



Palestine Technical University – Kadoorie
College of Engineering and Technology
Department of Computer Systems Engineering

Course name:

Graduation Project

Project title:

**DermAI: Intelligent Skin Cancer Detection
Using Convolutional Neural Network &
Transfer Learning Architectures**



Prepared By:

Maysam Rashed - 202110544
Raghad Mousleh - 202110283
Raghad Suliman - 202110010

Supervisor:

Dr.Rami Dib'i

Tulkarm, Palestine
14 Jan. 2026

ACKNOWLEDGMENT

We would like to express our sincere gratitude to our project supervisor, Dr. Rami, for his valuable guidance, constructive feedback, and continuous support throughout the development of this graduation project. His academic supervision and technical insights played a significant role in shaping the direction of this work and in overcoming the challenges encountered during its implementation.

This project was completed through sustained independent effort over an extended period, requiring effective time management, careful problem analysis, and persistence in addressing technical limitations and practical constraints. Access to academic resources and technical references was essential in supporting informed decision-making throughout the development process and in maintaining an engineering-oriented approach to problem solving.

We are deeply grateful to our families for their constant encouragement, patience, and understanding throughout a demanding academic journey. Their support provided the motivation and stability necessary to complete this work. We also extend our appreciation to everyone who offered sincere assistance, advice, or encouragement during the development stages, particularly when persistence and focus were required.

Finally, this project is dedicated to ourselves as a team. It reflects consistent effort, collaboration under pressure, and responsible work within real academic and engineering constraints. Beyond fulfilling graduation requirements, this work represents the knowledge, discipline, and professional growth developed throughout our engineering studies.

ABSTRACT

Skin cancer is one of the most prevalent forms of cancer worldwide, and early detection plays a critical role in improving patient survival rates and treatment outcomes. This project presents DermAI, a fully implemented web-based intelligent system designed to assist in the preliminary detection of skin cancer using artificial intelligence techniques. The system analyzes skin lesion images uploaded by users and classifies them into benign or malignant categories using a convolutional neural network (CNN) based on a fine-tuned ResNet50 architecture.

DermAI integrates a complete end-to-end pipeline, including dataset preparation, image validation, AI-based classification, and result presentation through an interactive web interface. To enhance transparency and clinical interpretability, the system incorporates Grad-CAM visual explanations that highlight the image regions influencing the model's predictions. In addition, DermAI provides location-based recommendations for nearby dermatology clinics to support timely medical consultation.

The system was developed with a strong emphasis on usability, security, and explainability, ensuring that users can easily upload images, interpret results, and access generated diagnostic reports. Experimental evaluation demonstrated stable performance across multiple metrics, validating the system's effectiveness as an AI-assisted screening tool. Overall, DermAI offers a practical, interpretable, and user-friendly solution that supports early skin cancer detection while complementing professional medical judgment.

الملخص

يُعد سرطان الجلد من أكثر أنواع السرطان انتشاراً عالمياً، ويُعتبر الكشف المبكر عاملاً حاسماً في تحسين فرص العلاج وزيادة معدلات النجاة. يقدم هذا المشروع DermAI، وهو نظام ذكي متكامل قائم على الويب تم تنفيذه فعلياً، يهدف إلى دعم الكشف الأولي عن سرطان الجلد باستخدام تقنيات الذكاء الاصطناعي. يقوم النظام بتحليل صور الآفات الجلدية التي يرفعها المستخدم وتصنيفها إلى حالات حميدة أو خبيثة بالاعتماد على شبكة عصبية التفافية (CNN) مبنية على نموذج ResNet50 بعد إعادة تدريبه وتخسيصه.

يتضمن نظام DermAI سلسلة معالجة كاملة تبدأ من إعداد البيانات والتحقق من جودة الصور، مروراً بعملية التحليل باستخدام الذكاء الاصطناعي، وانتهاءً بعرض النتائج من خلال واجهة ويب تفاعلية. ولتعزيز الشفافية وقابلية التفسير الطبي، تم دمج تقنية Grad-CAM لتوضيح المناطق المؤثرة في قرار النموذج. كما يوفر النظام ميزة اقتراح العيادات الجلدية القريبة اعتماداً على موقع المستخدم لدعم التوجّه السريع إلى الرعاية الطبية.

تم تصميم النظام مع التركيز على سهولة الاستخدام، الأمان، وقابلية التفسير، مما يتيح للمستخدمين رفع الصور، فهم النتائج، وتزيل التقارير التشخيصية بسهولة. أظهرت نتائج التقييم التجاري أداءً مستقرًا عبر عدة مقاييس، مما يؤكّد فعالية DermAI كأداة مساعدة في الفحص الأولي، مع التأكيد على أنه لا يعني عن التشخيص الطبي المتخصص.

TABLE OF CONTENTS

ACKNOLOGMENT.....	I
ABSTRACT.....	II
TABLE OF CONTENTS	IV
LIST OF FIGURS.....	X
LIST OF TABLES.....	IV
CHAPTER 1: INTRODUCTION.....	1
1.1 Introduction.....	2
1.2 Research Problem.....	2
1.3 Research Significance.....	3
1.4 Research Objectives.....	4
1.5 State of the Art.....	4
CHAPTER 2: LITERATURE REVIEW.....	6
2.1 Introduction.....	7
2.2 Overview of the Techniques.....	8
2.2.1 Artificial Intelligence (AI).....	8
2.2.2 Machine Learning (ML).....	9
2.2.3 Deep Learning Algorithms.....	10
2.2.4 Convolutional Neural Networks (CNN)	11
2.3 Chronological Analysis of Recent Advances in Skin Cancer Detection Using CNN.....	12
2.4 Comparative Analysis and Research Gaps	14
CHAPTER 3: SYSTEM ANALYSIS METHODOLOGY.....	16
3.1 Introduction.....	17
3.2 Dataset Acquisition and Preparation.....	18
3.2.1 Dataset Acquisition.....	18
3.2.2 Data Cleaning and Quality Filtering.....	18

3.2.3 Final Class Distribution.....	19
3.2.4 Data Preprocessing.....	20
3.2.5 Metadata Construction and Dataset Structuring.....	21
3.2.6 Observations and Challenges.....	22
3.3 Deep Learning Foundations for DermAI.....	22
3.3.1 Convolutional Neural Networks (CNNs)	22
3.3.2 Transfer Learning.....	24
3.4 Performance Metrics & Evaluation Framework.....	25
3.5 Comparative Study of Deep Learning Models.....	26
3.6 Selection of ResNet50 as the Final Model.....	29
3.7 Model Interpretability Using Grad-CAM.....	29
3.8 Decision Threshold Strategy.....	30
3.9 Cross-Validation Strategy for ResNet50.....	31
3.9.1 Rationale for Using K-Fold Cross-Validation.....	31
3.9.2 Cross-Validation Workflow.....	32
3.10 Used Libraries and Tools.....	37
3.11 Ethical Considerations.....	40
 CHAPTER 4: TECHNOLOGIES	
4.1 Functional Requirements.....	42
4.1 User Interface Design	43
4.1.1 Figma.....	43
4.1.2 Image Validation and Quality Assessment.....	43
4.2 Front-End Technologies.....	44
4.2.1 Front-End Framework.....	44
4.2.2 React Ecosystem (React, React DOM, React Router DOM).....	45
4.2.3 HTTP Communication Layer.....	46
4.2.4 Form Handling & Client-Side Validation.....	46
4.2.4.1 Formik.....	46
4.2.4.2 Yup.....	46
4.2.5 User Feedback & Notifications.....	47
4.2.6 UI Frameworks & Component Libraries.....	47

4.2.6.1 React Bootstrap.....	47
4.2.6.2 Bootstrap.....	47
4.2.7 Mapping & Geolocation Visualization.....	48
4.2.7.1 Leaflet.....	48
4.2.7.2 React Leaflet.....	48
4.2.8 Styling & Layout Technologies.....	48
4.3 Back-End Runtime & Framework.....	49
4.3.1 Node.js.....	49
4.3.2 Express.js Framework.....	50
4.3.3 RESTful API Architecture.....	50
4.3.4 Authentication & Security Modules.....	51
4.3.4.1 JSON Web Token (JWT).....	51
4.3.4.2 bcrypt.js.....	52
4.3.4.3 Crypto Module.....	52
4.3.5 File Handling & AI Communication Libraries.....	52
4.3.5.1 Multer.....	52
4.3.5.2 Axios.....	53
4.3.5.3 Form-Data.....	53
4.3.6 Report Generation Module.....	53
4.3.7 Email Service Integration.....	54
4.3.8 Database Layer.....	54
4.3.8.1 MongoDB Atlas.....	54
4.3.8.2 Mongoose ODM.....	55
4.3.9 Integrated Development Environment.....	55
4.3.9.1 Visual Studio Code.....	55
4.3.9.2 Thunder Client (API Testing Extension).....	55
4.3.10 Version Control.....	56
 CHAPTER 5: SOFTWARE DESIGN.....	57
5.1 Use Case Diagrams.....	58
5.2 Activity Diagrams.....	59

