EPIGUARD AI - EPITHILIUM DETECTION

A DESIGN PROJECT REPORT

Submitted by

SANJAY RAMAJAYAM M

SANTHOSH T

SUDEESH B

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

K.RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

JUNE 2025

EPIGUARD AI - EPITHILIUM DETECTION

A DESIGN PROJECT REPORT

Submitted by

SANJAY RAMAJAYAM M (811722104131)

SANTHOSH T (811722104134)

SUDEESH B (811722104160)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

K.RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

JUNE 2025

K.RAMAKRISHNAN COLLEGE OF TECHNOLOGY(AUTONOMOUS)
SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report titled **"EPIGUARD AI - EPITHILIUM DETECTION"** is the bonafide work of **SANJAY RAMAJAYAM M (811722104131), SANTHOSH T (811722104134), SUDEESH B (811722104160)** who carried out the project under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or anyother candidate.

SIGNATURE	**SIGNATURE**
Dr. A Delphin Carolina Rani M.E.,Ph.d.,	Mrs. A Thenmozhi M.E.,
HEAD OF THE DEPARTMENT	**SUPERVISOR**
PROFESSOR	**ASSISTANT PROFESSOR**
Department of CSE	Department of CSE
K. Ramakrishnan College of Technology(Autonomous)	K.Ramakrishnan College of Technology(Autonomous)
Samayapuram – 621 112	Samayapuram – 621 112

Submitted for the viva-voce examination held on …………….

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION

We jointly declare that the project report on **"EPIGUARD AI - EPITHILIUM DETECTION"** is the result of original work done by us and best of our knowledge, similar work has not been submitted to **"ANNA UNIVERSITY CHENNAI"** for the requirement of Degree of **BACHELOR OF ENGINEERING**. This project report is submitted on the partial fulfilment of the requirement of the award of Degree of **BACHELOR OF ENGINEERING**.

Signature

____

SANJAY RAMAJAYAM M

____

SANTHOSH T

____

SUDEESH B

Place: Samayapuram

Date:

ACKNOWLEDGEMENT

It is with great pride that we express our gratitude and indebtedness to our institution, "**K RAMAKRISHNAN COLLEGE OF TECHNOLOGY**" for providing us with the opportunity to do this project.

We are glad to credit the honorable Chairman, **Dr. K RAMAKRISHNAN, B.E.,** for having provided the facilities during the course of our study in college.

We would like to express our sincere thanks to our beloved Executive Director, **Dr. S KUPPUSAMY, MBA, Ph.D.,** for forwarding our project and offering an adequate duration to complete it.

We would like to thank **Dr. N VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave the opportunity to frame the project to full satisfaction.

We heartily thank **Dr. A DELPHIN CAROLINA RANI M.E., Ph.D.,** Head of the Department of COMPUTER SCIENCE AND ENGINEERING, for providing her encouragement in pursing this project.

We express our deep and sincere gratitude to our project guide, **Mrs. A THENMOZHI, M.E.,** Department of **COMPUTER SCIENCE AND ENGINEERING,** for her incalculable suggestions, creativity, assistance and patience, which motivated me to carry out this project.

We render my sincere thanks to the Course Coordinator and other staffmembers for providing valuable information during the course. We wish to express my special thanks to the officials and Lab Technicians ofour department who rendered their help during the period of the work progress.

ABSTRACT

In the digital era, the integration of artificial intelligence (AI) in healthcare has become a transformative approach to improving accessibility, affordability, and efficiency of medical services.. To address this problem, this project introduces Epiguard AI, an AI-powered web application designed to assist in the early detection and education of skin diseases. Epiguard AI combines computer vision and natural language processing (NLP**)** technologies to offer two main features: (1) image-based skin disease detection and (2) an educational FAQ-based chatbot. The system utilizes the publicly available HAM10000 dataset—containing over 10,000 dermoscopic images across seven skin disease categories—and applies MobileNetV2, a lightweight but efficient convolutional neural network architecture through transfer learning. The application is deployed using Streamlit, providing an accessible and interactive interface for users to upload images or interact with the chatbot. Through evaluation, the project demonstrates strong generalization capability, ease of use, and potential for future expansion. Planned enhancements include integrating large language models (LLMs) for dynamic chatbot responses and applying data balancing techniques to improve classification fairness across all disease categories.In conclusion, Epiguard AI is a compelling example of how lightweight machine learning frameworks can be effectively applied in dermatology to assist in preliminary diagnosis and public education. It highlights the potential of AI to bridge gaps in healthcare access and offers a scalable foundation for future health tech innovations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS		FULL FORM
AI	-	Artificial Intelligence
BCC	-	Basal-Cell skin Cancer
CNN	-	Convolutional Neural Network
FC Layer	-	Fully Connected Layer
IDLE	-	Integrated Development and Learning Environment
NMSC	-	No Melanoma Skin Cancer
Open CV	-	Open Computer Vision
SCC	-	Squamous-Cell skin Cancer
SPF	-	Sun Protection Factor
RGB	-	Red Green Blue
RAM	-	Random Access Memory
UV	-	Ultra Violet

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Epithilium celiac is a type of cancer that develops in the skin cells. It is one of the most common forms of cancer, with millions of new cases diagnosed each year. The most common types of Epithilium celiac are basal cell carcinoma, squamous cell carcinoma, and melanoma. Basal cell carcinoma is the most common type of Epithilium celiac, accounting for about 80% of all cases. It usually appears as a small, raised bump that is shiny or pearly in appearance. Squamous cell carcinoma is the second most common type of Epithilium celiac, accounting for about 16% of all cases. It typically appears as a red, scaly patch or wart-like growth. Melanoma is the deadliest type of Epithilium celiac, accounting for about 4% of all cases. It usually appears as a dark, irregularly shaped mole that may have an uneven border, uneven color, or change in size over time. The main cause of Epithilium celiac is exposure to ultraviolet (UV) radiation from the sun or other sources such as tanning beds. Other risk factors include having fair skin, a history of sunburns, a weakened immune system, and a family history of Epithilium celiac. Early detection and treatment of Epithilium celiac are critical for successful outcomes. Treatment options may include surgery, radiation therapy, chemotherapy, and immunotherapy.

1.2 OVERVIEW

Epithilium celiacs are cancers that arise from the skin. They are due to the development of abnormal cells that have the ability to invade or spread to other parts of the body. There are three main types of Epithilium celiacs: Basal-Cell (BCC) and Melanoma. The first two, along with a number of less common Epithilium celiacs, are known as no melanoma Epithilium celiac (NMEC) Basal-cell cancer grows slowly and can damage the tissue around it but is unlikely to spread to distant areas or result in death. It often appears as a painless raised area of skin that may be shiny with small blood vessels running over it or may present as a raised area with an ulcer. Squamous-cell Epithilium celiac is more likely to spread. It usually presents as a hard lump with a scaly top but may also form an ulcer. Melanomas are the most aggressive. Signs include a mole that has changed in size, shape, color, has irregular edges, has more than one color, is itchy or bleeds.

1.3 PROBLEM STATEMENT

Epithilium celiac, a type of epithilim celiac, poses a significant health challenge due to its potential to cause serious medical complications, physical disfigurement, and even mortality if not diagnosed in its early stages. Traditional diagnostic methods largely depend on the expertise of dermatologists, which introduces subjectivity and delays, especially in regions with limited access to specialized care. Furthermore, current methods often require invasive testing or physical contact, which may not always be feasible or comfortable for patients. The lack of scalable and automated detection systems contributes to the underdiagnosis and late diagnosis of many epithilim celiac cases. Moreover, there is an increasing incidence of misdiagnosis due to the similarity in appearance between malignant and benign lesions. In rural or resource-constrained settings, the situation is more critical due to a shortage of dermatological tools and professionals. Therefore, there is a pressing need for a robust, accurate, and cost-effective solution that can automatically identify signs of Epithilium celiac from dermoscopic images or live video feeds using advanced technologies like deep learning.

