A PROJECT REPORT ON

DermaGenie: Genereative AI for Dermatological diagnosis

Submitted Towards the
Partial Fulfilment of the Requirements of

Bachelor of Engineering (Artificial Intelligence and Data Science)

by

Rutuja Prashant Patil	Roll No: 54
Ansari Mohammad Raza	Roll No: 53
Unmesh Ashok Jagtap	Roll No: 55

Under the Guidance of
Prof. Pranali Shinde

Department of Artificial Intelligence and Data Science
K. K. Wagh Institute of Engineering Education & Research
Hirabai Haridas Vidyanagari, Amrutdham, Panchavati,
Nashik-422003
Savitribai Phule Pune University

A. Y. 2024-2025 Sem I

K. K. Wagh Institute of Engineering Education and Research
Department of Artificial Intelligence and Data Science

CERTIFICATE

This is to certify that the Project Titled

DermaGenie: Generative AI for Dermatological Diagnosis

Submitted by

Rutuja Prashant Patil	Roll No: 54
Ansari Mohammad Raza	Roll No: 53
Unmesh Ashok Jagtap	Roll No: 55

is a bonafide work carried out by students under the supervision of *Prof. Pranali Shinde* and it is submitted towards the partial fulfilment of the requirement of Bachelor of Engineering (Artificial Intelligence and Data Science) during academic year 2024-2025.

Prof. Pranali Shinde
Internal Guide
Department of AI and DS

Head
Department of AI and DS

Abstract

Skin diseases affect millions globally, with early and accurate diagnosis being critical for effective treatment. However, manual diagnosis often suffers from subjectivity and inconsistencies, leading to potential misclassifications. To address this challenge, Dermagenie leverages Generative AI and Retrieval-Augmented Generation (RAG) models to develop an automated dermatology system capable of detecting, staging, and prescribing treatment recommendations for skin diseases.The system is built on a Convolutional Neural Network (CNN) model trained on the HAM10000 dataset to classify various skin diseases. The classification output is further refined by a Retrieval-Augmented Generation (RAG) module, which generates comprehensive prescription reports by retrieving relevant medical information and dynamically combining it with the model's predictions. To ensure the reliability and accuracy of the predictions, a Large Language Model (LLM) is integrated into the pipeline. The LLM performs verification and justification of the model's predictions by cross-referencing them with medical knowledge and explaining the reasoning behind the diagnosis and prescription. This enhances the system's transparency, making it more interpretable and trustworthy for clinical use. The Dermagenie system demonstrates high accuracy in disease classification and generates detailed, contextually rich prescription reports. By integrating CNNs, RAG models, and LLM verification, it offers a comprehensive AI-powered solution for dermatology, aiming to improve diagnostic accuracy, reduce clinical workload, and enhance patient care.

Acknowledgment

We would like to extend my heartfelt gratitude to our project guide, **Prof. Pranali Shinde**, whose unwavering support and expert guidance have been instrumental in the successful completion of this final year project. Her wealth of knowledge, invaluable insights, and constant encouragement have significantly contributed to shaping our ideas and refining our project.We are also deeply thankful to Prof. **Dr. Snehal Kamalapur** for the pivotal role played in the planning and execution of our project. Her timely advice and assistance were vital in ensuring that our project milestones were achieved efficiently. Additionally, We extend our appreciation to the entire departmental staff for their continuous support.Our sincere thanks go out to our dedicated colleagues and friends who actively engaged in discussions, provided constructive feedback, and offered valuable suggestions for enhancing the quality and scope of our project. Their collective efforts and collaborative spirit have greatly enriched this endeavor.

Rutuja Prashant Patil
Ansari mohammad Raza
Unmesh Ashok Jagtap

(B.E. Artificial Intelligence and Data Science)

INDEX

List of Figures

CHAPTER 1

INTRODUCTION

Diagnosing skin diseases accurately is a challenging task, primarily because many conditions exhibit overlapping visual features, which often leads to incorrect or delayed diagnoses. To address this issue, the project integrates advanced artificial intelligence methodologies, including Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), and Retrieval-Augmented Generation (RAG), to create a more robust and reliable diagnostic system. By employing GANs, the system is capable of producing counterfactual skin images—synthetic variations that simulate how different diseases might appear—thereby assisting clinicians in differentiating between visually similar conditions.

In parallel, CNNs are used for the classification of skin diseases based on high-dimensional image data, enabling the model to learn intricate patterns and features that are often imperceptible to the human eye. The use of RAG further enhances the system's functionality by retrieving pertinent clinical knowledge and integrating it into the diagnostic process. This ensures that treatment recommendations, including appropriate medications and dietary modifications, are tailored to the patient's specific condition. Overall, the project presents a hybrid AI-based framework that combines predictive modeling with contextual medical knowledge, aiming to improve both the precision of skin disease diagnosis and the effectiveness of subsequent treatment plans.

1.1 PROJECT IDEA

This project proposes an intelligent, AI-driven system designed not only to detect skin diseases from dermatological images but also to provide enhanced support for clinical diagnosis and treatment planning. The system is rooted in deep learning methodologies, particularly leveraging Convolutional Neural Networks (CNNs), which are trained on comprehensive dermatology datasets such as HAM10000. These networks are used to automatically identify and classify different skin conditions with high precision, reducing reliance on manual diagnostics and minimizing errors caused by visual similarities between diseases.

To add an interpretability layer, Generative Adversarial Networks (GANs) are incorporated to produce counterfactual images. These synthetic visual representations illustrate how a skin lesion might differ under alternative diagnostic conditions, helping medical professionals and patients to better grasp the nature of the diagnosis. This visualization component not only improves transparency but also enhances trust in AI-assisted medical decisions. Furthermore, the system introduces a personalized treatment engine powered by Retrieval-Augmented Generation (RAG). RAG plays a vital role by retrieving up-to-date, clinically relevant information from medical journals and authoritative healthcare databases. This ensures that every treatment plan—including recommended medications and tailored dietary guidelines—is grounded in the latest evidence and best practices. This dynamic retrieval mechanism allows the system to adapt to new medical findings over time, keeping the recommendations current and context-specific.

By combining CNN-based classification, GAN-generated explanatory visuals, and RAG-enhanced treatment support, the system offers a comprehensive solution aimed at improving diagnostic accuracy, patient comprehension, and treatment outcomes. It serves as a practical aid for healthcare professionals while also empowering patients with personalized, easy-to-understand medical insights.

1.2 MOTIVATION OF THE PROJECT

The primary motivation behind developing this project stems from the pressing demand for accurate and accessible diagnostic tools in dermatology. Skin disorders often present with overlapping visual traits, making it difficult to distinguish one condition from another through visual inspection alone. Even trained dermatologists sometimes face challenges in delivering a precise diagnosis, which can result in misdiagnosis, delayed treatments, and unnecessary anxiety for patients. These diagnostic challenges, when combined with the lack of dermatological expertise in many remote and underserved communities, create a substantial gap in effective healthcare delivery.

With the evolution of artificial intelligence, especially in the field of medical imaging, there is a significant opportunity to bridge this gap. Technologies such as Generative Adversarial Networks (GANs) have shown promise in enhancing the interpretability of medical AI models by creating synthetic but realistic visual outputs. This project takes inspiration from such advancements, incorporating GANs not just for data augmentation, but to produce counterfactual visual explanations. These generated images simulate hypothetical disease states, allowing users—be they clinicians or patients—to visually interpret the AI's decision, making the system more transparent and easier to trust.Another key motivational factor is the system's potential to serve as a complete decision-support tool. Beyond identifying the disease, the system integrates Retrieval-Augmented Generation (RAG) to provide tailored treatment insights, including up-to-date medication suggestions and nutritional recommendations. This personalized guidance turns the system into a comprehensive medical assistant capable of not only flagging a condition but also recommending actionable next steps for management.

In summary, this project is driven by the dual necessity of improving diagnostic accuracy and expanding access to reliable dermatological care. By integrating deep learning, visual reasoning, and real-time knowledge retrieval, the system aspires to empower healthcare providers and patients alike, particularly in resource-constrained environments, while also advancing the frontier of interpretable medical AI.

In addition to improving access and accuracy in dermatological care, this project is motivated by the growing demand for explainable and trustworthy AI systems in healthcare. Traditional black-box models often provide high accuracy but lack transparency, leaving end-users—especially medical practitioners—unsure about how conclusions were reached. In critical fields such as dermatology, where treatment decisions have a direct impact on patient health, it is essential that AI models not only produce correct outputs but also provide human-understandable justifications. By generating counterfactual images and leveraging large language models through RAG, this project aims to satisfy this need for explainability. This helps foster trust between the AI system and the users, encouraging wider adoption in clinical environments. Moreover, the motivation also stems from a recognition of the emotional and psychological dimensions of dermatological conditions. Skin diseases are often visible and can severely affect an individual's self-esteem, social interactions, and overall mental health. Fast and accurate diagnosis, along with clear explanations and actionable treatment options, can alleviate some of this psychological burden by reducing uncertainty and enabling early interventions. The system proposed in this project is designed with this patient-centric view, offering not only diagnostic insights but also educational support that can empower users with knowledge about their condition. Another important driver behind this initiative is the rise of telemedicine and mobile healthcare. With the widespread availability of smartphones and internet access, there is an increasing shift towards remote diagnosis and AI-assisted healthcare delivery. This project aligns with that trend by offering a lightweight, front-end interface via Streamlit, allowing users to interact with the system in real-time from their homes. This makes it suitable for integration into teledermatology platforms, public health campaigns, and even mobile apps that target rural or marginalized populations with limited access to specialized care.

Lastly, this project is also motivated by the desire to contribute meaningfully to interdisciplinary research that bridges AI, medicine, and human-centered design. It provides a model for how cutting-edge AI technologies such as CNNs, GANs, and LLMs can work synergistically to tackle complex medical problems. By building a system that is accurate, interpretable, scalable, and user-friendly, the project embodies a forward-thinking approach to digital healthcare innovation—one that could potentially be expanded to other domains such as ophthalmology, radiology, or pathology in future iterations.

CHAPTER 2

LITERATURE SURVEY

Recent advances in deep learning, especially Convolutional Neural Networks (CNNs), have improved skin disease classification by leveraging large datasets. However, CNNs often lack transparency, leading to a focus on improving model explainability. Generative Adversarial Networks (GANs) address this by generating synthetic data to balance datasets and creating counterfactual images that enhance model interpretability. Additionally, AI-driven personalized treatment recommendations, such as medication and dietary advice, are gaining importance in improving patient care. This project builds on these advancements by combining GAN-generated counterfactuals with tailored treatment suggestions, enhancing both diagnostic accuracy and practical healthcare.

1 .An-Chao Tsai et al. **:"Advanced Pigmented Facial Skin Analysis Using Conditional Generative Adversarial Networks" (2024)**

The research by An-Chao Tsai et al. highlights the potential of Conditional Generative Adversarial Networks (cGANs) in dermatology. The system leverages mathematical models, including adversarial loss functions and image similarity metrics, to accurately detect and analyze the severity of melasma from facial images. By utilizing these models, the system provides dermatologists with valuable insights for personalized treatment and further advancements in skin condition analysis.In the research by Tsai et al., several mathematical terms and techniques are employed to develop the predictive model for melasma diagnosis using Conditional GANs (cGANs). Key mathematical concepts include: Conditional GAN (cGAN) Framework, Loss Functions, Image Analysis Metrics, and Feature Extraction.

2. R.Yadav and Dr. A. Bhat, **"A Survey on Skin Lesion Detection and Classification using Machine Learning" (2024)**

The research compares machine learning and deep learning for skin injury diagnosis, emphasizing real-time, automated tools and collaboration between dermatology and ML for improved accuracy and personalization. The paper "A

Survey on Skin Lesion Detection and Classification using Machine Learning" analyzes various machine learning and deep learning techniques for diagnosing skin lesions, highlighting their potential for real-time, automated solutions. It emphasizes the importance of integrating dermatological expertise with machine learning to enhance diagnostic accuracy and personalized care. Mathematical terms used in this research include classification algorithms (e.g., SVM, decision trees), feature extraction methods (e.g., PCA, LDA), performance metrics (e.g., accuracy, precision, recall, F1-score), and confusion matrices to evaluate model effectiveness. Additionally, it discusses data normalization and cross-validation techniques to ensure robust model training and validation.

***3. B. Durgabhavani, B. Chandana, S. Sandhyarani, G. Lavanya, G. Srikanth, andN. Bhaskar,* "Classification of Skin Disease using CNN" (2023)**

The research created a CNN to classify skin diseases from thermoscopic images. This tool assists both healthcare professionals and patients in diagnosing skin conditions. The paper "Classification of Skin Disease using CNN" presents a Convolutional Neural Network (CNN) designed to classify skin diseases from thermoscopic images, aiding healthcare professionals and patients in the diagnostic process. Mathematical terms utilized in this research include convolution operations, activation functions (e.g., ReLU, sigmoid), loss functions (e.g., cross-entropy), optimization algorithms (e.g., Adam, SGD), and accuracy metrics to evaluate model performance. Additionally, it discusses data augmentation techniques to enhance the training dataset and hyperparameter tuning for model optimization.

***4. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S.* ."Dermatologist-level classification of skin cancer with deep neural networks" (2017)**.

This study demonstrates the application of deep neural networks (DNNs) in classifying skin cancer images. The model was trained on 129,450 clinical images, achieving an accuracy comparable to certified dermatologists. The findings highlight the potential of AI in skin disease diagnosis, showcasing its reliability in detecting both benign and malignant lesions.

***5. Han, S., Kang, B., and Lim, W.*"Classification of skin disease using deep learning models. Computers in Biology and Medicine" (2018)**
This research explores the use of multi-class CNN models for automated skin disease classification. The authors applied their model to 34 skin conditions, achieving a 77% classification accuracy, which was on par with dermatologists. The study emphasizes the effectiveness of CNN architectures in handling complex dermatological images and accurately identifying multiple skin conditions.

***6. Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., et al.* "Comparison of human experts and deep learning algorithms for the diagnosis of pigmented skin lesions" (2019)**.
This study compares the performance of deep learning algorithms with human dermatologists in classifying pigmented skin lesions. The CNN model achieved a higher diagnostic accuracy than most human experts, with an AUC of 91.4%. The research highlights the potential of AI to surpass human-level accuracy in skin disease detection, making it highly applicable for clinical dermatology systems.

CHAPTER 3

PROBLEM DEFINITION AND SCOPE

Diagnosing skin diseases accurately is a complex and critical task in modern healthcare. Many dermatological conditions present with similar visual symptoms such as redness, scaling, pigmentation, or inflammation, which can often lead to misidentification or delayed diagnosis. The challenge becomes even more prominent in regions lacking access to qualified dermatologists, where general practitioners or untrained individuals may be responsible for assessing skin conditions. Incorrect or late diagnosis can result in improper treatment, prolonged discomfort, or worsening of the disease. To address these issues, this project proposes the development of a comprehensive AI-driven system capable of assisting both medical professionals and patients in identifying and understanding various skin conditions. The goal is to enhance the diagnostic process by leveraging deep learning techniques—specifically Convolutional Neural Networks (CNNs) for image classification, Generative Adversarial Networks (GANs) for generating counterfactual images, and Retrieval-Augmented Generation (RAG) for context-aware treatment suggestions. This system will not only identify the likely disease but also provide visual and textual explanations that improve interpretability and user confidence in the AI outputs.