5.2.1 Signup Activity Diagram.....	59
5.2.2 Login Activity Diagram.....	60
5.2.3 Logout Activity Diagram.....	61
5.2.4 Forgot Password Activity Diagram.	62
5.2.5 Upload Image Activity Diagram.....	63
5.2.6 “View Nearby Clinics” button Activity Diagram.....	64
5.2.7 “View Grad-Cam” button Activity Diagram.....	65
5.2.8 “My Reports” button Activity Diagram.....	66
5.2.9 Imag-Valdition Activity Diagram.....	67
5.3 Sequence Diagrams.....	68
5.3.1 Signup Sequence Diagram.....	68
5.3.2 Successfully Activate Account Sequence Diagram.....	69
5.3.3 Login Sequence Diagram.....	70
5.3.4 Password Reset Workflow Sequence Diagram.....	71
5.3.6 Upload Image Sequence Diagram.....	72
5.3.7 “View Nearby Clinics” button Sequence Diagram.....	73
5.3.8 “View Grad-Cam” button Sequence Diagram.....	74
5.3.9 “My Reports” button Sequence Diagram.....	75
5.3.10 Imag-Valdition Sequence Diagram.....	76
5.4 Relational Schema Diagram.....	77
 CHAPTER 6: RESULTS AND PERFORMANCE ANALYSIS.....	78
6.1 Introduction.....	79
6.2 Dataset Statistics After Preparation.....	79
6.3 Evaluation Metrics Overview.....	79
6.4 Results of the Comparative Model Study.....	80
6.4.1 Comparative Performance Analysis.....	80
6.4.2 ResNet50 Performance Analysis.....	83
6.5 Cross-Validation Results Analysis.....	84
6.5.1 Cross-Validation Performance Metrics.....	84
6.5.2 ROC Curve Analysis Across Folds.....	85

6.5.3 Accuracy and Loss Curves Analysis.....	88
6.5.4 Cross-Validation Findings and Fold Selection.....	90
6.6 Final Training Analysis.....	91
6.6.1 Final Results Analysis.....	91
6.6.2 Confusion Matrix Interpretation.....	93
6.6.3 Threshold Optimization for the Final Model.....	94
6.6.4 Grad-CAM Visual Explanation Analysis.....	96
6.7 Discussion of Results.....	97
6.8 Chapter Summary.....	97
 CHAPTER 7: SYSTEM USER INTERFACES AND FUNCTIONAL WORKFLOW	98
7.1 Home and User Entry Interface.....	99
7.2 User Authentication and Account Management.....	99
7.3 Skin Image Upload Interface.....	102
7.4 AI Analysis Result Interface.....	102
7.5 Model Explainability and Visualization (Grad-CAM).....	103
7.6 Report Generation and Download.....	103
7.7 Nearby Clinics Locator Interface.....	104
 CHAPTER 8: FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS.....	105
8.1 Introduction.....	106
8.2 Functional Requirements	106
8.2.1 User Authentication and Account Management.....	106
8.2.2 Skin Lesion Image Upload.....	107
8.2.3 AI-Based Skin Lesion Analysis.....	107
8.2.4 Grad-CAM Heatmap Visualization.....	107
8.2.5 Analysis Result Presentation.....	108
8.2.6 Diagnostic Report Generation.....	108
8.2.7 Location-Based Clinic Recommendation.....	109
8.2.8 Error Handling and User Feedback.....	109
8.3 Non-Functional Requirements.....	109

8.3.1 Security and Privacy.....	109
8.3.2 Usability.....	110
8.3.3 Performance.....	110
8.3.4 Reliability and Availability.....	110
8.3.5 Maintainability.....	111
8.4 Conclusion.....	111
CHAPTER 9: FUTURE WORK AND SYSTEM ENHANCEMENTS.....	112
CHAPTER 10: References and sources.....	115

LIST OF FIGURS

Figure 2.1: Relationship between Artificial Intelligence, Machine Learning, and Deep Learning.....	8
Figure 2.2: Key Domains and Applications of Artificial Intelligence.....	9
Figure 2.3: Machine Learning Workflow	10
Figure 2.4: Deep Learning Algorithms.....	11
Figure 2.5: CNN Architecture for Skin Cancer Classification from Dermoscopy Images.....	12
Figure 3.1: Initial Dataset Scan Before Cleaning.....	18
Figure 3.2: Dataset Verification and Cleaning Summary.....	19
Figure 3.3: Class distribution of skin lesion dataset showing significant imbalance between benign and malignant cases.	19
Figure 3.4: Metadata Table Showing Sample Image Paths and Assigned Labels.	21
Figure 3.5: CNN Model Structure.	23
Figure 3.6: Transfer learning architecture with frozen base layers and trainable classification layers.	24
Figure 3.7: Confusion matrix of model predictions on benign (0) and malignant (1) skin lesions.	26
Figure 3.8: System architecture and model training pipeline.	28
Figure 3.9: Grad-CAM visualizations shown for ten image samples.	30
Figure 3.10: 3-Fold Cross-Validation Dataset.	33
Figure 3.11: Final Training Dataset Distribution (Train/Val/Test).	36
Figure 4.1: Figma Logo.	43
Figure 4.2: MongoDB Logo.	54
Figure 4.3: Mongoose Logo.	55
Figure 4.4: VS Code Logo.	55
Figure 5.1: User Use Case Diagram.	58
Figure 5.2: Signup Activity Diagram.	59
Figure 5.3: Login Activity Diagram.	60
Figure 5.4: Logout Activity Diagram.	61
Figure 5.5: Forgot Password Activity Diagram.	62

Figure 5.6: Upload Image Activity Diagram.	63
Figure 5.7: “View Nearby Clinics” button Activity Diagram.	64
Figure 5.8: “View Grad-Cam” button Activity Diagram.	65
Figure 5.9: “My Reports” button Activity Diagram.....	66
Figure 5.10: Imag-Valdition Activity Diagram.	67
Figure 5.11: Signup Sequence Diagram.	68
Figure 5.12: Successfully Activate Account Sequence Diagram.	69
Figure 5.13: Login Sequence Diagram.	70
Figure 5.14: Password Reset Workflow Sequence Diagram.	71
Figure 5.15: Upload Image Sequence Diagram.	72
Figure 5.16: “View Nearby Clinics” button Sequence Diagram.	73
Figure 5.17: “View Grad-Cam” button Sequence Diagram.	74
Figure 5.18: “My Reports” button Sequence Diagram.	75
Figure 5.19: Imag-Valdition Sequence Diagram.	76
Figure 5.20: Relational Schema Diagram.	77
Figure 6.1: Comparative performance of the evaluated models on the DermAI sample dataset.	82
Figure 6.2: Confusion matrix of the ResNet50 model evaluated on the DermAI validation dataset.	83
Figure 6.3: ROC curve and AUC value for Fold 1 obtained during cross-validation of the fine-tuned ResNet50 model.	85
Figure 6.4: ROC curve and AUC value for Fold 2 obtained during cross-validation of the fine-tuned ResNet50 model.	86
Figure 6.5: ROC curve and AUC value for Fold 3 obtained during cross-validation of the fine-tuned ResNet50 model.	87
Figure 6.6: Training and validation accuracy and loss curves for Fold 1 obtained during cross-validation of the fine-tuned ResNet50 model.	88
Figure 6.7: Training and validation accuracy and loss curves for Fold 2 obtained during cross-validation of the fine-tuned ResNet50 model.	89
Figure 6.8: Training and validation accuracy and loss curves for Fold 3 obtained during cross-validation of the fine-tuned ResNet50 model.	89

Figure 6.9: Comparison of final performance metrics (Accuracy, Precision, Recall, and F1-score) for the DermAI model on the validation and independent test sets.	92
Figure 6.10: ROC curves of the final trained DermAI model evaluated on the validation and independent test sets.	92
Figure 6.11: Confusion Matrices of the Final DermAI Model on Validation and Test Sets.	93
Figure 6.12: Effect of decision threshold variation on accuracy, precision, recall, and F1-score for the validation set. The dashed red line indicates the optimal threshold selected based on the maximum F1-score.	95
Figure 6.13: Example of Grad-CAM visualization showing the original dermoscopic image (right) and the corresponding heatmap (left), highlighting lesion regions that contributed most to the model's malignant classification decision.	96
Figure 7.1: Home page DermAI.	99
Figure 7.2: Sing-in and create account in DermAI.	99
Figure 7.3: Once an account is created, this email will be sent to confirm the account..	100
Figure 7.4: Forgot password in DermAI.	100
Figure 7.5: The message that the user receives when they log into their email to reset their password.	101
Figure 7.6: Reset Password Page in DermAI.	101
Figure 7.7: Upload image page in DermAI.	102
Figure 7.8: Analysis Result page in DermAI.	102
Figure 7.9: Grad-CAM Interface in DermAI.	103
Figure 7.10: Clicking the "Download Report" button will directly download the PDF file to your device.	103
Figure 7.11: When you click the "Find Clinics" button, a map will appear showing the locations of clinics near the user.	104

LIST OF TABLES

Table 2.1: Comparative Analysis and Research Gaps.....	14
--	----

Table 6.1 presents the comparative performance of the evaluated models in terms of accuracy, precision, recall, and F1-score, highlighting differences in their ability to detect malignant skin lesions.81

Table 6.2: Cross-validation performance metrics across three folds.....	85
---	----

Table 6.3: Final Performance Metrics of the DermAI Model on Validation and Test Sets.	91
--	----

Table 6.4: Performance comparison of the DermAI model on the validation set across different decision thresholds, illustrating the impact of threshold selection on accuracy, precision, recall, and F1-score.	94
---	----

CHAPTER 1

INTRODUCTION

1.1 Introduction

The rapid advancement of artificial intelligence particularly deep learning has enabled the development of innovative tools capable of analyzing medical images for early disease detection. Skin cancer is one of the most common types of cancer worldwide, and it poses a significant challenge in its early stages due to the difficulty of distinguishing between benign and malignant skin lesions. Many patients also delay seeking medical consultation when initial symptoms appear, leading to late diagnosis despite the fact that early detection significantly increases survival rates.