1.4 OBJECTIVES

The primary objective of this document is to present the end-to-end development process, implementation strategy, and evaluation results of Epiguard AI, a machine learning-powered web application aimed at enhancing early detection and public education regarding skin diseases. In response to the challenges faced by many individuals—particularly those with limited access to dermatologists—this project seeks to demonstrate how artificial intelligence can be leveraged to provide a fast, affordable, and user-friendly solution for initial skin disease assessment. The report outlines the motivations behind the project, emphasizing the importance of early diagnosis in preventing complications and reducing healthcare burdens. It also details the technical design of the system, which combines deep learning for image classification with natural language processing for an interactive educational chatbot. By using the HAM10000 dataset and the MobileNetV2 model architecture, the project aims to classify common skin conditions with reasonable accuracy, while also offering users medically relevant information via a rule-based chatbot. This document further serves to evaluate the system's performance, identify areas for improvement—such as class imbalance and minority class detection—and propose future enhancements, including the integration of large language models for more dynamic user interaction. Ultimately, this report is intended not only as a technical summary but also as a proof of concept showcasing the potential of AI in improving dermatological care and public health education.

.

1.5 IMPLICATIONS

The successful implementation of this project has far-reaching implications for the medical and technological community. Firstly, it can revolutionize dermatological screening by providing an AI-assisted tool capable of offering instant diagnostic feedback, thus minimizing the need for immediate specialist intervention. This not only speeds up diagnosis but also enables early detection, which is crucial for effective treatment and survival in cases of epithilim celiac. In rural and underserved regions, the deployment of such a system could bridge the healthcare accessibility gap by empowering general physicians and healthcare workers with a reliable diagnostic assistant. Moreover, the use of deep learning enhances diagnostic accuracy, reducing human errors and increasing confidence in automated assessments. Economically, the proposed system could contribute to significant cost savings by reducing the need for unnecessary biopsies and follow-up procedures, thus lowering the burden on healthcare infrastructure. Technologically, this project serves as a foundation for further research and development in medical image analysis, potentially expanding to other forms of cancer or skin diseases. On a societal level, increased awareness and early detection made possible by such systems could lead to better health outcomes and reduced mortality rates. However, the integration of AI in healthcare also raises concerns about data privacy, algorithmic bias, and the need for regulatory approval, which must be addressed for widespread adoption.

CHAPTER 2

LITERATURE SURVEY

1.Study of Breast Cancer and Epithilium celiac Diagnoses Using Deep Learning Method, Burcu Bilgic - 2024.

A study was conducted on deep learning method that made life important in the early diagnosis of breast cancer and Epithilium celiac, which was proposed by Burcu Bilgic. Breast cancer and Epithilium celiac data were classified as benign and malignant by deep learning method. While working with the deep learning method, the classification was made using the Convolutional Neural Network (CNN) algorithm. In this classification, the data are divided into benign cancer sets and malignant cancer sets. Finally, the data provided by the logistic regression method were analyzed and success charts were created and both types were compared. As a result, accuracy and loss graphs of both cancer types were formed. The aim of the study was to compare breast cancer and Epithilium celiac with the deep learning method and some breast cancer and Epithilium celiac diagnoses are confused, may be in further studies, the basis of differentiating the diagnosis of these two types of cancer will give us a enhanced system.

2.Ultrawideband, Stable Normal and Cancer Skin Tissue Phantoms for Millimeter-Wave Epithilium celiac Imaging, Amir Mirbeik-Sabzevari, Negar Tavassolian – 2024.

Amir Mirbeik et al introduces new, stable, and broadband skin-equivalent semisolid phantoms for mimicking interactions of millimeter waves with the human skin and skin tumors. Realistic skin phantoms serve as an invaluable tool for exploring the feasibility of new technologies and improving design concepts related to millimeter-wave Epithilium celiac detection methods. Normal and malignant skin tissues are separately mimicked by using appropriate mixtures of deionized water, oil, gelatin powder, formaldehyde, TX-150 (a gelling agent, widely referred to as “super stuff”), and detergent. The dielectric properties of the phantoms are characterized over the frequency band of 0.5-50 GHz using a slim-form open-ended coaxial probe in conjunction with a millimeter-wave vector network analyzer. The measured permittivity results show excellent match with ex vivo, fresh skin (both normal and malignant) permittivities determined in our prior work over the entire frequency range. Work by Amir Mirbeik et al results in the closest match among all phantoms reported in the literature to surrogate human skin tissues. The stability of dielectric properties over time is also investigated. The phantoms demonstrate long-term stability (up to 7 months was investigated). In addition, the penetration depth of millimeter waves into normal and malignant skin phantoms was calculated.

3.Deep Learning on Edge Device for EarlyPrescreening of Epithilium celiacs in Rural Communities,John Raiti, Yuntao Wang, Tommy Kuan-Wei Ho, Chen Ma, Chee Jen Ngeh - 2023.

Epithilium celiac is a possible curable disease when detected in the early stage, but Epithilium celiac diagnosis is difficult for people in developing countries where residents lack access to proper healthcare. In this review, John Raiti et al present a low-cost, easy-to-use, and internet-free prescreening solution to detect cancer earlier in rural areas where medical resources are scarce. John Raiti et al deliver a prototype of a device that can classify the skin anomaly into seven major categories and calculate the area segmentation. The prototype John Raiti et al designed includes a Raspberry Pi 3B+, Pi camera, magnifying camera attachment, a convolution neural network powering Epithilium celiac recognition, another network for Epithilium celiac boundary segmentation, and an interactive touchscreen user interface in a custom enclosure. John Raiti trained a MobileNetV2 for Epithilium celiac recognition and a U-Net for Epithilium celiac boundary segmentation on the Epithilium celiac MNIST dataset collected by The International Skin Imaging Collaboration and used it as our Epithilium celiac recognition model.

4. Cloud Infrastructure for Epithilium celiac Scalable Detection System,Pavels Osipovs, Dmitrijs Bliznuks, Ilona Kuzmina - 2023.

Epithilium celiac diagnostics is one of the medical areas where early diagnostic a allows achieving patients' high survival rate. Typically, Epithilium celiac diagnostic is performed by dermatologist, since the amount of such specialists is limited, mortality rate is high. By creating the low cost and easy to use diagnostic device, it is possible to bring Epithilium celiac diagnostic to primary care physicians and allow to check much more persons and diagnose Epithilium celiac on the early stages. There are several existing devices, that provide Epithilium celiac diagnostics. Most of them process the skin images locally and have limited diagnostic capabilities, some of them send images to dermatologists for manual analysis to achieve higher diagnostic quality. Therefore, there is a lack of diagnostic quality or response time. To be able to use the latest diagnostic algorithms and still have fast acting automated diagnostic system, pavels Osipovs et al proposed distributed cloud-based system. In that system, diagnostic device was used only for image acquisition under special multispectral illumination (405nm, 535nm, 660nm and 950nm). Obtained skin images were sent further to cloud system for analysis and diagnostic results visualization. By means of proposed approach, images could be processed by using the same Matlab algorithms that Epithilium celiac research team is using. That will eliminate the need of adopting each algorithm to a specific architecture of diagnostic device. Moreover, the proposed system keeps relation between multiple skin analysis from each patient and could be used to track skin lesions changes in time. Proposed cloud system have architecture that allows fast scaling according to real-time requirements.

5. In Vitro Dielectric Properties of Rat Skin Tissue for Microwave Epithilium celiac Detection,Cemanur Aydinalp, Sulayman Joof, Tuba Yilmaz, Nural Pastaci Ozsobaci, Fatma Ates Alkan, Ibrahim Akduman - 2022

Cemanur Aydinalp et al proposed that dermal tissue characterization based on dielectric properties can be utilized as a non-invasive method for diagnosis of Epithilium celiacs. To enable such technology, there was a need to develop techniques that can rapidly and accurately collect the dielectric properties of the skin tissues. Therefore, the measurement techniques and tools has to be optimized for Epithilium celiac detection. To this end, Cemanur Aydinalp et al presented dielectric property measurements with open-ended coaxial probes that had small apertures customized for detection of Epithilium celiac. Relative permittivity and conductivity of rat skin tissue was characterized with open-ended coaxial probes with outer diameters of 0.9mm and 0.5mm between 0.5GHz-6GHz and the measurement results were compared with the traditional probes had diameter of 2.2mm. The results agreed well with the reported literature data.

6. An Integrated Deep Learning Model for Epithilim celiac Detection Using Hybrid Feature Fusion Technique, S.Sundararajan,A.Rahman – 2021.

A study was conducted to enhance the accuracy of epithilim celiac classification using deep learning methods. The authors proposed a hybrid model that integrates features from two powerful CNN architectures: InceptionV3 and DenseNet121. These features were fused using a weighted sum strategy to leverage their individual strengths. The model was trained on the HAM10000 dataset, which includes a wide range of skin lesion types, and achieved a classification accuracy of 92.27%. The study focused on differentiating between benign and malignant lesions and showed that the hybrid feature fusion significantly improved the diagnostic performance. The research emphasized that combining multiple CNN models can offer better robustness and may lead to clinically usable solutions for early cancer detection.