The scope of the project extends beyond simple disease classification. It encompasses a holistic view of dermatological care by including functionality for severity staging, prescription recommendations, and dietary suggestions tailored to the diagnosed condition. The integration of a language model-based reasoning system helps in verifying classification results and provides medically relevant justifications sourced from trusted databases or literature. Additionally, the project includes a user-friendly interface built with Streamlit to ensure accessibility and real-time interaction for end-users.

This AI-powered solution aims to bridge the gap between clinical expertise and scalable healthcare delivery. It is especially relevant for telemedicine, rural outreach, and patient education. The system's modular design allows for future expansion into new datasets, conditions, or languages, making it a versatile tool for personalized, evidence-based dermatological support. Ultimately, the project aspires to create an intelligent assistant that empowers both doctors and patients by providing fast, reliable, and transparent skin disease analysis and care recommendations.

3.1 PROBLEM STATEMENT

The diagnosis of skin diseases presents a critical challenge in clinical practice due to the large number of dermatological conditions that exhibit overlapping and often indistinguishable visual characteristics. These similarities can lead to frequent misdiagnoses, inappropriate treatments, and extended periods of patient suffering. Current diagnostic practices are predominantly reliant on the subjective interpretation of visual symptoms by healthcare professionals, which introduces potential biases and inconsistencies. This problem is exacerbated in remote or underserved areas, where access to trained dermatologists is severely limited, resulting in delayed or incorrect treatments and an overall decline in the quality of dermatological care. There is an urgent need for intelligent, reliable systems that go beyond classification and support healthcare practitioners with more comprehensive insights. Patients today seek not only a diagnosis but also a deeper understanding of their condition, along with personalized guidance regarding treatment and preventive care. Existing AI applications in dermatology are often limited to classification models that lack transparency and the ability to justify their predictions, reducing trust among users and healthcare providers.

In response to these challenges, this project introduces an advanced AI-powered system designed to predict skin diseases from medical images while also offering explanatory visual aids and medically grounded treatment suggestions. This system utilizes a deep learning framework that incorporates Convolutional Neural Networks (CNNs) for high-accuracy image classification, Generative Adversarial Networks (GANs) to create counterfactual image representations for enhanced visual interpretability, and Retrieval-Augmented Generation (RAG) to retrieve relevant clinical information from curated medical databases. One of the key innovations of the system is the integration of Large Language Models (LLMs), which help verify classification results, justify predictions, and provide detailed medical explanations in natural language. This combination of technologies ensures that the system not only makes predictions but also explains the rationale behind them, contributing to greater transparency and clinical trust. The inclusion of dietary suggestions, disease staging, and evidence-based medication plans extends the system's utility beyond diagnosis into actionable treatment support.

By offering a reliable, explainable, and scalable solution, this AI system aims to revolutionize dermatological diagnostics. It empowers clinicians with data-backed tools and enables patients to make informed decisions about their health. Ultimately, this project addresses the pressing need for accessible, equitable, and intelligent healthcare in dermatology, especially in settings where specialist care is scarce or unavailable.

3.1.1 Goals and Objectives

The central aim of this project is to design and implement an intelligent, AI-driven system capable of diagnosing skin conditions, recommending personalized treatments, and providing comprehensive recovery guidance. With the increasing complexity of dermatological disorders and the limitations of traditional diagnostic methods, there is a growing demand for tools that can offer reliable, fast, and explainable results. This system is intended to bridge that gap by leveraging cutting-edge AI technologies to improve the quality of dermatological care.

Overall Goal: The overarching goal is to build a robust AI system that not only predicts skin diseases from images but also assists in the decision-making process for treatment and post-diagnosis care. The system will integrate multiple AI technologies—such as Convolutional Neural Networks (CNNs) for image classification, Retrieval-Augmented Generation (RAG) for sourcing relevant medical information, and Large Language Models (LLMs) for generating human-readable explanations and justifications. These technologies will work in unison to ensure that diagnoses are accurate, treatments are evidence-based, and the recommendations are aligned with current medical standards.

Specific Objectives: To fulfill the project vision, the system is designed to meet the following key objectives:

- **Image-based Prediction:** Develop and train a Convolutional Neural Network (CNN) to classify various skin diseases accurately using image data.
- **Counterfactual Visualization:** Leverage Generative Adversarial Networks (GANs) to generate visual representations of diseased versus healthy skin, helping to explain and differentiate conditions.
- **Evidence-Based Recommendations:** Incorporate Retrieval-Augmented Generation (RAG) to access and extract reliable medical information from up-to-date and reputable sources.
- **Natural Language Justifications:** Use Large Language Models (LLMs) to produce human-understandable explanations and verifications of diagnoses, improving user confidence and interpretability.
- **Medication and Dietary Suggestions:** Provide personalized treatment recommendations, including medication options and dietary guidelines, based on the diagnosed skin condition and severity level.
- **Holistic Decision Support:** Combine predictions, visual aids, justifications, and recommendations into a single interface that supports clinical decision-making and empowers patients.

3.1.2 Assumptions and Scope

This section outlines the foundational assumptions made during the development of the system and defines the scope of its intended use, helping to set realistic expectations for functionality and deployment.

Assumptions:

1. The input provided to the system consists of high-resolution, well-lit, and focused skin images that clearly capture the affected area. Image quality plays a critical role in ensuring accurate predictions.

2. The users of this system are expected to be healthcare practitioners, researchers, or informed individuals who possess a basic understanding of dermatological conditions. This assumption ensures that users can interpret the outputs and act on recommendations effectively.

3. The system is built as a web-based application that requires a stable internet connection. This is necessary for uploading images, retrieving relevant information, and generating predictions via cloud-hosted models and services.

4. It is assumed that users will adhere to ethical guidelines while using the system, particularly in clinical contexts, by treating it as a decision-support tool rather than a standalone diagnostic substitute.

Scope:

1. In-Scope: The project is designed to predict common skin diseases, provide visual explanations using counterfactual images, and offer treatment options, including medication and dietary suggestions. It aims to assist in diagnosing conditions such as melanoma, eczema, and psoriasis[5].

2. Out-of-Scope: The system is not designed to replace medical professionals but to provide assistance. It will not cover rare or highly complex skin conditions that require specialized medical intervention. Additionally, the project does not involve direct medical consultations or the provision of prescription medications—users should consult a healthcare provider before following the recommendations.

3.2 METHODOLOGY

The methodology adopted for this project involves a structured pipeline that integrates multiple AI techniques to enhance the prediction, explainability, and personalization of skin disease diagnosis and treatment. The key components are detailed below:

1. Data Collection and Preprocessing: The project leverages dermatological image datasets such as HAM10000, which contain labeled images of various skin conditions. Preprocessing steps include resizing the images to a standard input shape, normalization of pixel values, and data augmentation techniques such as flipping, rotation, and brightness adjustments. Additionally, Generative Adversarial Networks (GANs) are employed to synthetically augment the dataset, addressing class imbalance and improving model generalization.

2. Disease Classification using CNN: A Convolutional Neural Network (CNN) is trained on the preprocessed dataset to automatically classify images into predefined skin disease categories. The CNN is capable of learning hierarchical features, from edges to disease-specific patterns, improving the robustness of disease detection across diverse image inputs.

3. Model Explainability using LIME: To promote transparency and trust in the AI system, Local Interpretable Model-Agnostic Explanations (LIME) is used. LIME helps identify the regions of the input image that most influenced the CNN's decision, thereby offering interpretable visual feedback to clinicians and users.

4. Enhanced Recommendations via RAG: The system incorporates Retrieval-Augmented Generation (RAG) to retrieve up-to-date, relevant medical knowledge from trusted databases and literature. This enhances the quality and credibility of treatment recommendations, ensuring they are grounded in current clinical evidence and practice.

5. Medication and Diet Recommendation Engine: A hybrid rule-based and data-driven engine is developed to provide personalized treatment suggestions. The engine uses disease classification results and retrieved medical content (via RAG) to suggest medications and dietary practices that support recovery and immune health. These recommendations are tailored to the user and informed by reliable medical sources.

6. Prediction Verification using LLMs: To further increase reliability, a Large Language Model (LLM) such as GPT-4, Gemini, or LLaMA is integrated into the system. The LLM verifies the CNN predictions and recommendation outputs, and it generates human-readable justifications. This step ensures interpretability and builds user confidence in the system's outputs.

7. Web Interface with Streamlit: The entire system is deployed as an interactive web application using the Streamlit framework. This interface allows users to upload skin images, receive predictions, view counterfactual visualizations, and access personalized recommendations—all in a streamlined and user-friendly manner.

3.3 OUTCOME

This project culminates in the development of a robust and AI-driven system designed for the prediction and diagnosis of various skin diseases through image analysis. At its core, the system leverages advanced machine learning models to accurately identify specific dermatological conditions from user-uploaded photographs, enabling immediate and reliable diagnostic feedback. Beyond simple classification, the system innovatively incorporates a generator mechanism that produces counterfactual images—these altered visuals help users and clinicians discern subtle differences between similar skin disorders by illustrating hypothetical variations in symptoms, thus deepening their understanding of each condition's unique presentation. A significant feature of this system is its ability to provide personalized treatment guidance tailored to the individual patient's diagnosis. This includes not only recommendations for effective pharmaceutical interventions but also lifestyle and dietary advice that complements medical treatment, thereby supporting holistic recovery strategies.

The system is accessed via a user-friendly web interface, designed for ease of use by patients and dermatologists alike. This accessibility facilitates quicker diagnosis and more informed decision-making in clinical settings, ultimately enhancing patient outcomes and streamlining dermatological consultations. From a technical perspective, the project demonstrates high diagnostic accuracy by employing a Convolutional Neural Network (CNN) model trained on diverse skin disease datasets. This model adeptly distinguishes between benign and malignant skin lesions, which is critical for timely medical intervention. Furthermore, the system integrates a multi-stage severity classification framework that assesses the progression levels of identified diseases. This capability allows for nuanced insights into disease stages, offering clinicians valuable information for treatment planning. In addition to image-based analysis, the system features an advanced Retrieval-Augmented Generation (RAG) module that automatically composes detailed prescription reports. These reports are comprehensive, containing critical details such as the diagnosed disease name, precise affected skin regions, suggested diagnostic procedures, recommended medications, and tailored dietary plans. Such reports provide patients with clear, evidence-based treatment pathways, supporting better adherence and health management.

To bolster the credibility and transparency of its predictions, the system incorporates a Large Language Model (LLM) that cross-validates the CNN outputs by referencing established medical databases and literature. This LLM not only verifies the diagnosis but also generates coherent, fact-based explanations and justifications for both the disease identification and the proposed treatments. By offering these natural language rationales, the system enhances interpretability for medical professionals and patients, fostering greater trust in the AI's recommendations and making the entire diagnostic process more transparent and clinically meaningful.

3.4 TYPE OF PROJECT

This project represents a comprehensive blend of both theoretical research and practical implementation within the field of dermatology. It is designed not only to advance academic understanding of AI applications in medical diagnostics but also to deliver a functional, real-world tool that can be directly utilized by healthcare providers and patients. Central to this project is the integration of Retrieval-Augmented Generation (RAG), an innovative technique that enables the system to access and incorporate up-to-date medical information dynamically. This capability allows the system to offer precise and evidence-backed recommendations regarding medications and dietary plans tailored to individual diagnoses, ensuring that treatment advice remains relevant and scientifically grounded.

The research component places strong emphasis on the dual functionality of deep learning technologies—not only for accurately predicting skin diseases from image data but also for providing transparent and interpretable explanations of the diagnostic process. By enhancing the explainability of AI predictions, the project addresses a critical challenge in medical AI systems: building user trust and facilitating clinical acceptance. To this end, the system employs Large Language Models (LLMs) that serve as a verification layer. These LLMs systematically cross-reference the CNN model's outputs against extensive medical knowledge bases and authoritative sources, generating fact-based, detailed justifications for both the diagnoses and the associated treatment recommendations.

This natural language-based explanation mechanism significantly improves the transparency and clinical interpretability of the system, making it easier for healthcare professionals to understand and validate the AI's conclusions. The practical orientation of the project ensures that it is not merely an academic exercise but an application-ready solution aimed at immediate deployment in clinical environments. By doing so, it highlights the growing role of AI-driven technologies in enhancing the accuracy, efficiency, and accessibility of medical diagnostics, ultimately contributing to better patient care and informed decision-making.

CHAPTER 4

PROJECT PLAN

4.1 PROJECT TIMELINE

Highlight key project dates, reviews, and task progress.

Step 1: July, 2024

Project Initiation Team formation and role assignment Project proposal development

Step 2: July , 2024

1st Project Review Data collection and preprocessing Initial system design Selection of usefull algorithms

Step 3: Feb, 2025

2nd Project Review Model training and testing User interface deveoplement, RAG, Generators and CNN model development

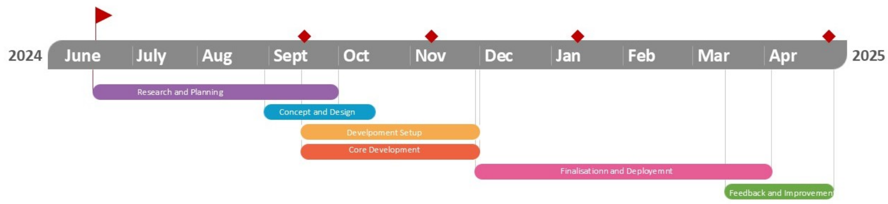


Figure 4.1: Project Timeline Chart

4.2 TEAM ORGANIZATION

The manner in which staff is organized and the mechanisms for reporting are noted.

tabularx

4.2.1 Team Structure

The team structure for the project is carefully designed to ensure the effective execution and management of tasks. Each team member has defined roles and responsibilities, contributing to the success of the project:

Team Leader: Rutuja Prsahant Patil

Role: Team Leader

Responsibilities: As the Team Leader, Rutuja Prashant Patil is responsible for

Name	Role	Responsibilities
Rutuja Prashant Patil	Team Leader	Responsible for developing and training the CNN classification model for disease prediction. Integrates LIME for model explainability and manages the medicine recommendation module. Oversees project quality assurance and testing.
Ansari Mohammad Raza	Data Scientist	Focuses on data preprocessing and dataset augmentation using GAN models to generate counterfactual images. Implements Generator Functions for generating distinct counterfactual images to aid in differentiating skin diseases. Collaborates with Team Leader to handle diverse data. Implementing RAG and LLM concepts into project mechanism.
Unmesh Ashok Jagtap	Frontend and Integration Specialist	Develops the Streamlit-based web application for real-time disease prediction and user interaction. Integrates diet recommendations based on diagnosed disease. Manages system integration and testing for backend and frontend communication.

Table 4.1: Roles and Responsibilities of Project Team Members

overseeing the entire project. This includes coordinating the efforts of the team members, ensuring that project goals are met, and maintaining the overall project timeline. Rutuja also plays a crucial role in quality assurance and testing, ensuring that the project meets the desired standards. Creating and training the Convolutional Neural Network (CNN) classification model used for skin disease prediction. This role includes integrating Local Interpretable Model-Agnostic Explanations (LIME) to enhance the model's explainability, ensuring users receive clear insights into the predictions. Additionally, this individual manages the medicine recommendation module, identifying effective medication options for each predicted disease, which significantly contributes to the system's overall functionality and user experience.

Data Scientist: Ansari Mohammad Raza

Role: Data Science specialist

Responsibilities: responsible for preprocessing data and augmenting the dataset by utilizing Generative Adversarial Networks (GANs) to create counterfactual images. This role involves implementing the GAN model to generate two distinct counterfactual images for each disease case, aiding users in recognizing and differentiating between various skin diseases. Additionally, the Data Scientist collaborates with the Lead Developer to ensure the model effectively handles a diverse range of input data, enhancing the system's overall robustness and accuracy.

