profoundly thankful.

Our heartfelt thanks go to our families, whose love, patience, and unwavering support have been the cornerstone of our achievements. Without their constant encouragement, this project would not have been possible.

We also wish to express our appreciation to all our professors, whose assistance and insights have been crucial in various aspects of this work.

This project is dedicated to every patient struggling to receive medical treatment. We sincerely hope that our efforts contribute meaningfully to their journey towards better health and well-being.

Thank you all for making this endeavor a reality.

Contents

[Abstract 2](#_Toc197463970)

[Acknowledgments 3](#_Toc197463971)

[Introduction 5](#_Toc197463972)

[Literature Review / Related Work 6](#_Toc197463973)

[Proposed system 6](#_Toc197463974)

[Implementation 6](#_Toc197463975)

[Testing & Evaluation 6](#_Toc197463976)

[Results & Discussion 6](#_Toc197463977)

[Conclusion & Future Work 6](#_Toc197463978)

[References 6](#_Toc197463979)

[Appendices (Optional) 6](#_Toc197463980)

Chapter 1

Introduction

* 1. Introduction

Skin cancer is one of the most common types of cancer that begins with the uncontrolled reproduction of skin cells. It can occur because of the ultraviolet radiation from sunshine or tanning beds, and it causes skin cells to multiply and form malignant tumors. Skin cancer is one of the primary reasons for deaths worldwide. According to statistics published by, 97,160 Americans were diagnosed with skin cancer in 2023, which is 5.0% of the total cancer cases reported in the United States, and 7990 people died because of skin cancer which is 1.3% of the total deaths because of skin cancer in the United States. Melanoma is one of the most common and dangerous types of skin cancer that can spread quickly to other body parts. Approximately 21 out of 100,000 melanoma cases were diagnosed in the United States between 2016 and 2020. The death rate because of melanoma was 2.1 per 100,000 diagnosed cases, and 1,413,976 people were living with melanoma in 2020. The five-year survival rate of skin melanoma is 93.5% which is relatively high. The five-year survival rate is 99.6% when skin melanoma is diagnosed at an early stage. There are more chances of survival when skin

melanoma is localized, which means it does not spread to other body parts, but only 77.6% of skin melanomas are diagnosed at the local stage. The number of deaths because of skin melanoma can be reduced if it is detected at its early stages.Diagnosis in dermatology is largely based on visual inspection of a lesion on the suspicious skin area. Therefore, diagnostic ability and accuracy depends greatly on the experience and training of dermatologists or general practitioners, in areas where dermatological services are not readily available. When dermatologists get no access to additional technical support, they have an approximately 65%- 70% accuracy rate in skin cancer diagnosis. If the lesion is suspicious, the visual inspection is supplemented with different diagnostic tools (e.g. dermoscopy, confocal microscopy or optical coherence tomography) providing the ability to explore the skin in vivo, in depth and at a higher resolution. However, access to these instruments remains limited due to time, logistical and cost concerns. Even when this technical support is feasible, dermatologists rarely achieve average rates greater than 85%. The situation is even worse if we consider that there is a shortage of dermatologists whilst diagnostic accuracy of non-expert clinicians is sensibly

below what is observed with dermatologists, reaching estimated rates between 20 and 40%. Thus, new diagnostic tools assisting dermatologists or general practitioners to accurately diagnose skin lesions should be developed, evaluated and optimized. Analyzing cancers poses significant challenges, demanding intensive examination. Histopathology confirms over 50% of lesions, while follow-up examination, expert consensus, or confirmation by in-vivo confocal microscopy determines the remainder. The scarcity of experts, particularly radiologists, presents a bottleneck. This presents a great potential for the use of machine learning and deep learning methods for automatic skin lesion classification. The initial step in dermatological

diagnosis involves conducting a skin biopsy to determine whether a skin lesion is malignant or benign. This process, from securing a dermatologist appointment to receiving biopsy results, typically spans a week or longer. This project endeavors to shorten this timeline.

The approach uses Convolutional Neural Network (CNN) to classify 9 types of skin lesions:

● Actinic keratosis and intraepithelial carcinoma: common non-invasive variants of squamous cell carcinomas. They are sometimes seen as precursors that may progress to invasive squamous cell carcinoma.

● Basal cell carcinoma: a common version of epithelial skin cancer that rarely metastasizes but grows if it isn’t treated.

● Benign keratosis: contains three subgroups (seborrheic keratoses, solar lentigo, and lichen-planus like keratoses (LPLK)). These groups may look different but are biologically similar.

● Dermatofibroma: a benign skin lesion that is regarded as a benign proliferation or an inflammatory reaction to minimal trauma.

● Melanoma: a malignant neoplasm that can appear in different variants. Melanomas are usually, but not always, chaotic, and some criteria depend on the site location.

● Melanocytic Nevi: these variants can differ significantly from a dermatoscopic point of view but are usually symmetric in terms of distribution of color and structure.

● Vascular Lesions: generally categorized by a red or purple color and solid, well-circumscribed structures known as red clods or lacunes.5

● Squamous cell carcinoma: a common form of skin cancer originating from squamous cells in the epidermis. It often appears as a rough, scaly patch, an ulcer, or a red, firm bump. While it can metastasize if not treated, early detection usually allows for effective treatment.

● Seborrheic keratoses: benign skin growths that are often mistaken for warts or skin cancer. They appear as waxy, raised, and usually pigmented lesions that can vary in color from light tan to black. Though they are harmless, they may sometimes be removed for cosmetic reasons or if they become irritated.