In this context, DermAI aims to provide an intelligent diagnostic tool based on Convolutional Neural Networks (CNNs) for accurate analysis of skin images. The system allows users patients or physicians to upload an image of a suspicious skin lesion, which is then analyzed by a pretrained model to classify it as benign or malignant. The system generates an initial diagnostic report that guides users toward appropriate medical follow-up when necessary.

DermAI is designed to support, not replace, clinical diagnosis by offering a fast and user-friendly mechanism for obtaining a preliminary assessment while ensuring full protection of image privacy and user data. The system relies on a large and diverse dataset of skin cancer images, enabling the model to learn from various patterns and improve its accuracy over time. It also incorporates visual explanation techniques such as Grad-CAM to highlight the regions that influence the model's decision, enhancing interpretability for both users and clinicians.

Through this project, DermAI aims to deliver a smart and reliable tool that supports early skin cancer detection and provides users with an instant report that assists in making timely medical decisions.

1.2 Research Problem

Despite the significant advancements achieved in computer-aided dermatological systems, several challenges were still present in the accurate classification and diagnosis of skin cancer. Earlier studies identified two major limitations in previous approaches. The first was the lack of sufficient and diverse medical image datasets, which negatively impacted the generalization capability of machine learning models across different skin types and lesion variations. The second limitation was related to the imaging process, as traditional systems relied heavily on dermoscopic images or invasive diagnostic methods such as biopsies and microscopy, making them less practical for accessible and early screening.

Another challenge that emerged was the limited generalizability of these algorithms in real-world clinical environments. Deep-learning models were highly sensitive to the characteristics of their training datasets, which were often restricted in diversity. Variations in demographic factors—such as age, skin tone, and genetic background—as well as environmental factors like ultraviolet (UV)

exposure significantly influenced the appearance and frequency of skin cancers, and thus affected how algorithms learned to represent skin lesions.

This issue was known as dataset shift, where models produced inaccurate results when tested on data that differed from the training distribution, especially when encountering rare or previously unseen lesion types. For instance, in the ISIC 2019 skin cancer challenge, even the top-performing models showed substantial performance degradation when evaluated on the more realistic BCN20000 dataset compared to the cleaner, more controlled HAM10000 dataset. Some models correctly classified only 11% of new benign cases, while misclassifying 47% of them as malignant.

These findings demonstrated that existing AI-based systems were not yet ready for safe clinical deployment, particularly in settings with diverse imaging conditions and skin characteristics. Therefore, there was a clear need to develop more robust and adaptable systems trained and evaluated on diverse, real-world datasets to ensure consistent and reliable performance.

1.3 Research Significance

The significance of this research stems from the technical challenges involved in classifying skin cancer images using AI-based algorithms. As computer vision increasingly enters the medical domain, there is a growing demand for models that can achieve high diagnostic accuracy while dealing with real-world data that are often imbalanced, inconsistent in quality, and affected by variations in lighting, imaging devices, and skin types.

Despite the remarkable progress of deep learning, reliably distinguishing between benign and malignant lesions remains difficult due to the visual similarity across many skin conditions and the strong dependence of model performance on dataset quality. This research gains its importance by addressing these challenges through the development of a robust skin-lesion classifier and by examining the effects of transfer learning and ensemble methods on performance enhancement.

The project is also significant from a practical standpoint, as it demonstrates how AI models can be integrated into an accessible user interface to support—but not replace—clinical decision-making. Overall, this research contributes both scientifically and practically by proposing methods that improve accuracy and stability in skin-cancer classification and by showcasing how AI can be effectively incorporated into early-detection tools.

1.4 Research Objectives

This research aimed to develop an accurate and intelligent system for the early detection of skin cancer using dermoscopic images. To achieve this goal, the study pursued the following objectives:

- 1.** To develop a deep convolutional neural network (CNN) model capable of accurately distinguishing between benign and malignant skin lesions.
- 2.** To compare the performance of the developed model with well-established pre-trained architectures, including ResNet50, DenseNet, and EfficientNet, in terms of accuracy, computational efficiency, and generalization on skin cancer datasets.
- 3.** To investigate the effect of ensemble learning techniques and evaluate whether combining multiple models could enhance diagnostic stability and accuracy, especially when dealing with visually similar skin lesions.
- 4.** To address challenges related to data quality and class imbalance by applying preprocessing techniques such as data augmentation and stratified splitting to ensure fair representation of various lesion types and skin tones.
- 5.** To develop an accessible and secure web-based interface (DermAI) that enables users to upload dermoscopic images and obtain instant AI-based predictions, supporting clinical decision-making while ensuring privacy and ease of use.

1.5 State of the Art

In recent years, the field of skin cancer detection using deep learning has witnessed significant advancements. Convolutional Neural Networks (CNNs) demonstrated dermatologist-level performance in classifying skin lesions, as shown in studies such as Esteva et al. (2017), where training on large dermoscopic datasets enabled models to achieve high diagnostic accuracy. This established the feasibility of artificial intelligence as a supportive tool in clinical decision-making.

Modern research has relied heavily on pre-trained architectures such as ResNet, DenseNet, and EfficientNet, which were fine-tuned on specialized datasets like HAM10000 and ISIC. These models benefited greatly from transfer learning, achieving high performance while reducing computational requirements. Several studies confirmed that transfer learning was essential for improving accuracy in skin cancer classification tasks.

Despite these advancements, CNN-based systems still faced limitations. Their performance often declined when evaluated on data that differed from the training distribution—particularly when variations appeared in skin tones, image quality, or rare lesion types. Moreover, many studies focused mainly on accuracy or AUC metrics, with limited attention to clinical usability or model interpretability for dermatologists.

Ensemble learning emerged as a promising approach to enhance robustness, reduce variance, and improve diagnostic stability. Research showed that combining multiple models—through methods such as bagging or stacking—consistently improved accuracy when dealing with heterogeneous or noisy medical images.

However, a noticeable gap remained between academic research and real-world implementation. Most models were not deployed within user-friendly platforms that allow direct image uploading and automated prediction, which limited their adoption in clinical environments.

To address these limitations, the present project built upon recent advancements by developing a CNN-based classifier enhanced with ensemble techniques and deploying it through a practical, web-based diagnostic platform, enabling instant and accessible early detection of skin cancer.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

The rapid growth of artificial intelligence (AI) and deep learning in recent years has revolutionized various sectors, including healthcare. In particular, AI has shown remarkable potential in enhancing medical diagnostics, providing significant improvements in accuracy and efficiency. A key area that has garnered attention in the medical community is the early detection of skin cancer, which remains a significant global health concern. The ability of AI-powered systems to analyze vast amounts of data and images with high precision has created new opportunities for the development of intelligent diagnostic tools (Esteva et al., 2017; Garbe & Leiter, 2019).

The current literature highlights the use of deep learning models, particularly convolutional neural networks (CNNs), in achieving dermatologist-level performance in classifying skin lesions. Esteva et al. (2017) demonstrated the capability of CNNs to classify skin cancer images with accuracy comparable to experienced dermatologists, marking a significant step forward in automated dermatological analysis. Similarly, Tschandl et al. (2019) emphasized the benefits of combining human expertise with AI to enhance diagnostic outcomes (Esteva et al., 2017; Tschandl et al., 2019).

However, while these advancements are promising, the literature also identifies several limitations and challenges that need to be addressed. One of the primary concerns is the lack of generalizability of AI models in real-world clinical settings. Many studies rely on curated datasets with controlled imaging conditions, which do not adequately reflect the diversity of patients' skin types, lesion variations, and environmental factors encountered in clinical practice. Adamson and Smith (2018) highlight this challenge, noting that AI systems must be trained and validated on diverse datasets to ensure reliable performance across populations (Adamson & Smith, 2018; Tan et al., 2023).