7. Deep Learning on Edge Device for Early Prescreening of Epithilium Celiacs in Rural Communities ,John Raiti, Yuntao Wang, Tommy Kuan-Wei Ho, Chen Ma, Chee Jen Ngeh - 2019.

This study proposed a low-cost, portable device for the early detection of epithilim celiacs (Epithilium Celiac) in resource-poor environments. The device incorporates a Raspberry Pi, Pi camera, and touchscreen interface powered by deep learning models such as MobileNetV2 and U-Net. The MobileNetV2 model was used for classification while U-Net handled lesion boundary segmentation. The model was trained on the ISIC (Skin Imaging) dataset and tested in real-time scenarios. The research highlighted the importance of making advanced healthcare accessible to rural populations by deploying efficient models on low-power devices. The study confirmed that AI models can support pre-screening tasks with satisfactory accuracy in underserved regions.

8. An Interpretable Deep Learning Approach for Epithilim celiac Categorization, A.Sharif,M.Raza,S.Akram – 2018.

This study focused on the development of an explainable and interpretable AI framework for the classification of skin lesions, including Epithilium Celiac. Using pre-trained CNN architectures such as XceptionNet and EfficientNetV2, the system was trained on publicly available dermoscopic datasets. To increase trust and transparency in the model, the authors employed Grad-CAM and LIME techniques for visual explanations of the predictions. The model achieved 88.72% accuracy in differentiating between benign and malignant skin conditions. The paper emphasized the importance of interpretability in medical AI applications and proposed that combining high accuracy with explainability can promote clinician adoption.

9. Cloud Infrastructure for Epithilium Celiac Scalable Detection System

Pavels Osipovs, Dmitrijs Bliznuks, Ilona Kuzmina - 2018.

The researchers introduced a scalable, cloud-based architecture for the detection of epithilim celiac (Epithilium Celiac). The system separates the image acquisition process from computation by capturing multispectral images (405nm, 535nm, 660nm, and 950nm) locally and sending them to the cloud for analysis. By applying MATLAB-based diagnostic algorithms, the system offers real-time analysis and stores patient history for tracking changes over time. This setup ensures fast diagnostic turnaround and supports a large number of simultaneous users. The paper demonstrated how cloud computing could make high-performance diagnostic tools available even in low-resource clinical environments.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

In existing system skin electrical resistance changes could serve as a diagnostic and therapeutic biomarker associated with physiological changes in patients with malignant versus benign breast cancer lesions.Correlation coefficients were calculated to determine the technique's intrasession and intersession reproducibility.Skin resistance data were used to train a machine learning random forest classification algorithm to diagnose cancer lesions.Skin resistance in patients with breast cancer may serve as a convenient screening tool for cancer and evaluation of therapy

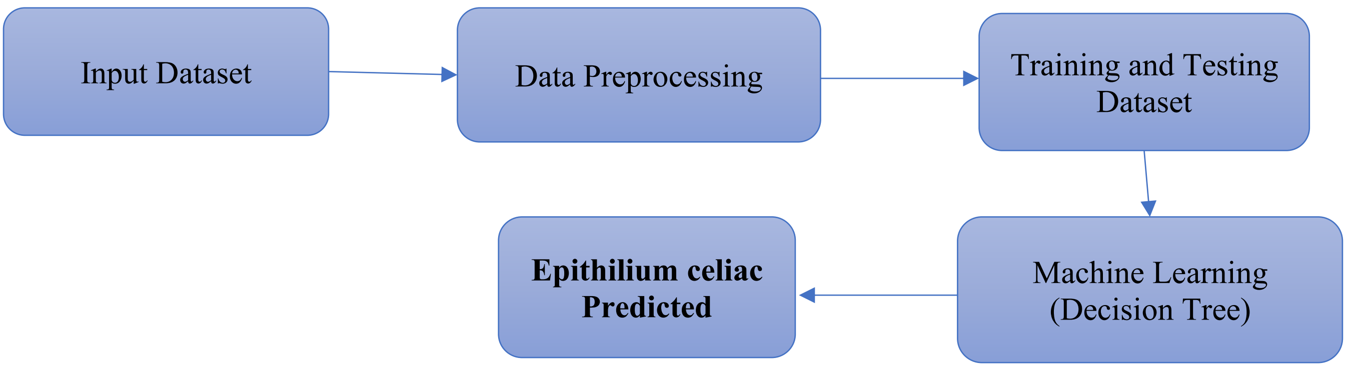

Fig 3.1 EXISTING BLOCK DIAGRAM

3.1.1 DRAWBACKS

- Epithilium celiac can lead to the formation of disfiguring scars or skin grafts. In some cases, the surgery required to remove the cancerous tissue may be extensive and leave a large scar.
- Epithilium celiac can cause physical discomfort and pain, and may affect a person's ability to engage in normal activities. Treatment can be time-consuming and can interfere with daily life.
- If Epithilium celiac is not detected early, it can spread to other parts of the body, including vital organs, which can be life-threatening.

3.2 PROPOSED SYSTEM

In this paper we approach a new method of Deep learning Algorithm Convolutional Neural Network for the detection of Epithilium celiaca large brownish spot with darker speckles. A skin that changes in color, size or feel or that bleeds. A small lesion with an irregular border and portions that appear red, pink, white, brown or black. A painful lesion that itches or burns. The proposed system overcomes the existing contact methods and replaces it by high definition imaging cameras.

3.2.1 ADVANTAGES:-

- The Feature Extraction gives us a broad view about the image which is captured and help us to process the image for preprocessing.
- The given system has overcomes the errors and has higher efficiency than the current image processing Methods.
- By the usage of multiple hidden layers such as conv2D, maxpoolD, Flatten and Dense the area of the oil spill is detected, the area is calculated and it can be viewed by the user in the shell of python.

3.3 PROPOSED SYSTEM ARCHITECTURE:-

The proposed architecture diagram illustrates the workflow of the Epiguard AI skin disease detection and chatbot system. The process begins with the collection and division of the HAM10000 dataset into training and testing sets. These images undergo dataset preprocessing, which includes resizing, normalization, and augmentation. Following preprocessing, the data is processed using machine learning libraries such as TensorFlow, Keras, and OpenCV for model training, while PIL is used for image handling, and NLP matching handles chatbot functionality. After training, the system is prepared for testing. When a new image is input by the user, it is passed through the MobileNetV2 algorithm, a pre-trained and fine-tuned deep learning model used for efficient and accurate image classification.This pipeline ensures a smooth integration of deep learning for computer vision and natural language processing for user interaction, forming a complete intelligent dermatology assistant system