The main objective is to mitigate skin cancer-related mortality by utilizing advanced image classification technology for skin lesion classification. An accurate model could be a big aid in early skin cancer detection, thereby reducing unnecessary biopsies and complications. Accessibility is a key focus, and we aimed to achieve this through the development of an intuitive mobile application. Users, including dermatologists, can upload skin lesion images. The model swiftly analyzes it and returns accurate results. The app features an educational section dedicated to providing information about various types of skin cancers, promoting awareness and proactive skin health management

* 1. Background and motivation for the project.

Skin cancer is one of the most prevalent forms of cancer worldwide, often caused by prolonged exposure to ultraviolet radiation from the sun or tanning beds. It poses a major public health concern due to its increasing incidence and the risk of mortality, especially in cases like melanoma — a highly aggressive type of skin cancer. In the United States alone, over 97,000 new skin cancer cases were reported in 2023, with nearly 8,000 deaths. Although early diagnosis can significantly improve survival rates (up to 99.6% for early-stage melanoma), only about 77% of cases are detected early.

Traditionally, skin cancer diagnosis heavily depends on visual inspection by dermatologists. However, the accuracy of diagnosis varies widely based on the experience of the clinician and the availability of diagnostic tools like dermoscopy or confocal microscopy. Even with advanced tools, dermatologist accuracy often does not exceed 85%, while non-expert clinicians may achieve as low as 20–40% accuracy. Furthermore, due to the limited availability of dermatologists and the time-consuming nature of traditional biopsy procedures, delays in diagnosis are common.

These challenges create a strong need for intelligent, accessible, and automated diagnostic tools that can assist medical professionals in identifying skin cancer more quickly and accurately. This project is motivated by the potential of deep learning, particularly Convolutional Neural Networks (CNNs), to automate skin lesion classification. By training a CNN model on a large dataset of labeled dermoscopic images, we aim to enable early and reliable skin cancer detection.

The project’s broader motivation is to reduce diagnostic delays, avoid unnecessary biopsies, and ultimately lower the mortality rate associated with skin cancer. Additionally, we aim to make this solution widely accessible through a web application, enabling users—whether medical professionals or the general public—to upload skin images and receive instant diagnostic feedback. The platform also includes an educational section to raise awareness about different skin cancer types and promote early detection behaviors.

* 1. Importance of the problem being addressed.

Skin cancer, particularly melanoma, is a life-threatening disease that requires early and accurate diagnosis to ensure effective treatment and high survival rates. Despite the technological advances in dermatology, diagnostic accuracy still heavily depends on the availability and expertise of dermatologists, which is a significant limitation in many regions—especially underserved and rural areas. This leads to late diagnoses, higher treatment costs, and lower survival rates.

The importance of this problem is heightened by several key factors:

* **High Incidence and Mortality**: Skin cancer cases are rapidly increasing worldwide. Melanoma, while less common than other types, is responsible for the majority of skin cancer deaths due to its aggressive nature and ability to spread.
* **Limited Access to Dermatological Services**: Many people, especially in remote or low-resource settings, do not have regular access to trained dermatologists or diagnostic equipment, which creates delays in detection.
* **Diagnostic Limitations**: Even skilled professionals can misdiagnose skin lesions, and visual diagnosis accuracy without supporting tools ranges between 65–70%, with non-specialists performing even worse.
* **Time and Cost of Biopsy Procedures**: Traditional diagnostic methods like biopsies can be time-consuming, expensive, and uncomfortable for the patient. Reducing unnecessary biopsies through better initial screening could improve patient experience and reduce healthcare burdens.

By introducing an automated skin lesion classification model using deep learning (CNN), this project addresses a critical gap in healthcare. It offers a scalable, accessible, and low-cost diagnostic aid that can be integrated into teledermatology platforms and mobile/web applications. This approach empowers both healthcare providers and patients, enhances early detection, and ultimately contributes to reducing the global skin cancer burden.

* 1. Problem Statement

- Clear definition of the problem your project addresses.

Early and accurate detection of skin cancer remains a major challenge in the medical field. Traditional diagnosis relies heavily on expert dermatologists performing visual inspections or dermoscopy, which often results in diagnostic variability. In many regions, there is a shortage of specialists, and access to advanced imaging equipment is limited or unavailable. This causes delays in diagnosis, misclassification of skin lesions, and, in some cases, avoidable deaths.

Moreover, existing diagnostic procedures such as skin biopsies are time-consuming, costly, and sometimes unnecessary if the lesion is benign. These limitations highlight the need for a faster, more accessible, and reliable method to screen and classify skin lesions.

- Justification for why this problem is worth solving.

* **High Mortality Risk:** Melanoma, a deadly form of skin cancer, spreads rapidly if not detected early. Reducing time to diagnosis can save lives.
* **Lack of Expertise in Remote Areas:** Many patients in rural or underdeveloped areas don’t have access to dermatologists, leading to late or incorrect diagnosis.
* **Overloaded Healthcare Systems:** Automating lesion classification helps reduce the burden on specialists and limits the number of unnecessary biopsies.
* **Technological Potential:** The rise of deep learning and computer vision presents an opportunity to build powerful models that match or exceed expert-level accuracy in classifying skin lesions.
* **Global Relevance:** Skin cancer is a worldwide issue affecting millions. A scalable and user-friendly AI solution can be used globally, increasing accessibility and awareness.

**Therefore**, the project aims to develop a deep learning-based system, powered by Convolutional Neural Networks (CNNs), to accurately classify skin lesion images into multiple types, helping in early detection and enabling fast, cost-effective, and reliable support for dermatological diagnosis through a web-based application.

* 1. Objectives

- Main Objective: The primary goal of the project.