Another important aspect discussed in the literature is the role of ensemble learning in improving the robustness of AI models. Ensemble methods, such as bagging and boosting, have been shown to enhance the accuracy and stability of skin cancer classification, particularly when dealing with noisy or imbalanced datasets (Abayomi-Alli et al., 2021). Despite these proven benefits, ensemble learning remains underutilized in the field of automated skin cancer detection, representing a gap that future research could address (Abayomi-Alli et al., 2021).

Furthermore, the literature stresses the need for practical deployment of AI-powered diagnostic tools in clinical environments. As Codella et al. (2019) argue, integrating AI systems into user-friendly web-based platforms can facilitate their adoption and improve accessibility for both patients and healthcare providers. This aligns with the objectives of the DermAI project, which aims to develop an intelligent, user-friendly web platform for the early detection of skin cancer (Codella et al., 2019).

Overall, this literature review underscores the significant progress made in leveraging deep learning for skin cancer detection, while also highlighting critical gaps and challenges. Addressing these gaps through innovative approaches, such as

ensemble learning and real-world data integration, can pave the way for more reliable and accessible diagnostic tools, ultimately improving patient outcomes and contributing to more effective healthcare systems (Esteva et al., 2017; Abayomi-Alli et al., 2021; Codella et al., 2019).

2.2 Overview of the Techniques

With the rapid advancements in artificial intelligence (AI), intelligent algorithms have become increasingly capable of solving complex medical problems particularly in image analysis and disease diagnosis. This section provides an overview of the primary technologies employed in the DermAI system, beginning with general AI concepts and narrowing down to the core method used: Convolutional Neural Networks (CNNs). It also highlights alternative and complementary techniques that have been explored in related literature and diagnostic systems.

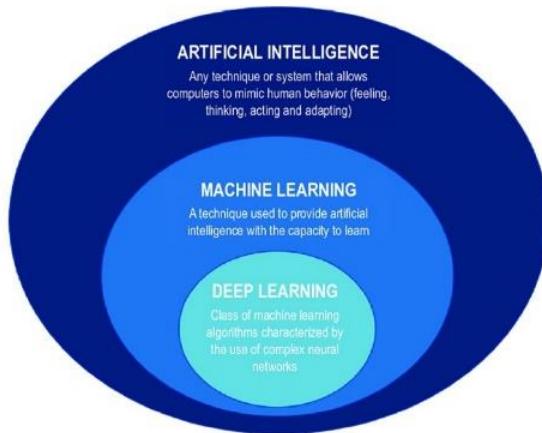


Figure 2.1: Relationship between Artificial Intelligence, Machine Learning, and Deep Learning.

2.2.1 Artificial Intelligence (AI)

Artificial Intelligence (AI) is a broad field in computer science that aims to create systems capable of performing tasks that typically require human intelligence. These tasks include problem-solving, learning, pattern recognition, natural language processing, and decision-making. In the medical field, AI has emerged as a transformative technology, providing innovative solutions to longstanding challenges in diagnostics, treatment planning, and patient monitoring.

AI systems are broadly categorized into rule-based systems and learning-based systems. While traditional rule-based systems rely on predefined logic and expert written rules, modern AI particularly machine learning (ML) and deep learning leverages large datasets to automatically learn complex patterns and make predictions. This paradigm shift has enabled the development of more adaptive and

accurate systems, especially in areas involving unstructured data such as images and medical records.

In the context of dermatology, AI plays a crucial role in enhancing diagnostic accuracy and efficiency. By analyzing large collections of skin images, AI models can learn to identify subtle visual cues that differentiate between various types of skin lesions. This capability is particularly important in the early detection of melanoma and other skin cancers, where timely diagnosis can significantly improve treatment outcomes.

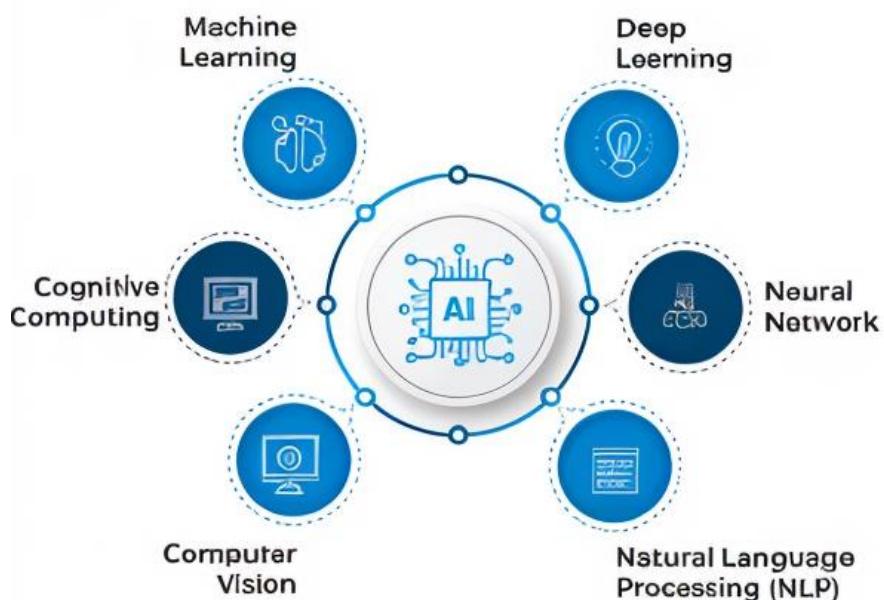


Figure 2.2: Key Domains and Applications of Artificial Intelligence.

2.2.2 Machine Learning (ML)

Machine Learning (ML) is a core discipline within artificial intelligence that empowers systems to learn from medical image data and make diagnostic predictions without manual programming for each condition. Instead of relying on predefined rules, ML algorithms detect patterns within image features and use these insights to classify or assess new, unseen cases, improving their accuracy over time with more exposure to data.

In the context of skin cancer detection, ML workflows typically involve three main stages: input data (such as labeled dermoscopic or clinical skin images), learning algorithms (such as supervised models trained to distinguish between benign and malignant lesions), and evaluation metrics (such as sensitivity, specificity, and AUC). Supervised learning is the most commonly used approach in medical

diagnosis, where the model is trained using annotated datasets to recognize disease patterns. In contrast, unsupervised methods may assist in discovering hidden structures within unlabeled data, while reinforcement learning is emerging for optimizing diagnostic decisions in interactive systems.

ML techniques have shown great promise in dermatology by enabling fast, scalable, and accurate lesion classification. They serve as the foundation for advanced deep learning models and contribute to building intelligent diagnostic systems that support early intervention and improve clinical outcomes.

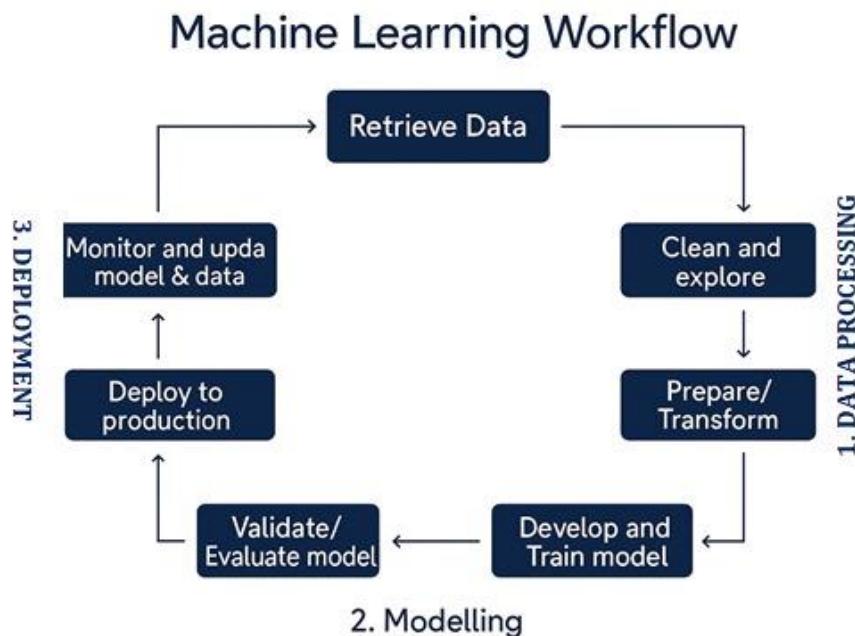


Figure 2.3: Machine Learning Workflow.

2.2.3 Deep Learning Algorithms

Deep learning is an advanced branch of machine learning based on artificial neural networks composed of multiple layers that mimic the structure and function of the human brain. These networks consist of interconnected nodes (neurons) and contain multiple hidden layers, enabling them to extract complex patterns from large datasets through multiple levels of abstraction and representation.

Deep learning is used to understand unstructured data such as images, text, and audio without the need for manual feature engineering. It is based on the idea that high-level concepts can be defined through lower-level concepts, which in turn form higher-level abstractions.