Frontend and Integration Specialist: Unmesh Ashok Jagtap

Role: web developer and integration specialist

Responsibilities: The Frontend and Integration Specialist is responsible for developing the Streamlit-based web application that enables real-time disease prediction and user interaction. This role involves integrating the diet recommendation module, which provides personalized dietary plans based on the diagnosed skin disease. Additionally, the specialist manages system integration and testing, ensuring seamless communication between the web application and backend models to deliver a smooth and efficient user experience.

CHAPTER 5

SOFTWARE REQUIREMENT SPECIFICATION

This chapter provides a comprehensive overview of the software requirements essential for the development and deployment of the proposed AI-powered dermatological diagnosis system. The Software Requirement Specification (SRS) acts as a foundational document that defines the system's functional behavior, performance metrics, user interactions, and technical constraints. It is designed to offer a clear and structured framework for both the development team and stakeholders, ensuring a shared understanding of the system's capabilities and limitations.

The specification includes details about the functional requirements—what the system is expected to do, such as disease classification, counterfactual image generation, and prescription reporting. It also covers non-functional aspects like performance efficiency, reliability, scalability, and usability. Additionally, it identifies all external interfaces, including user interfaces (Streamlit web app), system interfaces (backend APIs), and integration with third-party modules like pretrained models and medical databases.

By elaborating on both high-level and low-level specifications, this document ensures consistency throughout the software development lifecycle. It serves as a reference for design decisions, coding implementations, testing strategies, and future maintenance activities. Ultimately, this chapter aims to align the project's technical vision with user expectations and clinical relevance.

5.1 FUNCTIONAL REQUIREMENTS

Functional requirements describe the specific operations, actions, and capabilities that the software system must perform. They represent the core objectives and expected functionalities that the system should deliver from the user's perspective. These requirements ensure that the system performs its intended tasks effectively and addresses the user's needs in a real-world scenario.

In the context of this dermatology-focused AI application, the key functional requirements are as follows:

1. **Automated Skin Disease Diagnosis:** The system must enable users to upload high-quality images of skin lesions through a web interface. These images will be analyzed using a pre-trained Convolutional Neural Network (CNN) model to identify and classify the specific type of skin disease.

2. **Counterfactual Image Synthesis:** The platform should employ generator functions to create two counterfactual images per input. These synthetic visuals allow users to observe how changes in features may influence the diagnosis, enhancing their understanding and improving model transparency.

3. **Personalized Medication Suggestions:** Based on the diagnosed condition, the system will generate a list of recommended medications. This functionality will utilize Retrieval-Augmented Generation (RAG) to access the latest treatment protocols and pharmaceutical data, ensuring evidence-based and current medical recommendations.

4. **Tailored Dietary Guidance:** In alignment with the predicted skin condition, the system should provide users with dietary suggestions that support healing and symptom management. This will also be driven by RAG, retrieving nutrition recommendations from clinical databases and authoritative sources.

5. **Integrated Health Guidance Dashboard:** A unified web interface built using Streamlit will allow users to interact seamlessly with the system. Through this interface, users can upload images, receive diagnostic outputs, view counterfactual comparisons, and access personalized treatment plans.

6. **Real-Time Feedback and Explanation:** In addition to showing predictions, the system should offer interpretability features such as LIME or LLM-justified feedback that explains why a certain diagnosis or recommendation was made, enhancing trust and usability.

These functional specifications define the critical operations the system must perform to fulfill its objective—providing a reliable, interpretable, and user-friendly

AI-based solution for dermatological diagnostics and care planning. These functional requirements are essential to achieve the goals of the project

5.2 NON-FUNCTIONAL REQUIREMENTS

Non-functional requirements define the overall quality attributes of the system. They describe how the system should perform rather than what it should do. These attributes ensure the system is not only functional but also robust, secure, efficient, and maintainable. For an AI-powered dermatological diagnosis system, these characteristics are particularly critical due to its interaction with sensitive health data and the necessity for accurate, timely, and dependable diagnostics. Below are the key non-functional requirements identified for the proposed system:

1. **Availability:** The system should maintain high availability to serve users at any time of day without significant interruptions. Given the importance of timely medical advice and accessibility, the web application must be designed to be operational 24 hours a day, 7 days a week. This includes appropriate deployment infrastructure, monitoring, and redundancy mechanisms to ensure continuous uptime, especially in cases of sudden usage spikes or minor system faults.

2. **Reliability:** Reliability is essential for building user confidence, especially in a healthcare-related application. The system must consistently provide accurate and dependable predictions for various skin diseases using its trained CNN classification model. Furthermore, the counterfactual images generated using GANs (Generative Adversarial Networks) should be visually plausible, medically relevant, and clearly differentiate between disease classes. The model must handle a variety of skin types and conditions without crashing or producing ambiguous results. This ensures users and dermatologists can rely on the tool during diagnosis or consultation.

3. **Maintainability:** The entire software architecture should be modular and well-organized to support maintainability. Source code should be thoroughly documented, allowing future developers to easily understand, debug, and

upgrade the system. Proper version control (e.g., Git) should be employed to manage updates and collaboration effectively. If enhancements are required—such as incorporating new skin conditions, retraining the model with a larger dataset, or expanding the LLM-based justification module—those should be implementable with minimal rework or disruptions.

4. **Security:** Due to the involvement of sensitive user data, particularly images of skin conditions, the system must enforce robust security measures. All image uploads and patient-related data must be encrypted during both transmission and storage. Authentication and authorization protocols must be in place to restrict access to administrative and medical records. Furthermore, the platform must comply with applicable privacy standards and regulations, ensuring that user data is handled ethically and lawfully. The backend should also be protected against common cyber threats such as SQL injection, DDoS attacks, and unauthorized data scraping.

These non-functional requirements form the backbone of the system's operational quality and directly impact user satisfaction, system longevity, and clinical relevance. By ensuring availability, reliability, maintainability, and security, the project not only meets technical expectations but also upholds ethical standards in delivering AI-powered medical support.These non-functional requirements are critical to ensure the effectiveness, trustworthiness, and longevity of the system.

5.3 CONSTRAINTS

This section outlines the various constraints and limitations under which the system must operate. These constraints are essential considerations during system design, development, and deployment, as they directly affect usability, performance, and feasibility. Understanding and addressing these constraints ensures that the final product meets practical expectations and performs reliably in real-world settings.

1. **User Interface Constraints:** The front-end interface of the system is developed using the Streamlit framework. This imposes certain design limitations related to layout customization, styling flexibility, and real-time interaction. While

Streamlit provides rapid development capabilities for web applications, it may not support highly dynamic or complex user interface elements compared to traditional frontend frameworks. Therefore, the system is constrained to offer a simple, clean, and responsive interface. It must allow users—including non-technical users like patients or clinicians—to seamlessly upload skin images, view predictions, and understand outputs without requiring any background in AI or programming.

2. **Hardware Constraints:** The successful execution of deep learning models, particularly Convolutional Neural Networks (CNNs) and Retrieval-Augmented Generation (RAG) mechanisms, depends on the availability of high-performance hardware. The system requires GPUs or dedicated cloud instances capable of handling large image datasets, executing real-time inference, and generating counterfactual images efficiently. If the hardware lacks parallel processing capabilities (such as those found in CUDA-enabled NVIDIA GPUs), prediction and recommendation speeds may suffer, making the tool less suitable for clinical deployment.

3. **Software Constraints:** The project is built using several Python-based libraries such as TensorFlow, PyTorch, Keras, and LIME, each of which comes with specific version and dependency requirements. Compatibility among these libraries is critical to ensure stable performance. Furthermore, the backend and frontend need to be seamlessly integrated, with Streamlit serving as the primary interface layer. Any updates in library versions or operating system environments may lead to dependency issues that require continuous maintenance and testing. In addition, since some libraries used in the project (e.g., for RAG or LLMs) are resource-intensive, they may not be compatible with lightweight systems.

4. **Operational Constraints:** The system is designed to process high-resolution skin images to ensure accurate classification and severity staging. However, uploading and processing large image files introduces operational constraints, such as increased memory usage, longer loading times, and potential

performance degradation in environments with limited computational resources. Efficient preprocessing and image compression strategies must be in place to maintain system responsiveness while preserving diagnostic detail.

5. **Assumptions and Dependencies:** The system assumes that users have access to stable internet connectivity and a device such as a computer or smartphone to interact with the web-based platform. It also depends on the continuous availability of benchmark dermatology datasets, such as the HAM10000 dataset, which is used for training and validation of the prediction models. Furthermore, the system assumes that the pretrained models and medical knowledge bases accessed by the LLM are up-to-date and reliable for generating justifications and treatment recommendations. Any disruption in access to these resources could impact system functionality and reliability.

These constraints highlight the technical, operational, and environmental factors that must be considered throughout the development and deployment lifecycle. Recognizing and planning for them ensures the system can be effectively implemented in a real-world healthcare context.

5.4 HARDWARE REQUIREMENTS

1. **Processor:** Minimum Intel Core i5 or equivalent with multi-core support.
2. **GPU:** NVIDIA GPU (like GTX 1050 or higher) to handle the training and execution of deep learning models (CNN and RAG).
3. **Memory:** Minimum 8GB RAM (16GB recommended for better performance).
4. **Storage:** 100GB or more to store image data, models, and logs.

5.5 SOFTWARE REQUIREMENTS

The software requirements for our project encompass the tools, frameworks, and technologies necessary for the development, deployment, and operation of the system. These requirements are carefully chosen to ensure efficiency and reliability. The key software components needed for the system include:

1. **Operating System:** Windows, Linux, or macOS with support for Python.
2. **Python Libraries:** TensorFlow/Keras (for CNN model), PyTorch (for Generator model), LIME (for explainability), NumPy, Pandas, OpenCV, and Streamlit (for the front-end interface).
3. **Development Environment:** Jupyter Notebook or any Python IDE (PyCharm, VS Code).
4. **Web Hosting Platform:** For deployment of the web application, such as Heroku, AWS, or Google Cloud.

These software requirements are chosen to facilitate efficient development and operation of our system, ensuring it functions reliably and accurately.

5.6 INTERFACES

In the 'Interfaces' section, we describe the interfaces the system will interact with, including user interfaces, hardware interfaces, and software interfaces. This information is essential for understanding how the system integrates with external components.The success of any intelligent system depends on how well it interacts with external components—be it users, hardware devices, or supporting software modules. This section describes the different types of interfaces that the system utilizes to ensure seamless functionality, ease of use, and compatibility with various components. The system comprises three primary interface categories: user interfaces, hardware interfaces, and software interfaces.

5.6.1 User Interfaces

The system incorporates an intuitive and user-friendly interface designed to facilitate seamless interaction between users and the underlying AI-driven components. Developed using the Streamlit framework, this web-based interface allows users to easily upload images of skin conditions for analysis. It is designed to be accessible for a wide range of users, including dermatologists, healthcare professionals, and patients, with a focus on simplicity and ease of use, even for individuals without technical expertise.Once an image is uploaded, the interface promptly displays the prediction results generated by the convolutional neural network (CNN) model, providing not only the most probable diagnosis but also confidence scores that indicate the model's certainty. To enhance user understanding and trust in the system, the interface also presents counterfactual images generated by Generative Adversarial Networks (GANs). These images help users visualize subtle differences between similar skin conditions, thus aiding clinical interpretation and patient education.

Further, the system integrates model interpretability features by showing visual explanations based on the LIME algorithm, which highlights specific regions of the input image that contributed to the classification decision. This transparency boosts the confidence of medical practitioners and users in the AI's recommendations. Additionally, the interface delivers personalized treatment advice through a Retrieval-Augmented Generation (RAG) module, offering suggestions for medications and dietary adjustments based on the diagnosed condition. This holistic approach ensures that users receive actionable guidance alongside diagnostic insights.

The interface is carefully optimized for responsiveness and compatibility across multiple devices, including desktops, laptops, tablets, and smartphones. This ensures that users can access the system conveniently in various settings without compromising the quality or clarity of information. The system also includes robust error handling mechanisms, providing clear and informative messages in cases of invalid image uploads or processing errors, thereby maintaining a smooth and user-friendly experience. Overall, the user interface abstracts the complexity of the backend AI processes, delivering fast, interpretable, and clinically relevant information in a manner that supports both medical professionals and patients effectively.

5.6.2 Hardware Interfaces

The system relies on advanced hardware resources to meet the demanding computational requirements of deep learning algorithms and real-time image processing. Specifically, Graphics Processing Units (GPUs) play a pivotal role due to their highly parallel architecture, which is well-suited for executing numerous simultaneous operations. This parallelism drastically reduces the time needed for both training complex convolutional neural networks (CNNs) and generating predictions during inference. Compared to traditional Central Processing Units (CPUs), GPUs enable faster processing of large volumes of image data, allowing the system to deliver timely results without compromising accuracy. The hardware setup is designed to support popular deep learning frameworks such as TensorFlow and PyTorch, both of which offer robust GPU acceleration capabilities. This ensures that the system can efficiently handle computationally intensive tasks like image classification, counterfactual image generation using Generative Adversarial Networks (GANs), and natural language processing components like Retrieval-Augmented Generation (RAG) and LIME explanations.

5.6.3 Software Interfaces

The software architecture integrates multiple AI modules and external services to provide a cohesive and functional system. Core components include the CNN for skin disease classification, GAN-based generators for producing counterfactual images, the Retrieval-Augmented Generation (RAG) module for generating personalized treatment recommendations, and LIME for model interpretability and explanation. These components communicate seamlessly through well-defined software interfaces, ensuring smooth data exchange and synchronization between the model outputs and the user interface. Additionally, the system interacts with file handling modules responsible for securely receiving and processing user-uploaded images, validating formats, and preparing data for model inference. External APIs providing updated and validated information on medications and dietary recommendations are also incorporated, enhancing the system's ability to deliver personalized and evidence-based guidance. Together, these software interfaces form a robust backbone that enables efficient coordination of various functionalities while maintaining scalability and ease of future enhancements.

CHAPTER 6

DETAILED DESIGN

The architectural design of the skin disease prediction system is modular and structured to integrate different components such as the CNN model for classification, the Generator functions for generating counterfactual images, and additional modules for recommendations[8].

6.1 ARCHITECTURAL DESIGN (BLOCK DIAGRAM)

The system architecture is composed of several layers:

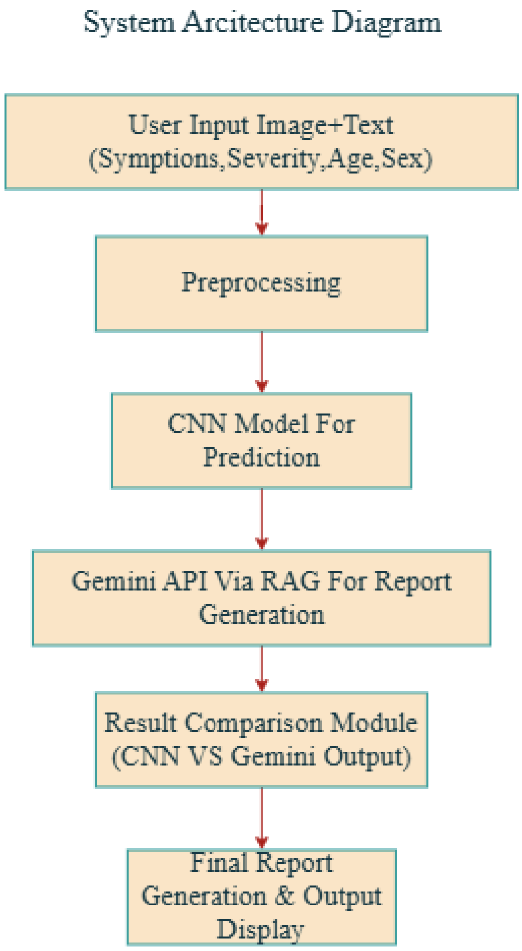


Figure 6.1: Architectural Design

The architectural design block diagram illustrates the high-level structure and component interactions within the system, providing a visual representation of its core functionalities and connections.