To develop a deep learning-based skin cancer classification system using Convolutional Neural Networks (CNNs) that accurately identifies various types of skin lesions from medical images, and integrate it into a user-friendly web application to assist in early diagnosis and reduce unnecessary biopsies.

- Specific Objectives: Breakdown of tasks required to achieve the main goal.

1. **Data Collection and Preparation:**
   * Collect and clean a comprehensive dataset of dermoscopic skin lesion images (e.g., HAM10000).
   * Perform data preprocessing steps such as resizing, normalization, and augmentation to handle class imbalance.
2. **Model Development:**
   * Design and train a CNN-based model .
   * Tune hyperparameters to optimize performance and avoid overfitting.
3. **Model Evaluation:**
   * Assess the model using key performance metrics like accuracy, precision, recall, and F1-score.
   * Use confusion matrix and ROC curve for in-depth performance analysis.
4. **Web Application Integration:**
   * Develop a front-end interface that allows users (doctors/patients) to upload skin lesion images.
   * Build a back-end API to serve predictions from the trained model.
5. **User Education and Awareness:**
   * Include an informative section in the web app to educate users about different types of skin cancer.
6. **Testing and Validation:**
   * Perform testing using unseen test data to validate real-world performance.
   * Collect feedback to improve usability and accuracy.
7. **Deployment and Accessibility:**
   * Deploy the web application online to ensure wide accessibility.
   * Optimize the system for performance and user experience across devices.
   1. Brief overview of the proposed solution.

The proposed solution involves developing an intelligent system based on **Convolutional Neural Networks (CNNs)** for the **automated classification of skin lesions** using dermoscopic images. The system will be trained on a large, labeled dataset such as **HAM10000**, which contains images of 7 to 9 different types of skin conditions, including both benign and malignant lesions.

The process starts with **data preprocessing**, which includes resizing images, normalization, and applying **data augmentation techniques** to handle class imbalance. Then, a **CNN-based model** is built and trained to classify images into the correct skin lesion category.

To make this system accessible, it will be integrated into a **user-friendly web application**. This app will allow users—such as dermatologists or general users—to **upload skin lesion images**, and the model will return **instant classification results** along with confidence scores.

The application will also include an **educational section** to raise awareness about different types of skin cancer, symptoms, and when to seek medical help. This solution aims to assist healthcare professionals in **early detection**, reduce diagnostic workload, and ultimately help in saving lives by **speeding up diagnosis and improving accuracy**.

Chapter 2

Literature Review / Related Work

- Summary of existing research and technologies related to your project.

There are many research papers on the subject of using ML and DL for skin cancer and skin lesion classification. Several apps are also available for skin lesion classification. At the start of our project we set out to research and learn from them.

One major application is called Medical Dermatology and is one of the top applications in this domain. It relies on patient symptoms and data like gender, age, etc to make a classification for the skin lesion. The drawback of the app is that it has a mediocre interface that is quite unattractive for users.

Another application, called Skin Vision, is published but not widely used. It has an appealing and creative user interface that helps the patient to get good and satisfactory results.

An application called Aysa comes with a lot of features. It definitely has the best and most creative user interface and is totally free. It saves the patient’s history and includes a lot of medical knowledge questions to reach the best, most accurate result.

An application called Visus is another one that provides image classification for 7 types of skin lesions, but it has only modest accuracy.

Another application called AI Dermatologist includes an amazing user interface to classify image lesions, but with limited features. It’s also not free.

Yet another is called SmartSkin. This one has a pretty bad interface, limited features, and low accuracy.

Lots of other applications that target skin lesions classification exist but are not generally that reliable, like All Skin Dermatologist, Skin Diseases, AISkin, Scanoma and more.

- Gaps in current solutions that your project aims to fill.

Despite the advancements in medical imaging and diagnostics, several critical gaps remain in existing skin cancer detection systems:

**1. Limited Accessibility to Dermatologists:**

* In many rural or underdeveloped areas, there is a shortage of trained dermatologists.
* Patients may have to wait weeks for appointments, delaying critical diagnosis and treatment.

✅ **Our Solution:**

* A web-based diagnostic system accessible from any device allows for **instant screening** by uploading a skin image.

**2. Diagnostic Inaccuracy by Non-Specialists:**

* General practitioners often have **low diagnostic accuracy (20–40%)** compared to dermatologists (~65–85%).
* Manual diagnosis is subjective and varies based on experience.

✅ **Our Solution:**

* A CNN model trained on thousands of labeled images provides **objective and consistent predictions** across users.

**3. Overdependence on Expensive Equipment:**

* Techniques like **dermoscopy, confocal microscopy, or OCT** are effective but **costly** and not widely available.

✅ **Our Solution:**

* Uses only **image uploads from standard cameras or smartphones**, reducing reliance on specialized equipment.

**4. Time-Consuming Biopsy-Based Confirmation:**

* Diagnosis through biopsy can take a **week or more**, causing stress and delaying treatment.

✅ **Our Solution:**

* Enables **preliminary screening in seconds**, helping doctors prioritize urgent cases more effectively.

**5. Lack of Educational Support for Patients:**

* Most systems do not inform users about the **types of lesions, warning signs, or prevention** tips.

✅ **Our Solution:**

* Integrates an **educational module** within the app to raise awareness and promote proactive skin care.

- Summary

Skin cancer is among the most common and potentially deadly forms of cancer worldwide. Early diagnosis significantly increases survival rates, especially for aggressive types such as melanoma. However, current diagnostic methods heavily rely on specialist expertise, expensive imaging tools, and invasive procedures like biopsies — all of which are not always accessible or affordable, particularly in underserved areas.