Key deep learning models include Convolutional Neural Networks (CNNs) for image processing, Recurrent Neural Networks (RNNs) for sequential data, and Transformers for natural language processing. This approach allows for the extraction of relevant features in both supervised and unsupervised manners.

Successful deep learning relies on large amounts of data, high computational power (such as GPUs), and sophisticated network architectures. These elements have enabled significant advances in many fields, including medical diagnosis (e.g., skin cancer detection), image recognition, language processing, and autonomous vehicles.

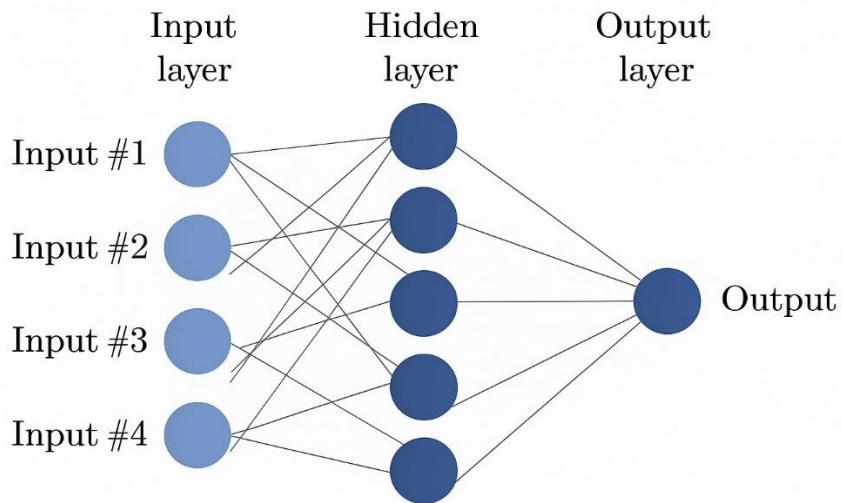


Figure 2.4: Deep Learning Algorithms.

2.2.4 Convolutional Neural Networks (CNN)

The Convolutional Neural Network algorithm is a multilayer perceptron tailored for the recognition of two-dimensional visual data. Input, convolution, sample, and output layers are always present. Moreover, in deep network architecture, both the convolution layer and the sample layer may have more than one iteration. Most commonly, CNN is employed to detect invariants of distortion in two-dimensional graphics, such as displacement and zoom. When we utilize CNN, the feature detection layer learns from the training data without the need for any explicit feature extraction. CNN algorithm is also used to classify skin cancer. Along with this, transfer learning models are built for classifying skin cancer images containing CNN layers. These models help to decide whether the skin cancer is malignant or benign.

For the CNN algorithm, convolution and sampling are the two most important steps.

i. Convolution process:

De-convolution of the input picture using a trainable filter F_x (the first stage is the input image, the input of the post-convolution is the image features of the current layer and is named; Feature Map), followed by the addition of a bias b_x , yields a convolution layer C_x .

ii. Sampling process:

After each neighborhood's n pixels have been pooled into a single pixel, the resulting pixel is then subjected to a scalar weighting of $Wx + 1$, given a bias of $b_x + 1$, and activated to generate a narrow n -times feature map $S_x + 1$. CNN is excellent at solving facial recognition issues, however training such a network requires a large number of annotated photos.

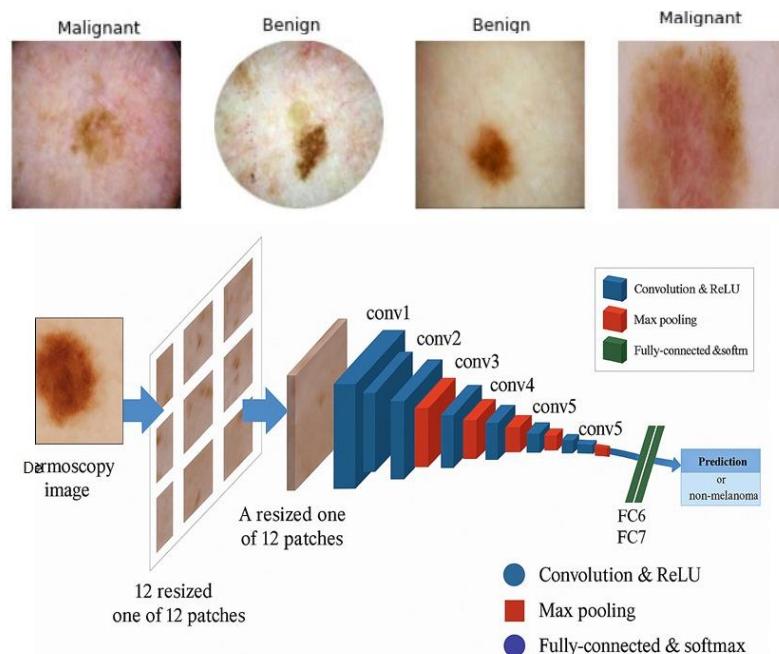


Figure 2.5: CNN Architecture for Skin Cancer Classification from Dermoscopy Images.

2.3 Chronological Analysis of Recent Advances in Skin Cancer Detection Using CNN

This section of the research presents a comprehensive review and analysis of key scientific studies that have explored the use of artificial intelligence particularly convolutional neural networks (CNNs) in the diagnosis of skin cancer. The purpose of this review is to trace the scientific progression in the field, highlight the methodologies applied in each study, and identify the major findings and challenges encountered.

In 2020, a team of researchers published a study titled “Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check,” which focused on evaluating the real-world performance of artificial intelligence in diagnosing skin cancer. The study conducted a systematic review of research using CNNs to classify skin images as benign or malignant. It identified a significant gap between the excellent

performance observed in laboratory settings and the reduced accuracy in clinical practice, especially when dealing with low-quality or non-representative images (e.g., dark skin tones). The study further discussed problems related to data bias, lack of generalizability, and difficulties in integrating AI systems into healthcare environments. It concluded that AI should be used in support of, rather than as a substitute for, clinical expertise.

In a subsequent study titled “Artificial Intelligence-Driven Enhanced Skin Cancer Diagnosis,” the researchers proposed a diagnostic model that combines CNNs with Discrete Wavelet Transform (DWT) to enhance prediction accuracy. Using benchmark datasets such as ISIC2018 and HAM10000, the study applied preprocessing steps including grayscale conversion, noise removal, and the extraction of the LL component via DWT to reduce complexity while retaining essential features. The model achieved an accuracy of 94%, outperforming traditional models such as ANN and MLP. This result confirmed the effectiveness of combining DWT and CNN for high-accuracy skin lesion classification.

Following that, a more advanced study was published under the title “Explainable AI-Based Skin Cancer Detection,” which utilized an Xception-based CNN enhanced with Particle Swarm Optimization (PSO). The researchers modified the architecture by adding new fully connected layers to capture rich feature representations. They addressed the black box issue by integrating explainable AI techniques such as Grad-CAM and LIME to visualize the regions influencing the model’s decisions. The proposed system achieved 98.5% accuracy on ISIC data and 86.1% on HAM10000, showing strong generalization capability. The importance of this study lies in its contribution to building transparent AI tools that foster trust in clinical applications.

In 2024, Behara, Bhero, and Agee published a comprehensive review titled “AI in Dermatology: A Comprehensive Review into Skin Cancer Detection.” The study followed PRISMA guidelines and included over 90 selected research articles.

It examined the benefits and limitations of AI models, with a particular focus on clinical integration, data privacy, bias against underrepresented skin types, and ethical-legal considerations. The authors recommended the development of more inclusive, reliable, and ethically grounded AI systems that could be deployed in real-world healthcare settings, especially in underserved regions.

Finally, a significant study was published in Nature Medicine titled “Human–Computer Collaboration for Skin Cancer Recognition,” which examined the collaborative dynamic between physicians and AI in skin cancer diagnosis. The study involved 12 physicians with varying experience levels and found that AI-assisted decision-making (through features such as multiclass probability outputs and class activation maps) significantly improved diagnostic accuracy, particularly for junior physicians. However, the study also warned that incorrect AI outputs

could negatively influence even experienced doctors, emphasizing the need for trustworthy and well-calibrated AI systems in medical environments.

2.4 Comparative Analysis and Research Gaps

A structured comparison was conducted to summarize recent studies on AI-based skin cancer detection, highlighting their models, techniques, datasets, and key limitations. This overview clarifies current trends and exposes gaps that the proposed DermAI system aims to address.