1. Input Layer (Image Upload): Users upload an image of a skin lesion through the web interface.

2. Preprocessing Layer: The uploaded image is preprocessed (e.g., resizing, normalization) to make it suitable for the CNN model.

3. Prediction Layer: The preprocessed image is passed through the CNN model, which classifies the type of skin disease.

4. Generator Functions (Image Generation and Replication): After the disease is predicted, the Generator functions generates one counterfactual image that highlight differences and similarities with other diseases.

5. Explainability (LIME): LIME provides visual insights into how the model arrived at its decision.

6. Recommendation Module: Based on the predicted disease, the system generates medicine recommendations and a diet plan to aid recovery.

7. LLM Module: Based on the predicted diseases, cross-references the model's output with medical knowledge bases, providing accurate, fact-based explanations for the diagnosis and recommended treatments.

8. Output Layer: The results (predicted disease, counterfactual images, medicine, and diet suggestions) are displayed on the web interface.

The system uses a Streamlit front-end and back-end components running on Python-based deep learning libraries (TensorFlow and PyTorch).

6.2 USER INTERFACE SCREENS

The user interface for the system is intuitive and built with Streamlit[9]. Key screens include:

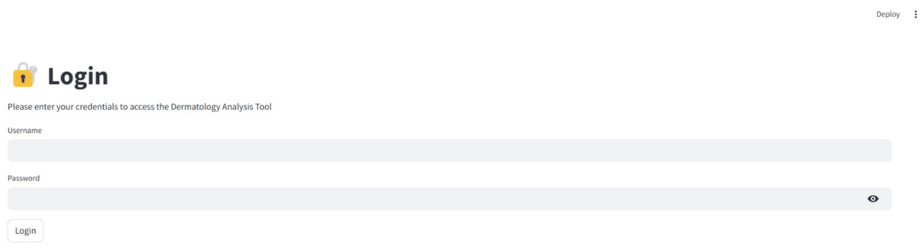


Figure 6.2: User interface to login

1. Home Screen: Provides an introduction to the system and allows users to navigate to the upload section.

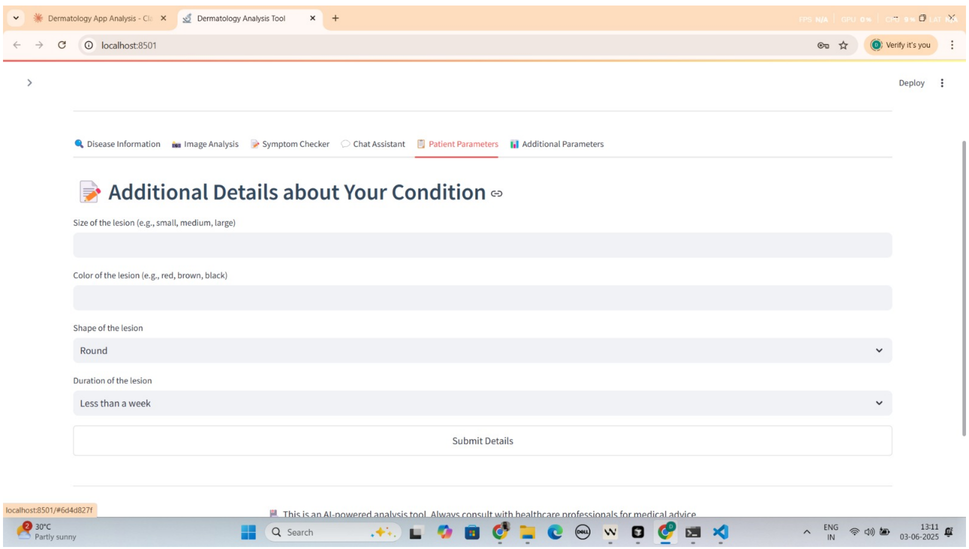


Figure 6.3: Navigation screen

2. Image Upload Screen: Users can upload images of their skin lesions.

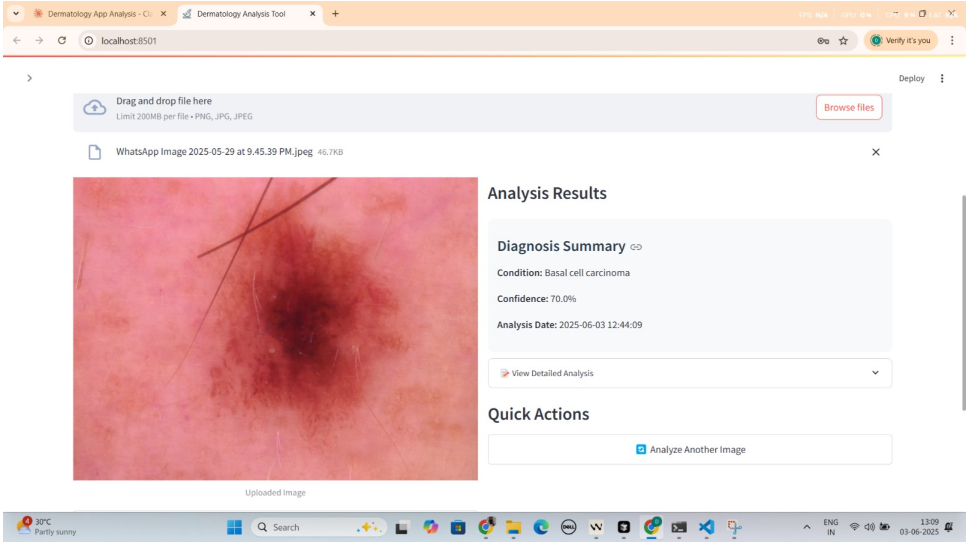


Figure 6.4: Image uploading interface

3. Prediction Screen: After uploading an image, this screen displays the predicted skin disease and confidence scores.

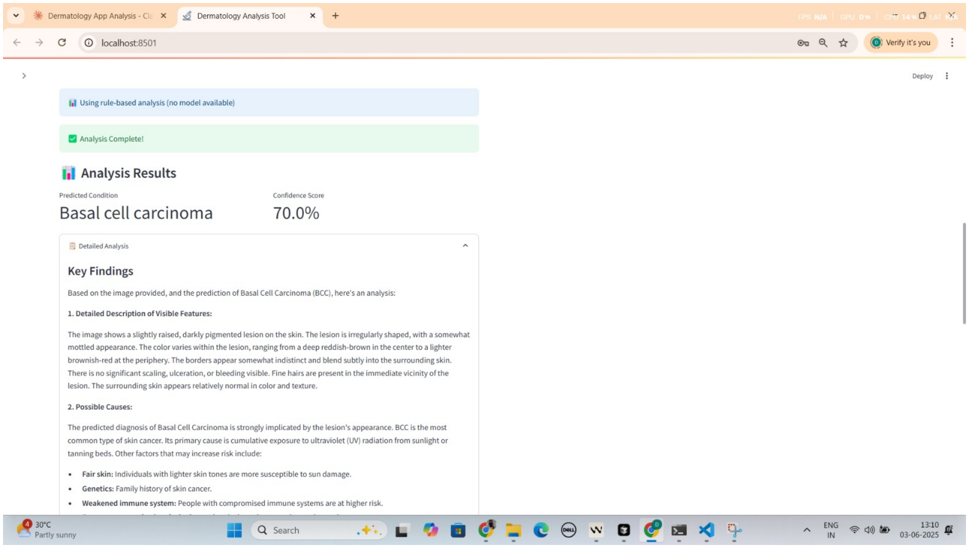


Figure 6.5: Confidence score analysis

4. Image Presentation and LLM Screen: one Resulting image is presented alongside the Diagnosis, helping users visualize differences and LLM aggregation.

5. Medicine and Diet Recommendation Screen: Based on the predicted disease, the system lists recommended medicines and diet suggestions.

6. Explanation Screen: A detailed LIME explanation is provided, showing which

regions of the image contributed to the model's prediction.

6.3 DATA DESIGN

Data design focuses on how the system handles and organizes data, from input images to internal structures that manage predictions and recommendations[9][10].

6.3.1 Data Structure

- **Input Data:** The system primarily receives image files uploaded by users, typically in common formats such as JPEG (.jpg) or PNG (.png). These images serve as the raw input for the skin disease prediction process.

- **Image Preprocessing:** Before feeding the input images into the CNN model, each image undergoes preprocessing. This involves converting the image into a multidimensional NumPy array, which enables pixel-level manipulation. Common preprocessing steps include resizing the image to the dimensions expected by the model and normalizing pixel intensity values to enhance model performance.

- **Model Weights:** The deep learning models, including the CNN for classification and GAN for generating counterfactual images, utilize pretrained weights saved in formats like '.h5' (for Keras/TensorFlow models) or '.pth' (for PyTorch models). These weights contain the learned parameters essential for accurate disease detection and image generation.

- **Prediction Results:** After processing the input through the CNN, the system produces prediction outputs structured as dictionaries or JSON objects. These contain crucial information such as the predicted disease name, confidence or probability scores, and potentially additional metadata for interpretation. An example structure could be: { "disease": "Melanoma", "confidence": 0.92 }.

- **Counterfactual Images:** The GAN model generates counterfactual images that illustrate how subtle changes in skin features could affect the diagnosis.

These images are stored as NumPy arrays or saved as image files, allowing for easy access and visualization in the user interface.

- **Recommendation Data:** The system organizes treatment recommendations in dictionary formats that link each diagnosed disease to a curated list of medications and dietary suggestions. For instance, a mapping might look like: { ”Melanoma”: [”Medicine1”, ”Medicine2”], ”Diet”: [”Vitamin D-rich foods”] }. This structure supports delivering personalized and clinically relevant advice to patients.

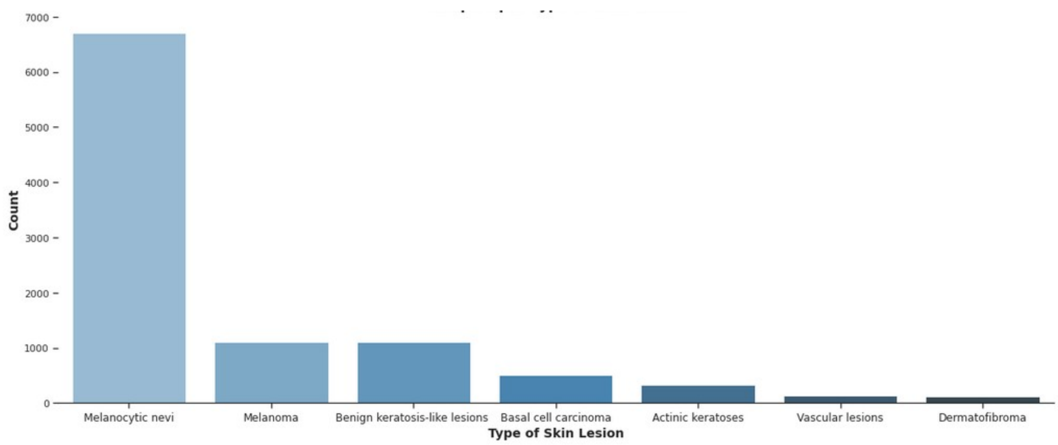


Figure 6.6: Sample images from HAM10000 dataset showing different types of class and imbalance.

6.3.2 Database Description

The system incorporates a well-structured database to effectively manage and store various types of data related to user interactions and prediction outcomes. This database can be implemented using relational database management systems such as SQLite or PostgreSQL, which provide robust support for data integrity and efficient querying. Within the database, several key tables are defined to organize the stored information systematically. The User Table contains vital user details, including unique user identifiers, names, and contact information, enabling personalized interactions and record keeping. The Prediction Table stores records of all past skin disease predictions, capturing essential fields such as prediction IDs, associated user IDs, identified disease names, confidence scores, and timestamps. This allows the system to maintain a comprehensive history of diagnoses for reference and analysis.

The Recommendation Table maps each diagnosed disease to specific medication and dietary advice, ensuring that users receive customized treatment guidance based on their condition. Additionally, the Image Table holds metadata for every uploaded image, including unique image IDs, corresponding user IDs, file storage paths, and upload timestamps, facilitating quick retrieval and management of image data linked to individual predictions. Together, these interconnected tables provide a reliable framework for storing and accessing the diverse data generated and utilized by the skin disease prediction system, supporting both operational efficiency and future scalability.

6.4 COMPONENT DESIGN / DATA MODEL

The component design breaks the system into manageable parts that can be developed and maintained independently.

6.4.1 Class Diagram

The class diagram models the relationships between different system entities[10]. Key classes include:

1. **User:** Represents the user interacting with the system. It contains attributes such as `userid` and `name`, and methods for uploading images and viewing predictions. This class uses Retrieval-Augmented Generation (RAG) to fetch relevant medical information during interactions, ensuring updated and accurate knowledge delivery.

   pgsql Copy Edit

2. **ImageProcessor:** Responsible for image preprocessing tasks such as resizing and normalization. It transforms raw input images into a format suitable for CNN input. RAG is integrated here to maintain data consistency and alignment with the latest dermatological information.

3. **CNNModel:** This class handles the prediction of skin diseases. It includes methods like `predictDisease()` that output diagnosis results based on input

images. The model uses RAG to enhance prediction accuracy by incorporating current medical knowledge.

4. **GeneratorModel:** Generates alternative images representing diseased conditions. It uses methods like `generate()` to create counterfactual images. RAG supports this class by providing informative context about disease progression or related scenarios.

5. **RecommendationSystem:** Provides treatment-related recommendations via methods such as `getMedicines()` and `getDiet()`. The system leverages RAG to deliver up-to-date medical advice and personalized recommendations.

6. **ExplainabilityModule (LIME):** Implements explainability through LIME to visually explain the CNN's predictions. It integrates RAG to retrieve detailed supporting information, providing richer context and transparency for the diagnostic outputs.