This project proposes an automated skin lesion classification system using a Convolutional Neural Network (CNN) to aid in the early detection of skin cancer. By training the model on a well-established dataset (HAM10000), the system learns to accurately classify skin lesions into one of nine distinct categories. The proposed solution is deployed via a user-friendly web application, allowing users to upload skin lesion images and receive instant predictions. The platform also includes educational content and contact access to dermatologists.

Through this project, we aim to bridge the gap between patients and expert-level diagnostics, improve early detection rates, reduce unnecessary biopsies, and ultimately contribute to lowering skin cancer mortality through accessible, fast, and intelligent technology.

Chapter 3

Proposed system

* 1. Approach used to solve the problem

To address the challenges of early and accurate skin cancer detection, we adopted a **deep learning-based approach** focused on automating the classification of dermoscopic skin lesion images. The methodology involves the following core steps:

**1. Data Acquisition and Understanding**

* We used the **HAM10000 dataset**, which contains over **10,000 dermoscopic images** labeled with 7–9 skin lesion types.
* Each image includes metadata like lesion type, patient age, gender, and anatomical location.

**2. Data Preprocessing**

* All images were resized to a standard shape (e.g., 128×128 pixels) to ensure uniformity during training.
* Normalization was applied to scale pixel values between 0 and 1.

**3. Data Augmentation & Balancing**

* The dataset is imbalanced (e.g., “Melanocytic Nevi” dominates), so we applied **augmentation techniques** such as:
  + Rotation
  + Horizontal flipping
  + Zoom
  + Random cropping
* **Oversampling** was also applied by replicating underrepresented classes to ensure balanced class distribution.

**4. Dataset Splitting**

* The dataset was split into:
  + **70% Training**
  + **15% Validation**
  + **15% Testing**

**5. Model Design and Training**

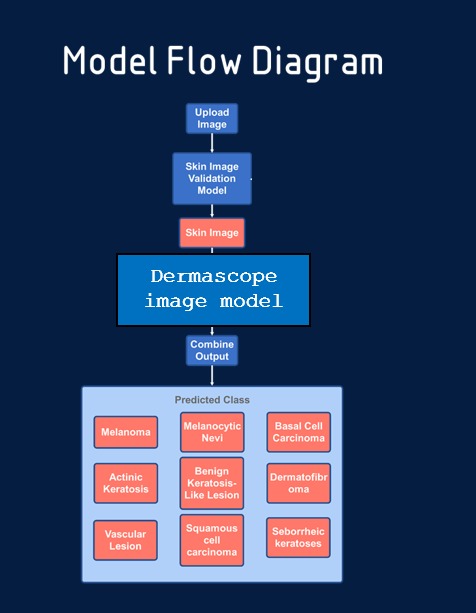
* A **Convolutional Neural Network (CNN)** architecture was implemented.
* The model includes layers such as:
  + Conv2D for feature extraction
  + MaxPooling2D for downsampling
  + Dropout to reduce overfitting
  + Dense and Softmax for classification
* Optimizer: **Adam**
* Loss Function: **Categorical Crossentropy**
* Trained for ~25 epochs

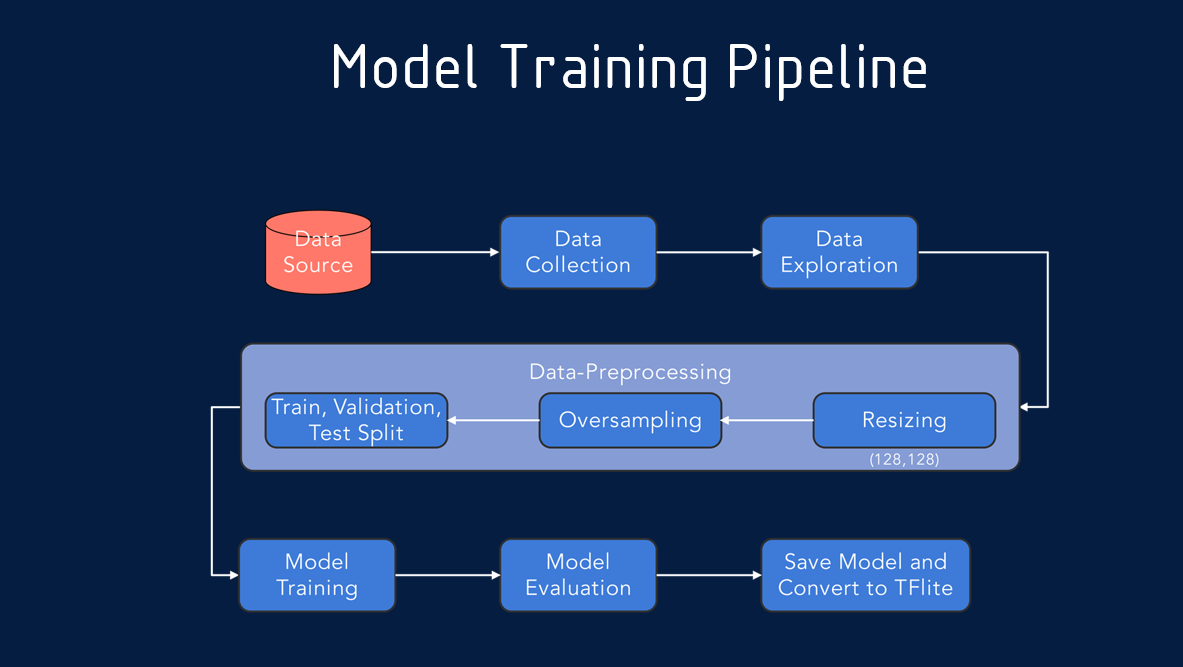
**6. Model Evaluation**

* Evaluated using metrics such as:
  + Accuracy
  + F1-Score
  + Confusion Matrix
  + Loss/Accuracy Curves

**7. Web Deployment**

* The trained model was integrated into a **web application**.
* Users can upload images and instantly receive diagnostic predictions.
* The app also includes:
  + Educational resources on skin cancer types
  + Contact info for dermatologists
  + Lab search functionality
  1. System architecture (diagrams preferred: UML, flowcharts, ER diagrams, etc.)