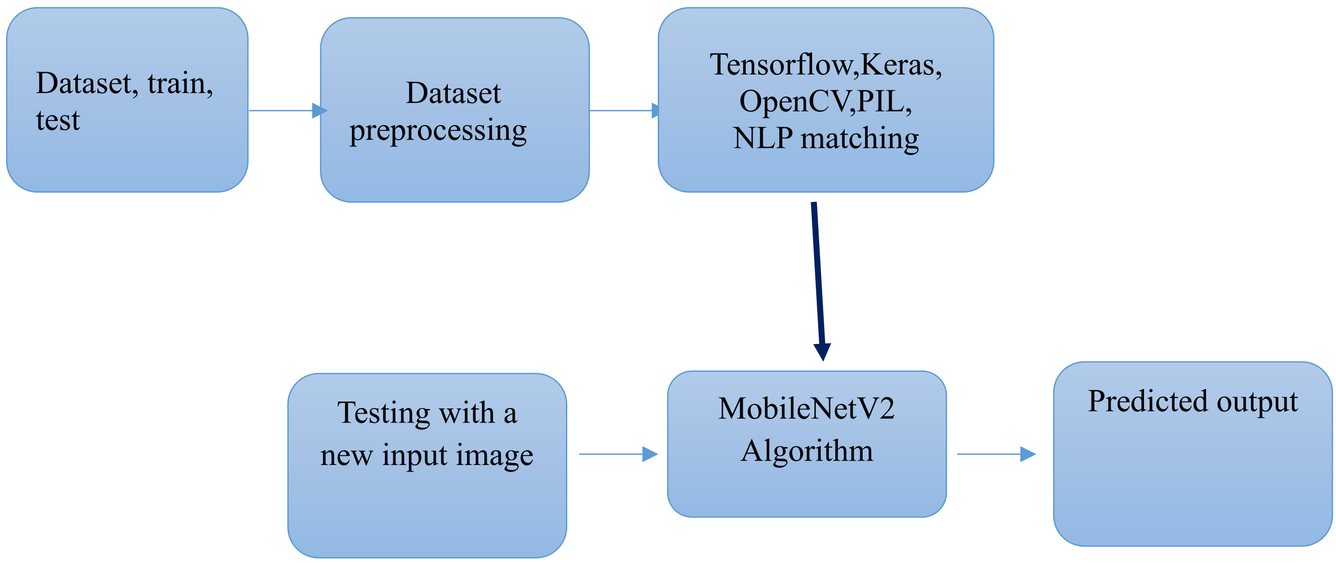

Fig. 3.2 PROPOSED ARCHITECTURE DIAGRAM

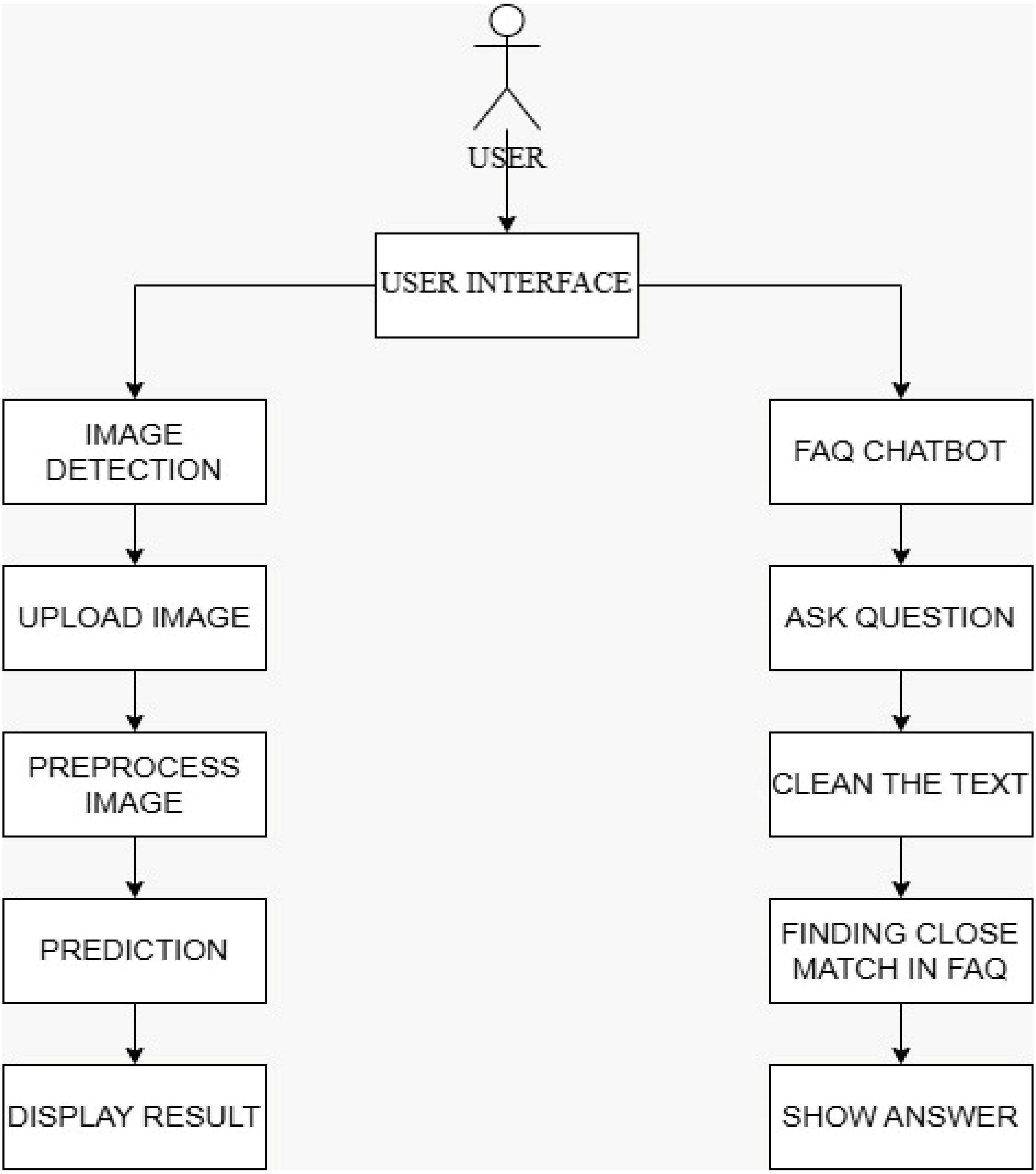

Fig. 3.3 FLOW CHART DIAGRAM

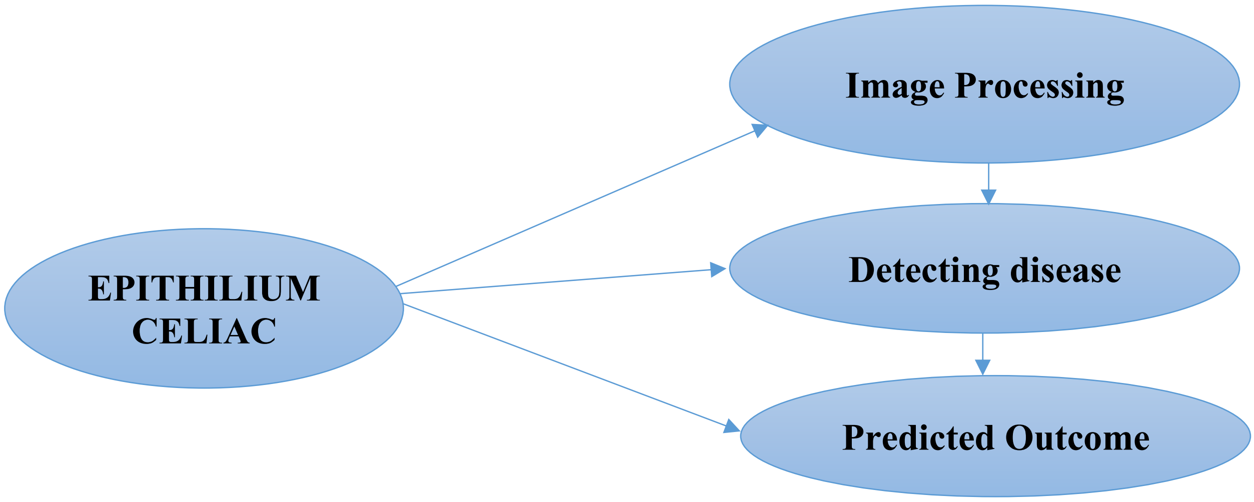

FIG.3.4 USE CASE DIAGRAM

CHAPTER 4

MODULES

4.1 MODULE DESCRIPTION

The Epiguard AI system is composed of five key modules working together to deliver an AI-powered dermatology assistant. The User Interface Module provides an intuitive web-based platform for users to upload skin images and ask health-related questions. The Image Processing and Prediction Module manages image preprocessing and uses a MobileNetV2 deep learning model to classify skin diseases and provide confidence scores. The Chatbot Module responds to user queries using a rule-based system that matches questions with a predefined FAQ dataset.

- User Interface Module
- Image Processing and Prediction Module
- Model Training and Evaluation Module
- Chatbot Module
- Deployment and Integration Module

4.1.1 USER INTERFACE MODULE

The User Interface (UI) module serves as the primary interaction point between the user and the Epiguard AI system. Built using Streamlit, this module provides a web-based platform that is both intuitive and responsive, designed to be accessible even to users with minimal technical knowledge. It allows users to upload images of skin conditions for analysis and input questions regarding skin diseases for chatbot responses. The UI displays the prediction results, including the detected skin condition and its associated confidence score, and also returns chatbot answers in a conversational format. This module plays a critical role in usability and adoption, as it determines how easily and effectively users can leverage Epiguard AI for skin disease detection and health information.