Study	Methodology	Accuracy	Limitations	Our Contribution
Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check	CNN + Systematic Review	Not specified	Bias in datasets, lack of real-world testing	Highlighting the need for real-world studies and clinician–AI collaboration
Artificial intelligence-driven enhanced skin cancer diagnosis	CNN + DWT	94%	Lack of real-world clinical validation	Integration of DWT to improve input preprocessing for CNN
Explainable AI-Based Skin Cancer Detection	Xception CNN + PSO + Grad-CAM	98.5%	Lower performance on new datasets	Use of explainability tools (Grad-CAM, LIME) to improve trust
AI in Dermatology: A Comprehensive Review	Systematic Literature Review	N/A	Not a direct model (theoretical)	Providing a structured framework and roadmap for future studies
Human–Computer Collaboration for Skin Cancer Recognition	Experimental doctor–AI collaboration	Varies (context-dependent)	AI errors reduce performance if misaligned	Emphasizing human–AI synergy and best practices for effective collaboration
DermAI(2026)	Transfer learning using a pretrained ResNet50 CNN fine-tuned for binary skin lesion classification (benign vs. malignant). Class weighting and data augmentation were applied to address class imbalance.	83.70%	Moderate recall for malignant lesions and evaluation limited to a single dataset.	A complete and stable skin cancer classification pipeline with balanced performance and integrated Grad-CAM visual explanations to enhance model transparency and clinical interpretability.

Table 2.1: Comparative Analysis and Research Gaps

This table reveals key variations in methodology, scope, and outcomes across recent AI research focused on skin cancer detection. While some models emphasize diagnostic accuracy through advanced deep learning architectures or optimization

techniques, others prioritize interpretability and real-world clinical integration. Despite these advancements, several common limitations remain evident across studies.

To examine the evolution and constraints of current automated skin cancer detection approaches, a comparative analysis of recent studies was conducted. The reviewed works employed diverse methodologies, ranging from conventional CNN architectures to enhanced models incorporating wavelet transforms and explainable AI techniques. Although many studies reported high classification accuracy under controlled experimental conditions, persistent challenges continue to limit real-world applicability, including limited generalizability, dataset bias, class imbalance, and reduced integration into clinical workflows.

Within this research context, the proposed DermAI system contributes a practical and structured approach by delivering a complete and stable binary classification pipeline, integrating Grad-CAM to enhance transparency, and emphasizing usability and accessibility within a web-based diagnostic framework. Nevertheless, in line with previous studies, certain limitations remain, such as moderate recall for malignant cases, residual class imbalance, evaluation on a single dataset, and reliance on a single deep learning architecture.

Based on this comparative review, the following research gaps were identified:

- Lack of model generalization to diverse skin lesion types, imaging conditions, and clinical environments.
- Inadequate representation of skin tone diversity in public datasets, leading to potential bias toward lighter skin types.
- Limited interpretability of deep learning models, reducing clinical trust and decision support.
- Insufficient exploration of real-time collaboration between AI systems and dermatologists.
- Overemphasis on accuracy metrics while overlooking usability, transparency, and explainability.

These findings highlight the need for balanced and inclusive AI-based diagnostic systems that achieve reliable performance while addressing real-world deployment constraints. Accordingly, DermAI aims to contribute toward this goal by incorporating explainable AI mechanisms, supporting clinical decision-making, and ensuring robust usability across diverse patient demographics and care environments.

CHAPTER 3

SYSTEM ANALYSIS METHODOLOGY

3.1 Introduction

Skin cancer is considered one of the most common types of cancer, arising from abnormal growth of skin cells due to DNA damage that leads to uncontrolled cell division. Early detection is a critical factor in improving treatment outcomes and reducing tumor progression, which makes the analysis of skin images an important tool in supporting medical decision-making. With the rapid advancement of artificial intelligence technologies, it has become possible to leverage deep learning models to provide accurate and rapid preliminary assessments based on image analysis.

This project introduces DermAI, a web-based artificial intelligence system capable of analyzing skin cancer images and classifying them into two categories: benign and malignant, based on convolutional neural networks (CNNs) and modern deep learning techniques. The system is designed to be user-friendly and to provide interpretable outputs that serve both medical specialists and patients.

This chapter presents the scientific methodology followed in developing the proposed deep learning model, starting with data collection and enhancement. The dataset was obtained from open-source medical repositories and underwent a comprehensive cleaning process, including quality inspection, removal of corrupted or duplicate samples, and handling of class imbalance. The preparation stage also included the application of several preprocessing techniques, such as image resizing, normalization, and data augmentation, in order to improve sample consistency and enhance the model's generalization capability.

Following this, a comparative study of several deep learning models used for skin cancer detection is presented, with the aim of identifying the most suitable and stable architecture according to a set of standardized evaluation criteria. This analysis leads to the selection of the final model and the description of its construction and training stages using transfer learning and fine-tuning strategies.

In conclusion, the chapter outlines the model evaluation methodology using commonly adopted performance metrics such as accuracy, sensitivity (recall), and F1-score, in addition to employing cross-validation to ensure the reliability and stability of the model across different data partitions.

3.2 Dataset Acquisition and Preparation

This section discusses the process of obtaining the raw data and preparing it for use in training the DermAI model. The procedure included several stages, starting with data collection, followed by cleaning, quality verification, and finally organizing the dataset into a clear and consistent structure suitable for training and testing purposes.

3.2.1 Dataset Acquisition

The dataset used in this project was obtained from reputable and publicly available sources in the field of skin disease diagnosis, most notably:

- International Skin Imaging Collaboration (ISIC Archive)
- HAM10000 Dataset (Kaggle)
- Additional Kaggle repositories containing skin cancer images

These repositories contain real-world images of skin abnormalities captured using dermatoscopes and various imaging devices, which resulted in considerable variation in image quality, lighting conditions, background, and resolution.

The initial number of images before any processing was approximately **20,000 images**, distributed across the two target classes:

- **Benign**
- **Malignant**

This stage was accompanied by several observations, most notably the clear variations in image quality and resolution, as well as the presence of duplicate samples across multiple sources.

```
Base directory: /content/drive/MyDrive/Dataset/Dataset
benign: 13543 files
malignant: 6189 files
Total files: ~19,732
```

Figure 3.1: Initial Dataset Scan Before Cleaning.

3.2.2 Data Cleaning and Quality Filtering

Following data collection, a comprehensive cleaning process was conducted to ensure the quality of each image before being introduced to the model. This process included the following steps:

- Detecting and removing corrupted images using image loading and integrity validation mechanisms
- Eliminating blurry or low-quality images
- Removing duplicate images using hashing techniques and visual inspection
- Verifying lighting and color quality to ensure suitability for analysis
- Standardizing the file format and extension for all accepted images

After completing all cleaning steps, the dataset size was reduced to **19,500 valid images** suitable for use in the training stages.

```
Scanning and cleaning dataset...
Dataset verification complete

Base directory: /content/drive/MyDrive/Dataset/Dataset
benign: 13291 files
malignant: 6210 files
Total files: ~19,501
```

Figure 3.2: Dataset Verification and Cleaning Summary.

3.2.3 Final Class Distribution

The class distribution analysis after the cleaning process revealed the following:

- **Benign: 13,290 images**
- **Malignant: 6,210 images**

This distribution indicates the presence of a class imbalance, as benign samples represent a significantly larger proportion than malignant ones, which is a common characteristic in skin cancer datasets.

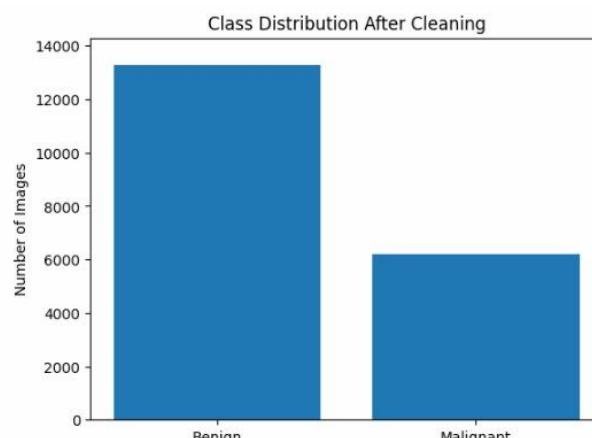


Figure 3.3: Class distribution of skin lesion dataset showing significant imbalance between benign and malignant cases.

3.2.4 Data Preprocessing

Data preprocessing represented an essential phase in preparing the dataset for model training, as it ensured consistency across samples and improved the model's ability to learn meaningful features. Several preprocessing steps were applied to all images before they were introduced into the training pipeline.

Image Resizing

Because the collected images came in varying resolutions and aspect ratios, all samples were resized to 224×224 pixels, aligning with the input requirements of pre-trained CNN architectures such as ResNet50, DenseNet, EfficientNet models. This step was performed using:

- OpenCV (cv2) for efficient resizing operations.
- PIL as an alternative for reading and adjusting image dimensions.

Resizing ensured a unified input shape and enabled stable batch processing during training.

Normalization

To improve training stability, pixel values were normalized by scaling them from the original 0–255 range to the [0, 1] range. This was done using simple NumPy operations and supported natively by TensorFlow/Keras. Normalization contributed to faster convergence and reduced training instability.

Data Augmentation

Given the natural imbalance in the dataset—where benign samples outnumber malignant ones—augmentation was used to increase variability in the training set and reduce overfitting. The following transformations were applied:

- Rotation
- Horizontal/vertical flipping
- Brightness adjustments
- Zoom and shift operations

These augmentations were implemented using Keras Image Data Generator, which applies transformations dynamically during training.