* 1. Algorithms or frameworks used.

To effectively solve the problem of automatic skin cancer classification, the project leverages a combination of machine learning techniques and modern deep learning frameworks. Below is a breakdown of the key components:

**1. Convolutional Neural Networks (CNNs)**

CNNs are the core algorithm used for classifying dermoscopic skin images. These networks are especially effective in image recognition tasks due to their ability to learn spatial hierarchies and extract visual features like edges, shapes, and textures.

* **Key layers used**:
  + Conv2D: for feature extraction from images.
  + MaxPooling2D: for dimensionality reduction.
  + Dropout: for preventing overfitting.
  + Flatten and Dense: for classification.
  + Softmax: for multi-class output probabilities.
* **Activation Function**: ReLU (Rectified Linear Unit)
* **Loss Function**: Categorical Crossentropy
* **Optimizer**: Adam (Adaptive Moment Estimation)

**🔹 2. Data Augmentation Techniques**

To address dataset imbalance and enhance generalization, image augmentation was applied using:

* Rotation
* Horizontal Flip
* Zoom
* Random Crop

**3. Frameworks and Libraries**

The following tools were used throughout the pipeline:

| **Library / Framework** | **Purpose** |
| --- | --- |
| **TensorFlow / Keras** | Building and training CNN model |
| **OpenCV** | Image loading, resizing, and preprocessing |
| **Scikit-learn** | Data splitting, evaluation metrics |
| **Matplotlib & Seaborn** | Visualization of data and model performance |
| **Pandas & NumPy** | Data manipulation and numerical operations |

**4. Web Development Frameworks (for Deployment)**

For real-time prediction and usability via a web interface:

| **Tool** | **Purpose** |
| --- | --- |
| **Flask API** | Backend API for handling model inference |
| **React.js** | Frontend for user interaction and image upload |
| **TensorFlow.js** | (optional) For in-browser ML processing |

Chapter 4

Implementation

* 1. Technologies, tools, and programming languages used.

**Backend Development:**

* **Programming Language:** Python (for Flask API), JavaScript (for Express.js in some backend modules)
* **Frameworks/Libraries:**
  + Flask (Python web framework for API development)
  + Express.js (Node.js framework for modular backend services)
  + Flask-RESTful (for structuring the API)
  + Flask-JWT/Flask-Login (for authentication)
  + Flask-CORS (for cross-origin resource sharing)
  + Flask-SQLAlchemy (for database integration)
  + Mongoose (MongoDB ODM for database interaction)
  + jsonwebtoken (JWT-based authentication)
  + bcrypt (password hashing)
  + multer (handling image uploads)
  + axios (HTTP requests for model integration)

**Frontend Development:**

* **Programming Language:** JavaScript
* **Frameworks/Libraries:**
  + React.js (for building the user interface)
  + Tailwind CSS (utility-first CSS framework for responsive design)
  + Axios (for HTTP communication with the backend)
  + react-router-dom (for navigation between pages)

**Machine Learning:**

* **Programming Language:** Python
* **Frameworks/Libraries:**
  + TensorFlow (for building and training CNN models)
  + Keras (high-level API for TensorFlow)
  + EfficientNetB0 (pre-trained model for image classification)
  + imblearn (for data oversampling)
  + OpenCV/PIL (for image preprocessing)
  + NumPy, Pandas (for data manipulation and analysis)
  + Matplotlib, Seaborn (for data visualization and EDA)

**Database:**

* **Database System:** MongoDB (NoSQL database hosted on MongoDB Atlas)
* **Tools:**
  + MongoDB Atlas (cloud database hosting)
  + Mongoose (for schema modeling and interaction)

**Deployment:**

* **Platforms:**
  + Render (for deploying the Flask API and backend services)
  + GitHub (for version control and collaboration)
* **Tools:**
  + Postman (for API testing)
  + Git (for version control)

**Development Environment:**

* **IDEs/Tools:**
  + Visual Studio Code (primary IDE for development)
  + Figma (for UI/UX design and prototyping)
  + Jupyter Notebook (for exploratory data analysis and model training)

**Methodologies:**

* Agile Framework (for iterative development and project management)
* Clean Architecture (for modular and maintainable code structure)
* RESTful API Design (for backend services)
  1. Key components/modules of the system.