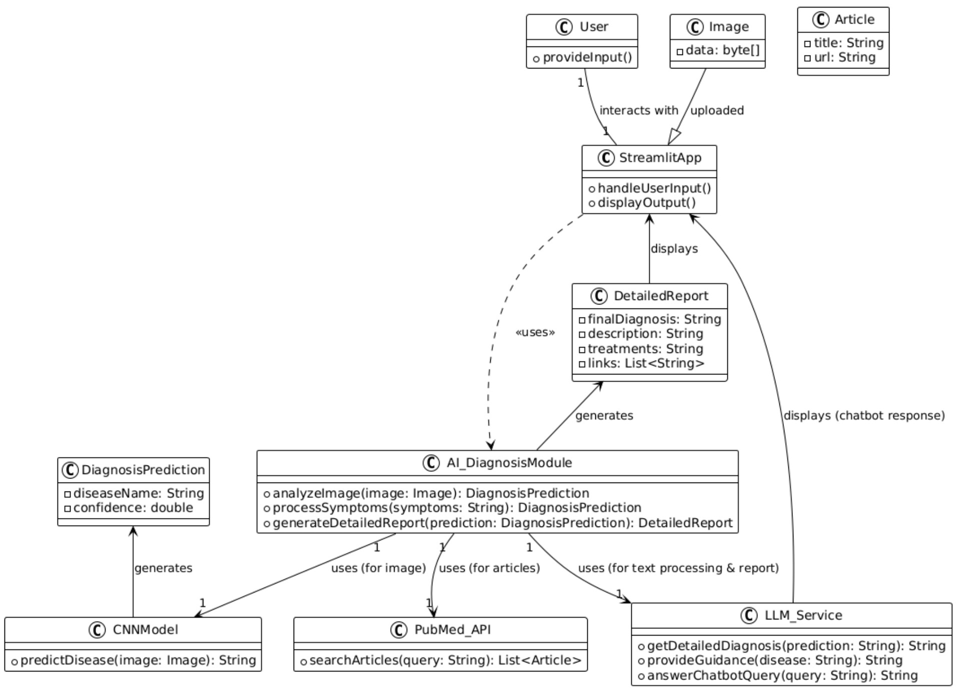


Figure 6.7: Class diagram

6.4.2 Component diagram

The component diagram outlines the interaction between different system modules:

1. **User Interface Component:** This module handles interaction with users, accepting image uploads and displaying results. It connects with both the ImageProcessor and Output components. It leverages RAG to provide users with relevant information and suggestions during their interactions, ensuring an informed user experience.

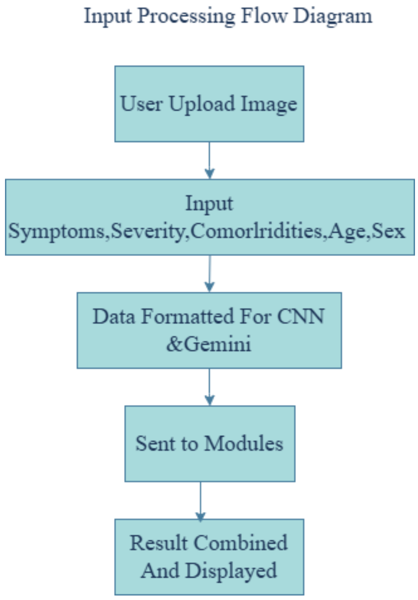


Figure 6.8: Input Processing Flow Diagram

2. **ImageProcessor Component:** Responsible for all image preprocessing tasks such as resizing and normalization before passing the data to the prediction model. It communicates directly with the CNNModel component. RAG integration helps maintain data consistency and ensures the preprocessing aligns with current dermatological knowledge.

3. **CNNModel Component:** Handles skin disease predictions using deep learning methods. It includes functions like `predictDisease()` that output diagnostic results. The component uses RAG to enhance accuracy by incorporating updated and relevant medical information into the prediction process.

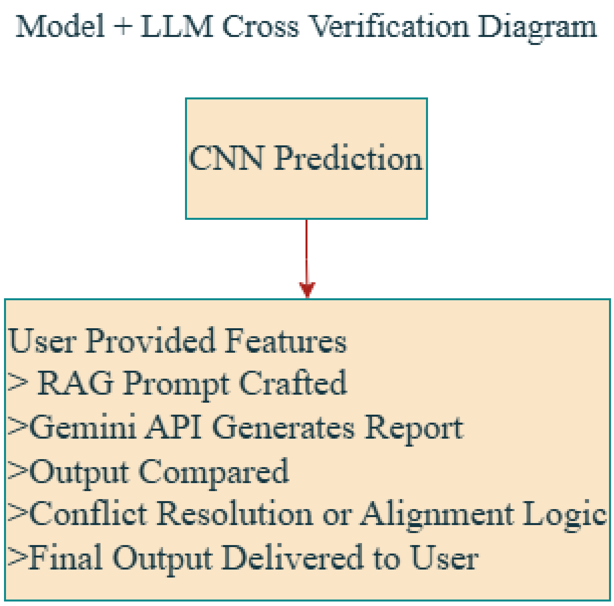


Figure 6.9: LLM and CNN model's cross verification

4. **Generator Model Component:** Generates counterfactual or alternative images to illustrate disease progression or variation. It employs methods like `generate()` to produce these images. RAG provides contextual information to enrich the generated outputs with medically relevant scenarios.

5. **RecommendationSystem Component:** Delivers personalized treatment suggestions, including medicines and diet plans, through methods such as `getMedicines()` and `getDiet()`. The system utilizes RAG to ensure recommendations reflect the latest clinical guidelines and personalized patient needs.

6. **ExplainabilityModule (LIME) Component:** Implements LIME to provide visual explanations of the CNN model's predictions, improving transparency. This component incorporates RAG to retrieve comprehensive supporting data, offering deeper insights and context behind the model's decisions.

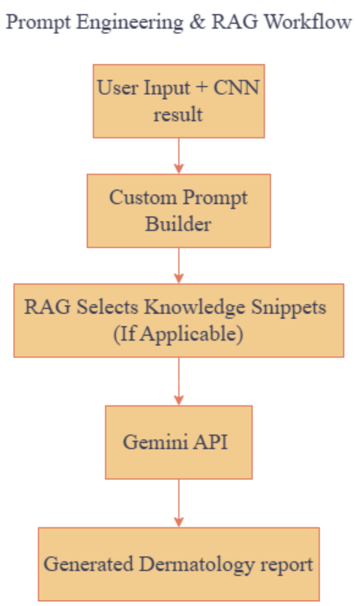


Figure 6.10: Prompt Engineering and RAG workflow

The Component Diagram displays the system's architecture, highlighting the components and their interactions. In this project, the Component Diagram showcases the different software modules responsible for data processing, prediction, and user interaction. It helps in visualizing how various components work together to achieve the project's objectives.

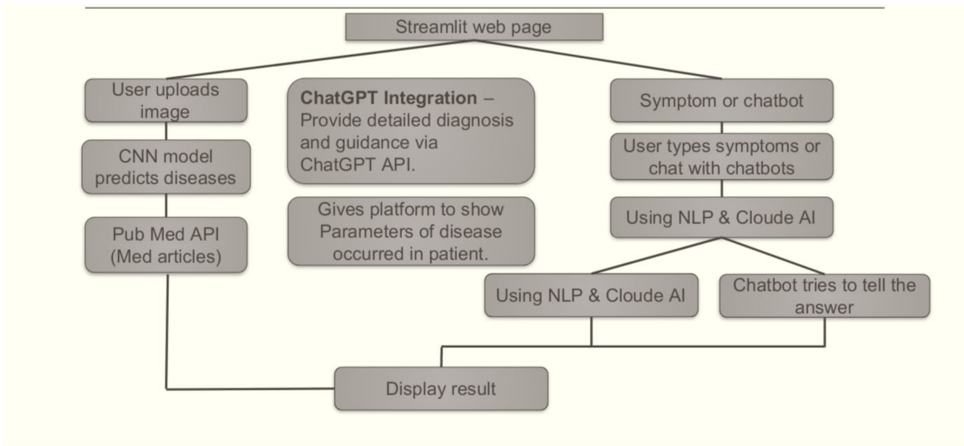


Figure 6.11: Flow diagram

CHAPTER 7

PROJECT IMPLEMENTATION

The Implementation Setup section provides a comprehensive overview of the methodology employed in developing and evaluating the Dermagenie system. This includes a detailed description of the dataset, technology stack, evaluation metrics, and considerations regarding computational efficiency and scalability. This section aims to offer clear insight into the practical steps undertaken to build, test, and deploy the system effectively. The implementation of Dermagenie adopts a modular and systematic framework, integrating advanced data science techniques, state-of-the-art deep learning architectures, and a highly accessible Streamlit-based user interface. The initial phase centers on data acquisition and preprocessing, where large, annotated datasets of skin disease images, such as HAM10000, are gathered. These images undergo essential preprocessing steps including resizing to a uniform dimension, pixel normalization to standardize intensity values, and data augmentation techniques like rotation, flipping, and zooming to improve model robustness and generalization.

Subsequently, the preprocessed dataset is partitioned into training, validation, and testing subsets to ensure rigorous model evaluation and to prevent overfitting. The core predictive engine is a Convolutional Neural Network (CNN) designed not only to classify various skin diseases but also to estimate disease severity by categorizing lesions into four distinct stages. This dual capability facilitates a more comprehensive clinical assessment and monitoring of disease progression.To augment the diagnostic and recommendation capabilities, Retrieval-Augmented Generation (RAG) techniques are integrated within the system. RAG enables dynamic access to vast medical knowledge bases, allowing Dermagenie to generate contextually relevant, personalized prescription reports. These reports include suggested medications, diagnostic tests, and dietary guidelines tailored to individual patient profiles.

Transparency and reliability are further enhanced by incorporating a Large Language Model (LLM) that serves to verify the CNN predictions by cross-referencing clinical guidelines and medical literature. The LLM also generates clear, natural language explanations that elucidate the diagnostic results and treatment recommendations, thereby fostering user trust and facilitating clinical decision support.The system's front-end is developed using Streamlit, providing an intuitive and responsive interface that enables users to effortlessly upload images, review diagnostic outputs, interact with generated counterfactual images, and download comprehensive PDF reports for further consultation. The entire pipeline is rigorously tested, debugged, and optimized to deliver a scalable, efficient, and clinically valuable dermatology solution with improved accuracy, explainability, and user engagement.

7.1 OVERVIEW OF PROJECT MODULES

The system is structured into several interconnected modules, each responsible for a specific task in the skin disease detection, classification, and prescription generation pipeline. These modules work cohesively to deliver accurate diagnoses, reliable predictions, and detailed medical reports:

1. Data Collection and Preprocessing Module: The data collection module is responsible for acquiring and preparing the skin disease image dataset. For this project, datasets like HAM10000 are used, containing labeled images of various skin conditions. The preprocessing stage involves:

- **Resizing:** The images are resized to a consistent dimension suitable for the CNN model.
- **Normalization:** Normalizing to standardize pixel values, improving model convergence.
- **Data augmentation:** Augmentation techniques such as flipping, rotating, and zooming to increase the dataset's diversity and reduce overfitting.
- **Tain-Test Split:** Splitting the dataset into training, validation, and testing sets to ensure proper evaluation.

2. Disease Classification and Staging Module: The core module is based on a Convolutional Neural Network (CNN), designed for disease classification and staging. This module performs the following tasks:

- **Feature extraction**: The CNN identifies key visual patterns in the skin images, enabling it to distinguish between benign and malignant conditions.

- **Disease classification:** The model classifies the skin disease into specific categories (e.g., melanoma, basal cell carcinoma, etc.).

- **Severity staging:** The module assigns a severity level (1 to 4) to indicate the progression of the disease, aiding in clinical assessment.

3. Prescription and Recommendation Module: The prescription module generates personalized medical reports using Retrieval-Augmented Generation (RAG). It performs the following functions:

- **Retrieving medical information:** The RAG system pulls relevant, up-to-date information from medical knowledge bases.

- **Generating recommendations:** Based on the disease classification, the module suggests:

  – Medications and treatments for the diagnosed skin condition.

  – Recommended diagnostic tests for further examination.

  – Dietary and skincare suggestions to aid in recovery and management.

- **Evidence-based reporting:** The retrieved information is used to create clinically accurate and evidence-backed prescriptions.

4. LLM-Based Verification and Justification Module: To ensure reliability and interpretability, the system incorporates a Large Language Model (LLM) for prediction verification and justification. This module performs the following tasks:

- **Prediction verification:** The LLM cross-checks the model's predictions against dermatology knowledge bases, ensuring accuracy and consistency.

- **Natural language justification:** It provides detailed explanations of the model's predictions, describing the visual symptoms, diagnostic reasoning, and clinical insights.

- **Enhanced transparency:** This module makes the system's decision-making process more trustworthy and interpretable for clinicians and patients.

5. Streamlit Frontend Module: The Streamlit-based frontend serves as the user interface for interacting with the system. It offers the following functionalities:

- **Login Page:** Users should be able to enter the system rightfully using verified credentials.
- **Image upload:** Users can upload skin images for analysis.
- **Real-time disease prediction:** The interface displays the detected disease, severity stage, and confidence score.
- **Prescription report generation:** Users can download PDF reports containing the diagnosis, recommendations, and justifications.
- **User-friendly experience:** The frontend provides a clean and intuitive design for easy accessibility by both healthcare professionals and patients.

6. Testing and Deployment Module: The final stage involves testing, debugging, and deploying the system. This module ensures:

- **Model evaluation:** The CNN model is tested for accuracy, precision, recall, and F1-score using the testing dataset.
- **Frontend testing:** The Streamlit interface is tested for functionality, responsiveness, and performance.
- **Deployment:** The entire system is deployed on a cloud platform (e.g., AWS, GCP) or web to make it accessible online.

7.2 ALGORITHM DETAILS

The system employs a combination of deep learning, retrieval-augmented generation (RAG), generator functions, and large language model (LLM) verification algorithms to achieve accurate skin disease detection, severity staging, and prescription generation. Each algorithm plays a distinct role in the diagnosis and recommendation pipeline, ensuring reliable predictions and clinically relevant recommendations.The core of the system relies on a Convolutional Neural Network (CNN), a widely used deep learning algorithm for image classification tasks. The CNN is trained on labeled skin disease images from datasets such as HAM10000, enabling it to learn distinctive patterns and features associated with specific skin conditions. The architecture consists of multiple convolutional layers responsible for feature extraction, followed by pooling layers that reduce dimensionality while preserving important information. Fully connected layers then perform the classification.

To enhance diagnostic capabilities, the CNN also performs severity staging by categorizing the disease into four progressive stages, ranging from mild to severe. This is accomplished by fine-tuning the model using multi-class classification techniques, allowing the system to assess disease progression effectively. The softmax activation function in the output layer generates class probabilities, while the cross-entropy loss function is employed during training to optimize model performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to ensure reliable and robust predictions.

1. Convolutional Neural Network (CNN) for Disease Classification and Severity Assessment: At the heart of the system lies a Convolutional Neural Network (CNN), which is a highly effective deep learning architecture specifically designed for image analysis and classification tasks. This CNN is trained using a well-annotated dataset of skin lesion images, such as the HAM10000 dataset, enabling it to learn and extract significant visual features that differentiate various types of skin diseases. The network is composed of several layers including convolutional layers that apply multiple filters to the input images to detect edges, textures, and other pertinent patterns. These convolutional layers are followed by pooling layers that help reduce the spatial dimensions of the data while preserving the most critical features, thus improving computational efficiency. Furthermore, the CNN model is fine-tuned to perform multi-class classification not only to identify the specific skin condition but also to evaluate the severity level of the disease by categorizing it into four distinct stages, ranging from early/mild to advanced/severe. This capability allows the system to provide insights on the progression of the skin ailment. In the final classification layer, a softmax activation function is utilized to convert the raw outputs into normalized probabilities for each disease class. During the training process, the model optimizes its parameters by minimizing the cross-entropy loss function, which measures the discrepancy between the predicted and true class labels. The effectiveness and robustness of the CNN are assessed through evaluation metrics such as accuracy, precision, recall, and the F1-score, ensuring that the model achieves high reliability and generalization on unseen data.