3.2.5 Metadata Construction and Dataset Structuring

The final stage of dataset preparation involved constructing a unified metadata structure to organize the cleaned and verified images prior to model training. Using a pandas DataFrame, all remaining samples were indexed and labeled to form a consistent and traceable representation of the dataset.

This process recorded the absolute file path of each image, assigned the appropriate class label (benign or malignant), and generated an encoded numerical label (0 for benign, 1 for malignant). The DataFrame was then shuffled to eliminate ordering bias and exported as a CSV file to ensure reproducibility and maintain a stable reference point throughout the development process.

This metadata table provided a central, structured overview of the entire dataset after cleaning, and served as the foundation for subsequent steps such as controlled dataset partitioning, batch generation, and performance evaluation.

Total samples: 19501			
Label distribution:			
label			
benign	13291		
malignant	6210		
Name: count, dtype: int64			
First 10 rows:			
	path	label	label_idx
0	/content/drive/MyDrive/Dataset/Dataset/maligna...	malignant	1
1	/content/drive/MyDrive/Dataset/Dataset/maligna...	malignant	1
2	/content/drive/MyDrive/Dataset/Dataset/maligna...	malignant	1
3	/content/drive/MyDrive/Dataset/Dataset/benign/...	benign	0
4	/content/drive/MyDrive/Dataset/Dataset/benign/...	benign	0
5	/content/drive/MyDrive/Dataset/Dataset/maligna...	malignant	1
6	/content/drive/MyDrive/Dataset/Dataset/benign/...	benign	0
7	/content/drive/MyDrive/Dataset/Dataset/benign/...	benign	0
8	/content/drive/MyDrive/Dataset/Dataset/benign/...	benign	0
9	/content/drive/MyDrive/Dataset/Dataset/benign/...	benign	0

Figure 3.4: Metadata Table Showing Sample Image Paths and Assigned Labels.

3.2.6 Observations and Challenges

During the preparation of the dataset, several practical challenges were observed that directly influenced the subsequent design decisions of the system. The images originated from multiple open-access medical archives, which introduced noticeable variations in lighting conditions, background textures, and camera devices. This heterogeneity made standardization more difficult and required additional verification to ensure that all samples adhered to a consistent visual structure.

Another significant challenge was the presence of duplicate or near-duplicate images across different repositories. Although a hashing-based mechanism successfully detected and removed many of them, the duplicates were not always identical, as some differed slightly in cropping or brightness. This required deeper inspection and reinforced the need for strict quality control.

The dataset also exhibited a clear class imbalance, with benign lesions occurring more frequently than malignant ones. While this reflects real medical distribution, it introduced the risk of bias during model training and motivated the use of augmentation and later rebalancing strategies.

Additionally, some images contained visual artifacts—such as ink markings, ruler shadows, or irrelevant objects—that complicated feature extraction. These issues highlighted the importance of careful preprocessing and influenced the design of the validation module introduced later in the system.

Overall, these challenges provided valuable insight into the variability and imperfections of real-world dermatological data, and they shaped the methodological choices applied throughout the model development pipeline.

3.3 Deep Learning Foundations for DermAI

The development of the DermAI model is built upon a set of well-established scientific foundations in the field of deep learning, particularly those related to medical image analysis. The model relies on convolutional neural network (CNN) architectures and transfer learning techniques, both of which have proven highly effective in classifying complex visual data such as images associated with skin cancer.

3.3.1 Convolutional Neural Networks (CNNs)

The DermAI system adopts convolutional neural networks (CNNs) as the primary architecture for analyzing images of skin cancer, given their suitability for complex

medical imaging tasks. CNNs process images by applying small sliding filters that traverse the input to extract essential visual features such as color variations, lesion boundaries, and surface textures. Through their sequential layered structure, CNNs construct a hierarchical representation of features—where the early layers capture simple patterns, while deeper layers learn more abstract and lesion-specific characteristics.

CNNs have demonstrated strong effectiveness in medical imaging applications, as they are robust to variations in illumination, capture resolution, and background conditions commonly found in real-world dermatological data. Additionally, CNNs eliminate the need for handcrafted feature engineering, which is particularly important in dermatology where images exhibit substantial variability in color, shape, and texture.

Moreover, CNNs are inherently compatible with visual interpretability techniques such as Grad-CAM, which highlight the regions most influential to the model’s decision. This enhances the system’s transparency and supports its integration into clinical environments. For these reasons, CNNs served as the foundation for all models evaluated within the DermAI system, including both the custom-built architecture and the transfer-learning-based models.

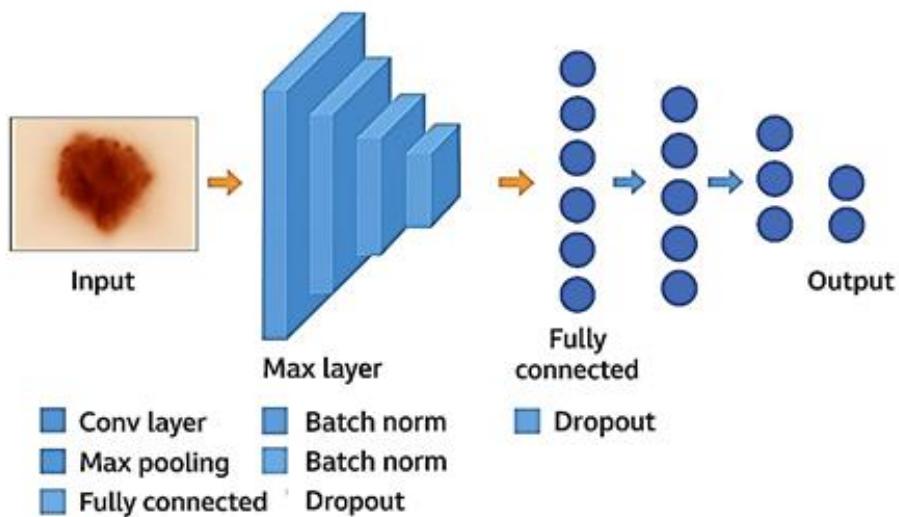


Figure 3.5: CNN Model Structure.

3.3.2 Transfer Learning

Transfer learning is a fundamental technique in the development of deep learning models, particularly in medical domains where data are often limited in quantity or exhibit high variability. This approach leverages deep models that have been pre trained on large-scale datasets, where the early layers have already learned general visual patterns such as edges, textures, and basic shapes. These learned representations are then repurposed by adapting the upper layers to fit the requirements of the new task.

This strategy offers several advantages over training a model from scratch, including improved training stability, faster convergence, and reduced risk of overfitting especially when working with medical images that display significant variability, such as those associated with skin cancer. Transfer learning also enables higher performance with fewer computational resources and decreases the need for large numbers of labeled images.

In the DermAI project, transfer learning was adopted to facilitate the development of a model capable of recognizing fine-grained dermatological patterns using pre trained deep layers. This approach provided strong visual representations from the earliest stages of training while allowing the final layers to be customized for the task of classifying cancerous lesions as benign or malignant. As a result, the model achieved greater stability and improved robustness in handling the high variability characteristic of dermatological imaging.

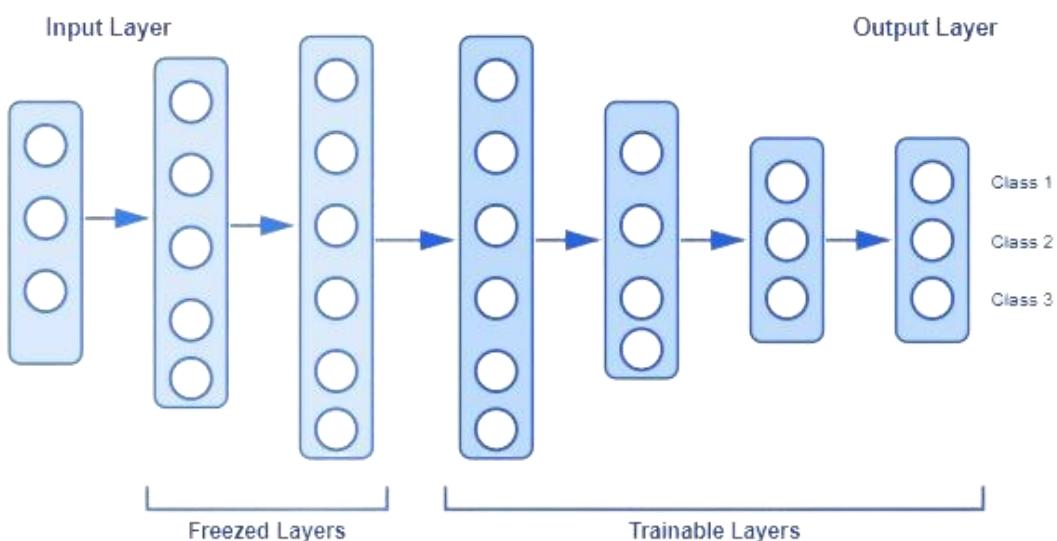


Figure 3.6: Transfer learning architecture with frozen base layers and trainable classification layers.