1. **Skin Photo Validation Model ("Model-1")**
   * **Purpose**: Determines whether the uploaded image is a skin photo or not.
   * **Functionality**: If the image is not identified as skin, the user is prompted to upload a valid skin image.
   * **Dataset**: Combines subsets from ImageNet (for "not skin" images), HAM10000, and Fitz17k (for "skin" images).
   * **Preprocessing**: Resizes images to 128x128 pixels.
   * **Architecture**: A basic CNN model trained for 20 epochs.
2. **Dermascope Image Classification Model ("Model-2")**
   * **Purpose**: Classifies dermascope images into seven types of skin diseases, including skin cancers.
   * **Functionality**: Uses a pre-trained EfficientNetB0 architecture with fine-tuning for high accuracy.
   * **Dataset**: HAM10000 dataset, which includes dermascopic images of skin lesions.
   * **Preprocessing**: Includes resizing, data augmentation, and normalization.
3. **Clinical Image Classification Model ("Model-3")**
   * **Purpose**: Classifies skin lesion images taken under normal conditions (non-dermascope).
   * **Functionality**: Similar to Model-2 but tailored for clinical images.
   * **Dataset**: Likely derived from HAM10000 or similar datasets.
4. **Web Application Interface**
   * **Features**:
     + User authentication (login/register).
     + Image upload and classification using the AI models.
     + Display of classification results.
     + Educational section about skin cancer types.
     + Functionality to locate the nearest medical labs and book appointments.
5. **API and Cloud Deployment**
   * **Technology**: Flask for backend API development.
   * **Functionality**: Handles requests from the frontend, processes images through the models, and returns results.
   * **Deployment**: Hosted on Render for accessibility.
6. **Database**
   * **Purpose**: Stores user data, test results, and clinic information.
   * **Technology**: MongoDB Atlas for cloud-based data management.
7. **Testing and Evaluation Modules**
   * **Types**: Unit testing, integration testing, functional testing, and end-to-end testing to ensure system reliability and accuracy.
   1. Challenges faced and how they were resolved.

**1. Class Imbalance in Datasets**

* **Challenge**: The HAM10000 dataset exhibited significant class imbalance, with Melanocytic Nevi (NV) dominating (67% of the data), while rare classes like Dermatofibroma (DF) and Vascular Lesions (Vasc) had fewer samples. This imbalance risked biasing the model toward majority classes.
* **Resolution**:
  + **Data Oversampling**: Duplicated samples from minority classes to balance the dataset.
  + **Data Augmentation**: Applied techniques like rotation, zooming, and flipping to artificially expand underrepresented classes.

**2. Model Generalization**

* **Challenge**: Initial models overfit the training data, performing poorly on unseen validation/test sets due to limited diversity in training samples.
* **Resolution**:
  + **Transfer Learning**: Used pre-trained EfficientNetB0, fine-tuned on the HAM10000 dataset to leverage learned features and improve generalization.
  + **Regularization**: Added dropout layers and L2 regularization to reduce overfitting.

**3. Computational Resource Constraints**

* **Challenge**: Training deep learning models (e.g., EfficientNetB0) required high computational power, which was limited by available hardware.
* **Resolution**:
  + **Optimized Training**: Reduced image resolution (128x128 pixels) and used batch processing.
  + **Cloud Deployment**: Leveraged cloud platforms (Render) for API hosting and model inference.

**4. Image Quality and Variability**

* **Challenge**: Images varied in lighting, angle, and resolution (e.g., clinical vs. dermoscopic), affecting model accuracy.
* **Resolution**:
  + **Preprocessing Pipeline**: Standardized images via resizing, normalization, and augmentation.
  + **Two-Stage Validation**: Implemented "Model-1" (skin vs. non-skin classifier) to filter invalid inputs before classification.

**5. Integration with the Web Application**

* **Challenge**: Deploying the TensorFlow model into a user-friendly web app while maintaining low latency.
* **Resolution**:
  + **Lightweight Models**: Converted models to TensorFlow Lite for faster inference.
  + **Modular Backend**: Used Flask to create a REST API, enabling seamless communication between the React frontend and ML models.

**6. Data Privacy and Security**

* **Challenge**: Handling sensitive medical data while ensuring compliance with privacy regulations.
* **Resolution**:
  + **Anonymization**: Removed patient identifiers from datasets.
  + **Authentication**: Implemented JWT-based user authentication and encrypted data storage in MongoDB.

**7. Real-World Usability**

* **Challenge**: Ensuring the app’s practicality for both patients and doctors (e.g., unclear UI, lack of doctor-patient interaction features).
* **Resolution**:
  + **Agile Feedback Loops**: Conducted iterative testing with stakeholders to refine features like appointment booking and nearest lab localization.
  + **Educational Component**: Added a skin cancer information section to enhance user engagement and awareness.

Chapter 5

Testing & Evaluation

* 1. Testing strategies (unit testing, integration testing, user testing).

**1. Unit Testing**

* **Purpose**: To validate individual components (e.g., functions, methods, or classes) in isolation.
* **Scope**:
  + Tests for specific functions like user authentication (handelSignUp, handelLogIn), image preprocessing, and model prediction logic.
  + Ensures each module behaves as expected under controlled conditions.
* **Example**:
  + Testing the AuthenticateUser middleware to verify it correctly validates JWT tokens.
  + Validating the resize\_and\_save function in data preprocessing to ensure images are resized without errors.
* **Tools**:
  + Frameworks like Jest or Mocha (for JavaScript/Node.js) or PyTest (for Python) could be used, though the document does not specify tools explicitly.

**2. Integration Testing**

* **Purpose**: To verify interactions between components (e.g., API endpoints, database connections, and model integration).
* **Scope**:
  + Tests the collaboration between the frontend (React), backend (Flask/Express), and database (MongoDB).
  + Validates workflows like uploading an image → processing it with the AI model → returning a classification result.
* **Example**:
  + Testing the /upload-image endpoint to ensure it communicates correctly with the skin cancer classification model and returns valid JSON responses.
  + Verifying that user registration data is correctly stored in the database.
* **Challenges**:
  + Mocking external dependencies (e.g., TensorFlow model calls) to isolate failures.
* **Documentation**:
  + The project mentions testing API endpoints like Register, Login, and AI Test (Section 8.2.3–8.2.5), including edge cases (e.g., empty fields).