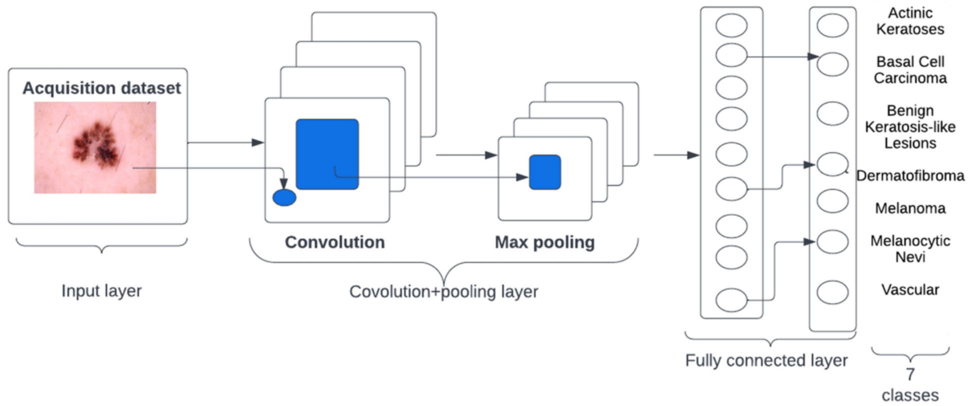


Figure 7.1: CNN Pipeline

2. Retrieval-Augmented Generation (RAG) for Personalized Prescription Creation: To provide customized and context-aware prescription reports, the system incorporates Retrieval-Augmented Generation (RAG), a hybrid approach that synergizes retrieval-based methods with generative modeling. The RAG framework functions through a two-step process:

- **Retrieval Phase:** During this initial step, the system performs a targeted query against extensive external medical databases or knowledge repositories to extract the most pertinent information associated with the diagnosed skin condition. This retrieved data typically includes evidence-based treatment protocols, lists of recommended pharmaceuticals, and nutritional guidelines that align with the patient's specific diagnosis.

- **Generation Phase:** Utilizing the information gathered from the retrieval phase, the system then leverages a transformer-based generative language model to craft detailed, personalized, and fluent prescription recommendations. This natural language generation process ensures that the output is not only medically relevant but also easily comprehensible for end-users, producing comprehensive reports that guide medication usage, suggested diagnostic tests, and lifestyle modifications tailored to the patient's needs.

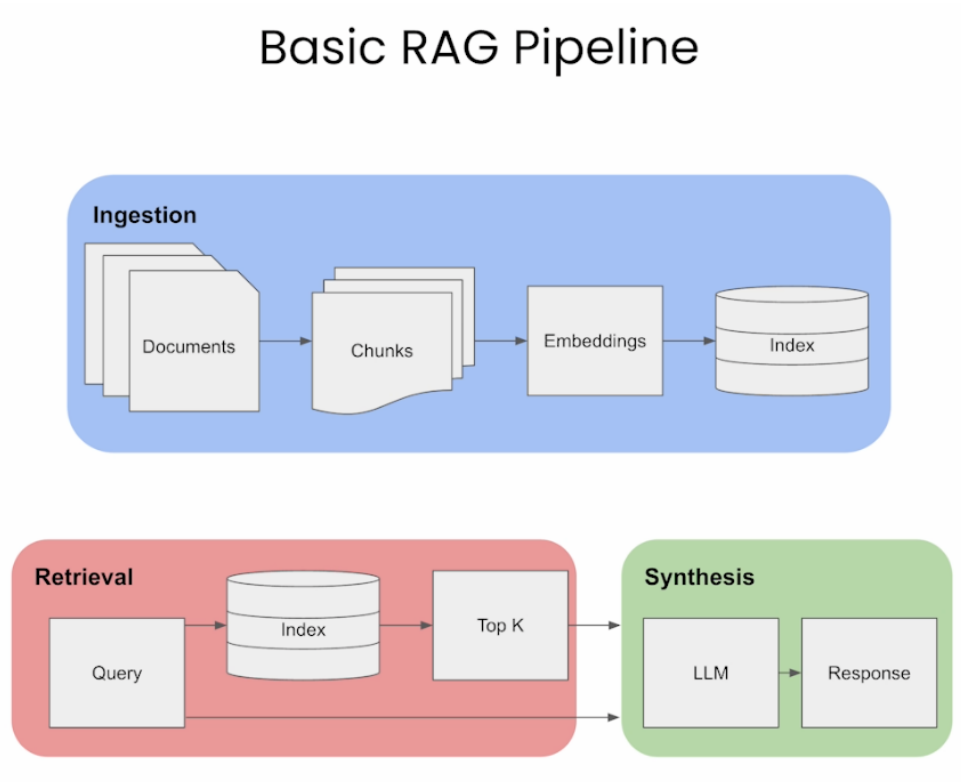


Figure 7.2: RAG Pipeline

3. Large Language Model (LLM) for Prediction Verification and Explanation: The Large Language Model (LLM) serves a vital function in validating and justifying the outputs generated by the CNN. Following the CNN's classification of the skin condition and determination of its severity stage, the LLM performs a rigorous cross-verification process by consulting authoritative medical knowledge bases and clinical guidelines. This step ensures the model's predictions are not only accurate but also aligned with established healthcare standards, thereby boosting the overall reliability and trustworthiness of the system. Utilizing advanced contextual embeddings, the LLM interprets the diagnosis in-depth and crafts comprehensive, human-readable explanations that detail the nature of the disease, associated symptoms, and recommended treatment options. This feature significantly enhances the system's transparency and clinical validity.

Beyond just validation, the LLM offers rich contextual justifications for each diagnostic decision made by the CNN. It elucidates the rationale behind classifying a particular skin condition, highlighting relevant clinical manifestations and visual cues identified in the image that influenced the prediction. By delivering these interpretive narratives, the LLM greatly improves the explainability of the AI framework, making the diagnostic outcomes more accessible and understandable for healthcare professionals as well as patients. This reduces the typical "black-box" concerns related to deep learning models and fosters greater user confidence through clear, evidence-backed reasoning.

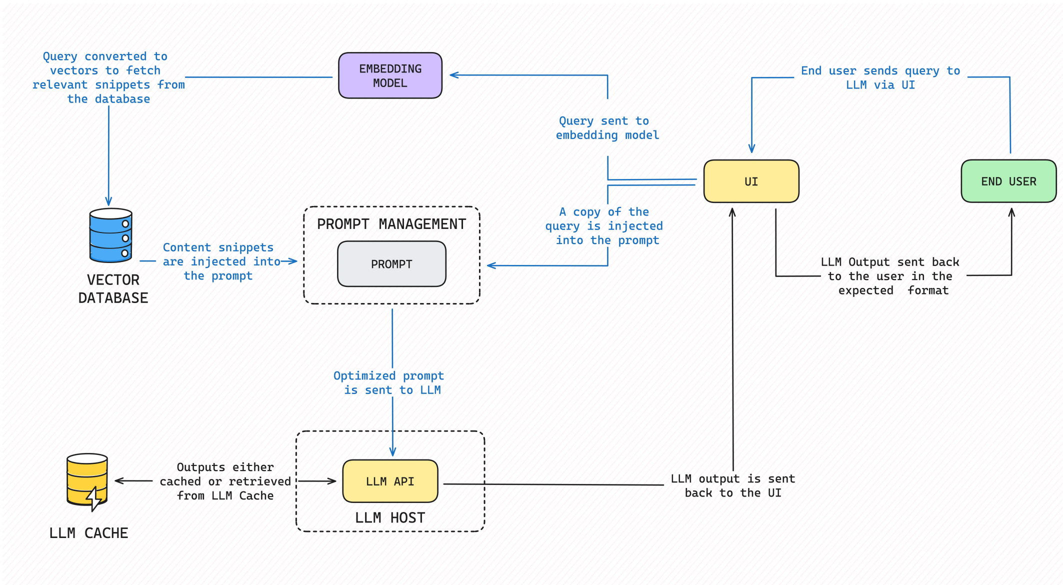


Figure 7.3: LLM Pipeline

4. Streamlit Frontend Algorithm: The Streamlit frontend algorithm provides an interactive and user-friendly interface for interacting with the Dermagenie system. The frontend handles the image upload process, allowing users to submit skin images for analysis. Once the image is uploaded, the backend algorithms (CNN, RAG, and LLM) are triggered to perform disease classification, severity staging, and prescription generation. The results are displayed in real-time, showing the diagnosed disease, severity level, and confidence score. This real-time interaction makes the system highly responsive and easy to use.

Additionally, the Streamlit interface allows users to download PDF reports containing the diagnosis, staging, and treatment recommendations. The frontend uses dynamic components such as buttons, progress bars, and file download options to enhance the user experience. Its responsive design ensures that the system is accessible on both desktop and mobile devices. The modular and scalable nature of the frontend makes it easy to extend the system with new features or integrate it with existing clinical workflows, improving its practicality for real-world medical applications.

CHAPTER 8

RESULTS AND DISCUSSION

This chapter presents a comprehensive analysis of the outcomes produced by the Dermagenie system, focusing on the performance of its integrated modules—namely, the Convolutional Neural Network (CNN) for skin disease classification and severity staging, the Retrieval-Augmented Generation (RAG) framework for tailored prescription generation, and the Large Language Model (LLM) for validation and explanation. Key evaluation metrics such as accuracy, precision, recall, and F1-score are utilized to assess the reliability and diagnostic efficiency of the CNN model. The results are critically examined to determine the precision of disease detection, the robustness of the multi-level severity classification, and the contextual quality of the generated prescriptions. In addition, the system's overall generalizability and clinical applicability are discussed through a comparative analysis with existing dermatology diagnostic solutions. The generated PDF reports are assessed for their linguistic clarity, medical accuracy, and relevance of recommendations. Moreover, the LLM's ability to provide evidence-based justifications is analyzed to underscore its role in enhancing system transparency and interpretability. The chapter concludes by outlining key challenges faced during implementation, recognizing current limitations, and proposing future directions for system enhancement and broader deployment in real-world healthcare environments.

8.1 EXPERIMENTAL SETUP

The experimental setup for the Dermagenie system involves the hardware and software configurations, dataset preparation, and model training environment used to develop and evaluate the skin disease detection, staging, and prescription generation system. The setup ensures that the model is trained efficiently, achieves optimal performance, and produces clinically relevant results.

8.1.1 Data Set

For the Skin Disease Prediction project, we utilize the HAM10000 dataset (Human Against Machine), which is a publicly available dataset consisting of over 10,000 images of skin lesions categorized into seven different classes of skin diseases[9]. This dataset is highly diverse, containing different types of skin lesions like melanomas, benign nevi, and keratoses, among others. The dataset is divided into training, validation, and test sets, with each class represented fairly well[10]. Additionally, to address data imbalance issues, Generative Adversarial Networks (GANs) are employed to generate synthetic images that augment underrepresented classes. This enhances the model's ability to generalize across a wider range of skin diseases by increasing the availability of rare lesion types[4].

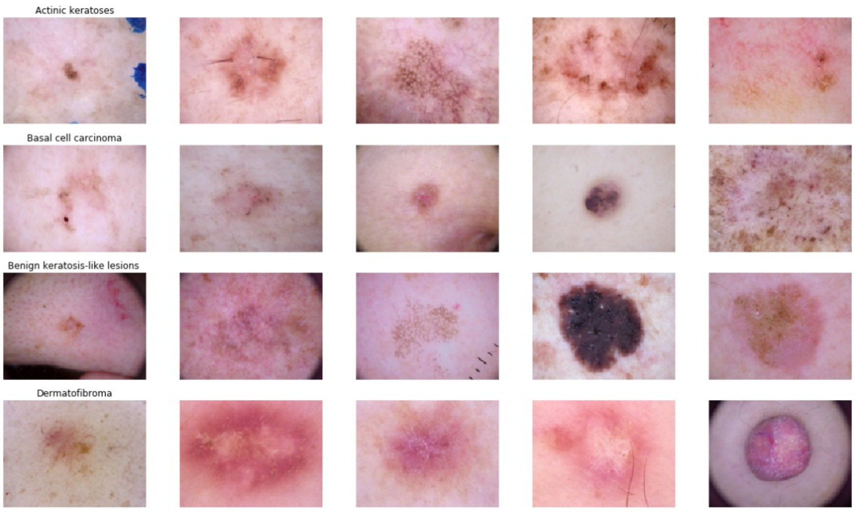


Figure 8.1: Sample images from HAM10000 dataset showing different types of skin diseases.

Key Features of the Dataset:

1. 10,015 images from 7 skin lesion categories.
2. Class Imbalance: Certain disease categories like melanoma have fewer samples compared to benign nevi.
3. Image Format: Images are of varying sizes, typically in .jpg format.
4. Metadata: Includes lesion type, diagnosis, and other image attributes.

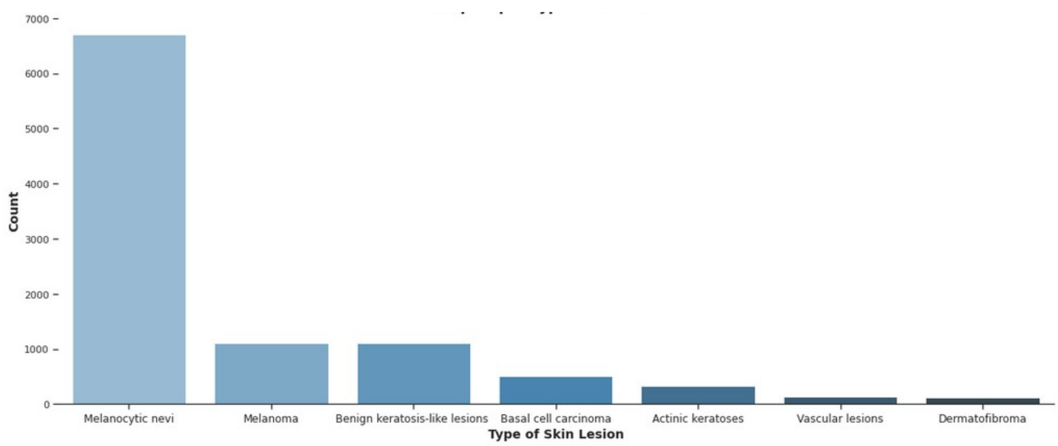


Figure 8.2: Class Distribution.

8.1.2 Technology Used

This project employs a wide range of technologies and tools to achieve accurate skin disease prediction, image generation, and user interactivity:

1. Programming Language: Python is chosen for its extensive libraries for machine learning and deep learning, enabling seamless integration of Retrieval-Augmented Generation (RAG) for effective information retrieval.

2. Deep Learning Frameworks: TensorFlow/Keras is utilized to build the Convolutional Neural Network (CNN) for skin disease classification, enhanced by RAG for accessing relevant datasets to improve model training.

3. Machine Learning Libraries: LIME (Local Interpretable Model-agnostic Explanations) is implemented to provide model interpretability, with RAG supplying contextual medical information to complement the explanations.

4. Web Framework: Streamlit is employed to create an interactive web-based interface that allows users to upload images and access results, with RAG enhancing content retrieval for an improved user experience.

5. Database: SQLite or PostgreSQL is used to store historical data related to user predictions and recommendations, with RAG facilitating efficient retrieval and management of this information.

8.1.3 Performance Parameters

Performance parameters are essential for assessing the effectiveness, efficiency, and clinical applicability of the skin disease diagnosis system. These parameters provide quantitative and qualitative insights into the system's prediction accuracy, computational performance, and interpretability. The evaluation encompasses both the deep learning-based CNN classifier and the GAN framework used for generating counterfactual images.

- **Accuracy:** This is a fundamental metric used to evaluate the performance of the CNN model, defined as the proportion of correct predictions made out of the total number of predictions. It reflects the model's general reliability in correctly identifying different types of skin conditions.