4.1.2 IMAGE PROCESSING AND PREDICTION MODULE

This module is at the core of Epiguard AI's diagnostic capability. It handles the entire pipeline from image input to disease classification output. When a user uploads a skin image, the system first preprocesses it by resizing it to a standard input dimension (224x224 pixels) and applies data augmentation techniques such as rotation, flipping, and zooming to enhance model robustness. The preprocessed image is then fed into a deep learning model—specifically, MobileNetV2, a lightweight yet powerful convolutional neural network architecture optimized through transfer learning. This model outputs a prediction that includes the most likely skin disease class and a confidence score indicating the model's certainty. By using pre-trained knowledge and fine-tuning for the specific HAM10000 dataset, the module achieves a reasonable balance between performance and computational efficiency, enabling real-time predictions in a web environment. This module is essential for delivering accurate medical insight at the user's fingertips.

4.1.3 MODEL TRAINING AND EVALUATION MODULE

The Model Training and Evaluation module governs the backend machine learning lifecycle of the project. Using the HAM10000 dataset—which consists of 10,000 dermoscopic images labeled with seven different skin disease types—this module performs structured data preprocessing, augmentation, and splitting into training, validation, and testing sets. The MobileNetV2 architecture is used for model development, leveraging transfer learning to expedite training while ensuring high accuracy. These metrics help in identifying both the strengths and limitations of the model—such as its tendency to perform better on majority classes like 'nevus' and less effectively on critical but underrepresented classes like 'melanoma'. This module also lays the groundwork for future enhancements, such as class balancing, advanced regularization, or use of alternative architectures.

4.1.4 CHATBOT MODULE

The Educational Chatbot module complements the diagnostic functionality of Epiguard AI by providing users with informative, conversational responses to common questions about skin conditions. Instead of relying on AI language generation models, this module uses a rule-based approach that matches user queries with predefined questions in a structured JSON FAQ dataset. To ensure accuracy and robustness, the module preprocesses user input by applying natural language processing steps such as lowercasing and stemming. Once the best match is found, the corresponding answer is presented to the user. This approach ensures fast and contextually relevant answers while keeping the system lightweight and easy to maintain. The chatbot is designed to be educational, helping users learn about skin diseases such as melanoma, basal cell carcinoma, and others. Future plans include integrating this module with large language models to support more flexible and intelligent conversations.

4.1.5 DEPLOYMENT AND INTEGRATION MODULE

The Deployment and Integration module brings all components of Epiguard AI together into a cohesive, functional application accessible to users via the web. This module ensures that the trained model, image processing pipeline, chatbot logic, and user interface all work in synchrony. Streamlit is used not only as a front-end interface but also as the deployment framework, allowing the entire application to run seamlessly on local servers or be hosted on Streamlit Cloud. This module also includes version control using Git and GitHub, facilitating collaborative development and updates. Scalability, maintainability, and performance optimization are key concerns addressed here, with future improvements potentially including containerization with Docker or migration to cloud platforms like AWS or GCP for broader accessibility and improved performance. This module is the bridge between development and real-world applic

CHAPTER 5

SOFTWARE DESCRIPTION

5.1. PYTHON

Python is the foundational programming language for the Epiguard AI project. Known for its clear syntax and powerful capabilities, Python is widely used in the field of artificial intelligence, data science, and web development. In this project, Python acts as the glue connecting all components — from handling image data and training deep learning models to building an interactive chatbot and deploying the web interface. Its large ecosystem of libraries, strong community support, and compatibility with major AI frameworks make it an ideal choice. Python simplifies complex tasks like image processing, machine learning pipeline construction, and user interface deployment, allowing the development team to focus on delivering functionality rather than managing low-level technical details.

5.2 TENSORFLOW & KERAS

TensorFlow is a robust open-source platform developed by Google for building machine learning and deep learning applications, while Keras is a high-level API that runs on top of TensorFlow, designed for ease of use and fast experimentation. MobileNetV2 is chosen for its lightweight yet effective structure, making it suitable for real-time prediction tasks even in resource-constrained environments like web applications. The model is trained using transfer learning, leveraging pre-trained weights from large datasets to recognize patterns in skin disease images with minimal training time. TensorFlow handles the backend computations efficiently, including gradient updates and model optimization, while Keras provides a user-friendly syntax to design and fine-tune the network. Together, they empower Epiguard AI to achieve reliable image classification performance.

5.3 STREAMLIT

Streamlit is a modern open-source Python framework designed to simplify the creation of interactive web applications for machine learning and data science. In the Epiguard AI project, Streamlit serves as the front-end interface that allows users to interact with the system effortlessly through a web browser. It enables the integration of two main features — image-based skin disease prediction and the chatbot for dermatological education — within a clean and responsive layout. With minimal code, Streamlit supports dynamic content updates, file uploads (such as medical images), and real-time display of prediction results, including disease name and confidence scores. It also supports chatbot interaction where users can input questions and receive educational responses. Streamlit's ability to deploy applications locally or on the cloud further enhances accessibility, making Epiguard AI available to a broader audience.

5.4 SCIKIT-LEARN

Scikit-learn is one of the most widely used machine learning libraries in Python, offering efficient tools for model evaluation, data analysis, and preprocessing. Within Epiguard AI, Scikit-learn plays a critical role in evaluating the performance of the deep learning model. It provides functions to compute a wide range of metrics including accuracy, precision, recall, F1-score, and confusion matrices. These metrics are particularly important for medical classification tasks where accuracy alone may not be sufficient due to class imbalance — for instance, correctly identifying rare diseases like melanoma is far more critical than classifying common conditions. Scikit-learn also supports visualization tools that help in interpreting the model's strengths and weaknesses. The insights gained from these evaluations guide further tuning, balancing strategies, and improvements in model architecture to ensure reliable diagnosis support.

5.5 OPENCV AND PIL (PILLOW)

OpenCV (Open Source Computer Vision Library) and PIL (Python Imaging Library, also known as Pillow) are essential libraries for image processing in the Epiguard AI pipeline. This includes resizing images to the required 224x224 pixel dimension, converting file formats, and applying data augmentation techniques such as rotation, flipping, and zooming. These preprocessing steps are crucial to improve the model's ability to generalize from limited datasets and to reduce overfitting. OpenCV's advanced functions like edge detection and color space transformations can also be extended in future development for enhanced image analysis capabilities.

5.6 DIFFLIB

Difflib is a Python module for comparing sequences, widely used in text processing applications. In Epiguard AI, Difflib powers the educational chatbot that responds to user questions regarding skin diseases. The chatbot is built using a rule-based approach, where a list of frequently asked questions (FAQs) and their answers are stored in a structured format (typically JSON). This method is lightweight and does not require heavy NLP models, making it suitable for quick deployment and low-resource environments. Although simple, this chatbot provides an efficient and responsive way to deliver accurate health-related information and can later be enhanced with more sophisticated Natural Language Processing (NLP) models or integrated with Large Language Models (LLMs) for greater conversational flexibility.

CHAPTER 6
TEST RESULT AND ANALYSIS

6.1 TESTING

The Epiguard AI project was tested in two main parts: the skin disease image detection model and the chatbot. For the image detection part, the model was trained using a dataset called HAM10000, which contains over 10,000 pictures of different skin diseases grouped into seven categories. The model used a deep learning method called MobileNetV2, which was already trained on a large image dataset. After training, the model was tested and reached about 74% accuracy, which means it correctly identified skin diseases most of the time. The accuracy during training was around 76%, and both training and validation losses decreased smoothly. This shows the model was learning well and was not overfitting.

However, the model did better on skin diseases that had more images in the dataset, especially the most common type called “nv.” It did not perform as well on rare diseases like melanoma or akiec, which had fewer images. This is a common issue in machine learning when one class has more examples than others. To fix this, improvements like adding more images for rare diseases or adjusting how the model learns can be used in the future.

The second part of the project was the chatbot, which helps answer questions about skin diseases. It uses a simple method to match the user’s question to a list of pre-written questions and answers. Even though it’s a basic system, it runs fast and doesn’t use a lot of resources. In the future, the chatbot can be improved by using more advanced language models that better understand what users are asking.

Overall, the project worked well and showed that AI can help in identifying skin diseases and educating users. There is still room for improvement, especially in handling rare diseases and making the chatbot smarter, but it’s a strong starting point for building useful healthcare tools.

6.2 TEST OBJECTIVES:

The objective of this project is to design and implement a comprehensive AI-powered health application, Epiguard AI, that serves two core functions: automated skin disease detection from images and an interactive educational chatbot. The detection system leverages deep learning, specifically a fine-tuned MobileNetV2 model, to classify dermatological conditions from user-submitted skin images.