**3. User Testing**

* **Purpose**: To evaluate the system’s usability and functionality from an end-user perspective.
* **Scope**:
  + Real-world testing of the web application by patients and doctors to ensure intuitive navigation and accurate results.
  + Feedback on UI/UX (e.g., image upload flow, clarity of cancer classification results).
* **Examples from Document**:
  + **AI Test Detection**: Users upload skin images; the system validates input and provides diagnostic feedback (Section 6.1.3).
  + **Nearest Medical Lab Feature**: Tested for location accuracy and usability (Section 8.2.7).
* **Methods**:
  + Manual testing of user flows (e.g., login → upload image → view results → book appointment).
  + Usability feedback sessions with stakeholders (doctors/patients).

**Key Testing Insights from the Document**

1. **Test Coverage**:
   * Focuses on critical paths (e.g., authentication, image processing) but lacks detail on automated test coverage or CI/CD integration.
2. **Defect Handling**:
   * Errors like invalid image uploads or empty form fields are explicitly tested (Section 8.2.3–8.2.5).
3. **Non-Functional Testing**:
   * Performance (e.g., response time for AI predictions) and reliability are mentioned as requirements (Section 6.2).
   1. Performance metrics (accuracy, speed, scalability, etc.).

**Performance Metrics:**

1. **Accuracy:**
   * **Model-1 (Skin Photo Validation - Basic CNN):**
     + Training Accuracy: **98%**
     + Validation Accuracy: **97.8%**
   * **Model-2 (Dermascope Image Classification - EfficientNetB0):**
     + Training Accuracy: **97%**
     + Validation Accuracy: **93%**
2. **Speed:**
   * The document does not explicitly mention processing speed (e.g., inference time per image), but it highlights that the models are designed for lightweight deployment (e.g., using TensorFlow Lite for Model-1) to ensure efficiency on low-resource devices.
3. **Scalability:**
   * The models are deployed via a **Flask API** and integrated into a **web application**, suggesting they are scalable for multiple users. The use of **EfficientNetB0** (a lightweight architecture) further supports scalability by reducing computational overhead.
4. **Generalization:**
   * The models exhibit strong generalization, with minimal overfitting or underfitting:
     + **Model-1:** Shows stable training and validation accuracy/loss curves.
     + **Model-2:** Validation accuracy (93%) closely follows training accuracy (97%), indicating good generalization despite the dataset's class imbalance.
5. **Additional Metrics:**
   * **Confusion Matrices & Classification Reports:** Provided for both models, showing precision, recall, and F1-scores for each class (e.g., Model-2’s performance across 7 skin lesion types).
   * **Data Augmentation & Oversampling:** Used to address class imbalance, improving model robustness.
   1. Comparison with existing solutions (if applicable).

* **User Interface:** Modern, intuitive, and user-friendly.
* **Image Classification:** Utilizes a chain of three deep learning models (CNN, EfficientNetB0) for high accuracy (93-98%), processing both dermoscopic and clinical images.
* **Data Sources:** Combines HAM10000, Fitz17k, and ImageNet datasets, with data augmentation and oversampling.
* **Functionality:** Offers AI-based classification, nearest lab detection, educational resources, and appointment booking.
* **Accessibility:** Free and accessible via a web application.
* **Target Users:** Patients, dermatologists, and general practitioners.
* **Strengths:** High accuracy, multi-model approach, educational content, and practical tools.
* **Limitations:** Requires internet connectivity for AI analysis.

Chapter 6

Results & Discussion

* 1. Introduction

The project aimed to develop a machine learning-based system for the early detection and classification of skin cancer using convolutional neural networks (CNNs). The system comprises three models:

- \*\*Model-1\*\*: Validates whether an uploaded image is a skin image.

- \*\*Model-2\*\*: Classifies dermoscopic images into seven types of skin lesions.

- \*\*Model-3\*\*: Classifies clinical images (non-dermoscopic) into nine categories of skin diseases.

The project also included the development of a web application to make the system accessible to users, providing features like image upload, classification results, educational resources, and clinic booking.

* 1. Summary of findings.

1. 1. Model Performance:
2. - Model-1 (Skin Validation): Achieved a validation accuracy of 97.8%, demonstrating high reliability in distinguishing skin from non-skin images.
3. - Model-2 (Dermoscopic Classification): Utilized EfficientNetB0 and achieved a validation accuracy of 93%, with strong performance across all seven lesion classes.
4. - The models exhibited good generalization, with no significant overfitting or underfitting.
5. 2. Dataset and Preprocessing:
6. - The HAM10000 dataset was used for training, with preprocessing steps like resizing, oversampling, and augmentation to address class imbalance and enhance model robustness.
7. 3. Application Features:
8. - The web app successfully integrated the models, allowing users to upload images, receive classification results, and access educational content.
9. - Additional functionalities like clinic booking and nearest lab localization were implemented to enhance usability.

6.3 Interpretation of Results

* **Early Detection**: The models achieved high accuracy in classifying skin lesions, facilitating early diagnosis.
* **Accessibility**: The web app provided a user-friendly interface for patients and doctors, bridging gaps in dermatological care.
* **Educational Value**: The app included resources to raise awareness about skin cancer types, symptoms, and prevention.

However, the success of the system depends on the quality of uploaded images and internet connectivity, which could affect real-world performance.