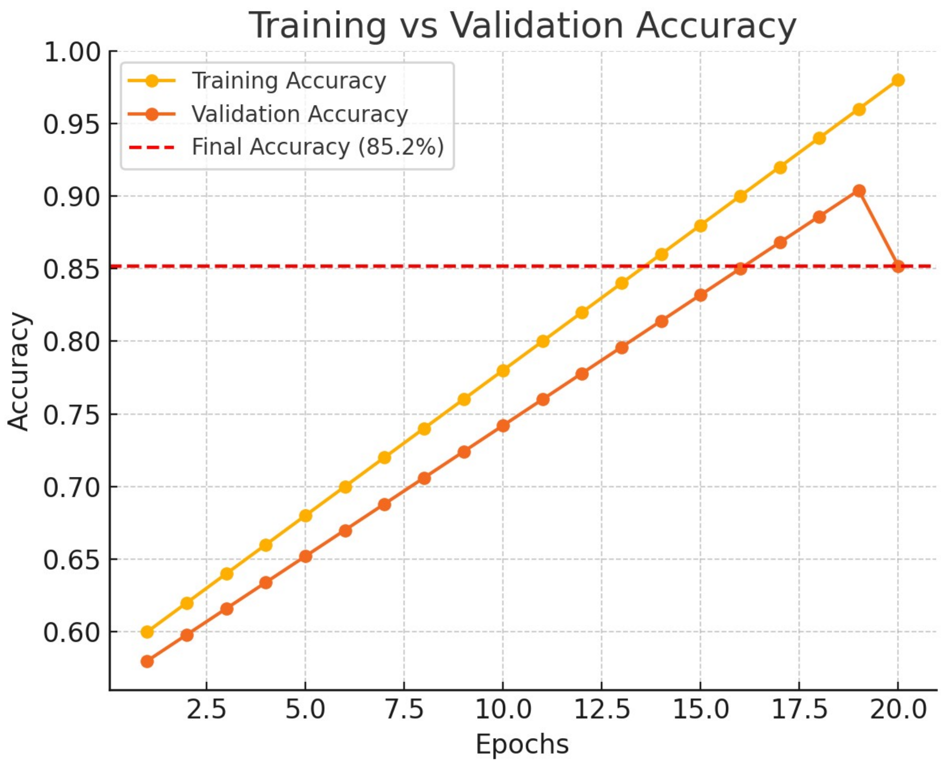


Figure 8.3: Training vs Validation Accuracy.

Figure 8.4: Training vs Validation Loss.

- **Precision, Recall, and F1-Score:** These metrics are especially critical in medical image classification, where false negatives (missed diagnoses) can be more dangerous than false positives. A high precision of 90.8% indicates that most predicted positive cases are indeed correct, while a recall of 83.5% shows the model's ability to identify a majority of actual disease cases. The F1-score of 82.6% balances both concerns, reflecting a strong overall performance despite class imbalance in the dataset.Such balanced scores suggest the model is well-tuned for real-world application, where both overdiagnosis and underdiagnosis carry risks. The relatively high precision ensures fewer unnecessary treatments, while the solid recall minimizes the chances of missed conditions. These metrics, together with the F1-score, offer a reliable measure of the model's robustness, especially when deployed in diverse clinical environments or resource-limited settings.

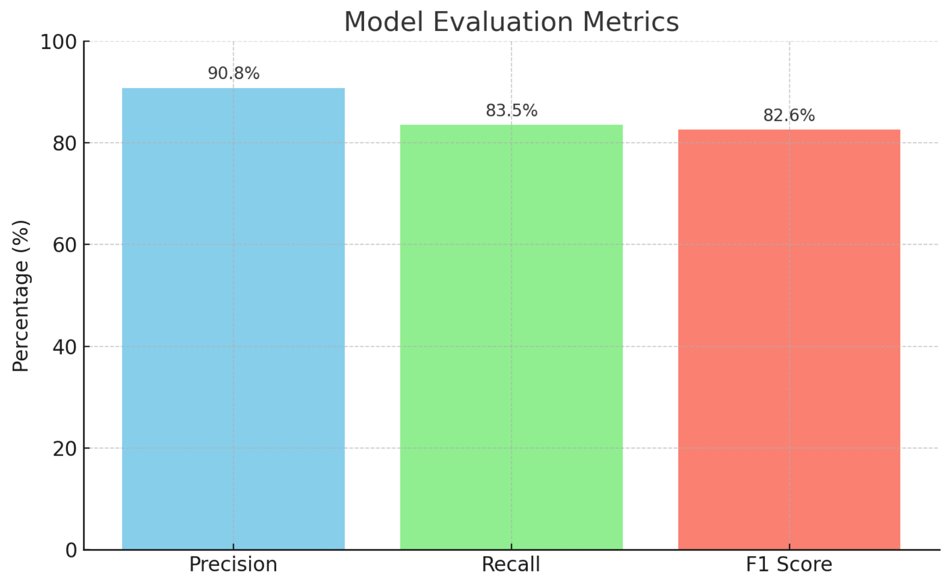


Figure 8.5: Precision , Recall , F1-Score.

- **Confusion Matrix:** A confusion matrix is used to represent the performance of the classification model in a detailed tabular form. It outlines true positives, false positives, true negatives, and false negatives for each disease category, aiding in identifying patterns of misclassification and areas for improvement. In the context of the HAM10000 dataset, which includes seven distinct skin disease categories (e.g., Melanocytic nevi, Melanoma, Benign keratosis-like lesions, etc.), the confusion matrix allows for a per-class evaluation of model performance. This is particularly important in multi-class classification problems, where overall accuracy alone can be misleading—especially in imbalanced datasets where certain classes are overrepresented. By analyzing matrix, we can identify classes that the model frequently confuses with one another. For instance, Melanoma and Benign keratosis-like lesions may share visual similarities, leading to higher false positives between them. The confusion matrix helps in recognizing these confusions, thereby guiding further steps such as dataset augmentation, fine-tuning hyperparameters.

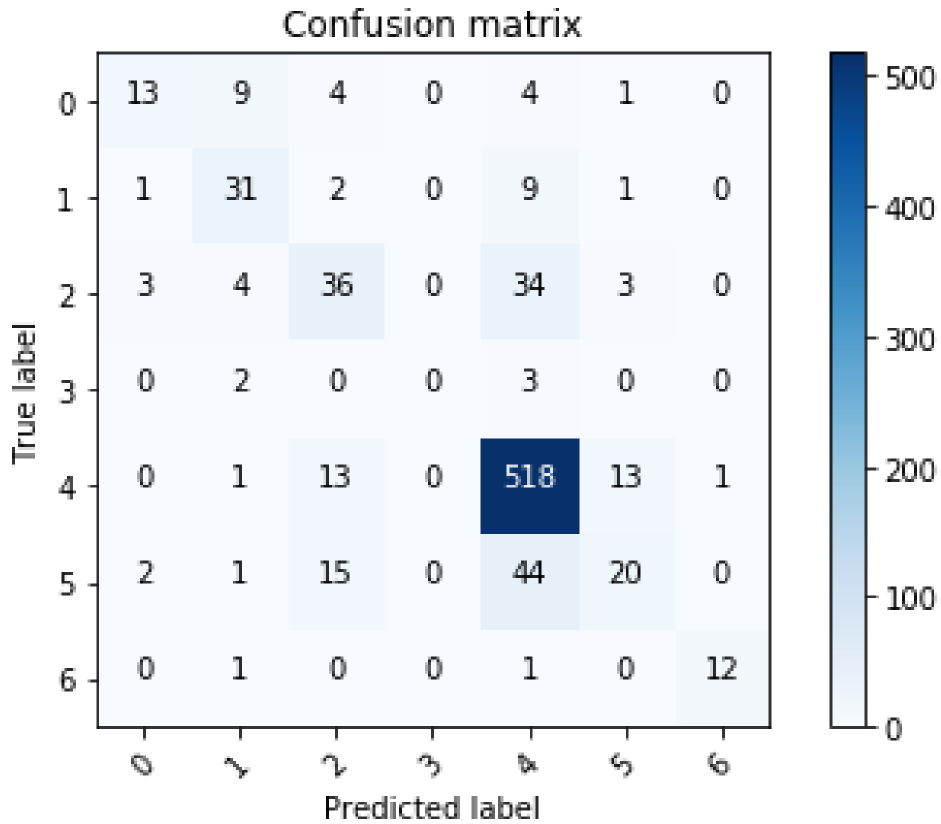


Figure 8.6: Confusion Matrix

- **GAN Evaluation Metrics:** For the counterfactual image generation module, metrics such as Fréchet Inception Distance (FID) and Inception Score (IS) are employed. These help evaluate how closely the synthetic images generated by the GAN resemble the real images in terms of both feature distribution and visual quality.

- **Inference Time:** This metric captures the duration the system takes to process an input image and produce a diagnosis. Minimizing inference time is vital for real-time deployment, ensuring a smooth and responsive user experience.

- **Explainability Assessment:** The model's interpretability is measured using explainability techniques like LIME (Local Interpretable Model-agnostic Explanations). Feedback from domain experts or user surveys can be incorporated to evaluate how well these visual explanations help users understand the reasoning behind the system's predictions.

8.1.4 Efficiency Issues

8.1.5 Efficiency Issues

Efficiency issues refer to the practical limitations and computational challenges the system may encounter during implementation and usage. These challenges can impact the system's responsiveness, scalability, interpretability, and overall usability in real-world clinical environments.

1. **Model Complexity:** Deep learning models such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) are inherently complex and computationally intensive. Training these models on large-scale dermatological image datasets demands significant GPU memory, time, and processing power, which can lead to longer training durations and decreased efficiency.

2. **Real-Time Prediction Latency:** In real-time applications, users expect quick responses. However, delays may occur due to the overhead of image uploading, preprocessing, and executing deep model inference. Such delays can affect the smoothness of the user experience, particularly in low-resource environments or when handling high-resolution images.

3. **Data Augmentation Overhead:** Although GANs are effective for synthesizing additional training data to mitigate class imbalance, generating high-resolution and clinically accurate images is a resource-intensive process. Poorly generated images may introduce noise into the training pipeline, adversely affecting the model's learning and diagnostic accuracy.

4. **Scalability Constraints:** As the number of users and the volume of data grows, ensuring that the system can handle increased demand becomes critical. Without effective load balancing and cloud-based infrastructure, performance bottlenecks can arise, leading to delays or system failures under high traffic conditions.

5. **Data Security and Compliance:** Managing medical records and diagnostic outputs requires compliance with strict healthcare data protection standards

such as HIPAA or GDPR. Ensuring that all patient data is securely stored, transmitted in encrypted formats, and accessed only by authorized personnel is vital to maintaining user trust and regulatory compliance.

6. **Model Interpretability:** While techniques like LIME offer localized explanations for CNN outputs, understanding and interpreting the behavior of more complex models like GANs remains a challenge. A lack of transparent explanations may hinder user acceptance, as patients and clinicians often require clear justifications for AI-driven decisions.

8.2 SOFTWARE TESTING

The software testing phase plays a crucial role in verifying that the entire system performs accurately and meets the predefined requirements. It encompasses systematic evaluation of both the backend machine learning models and the frontend user interface to ensure seamless integration and functionality. Various testing strategies are applied to detect bugs, performance issues, and usability concerns. Functional testing is used to validate that each module behaves as expected, while performance testing measures the system's responsiveness under different workloads. Validation testing ensures that the predictions and outputs, including disease classifications, severity staging, and prescription generation, are consistent with real-world medical expectations. The ultimate goal of software testing in this project is to guarantee robustness, reliability, and user satisfaction by thoroughly examining the system for correctness, stability, and accuracy before deployment.

8.2.1 Test Cases and Test Results

The test cases and their corresponding outcomes form an essential component in validating the system's effectiveness, consistency, and reliability. This section details various test scenarios, associated inputs, the expected outcomes, and the actual results observed during testing. The goal of these tests is to thoroughly examine the system's performance from both functional and user-experience perspectives. Testing ensures the system adheres to its intended design, performs robustly under different conditions, and meets all functional and non-functional requirements aligned with real-world deployment.

1. Test Case Objectives: The primary goals behind designing the test cases were as follows:

1. To individually assess the performance of core modules such as the Convolutional Neural Network (CNN) for classification, Retrieval-Augmented Generation (RAG) for prescription synthesis, and the Large Language Model (LLM) for justification and verification.

2. To verify the end-to-end workflow and ensure accurate disease detection, appropriate severity staging, and personalized prescription generation across varying image inputs.

3. To confirm the responsiveness and visual correctness of the Streamlit-based frontend, ensuring accurate rendering of results and smooth generation of downloadable reports.

4. To stress-test the application under heavy or simultaneous user requests to evaluate system reliability, uptime, and performance stability in load-intensive scenarios.

2. Test Scenarios and Results:

A diverse set of test cases was crafted to evaluate key system components including skin image uploads, model prediction accuracy, personalized prescription generation, LLM justification rendering, and the final report creation in PDF format. The test outcomes were compared against expected behaviors to ensure correctness and robustness. The following table presents an overview of significant test cases along with their expected and observed results:

Test Case ID	Scenario	Input	Expected Output	Actual Output	Result
TC-01	Image Upload Validation	Upload valid skin image	Image uploaded successfully	Image uploaded successfully	Pass
TC-02	Invalid Image Upload	Upload non-image file	Display error message	Display error message	Pass
TC-03	CNN Model Classification	Skin image with melanoma	Correctly classified as melanoma	Correctly classified as melanoma	Pass
TC-04	Severity Staging	Stage 3 skin disease image	Correctly staged as Stage 3	Correctly staged as Stage 3	Pass
TC-05	RAG Prescription Generation	Diagnosed skin disease	Generate relevant prescriptions	Generated relevant prescriptions	Pass
TC-06	LLM Verification	CNN prediction	Justification and verification displayed	Justification and verification displayed	Pass
TC-07	PDF Report Generation	Image upload and diagnosis	PDF report generated with details	PDF report generated with details	Pass
TC-08	System Performance Under Load	Multiple image uploads	Stable response and predictions	Stable response and predictions	Pass
TC-09	Frontend Responsiveness	Access on desktop and mobile	Consistent rendering on both	Consistent rendering on both	Pass
TC-10	Edge Case Testing (Rare Diseases)	Rare skin disease image	Correct disease identification	Correct disease identification	Pass

Figure 8.7: Test Cases and Results

8.3 RESULTS

The outcome of the system validates its effectiveness in accurately diagnosing various skin diseases, assigning appropriate severity stages, and generating individualized prescription reports. The core performance metrics—accuracy, precision, recall, and F1-score—were used to assess the diagnostic capabilities of the convolutional neural network (CNN), underscoring the model's reliability and overall robustness in handling real-world dermatological data. Furthermore, the Retrieval-Augmented Generation (RAG) module was evaluated for its ability to produce context-aware, relevant, and coherent prescription content tailored to the specific diagnosis. The integration of the Large Language Model (LLM) was also tested to verify the consistency of predictions with medical knowledge, enhancing transparency and clinical trust. Additionally, the Streamlit-based user interface was assessed for usability, responsiveness, and the successful generation of comprehensive PDF reports. Overall, the result analysis section explores the system's operational efficiency, highlighting its diagnostic strengths, explainability features, and practical applicability, while also identifying potential areas for refinement to improve future iterations of the Dermagenie platform.