This approach not only empowers users with early insights into potential skin issues but also bridges the gap in access to dermatological expertise, especially in remote or underserved areas. On the other hand, the chatbot component, built using a rule-based NLP system and an FAQ knowledge base, provides reliable and instant answers to common questions related to skin health. By simulating a conversational interface, it aims to educate users on symptoms, causes, and treatments in an engaging and accessible manner.

The project also focuses on scalability and real-world application by deploying the system through a user-friendly web interface using Streamlit. In addition to technical performance, such as achieving a balanced model accuracy and minimizing bias toward certain skin conditions, the project also emphasizes ethical considerations like data privacy, model transparency, and inclusivity in healthcare technology. Through this initiative, Epiguard AI aims to become a valuable digital health assistant that not only supports clinical workflows but also promotes proactive health management and informed self-care among the general public.

6.3 PROGRAM TESTING

Program testing in the Epiguard AI project was conducted to verify that each component of the application performed as expected. The testing approach included both unit testing and integration testing. Unit testing was carried out on specific modules such as the image uploader, preprocessing pipeline, and chatbot string-matching logic to ensure that each performed correctly in isolation. Integration testing ensured that all components—image classification, chatbot logic, and user interface—functioned together seamlessly within the web application. Additionally, user testing was performed to simulate real-world interactions, helping validate the end-to-end experience of submitting a skin image and receiving disease predictions, or asking health-related questions and receiving informative answers via the chatbot. During testing, scenarios with invalid inputs (e.g., unsupported file formats, blank questions) were also evaluated to confirm that the application handles errors gracefully.

6.4 TESTING AND CORRECTNESS

Correctness was evaluated through the use of well-established performance metrics. For the image classification model, the correctness of predictions was validated using metrics such as accuracy, precision, recall, and F1-score. The model achieved approximately 74% accuracy on the test dataset, indicating that it correctly predicted the skin disease in a majority of the test cases.

The classification model was particularly strong in detecting common classes like nevus (nv) but showed reduced performance on minority classes such as dermatofibroma (df) and vascular lesions (vasc). This imbalance influenced precision and recall metrics, especially for underrepresented categories. On the chatbot side, correctness was determined based on how accurately the chatbot responded to user queries. It correctly matched user input to predefined questions using string similarity, though it was limited in flexibility for variations in phrasing. These observations confirm that the system functions correctly in expected conditions but leaves room for further improvement.

6.5 ANALYSIS

From the analysis of testing results, several insights were gathered. The image classification model showed good generalization performance, with training and validation accuracies closely aligned (around 76% and 74% respectively), indicating minimal overfitting. However, class imbalance remains a significant issue, as shown by lower F1-scores for rare classes. To address this, techniques such as SMOTE (Synthetic

Minority Over-sampling Technique) or class weighting are proposed to improve performance on underrepresented categories.

Furthermore, the chatbot system, while functional, relies on a rule-based method that limits its ability to understand diverse natural language inputs. Future iterations could integrate transformer-based LLMs (e.g., GPT) for more dynamic and context-aware responses. Overall, while the current system demonstrates a strong baseline, continuous refinement in data balancing, model tuning, and natural language understanding is necessary to elevate the tool to a production-ready health AI assistant.

CHAPTER 7

RESULT AND DISCUSSION

7.1 RESULT

The Epiguard AI project successfully implemented a comprehensive and dual-function AI-driven web application aimed at supporting early detection and awareness of skin diseases. The application integrates two core modules: an image-based classification system for diagnosing dermatological conditions and an interactive chatbot designed to provide users with accurate and accessible dermatological information.

The image classification component leverages the HAM10000 dataset, a well-known and diverse dataset of dermatoscopic images containing over 10,000 labeled examples of skin lesions categorized into seven distinct classes, including melanoma, nevus, and actinic keratoses. The project adopted MobileNetV2, a state-of-the-art, lightweight convolutional neural network (CNN) architecture optimized for performance and efficiency, especially in web-based and mobile environments. Through the use of transfer learning and model fine-tuning, the image classifier achieved an overall accuracy of approximately 74%, which reflects a solid generalization capability across a wide range of common skin conditions.

The model's performance was validated using training and validation accuracy/loss curves, which demonstrated a stable learning process without signs of overfitting. The accuracy gap between the training and validation sets remained minimal, indicating that the model maintained good predictive consistency even on unseen data. Further analysis using metrics such as the confusion matrix and classification report confirmed that the classifier performed particularly well on majority classes like melanocytic nevi (nv), while performance on rarer classes such as dermatofibroma (df) and vascular lesions (vasc) indicated areas for potential improvement, such as through data augmentation or class balancing techniques.

Overall, the results of the Epiguard AI project demonstrate that a lightweight, accessible AI application can effectively assist in early skin disease screening and public dermatological education, especially in contexts where access to dermatologists is limited.

7.2 CONCLUSION

In conclusion, the Epiguard AI project demonstrates the practical application of machine learning and natural language processing in enhancing early access to dermatological information and pre-diagnosis services. This system provides an innovative solution to challenges faced by individuals in remote or underserved areas who may not have timely access to dermatologists. By allowing users to upload skin images for disease classification and interact with a chatbot for basic skin health information, the application addresses two critical pain points: early detection and public awareness.

While the current model performs well on majority classes such as “nv” (nevus), its performance on minority and critical classes like “mel” (melanoma) or “df” (dermatofibroma) still needs improvement. This limitation highlights the challenges of class imbalance in medical datasets and reinforces the importance of strategic improvements such as data balancing and class-specific tuning.

7.3 FUTURE ENHANCEMENT

- Model Improvement and Data Balancing: Incorporating techniques such as SMOTE (Synthetic Minority Oversampling Technique) or class weighting during training can improve model sensitivity to rare but critical skin diseases. Additionally, acquiring a more balanced or expanded dataset would be beneficial.
- Integration of Large Language Models (LLMs): Replacing or augmenting the rule-based chatbot with a context-aware generative LLM (like GPT-based models) can enable more natural, dynamic, and flexible responses to user queries, expanding beyond static FAQs.
- User Interface & Experience (UX) Enhancements: Improving the visual design, responsiveness, and usability of the Streamlit interface—especially for mobile devices—can make the platform more user-friendly for a broader audience.
- Real-Time API Deployment: Hosting the model and chatbot on cloud platforms (e.g., Streamlit Cloud or AWS) with real-time prediction APIs can improve scalability and accessibility for public use.
- Security & Privacy Features: Implementing user data encryption and compliance with healthcare data privacy regulations (like HIPAA or GDPR) will be essential for real-world deployment, especially when dealing with personal medical images.
- Multi-language Support and Localization: Adding support for multiple languages and culturally relevant medical FAQs can help reach users across different regions and literacy levels.