6.4 Limitations of the Proposed Solution

1. **Dataset Bias**: The HAM10000 dataset is imbalanced, with Melanocytic Nevi (67%) dominating, potentially biasing the model.
2. **Image Quality**: The system's accuracy relies on high-quality images, which may not always be available in real-world scenarios.
3. **Scope of Deployment**: The app's effectiveness is limited to regions with internet access and smartphone availability.
4. **Model Generalization**: While the models performed well on the test set, their performance on diverse, unseen data requires further validation.
5. **Ethical Considerations**: The system is a diagnostic aid and not a replacement for professional medical advice, which must be clearly communicated to users.

In conclusion, the project successfully demonstrated the potential of CNNs in skin cancer detection, but ongoing improvements and real-world validation are necessary for broader clinical adoption.

Chapter 7

Conclusion & Future Work

* 1. Summary of contributions.

The project "Skin Cancer Detection Using Machine Learning" represents a significant contribution to the field of medical diagnostics by leveraging advanced deep learning techniques to improve early detection and classification of skin cancer. Key contributions include:

1. **Development of Robust Models**:
   * **Model-1 (Skin Photo Validation)**: A binary CNN classifier to verify whether an uploaded image is a skin image, ensuring data quality for subsequent analysis.
   * **Model-2 (Dermascope Image Classification)**: A fine-tuned EfficientNetB0 model for classifying dermoscopic images into seven types of skin lesions with high accuracy (93% validation accuracy).
   * **Model-3 (Clinical Image Classification)**: A model designed for standard clinical images, expanding accessibility for users without specialized equipment.
2. **Comprehensive Dataset Utilization**:
   * Integration and preprocessing of diverse datasets (HAM10000, Fitz17k, ImageNet) to address class imbalance through oversampling and data augmentation.
3. **User-Centric Application**:
   * A web application with an intuitive interface for users to upload images, receive AI-driven classifications, and access educational resources about skin cancer.
   * Features like locating the nearest medical labs and booking appointments enhance practical usability.
4. **Technical Implementation**:
   * Adoption of modular architecture (Clean Architecture, RESTful API with Flask) for scalability and maintainability.
   * Deployment of lightweight models (e.g., TensorFlow Lite) for efficient performance on low-resource devices.
5. **Educational and Awareness Impact**:
   * Inclusion of detailed information about skin cancer types, symptoms, and prevention strategies to promote public health awareness.
   1. Possible improvements or extensions for future work.

To further enhance the project’s impact, the following improvements and extensions are proposed:

1. **Model Enhancements**:
   * **Expand Datasets**: Incorporate more diverse skin tones and rare lesion types to improve generalizability.
   * **Multi-Modal Inputs**: Combine image analysis with patient metadata (e.g., age, medical history) for more accurate diagnoses.
   * **Real-Time Processing**: Optimize models for faster inference to support real-time applications.
2. **Application Features**:
   * **Multi-Language Support**: Localize the app for non-English speakers to increase accessibility.
   * **Telemedicine Integration**: Enable direct consultations with dermatologists via the platform.
   * **Offline Mode**: Allow basic functionality without internet access for users in low-connectivity regions.
3. **Technical Upgrades**:
   * **Federated Learning**: Train models on decentralized data to preserve privacy while improving accuracy.
   * **Blockchain for Data Security**: Ensure tamper-proof storage of sensitive medical data.
4. **Clinical Validation**:
   * Partner with medical institutions for large-scale clinical trials to validate the system’s diagnostic accuracy.
   * Obtain regulatory approvals (e.g., FDA, CE) for deployment in healthcare settings.
5. **Community and Research**:
   * **Open-Source Collaboration**: Share models and datasets to foster global research collaboration.
   * **Longitudinal Studies**: Track user outcomes to assess the app’s impact on early detection rates.

References

(List all cited works in a standard format, e.g., APA, IEEE, or ACM.)

The project documentation does not explicitly list references in a standardized citation format (e.g., APA, IEEE). However, it mentions several key resources and datasets used, which can serve as references for further reading or citation. Below is a summary of potential references based on the content:

### \*\*Key References from the Project:\*\*

1. HAM10000 Dataset\*

- A publicly available dataset of dermatoscopic images for skin lesion classification.

- Source: [HAM10000 Archive](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T)

2. ImageNet Dataset\*

- Used for the "not skin" class in Model-1.

- Source: [ImageNet Website](https://www.image-net.org/)

3. EfficientNetB0 Architecture\*

- Pre-trained model used for skin lesion classification (Model-2).

- Reference Paper: Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946.

- Available: [arXiv Link](https://arxiv.org/abs/1905.11946)

4. Fitz17k Dataset

- Used for normal skin images in Model-1.

- Likely derived from the Fitzpatrick 17k dataset for skin tone analysis.

- Reference: Groh, M., et al. (2021). \*Fitzpatrick 17k Dataset\*.

5. Tools and Frameworks

- TensorFlow/Keras: For model training.

- Flask: For API deployment.

- React.js: For frontend development.

- MongoDB: For database management.

Additional References (Implied but Not Explicitly Listed):

- Agile Methodology: The project mentions using Agile frameworks; general references could include the \*Agile Manifesto\* or Scrum guides.

- Medical Applications: The "Related Work" section cites apps like \*Skin Vision\* and AI Dermatologist, which could be referenced for comparative analysis.

Suggested Citation Format (Example for HAM10000 Dataset):

> Tschandl, P., Rosendahl, C., & Kittler, H. (2018). \*The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions\*. Scientific Data, 5, 180161. DOI: [10.1038/sdata.2018.161](https://doi.org/10.1038/sdata.2018.161).

Appendices (Optional)

- Additional diagrams, code snippets, user manuals, or datasets.

- Survey questionnaires (if applicable).