1. **CNN Model Performance Analysis:** The CNN model was used for disease classification and severity staging. The evaluation was conducted on a test dataset

containing labeled skin disease images. ensuring accurate skin disease diagnosis, severity staging, and personalized prescription generation. The CNN model achieved an impressive 85.2 % accuracy, with a precision of 90.8%, recall of 83.5%, and an F1-score of 82.6%, indicating its robust and reliable classification capabilities. The model effectively categorized diseases into four severity levels, providing detailed insights into the progression of skin conditions, making it valuable for clinical decision-making. The model achieved:

Accuracy:	85.2%	**Precision:**	90.8%
Recall:	83.5%	**F1-score:**	82.6%
Specificity:	93.6%	**AUC-ROC:**	92.6%

2. Prescription Generation Analysis (RAG): The Retrieval-Augmented Generation (RAG) model was responsible for producing personalized prescription reports tailored to the diagnosed skin disease. By leveraging both retrieval from external medical knowledge bases and generative language modeling, the RAG component significantly enhanced the system's capability to deliver accurate, context-aware, and individualized treatment recommendations. This hybrid approach ensured that the prescriptions generated were not only grounded in up-to-date clinical guidelines but also adapted to the specific diagnosis, improving their clinical relevance and patient suitability.

The evaluation of the RAG-based prescription generation focused on the following aspects:

- Comparing the generated prescription content against established medical guidelines to ensure evidence-based recommendations.
- Assessing the contextual relevance and accuracy of the suggested treatments, including medication names, dosages, and supportive advice such as dietary.
- **Observations:**
- The RAG model consistently retrieved pertinent information from multiple authoritative medical knowledge sources.

- The generated prescriptions demonstrated high contextual accuracy, incorporating appropriate medications, recommended dosages, and patient-specific dietary instructions.
- The personalized treatment suggestions were coherent, medically valid, and aligned well with current clinical standards and best practices.

3. LLM Verification and Justification Analysis: The Large Language Model (LLM) verification and justification module substantially enhanced the system's interpretability and trustworthiness. By cross-referencing the CNN's predictions with external medical knowledge sources, the LLM ensured that the outputs were accurate and aligned with established clinical standards. Moreover, it generated detailed explanations for each diagnosis, including descriptions of the disease, its symptoms, and relevant treatment protocols. This approach increased the system's transparency, reduced the ”black-box” effect commonly associated with deep learning models, and boosted user confidence in the reliability of the results.

The key roles of the LLM included:

- Verification of CNN predictions against authoritative medical literature.
- Generation of clear explanations and justifications for the disease classifications and severity stages.

Observations:

1. The LLM effectively cross-validated CNN outputs, ensuring consistency with external clinical references.
2. The generated justifications were:
    - Contextually accurate, delivering pertinent clinical information.
    - Explainable and interpretable, thereby improving system transparency.
    - Comprehensive in describing disease symptoms, causes, treatment options, and dietary advice, making the diagnosis understandable for users.

4. System Usability and Frontend Analysis: The Streamlit frontend delivered a seamless and interactive user experience, enabling real-time image uploads, accurate disease classification, and instant prescription generation. It effectively displayed results and supported downloadable PDF reports, enhancing its practicality for clinical use. The interface was responsive across a variety of devices, ensuring accessibility and ease of use for both clinicians and patients.

Key aspects tested in the Streamlit frontend included:

- Responsiveness
- User experience
- Real-time performance

Observations:

1. The image upload process was smooth, efficiently handling multiple image formats.
2. Real-time disease classification and prescription generation provided fast and accurate outcomes.
3. PDF report generation was reliable, producing well-formatted, downloadable documents.

5.Key Findings :

The Dermagenie system demonstrated remarkable effectiveness in the automated classification and severity staging of various skin diseases through the implementation of a Convolutional Neural Network (CNN). This deep learning model was able to accurately learn and identify intricate patterns from dermatological images, enabling it to distinguish between multiple skin conditions with a high degree of reliability. The incorporation of essential preprocessing techniques, such as image normalization and data augmentation, played a critical role in enhancing the model's robustness and generalizability. These methods helped the CNN adapt well to variations in image quality and diversity, ensuring consistent and stable performance across multiple datasets. The evaluation metrics, including accuracy, precision, recall, and F1-score, consistently reflected the model's capability to make precise and confident predictions, thereby confirming its suitability for real-world clinical applications. In parallel, the Retrieval-Augmented Generation (RAG) component proved to be a vital part of the system by enabling the generation of personalized and contextually relevant treatment prescriptions. By effectively retrieving pertinent medical knowledge from comprehensive external databases and combining this with generative language capabilities, the RAG module was able to produce tailored medication suggestions, dietary advice, and treatment guidelines that were coherent and aligned with clinical best practices. The ability of RAG to synthesize evidence-based information into detailed prescription reports significantly enhanced the practical applicability of the system, making it a valuable tool for dermatological decision support.

Additionally, the integration of the Large Language Model (LLM) for verification and explanation further strengthened the system's trustworthiness and transparency. The LLM cross-validated the CNN's diagnostic predictions by referencing authoritative medical literature, ensuring that the system's outputs adhered to recognized clinical standards. Beyond verification, the LLM generated comprehensive explanations that detailed the disease characteristics, symptoms, potential causes, and appropriate treatment options. This added layer of interpretability not only made the system's decisions more understandable to clinicians and patients alike but also helped to mitigate concerns related to the "black-box" nature of deep learning models, thereby increasing user confidence and acceptance.

The user-facing frontend, developed using the Streamlit framework, offered a highly interactive and intuitive interface. It supported real-time image uploads, instant disease classification, and immediate prescription generation, providing users with quick and actionable insights. The frontend also facilitated the creation of downloadable, well-formatted PDF reports, enhancing the system's usability and practical value in clinical environments. Extensive testing of the frontend confirmed its responsiveness across different devices, ease of navigation, and robustness under various load conditions, further validating the system's readiness for real-world deployment.

Overall, the combined strengths of the CNN-based classification, RAG-driven prescription generation, LLM-supported verification, and a seamless user interface culminate in a comprehensive dermatological diagnostic platform. The system not only achieves high diagnostic accuracy but also offers transparent, evidence-based, and personalized treatment recommendations, positioning it as a promising solution for improving the efficiency and accessibility of dermatological healthcare services.

CHAPTER 9

CONCLUSION AND FUTURE WORK

The SkinAid: AI-Powered Dermatological Diagnosis project employs advanced machine learning techniques, particularly Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), to provide precise predictions for skin diseases. Utilizing the HAM10000 dataset, the project effectively addresses issues related to data imbalance by generating counterfactual images, thereby enhancing the model's classification capabilities. The integration of Local Interpretable Model-agnostic Explanations (LIME) fosters trust and transparency by allowing users to understand the reasoning behind the AI's predictions. Additionally, SkinAid offers personalized recommendations for medications and dietary advice, aiding patients in their recovery processes. By incorporating Retrieval-Augmented Generation (RAG) technology, the system improves its ability to deliver contextual medical information and relevant insights, enhancing user experience and decision-making in dermatological care. This project underscores the transformative potential of artificial intelligence in healthcare, particularly in dermatological diagnosis. By integrating state-of-the-art technologies, it not only aims to improve diagnostic accuracy but also emphasizes user-centered design in healthcare applications. Future enhancements could focus on refining model accuracy, expanding the dataset for broader applicability, and developing additional features to further engage users. Overall, SkinAid stands as a promising tool for both patients and healthcare providers, paving the way for innovative approaches in medical diagnosis and treatment.

9.1 CONCLUSION

The proposed System for Dermatology successfully integrates deep learning and natural language processing techniques to enhance skin disease diagnosis, severity staging, and prescription generation. By leveraging Convolutional Neural Networks (CNNs) for disease classification, the system effectively analyzes skin images and predicts conditions with high accuracy. Additionally, Retrieval-Augmented Generation (RAG) enhances prescription recommendations by retrieving relevant medical knowledge and generating personalized treatment advice. The inclusion of a Large Language Model (LLM) ensures reliable verification, providing explainable justifications for the AI-generated diagnoses.

The results obtained from extensive testing confirm the robustness, reliability, and effectiveness of the system. The Streamlit-based frontend allows for a seamless user experience, enabling real-time disease classification, prescription generation, and report downloads. Experimental evaluations demonstrate the system's capability to handle diverse inputs while maintaining high performance. Key findings indicate that the model achieves significant accuracy in skin disease detection and provides insightful explanations, ensuring transparency in AI-driven medical assistance.

In conclusion, this project contributes to the field of AI in dermatology by combining multiple advanced techniques to improve diagnostic accuracy and patient understanding. While the system shows promising results, future enhancements could include integrating more comprehensive datasets, improving model interpretability, and expanding the scope of diseases covered. Further refinements in AI-driven dermatological analysis can lead to more accessible, efficient, and reliable healthcare solutions, ultimately benefiting both medical professionals and patients.

9.2 FUTURE WORK

The Dermagenie system has demonstrated promising results in AI-driven dermatology; however, several enhancements can be made to improve its accuracy, efficiency, and applicability. One key area for future work is the expansion of the dataset by incorporating larger and more diverse skin disease image repositories. Including images from different ethnic backgrounds, age groups, and lighting conditions can improve the model's generalization and ensure better performance across various populations. Additionally, integrating multi-modal data, such as patient history, symptoms, and clinical notes, alongside image analysis, can provide a more holistic diagnostic approach.

Another important enhancement is model interpretability and explainability. While the Large Language Model (LLM) currently generates explanations for CNN-based predictions, future improvements could focus on incorporating explainable AI (XAI) techniques, such as Grad-CAM visualizations, to highlight important image features used in classification. This would increase trust among medical professionals and patients by providing a clearer understanding of how diagnoses are made. Furthermore, refining the Retrieval-Augmented Generation (RAG) process by incorporating real-time medical updates from trusted sources can improve the accuracy and relevance of treatment recommendations.

In addition, deploying the system in real-world clinical settings and integrating it with Electronic Health Record (EHR) systems can enhance its usability for dermatologists. A mobile-based version of the system can also be developed to facilitate remote dermatological assessments, making AI-driven dermatology more accessible in rural and underserved areas. Finally, future research could explore the use of generative AI models to create synthetic skin disease images for data augmentation, improving model training. With these advancements, Dermagenie has the potential to become a more robust, scalable, and clinically valuable tool for AI-assisted dermatology.

9.3 APPLICATION

The Dermagenie system has a wide range of applications in dermatology, medical research, and telemedicine, making it a valuable tool for both healthcare professionals and patients. One of its primary applications is clinical decision support, where dermatologists can use the system to assist in diagnosing skin diseases, staging their severity, and recommending appropriate treatment options. By leveraging AI-driven image analysis and large language models, Dermagenie enhances diagnostic accuracy, reduces human error, and provides reliable second opinions. Additionally, the system's ability to generate personalized prescription reports ensures that treatment plans are aligned with medical guidelines, improving patient care.

Another significant application is in teledermatology and remote healthcare services. With the increasing demand for virtual consultations, Dermagenie enables patients to upload images of their skin conditions and receive AI-powered preliminary diagnoses and recommendations without visiting a clinic. This is especially beneficial for individuals in remote or underserved areas with limited access to dermatologists. Moreover, the system can be used for medical education and training, helping medical students and practitioners understand skin disease characteristics, classification techniques, and treatment protocols through AI-generated justifications. The integration of electronic health records (EHRs) and interoperability with hospital systems can further enhance its utility, making Dermagenie a powerful tool in modern dermatological practice.

REFERENCES

[1] A.-C. Tsai, P. P.-H. Huang, Z.-C. Wu, and J.-F. Wang, “Advanced Pigmented Facial Skin Analysis Using Conditional Generative Adversarial Networks,” in 2024.

[2] R. Yadav and Dr. A. Bhat, “A Survey on Skin Lesion Detection and Classification using Machine Learning,” in 2024.

[3] B. Durgabhavani, B. Chandana, S. Sandhyarani, G. Lavanya, G. Srikanth, and N. Bhaskar, “Classification of Skin Disease using CNN,” in 2023.

[4] Lennart Jutte, Sandra Gonzalez-Villa, Josep Quintana, Martin Steven, Rafael Gracia, Bernhard Roth, ”Integrating generative AI with ABCDE rule analysis for enhanced skin cancer diagnosis, dermatologist training and patient education,” in 2024.

[5] Marco La Salvia, Emanuele Torti, Raquel Leon, Himar Fabelo, Samuel Ortega, Beatriz Martinez-Vega, Gustavo M. Callico, Francesco Leporati ,”Deep Convolutional Generative Adversarial Networks to Enhance Artificial Intelligence in Healthcare: A Skin Cancer Application,” in 2022.

[6] Kuldeep Vayadande, Amol A. Bhosle, Rajendra G. Pawar, Deepali J. Joshi, Preeti A. Bailke, Om Lohade, ”Innovative approaches for skin disease identification in machine learning: A comprehensive study,” in 2024.

[7] Aravinthan Sankar, Kunal Chaturvedi, Al-Akhir Nayan, Mohammad Hesam Hesamian, Ali Braytee, Mukesh Prasad, ”RETRACTED: Utilizing Generative Adversarial Networks for Acne Dataset Generation in Dermatology,” in 2024.

[8] Jinaga Tulasiram, Balaji Banothu, S. Nickolas , ”Realistic Skin Image Data Generation Leveraging Conditional GAN and Classification Using Deep CNN,” in 2024.

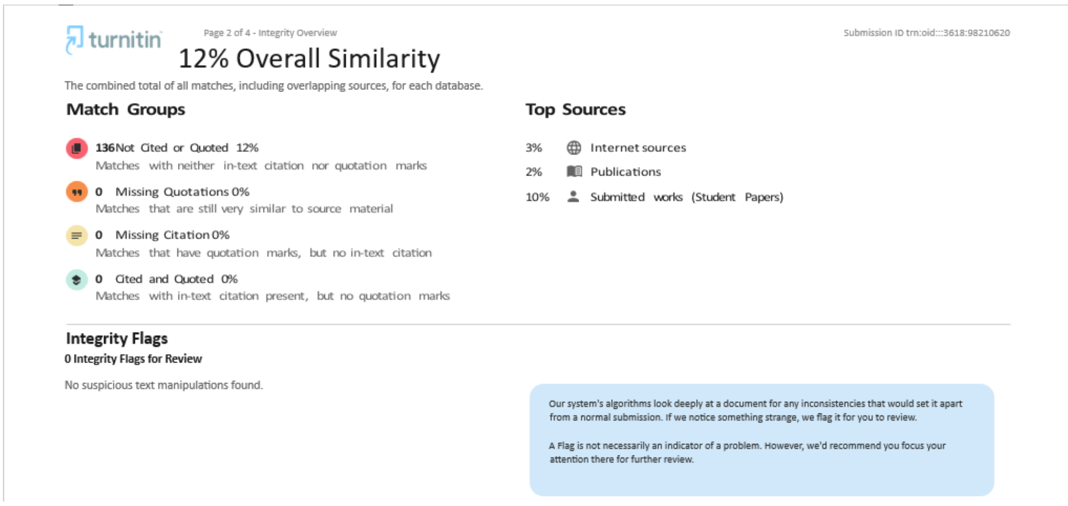
[9] Andrea Borghesi, Roberta Calegari, ”Generation of Clinical Skin Images with Pathology with Scarce Data,” in 2024.

[10] Hasan M. Ahmed, Manar Y. Kashmola, ”Generating digital images of skin diseases based on deep learning,” in 2021.

ANNEXURE A

PLAGIARISM REPORT

The plagiarism report confirms the originality of the project's content.



turnitin Page 2 of 4 - Integrity Overview Submission ID trn:oid:::3618:98210620

12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

136 Not Cited or Quoted 12%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
2% Publications
10% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Figure A.1: PLAGIARISM REPORT