APPENDIX – 1

SOURCE CODE

```
import streamlit as st
import numpy as np
import tensorflow as tf
from PIL import Image
import json
import difflib
import re


# Load model & data
MODEL_PATH = "models/skin_model_final_3.keras"
FAQ_PATH = "data/faq-chatbot-skin.json"


model = tf.keras.models.load_model(MODEL_PATH)


CLASS_NAMES = {
    0: "akiec", 1: "bcc", 2: "bkl", 3: "df", 4: "mel", 5: "nv", 6: "vasc"
}
CLASS_DESC = {
    "akiec": "Actinic Keratoses",
    "bcc": "Basal Cell Carcinoma",
```

```
    "bkl": "Benign Keratosis",
    "df": "Dermatofibroma",
    "mel": "Melanoma",
    "nv": "Melanocytic Nevi",
    "vasc": "Vascular Lesion"
}

# Load FAQ data
with open(FAQ_PATH, "r", encoding="utf-8") as f:
    faq_data = json.load(f)

# Function to clean user input for chatbot
def clean_text(text):
    return re.sub(r"[^\w\s]", "", text.lower())

def chatbot_response(user_input):
    cleaned_input = clean_text(user_input)
    if any(bad in cleaned_input for bad in faq_data.get("blacklist", [])):
        return "Please use polite language."

    questions = []
    q_to_a = {}
```

```
    for cat, qas in faq_data.items():
        if cat == "blacklist": continue
        for qa in qas:
            q_clean = clean_text(qa["question"])
            questions.append(q_clean)
            q_to_a[q_clean] = qa["answer"]

    matches = difflib.get_close_matches(cleaned_input, questions, n=1, cutoff=0.6)
    return q_to_a[matches[0]] if matches else "Apologies, I’m not sure I understand your
question yet."

# Streamlit UI
st.set_page_config("EpiGuard AI - Epithelium Detection", layout="wide")

# Sidebar
with st.sidebar:
    st.image("assets/epigaurd-logo.jpeg", caption="EpiGuard AI - Epithelium
Detection", use_container_width=True)
    st.title("Welcome to EpiGuard AI")

    st.markdown("""
    The EpicGaurd application is designed to help you detect skin diseases and provide
```

```
educational information about skin health.

    🌟 Main Features of EpiGaurd AI:
        Image Detection: Upload an image of your skin and let Epiguard AI provide a prediction about the type of skin disease.

        Chatbot FAQ: Get information related to skin diseases by asking our chatbot.
    """)

    # Choose feature (using radio button)
    feature = st.radio("Choose a Feature", ["📸 Image Detection", "💬 Ask the Chatbot"])

# Main Content (Tab Image Detection & Chatbot FAQ)
if feature == "📸 Image Detection":
    st.title("Epithelium Detection (Upload Image)")
    st.markdown("""
    Upload an image of your skin to detect the type of skin disease.
    The system will provide a prediction based on the uploaded image.
    """)
```

```
file = st.file_uploader("Select an image (jpg/png)", type=["jpg", "jpeg", "png"])

if file:
    try:
        # Process image
        img = Image.open(file).convert("RGB").resize((224, 224))
        st.image(img, caption="Uploaded Image", width=300)

        # Prepare image for prediction
        img_arr = np.expand_dims(np.array(img) / 255.0, axis=0)
        pred = model.predict(img_arr)
        class_id = int(np.argmax(pred))
        confidence = float(np.max(pred))

        class_code = CLASS_NAMES[class_id]
        st.success(f"🧠 Prediction: {CLASS_DESC[class_code]} (`{class_code}`)")
        st.metric(label="Confidence", value=f"{confidence*100:.2f} %")

        with st.expander("ℹ About this result"):
            st.markdown(f"""
            **{CLASS_DESC[class_code]}** is a type of skin disease that can be
identified by certain characteristics.
```

```
                This is only an initial prediction — please consult a dermatologist for an official diagnosis.
            """)
        except Exception as e:
            st.error(f"❌ Failed to process the image. Error: {str(e)}")

elif feature == "💬 Ask the Chatbot":
    st.title("Chatbot FAQ Skin")
    st.markdown("""
    Ask about skin diseases!
    Get information through our chatbot that will provide answers to your questions.
    """)

    if "chat_history" not in st.session_state:
        st.session_state.chat_history = []

    user_question = st.text_input("Ask something...")

    col1, col2 = st.columns([1, 5])
    with col1:
        if st.button("Send") and user_question:
```

```
        answer = chatbot_response(user_question)

        st.session_state.chat_history.append((user_question, answer))

with col2:

    if st.button("🔄 Reset Chat"):

        st.session_state.chat_history = []


for q, a in reversed(st.session_state.chat_history):

    st.markdown(f"**🧑 You:** {q}")

    st.markdown(f"**🤖 Bot:** {a}")

    st.markdown("---")
```

APPENDIX – B

SCREENSHOTS

Sample Output

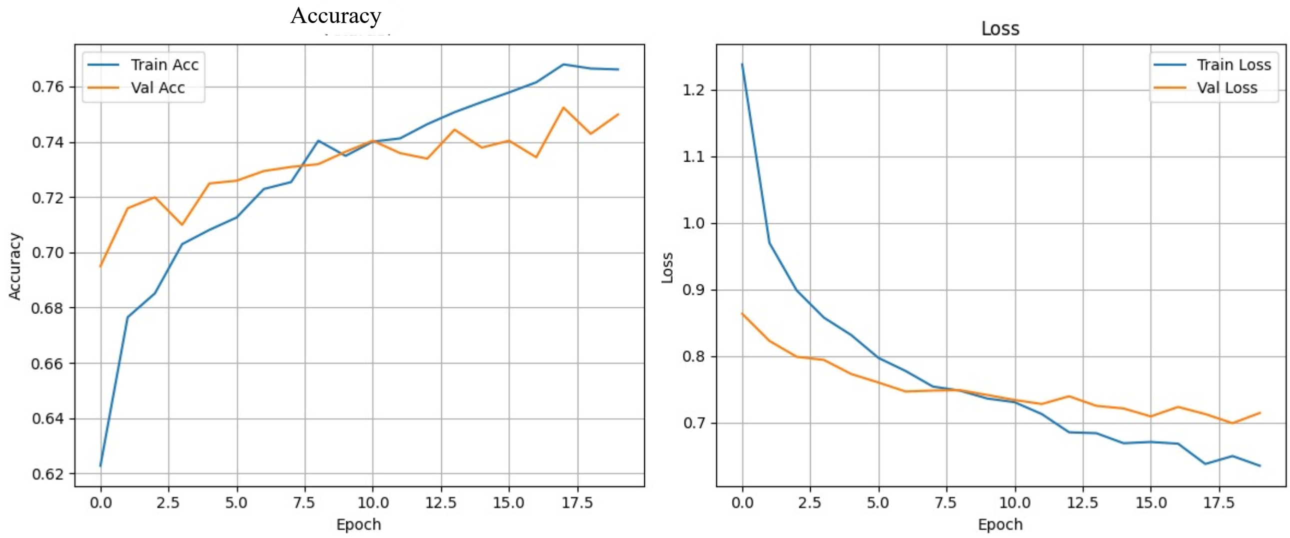

Fig B.1 : Accuracy and Loss

=== Classification Report ===

	precision	recall	f1-score	support
akiec	0.83	0.08	0.14	65
bcc	0.59	0.35	0.44	103
bkl	0.48	0.40	0.43	220
df	0.00	0.00	0.00	23
mel	0.48	0.39	0.43	223
nv	0.82	0.95	0.88	1341
vasc	0.72	0.46	0.57	28
accuracy			0.75	2003
macro avg	0.56	0.38	0.41	2003
weighted avg	0.72	0.75	0.72	2003

Fig B.2 : Classification Report

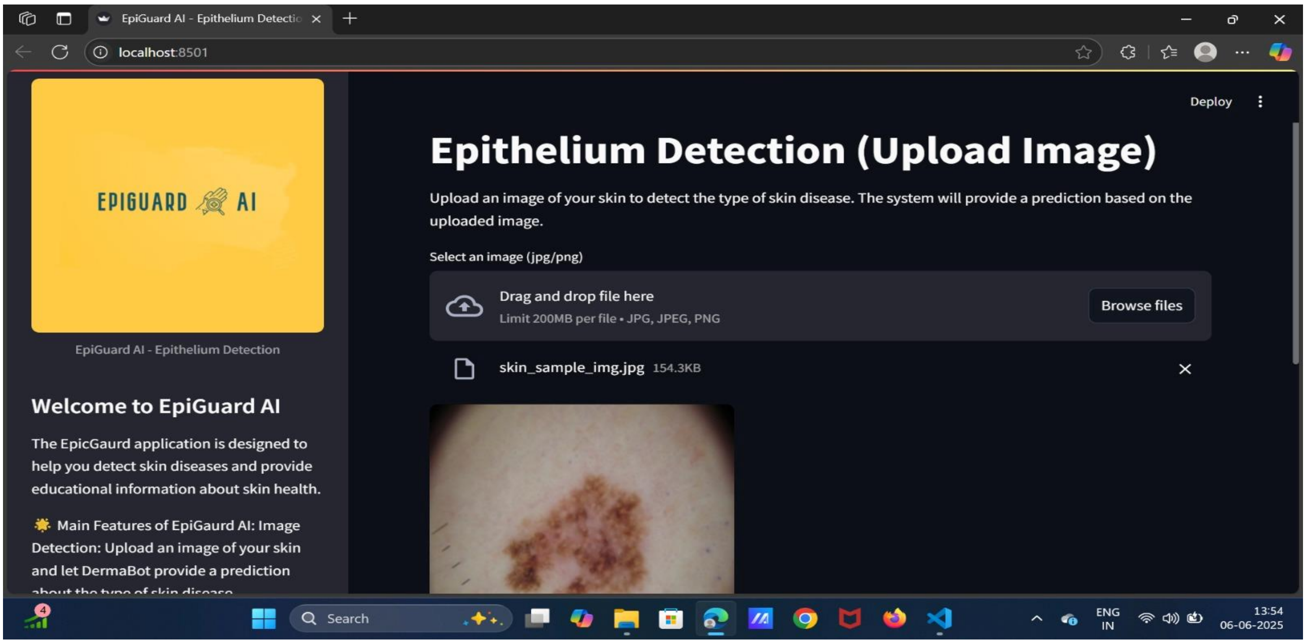

Fig B.3 : Image Upload

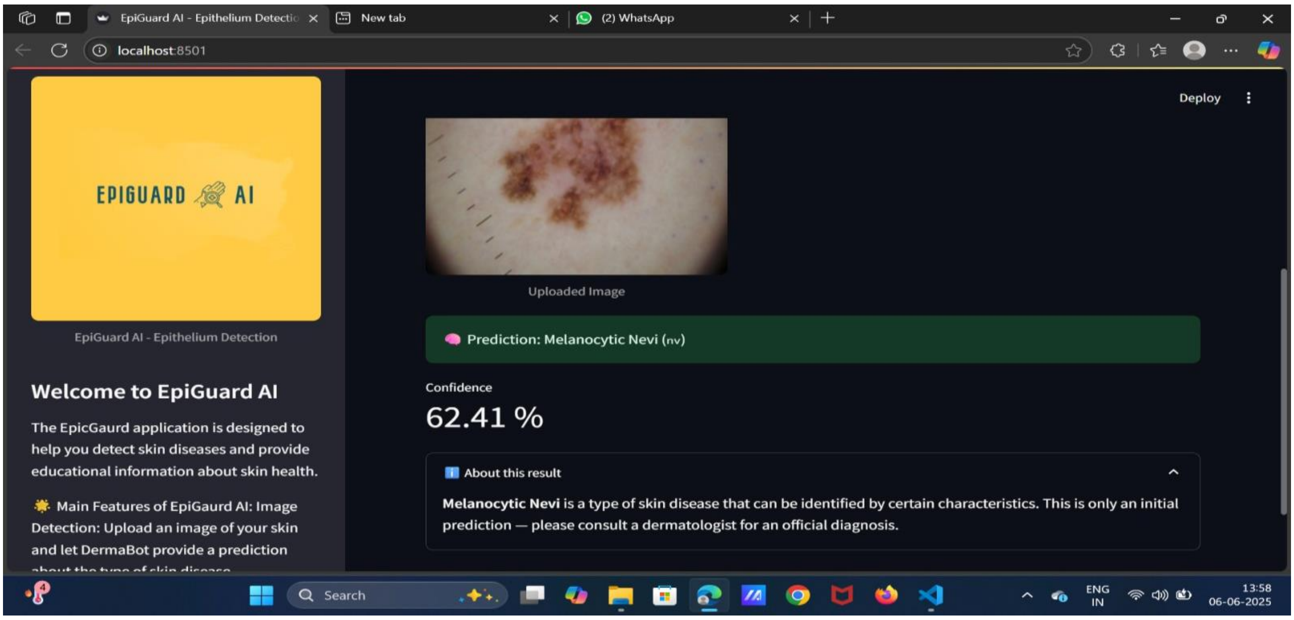

Fig B.4 : Epithelium Detection Page

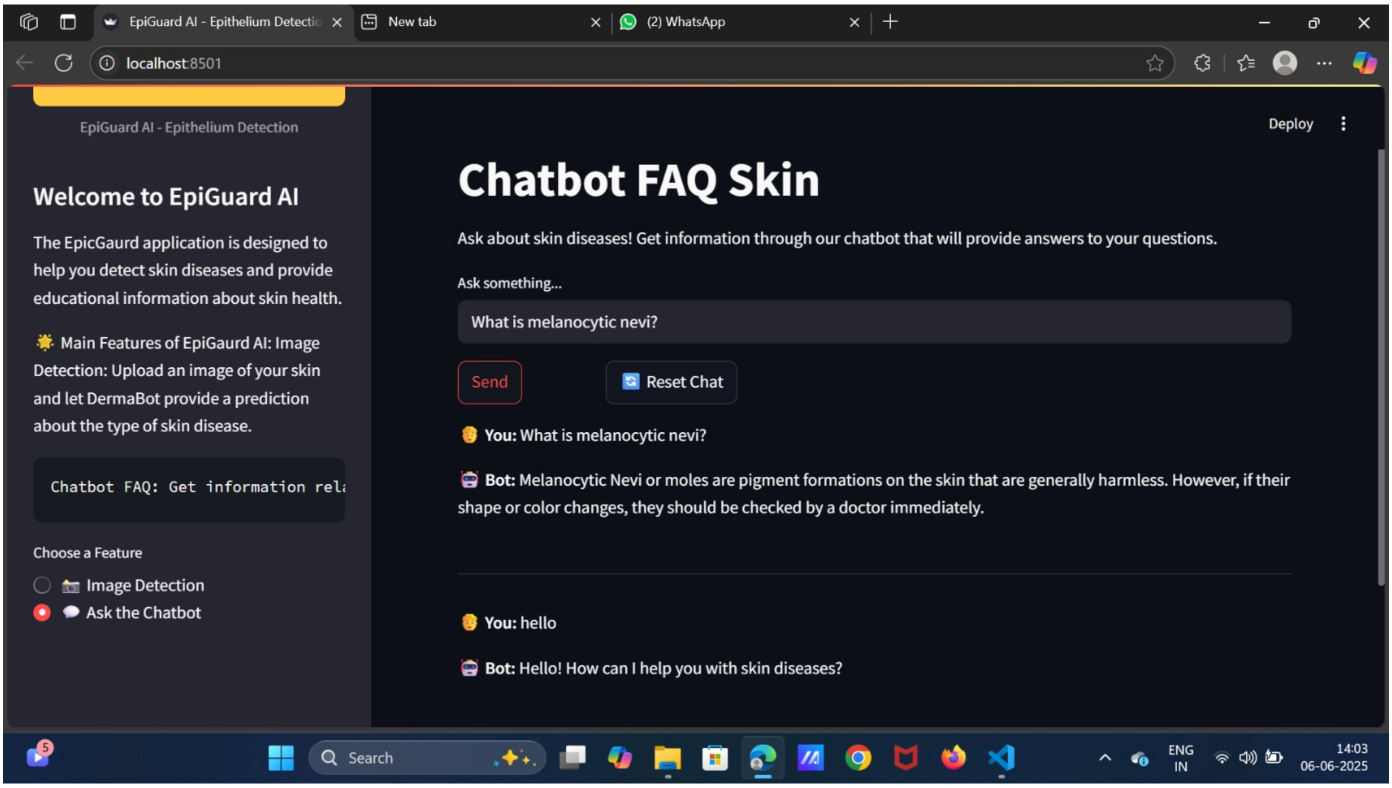

Fig B.5 : Chatbot FAQ Page

REFERENCES

[1] H. Gonzalez, C. Hagerling, and Z. Werb, ‘‘Roles of the immune system in cancer: From tumor initiation to metastatic progression,’’ Genes Develop., vol. 32, nos. 19–20, pp. 1267–1284, Oct. 2018.

[2] H. Kwon, W. Q. Malik, S. B. Rutkove, and B. Sanchez, ‘‘Separation of subcutaneous fat from muscle in surface electrical impedance myography measurements using model component analysis,’’ IEEE Trans. Biomed. Eng., vol. 66, no. 2, pp. 354–364, Feb. 2019.

[3] K. Kapur, B. Sanchez, A. Pacheck, B. Darras, S. B. Rutkove, and R. Selukar, ‘‘Functional mixed-effects modeling of longitudinal Duchenne muscular dystrophy electrical impedance myography data using statespace approach,’’ IEEE Trans. Biomed. Eng., vol. 66, no. 6, pp. 1761–1768, Jun. 2019.

[4] S. W. Duffy, D. Vulkan, H. Cuckle, D. Parmar, S. Sheikh, R. A. Smith, A. Evans, O. Blyuss, L. Johns, I. O. Ellis, J. Myles, P. D. Sasieni, and S. M. Moss, ‘‘Effect of mammographic screening from age 40 years on breast cancer mortality (U.K. Age trial): Final results of a randomised, controlled trial,’’ Lancetoncology, vol. 21, no. 9, pp. 1165–1172, Sep. 2020.

[5] M. M. D. M. Cardoner, H. Kwon, H. V. G. Pulido, J. Nagy, S. Rutkove, and B. Sanchez, ‘‘Modeling and reproducibility of twin concentric electrical impedance myography,’’ IEEE Trans. Biomed. Eng., vol. 68, no. 10, pp. 3068–3077, Oct. 2021.