3.4 Performance Metrics & Evaluation Framework

Evaluating the performance of the DermAI model requires a set of standardized metrics commonly used in medical image-classification tasks. These metrics provide a quantitative understanding of the model's diagnostic reliability and help identify strengths and potential limitations. Because skin-lesion datasets are often imbalanced, it is essential to use evaluation measures that capture both overall accuracy and the model's ability to detect malignant cases effectively.

The following metrics were adopted in this project:

Accuracy

Represents the overall proportion of correctly classified images. Although widely used, accuracy alone may be misleading in imbalanced datasets.

Precision

Indicates how many of the images predicted as malignant are truly malignant. High precision means fewer false alarms, which is important to avoid unnecessary patient anxiety.

Recall (Sensitivity)

Measures the model's ability to correctly identify malignant cases. Because missing a malignant lesion has serious clinical consequences, recall is considered one of the most critical metrics in dermatological applications.

F1-Score

The harmonic mean of precision and recall. This score provides a balanced assessment, especially when the dataset contains disproportional class frequencies.

AUC-ROC

Evaluates the model's discriminatory ability across different threshold settings. A higher AUC indicates more stable and reliable predictions, particularly valuable in medical scenarios where decision thresholds may vary.

Confusion Matrix

also generated as part of the evaluation framework to visualize correct and incorrect predictions across both classes. This matrix helps identify patterns such as the rate of false negatives—an important factor in medical diagnosis—and guides further tuning of the model to enhance clinical reliability.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Figure 3.7: Confusion matrix of model predictions on benign (0) and malignant (1) skin lesions.

3.5 Comparative Study of Deep Learning Models

As part of the model-selection stage in DermAI, a comparative benchmarking study was conducted to evaluate a diverse set of machine-learning and deep learning architectures. The purpose of this comparison was to establish a consistent and fair methodological framework that allows identifying the most suitable model family for the final diagnostic pipeline. All models were trained and evaluated under unified conditions, using the same dataset split, preprocessing pipeline, and evaluation criteria described previously.

All models were trained on a standardized subset of 5,000 dermoscopic images (3,407 benign and 1,593 malignant) sampled from the main DermAI dataset. The dataset underwent identical preprocessing resizing, normalization, augmentation and was kept fixed through an 80/20 split to ensure that every model experienced the same data distributions. This controlled setup allowed architectural differences to emerge clearly without interference from data variability.

To cover the spectrum of conventional and modern techniques, the following models were included:

Custom CNN: A handcrafted baseline designed to examine how well a simple convolutional hierarchy can extract lesion patterns without transfer learning.

ANN: A fully connected model operating on flattened pixels, used to highlight the importance of spatial feature learning in medical images.

KNN (with PCA): A classical machine-learning method included to contextualize the gap between traditional feature-agnostic classifiers and modern deep models.

ResNet50: A residual-learning architecture offering stable gradient flow and strong baseline feature representation.

VGG16: A historically influential CNN composed of uniform convolution blocks, included for foundational comparison.

EfficientNetB0: A compact, computationally efficient architecture balancing performance and speed through compound scaling.

InceptionV3: A multi-branch model designed to capture diverse feature scales useful for lesions with heterogeneous textures.

Xception: A depth wise-separable variant emphasizing channel-wise feature decomposition and efficiency.

DenseNet121: A densely connected architecture promoting feature reuse and strong gradient propagation.

All deep-learning architectures followed a unified transfer-learning strategy: freezing the convolutional backbone, appending an identical classification head, and training under shared hyper parameters and batch settings. This ensured that model differences reflected genuine architectural capability rather than differences in training configuration.

Evaluation relied on the same set of standardized metrics accuracy, precision, recall, F1-score, and confusion-matrix inspection computed on the fixed validation split.

These metrics were not used to report absolute numerical results, but rather to provide a consistent basis for assessing which architectures demonstrated stable

training dynamics, strong sensitivity to malignant cases, and reliable generalization patterns.

This comparative framework provided a clear, methodologically grounded understanding of how each model family performed within the constraints of medical image diagnosis. The insights extracted from this study informed the subsequent decision to advance the most suitable architecture—ultimately leading to the selection of ResNet50 for cross-validation and final deployment within the DermAI pipeline.

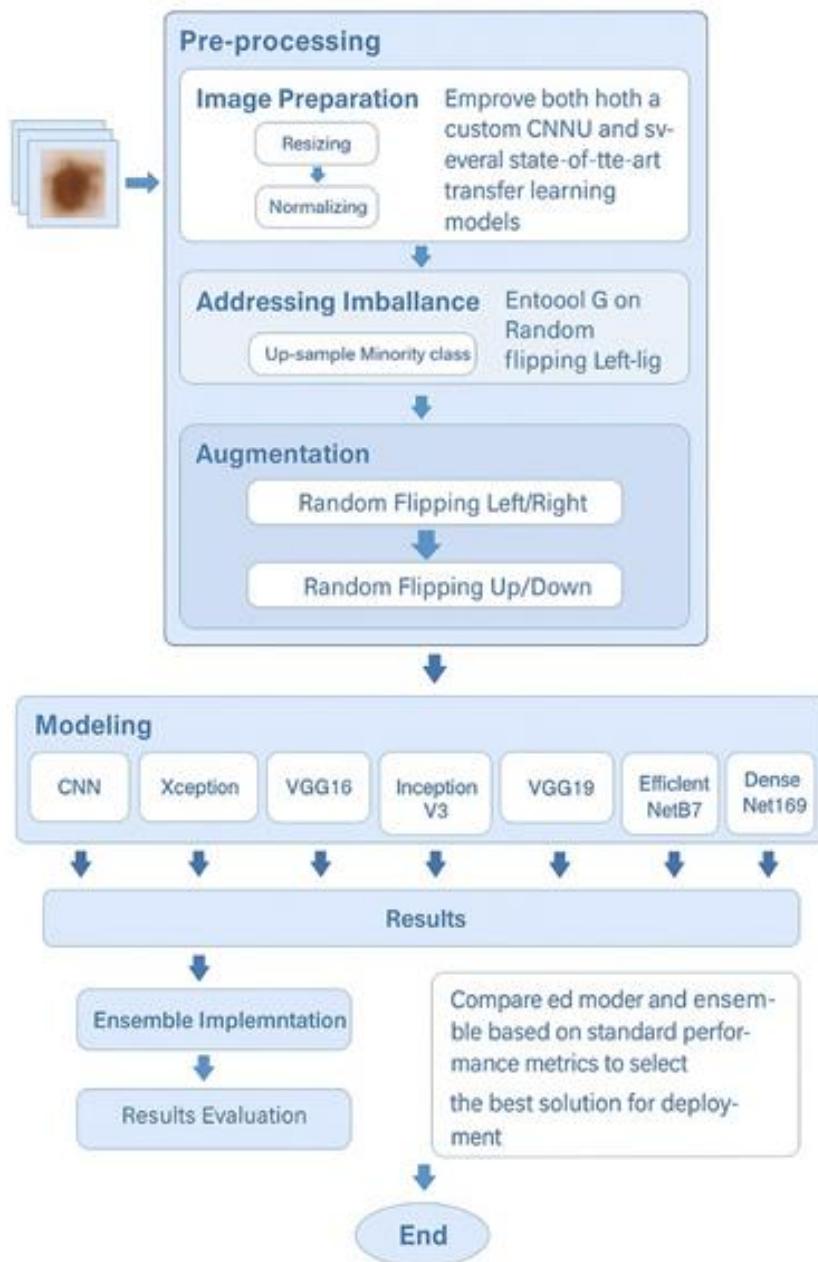


Figure 3.8: System architecture and model training pipeline.

3.6 Selection of ResNet50 as the Final Model

Based on the comparative evaluation conducted during the model-selection stage, ResNet50 was identified as the most suitable architecture to advance toward cross validation and final integration within the DermAI pipeline. The selection was guided by qualitative performance indicators such as training stability, reliable convergence, and consistent sensitivity to malignant cases rather than numerical reporting ensuring that the choice aligned with the diagnostic priorities of the system.

ResNet50 is a residual convolutional architecture that leverages skip connections to preserve gradient flow in deep networks. This design effectively addresses degradation issues seen in conventional deep CNNs and enables the extraction of complex, fine-grained features. Such capability is particularly valuable in dermoscopic imaging, where subtle variations in texture, color distribution, and lesion boundaries are essential for accurate classification.

In addition to its architectural stability, ResNet50 is one of the most widely adopted backbones in dermatology research and ISIC-related studies, reflecting its proven generalization ability across diverse skin-lesion datasets. Its moderate computational footprint, well-structured layer organization, and compatibility with common explainability tools (e.g., Grad-CAM) further support its suitability for clinical-oriented applications.

Overall, the comparative study reinforced ResNet50 as the model offering the most balanced combination of robustness, interpretability, and consistent learning behavior, making it the appropriate choice for subsequent cross-validation and deployment within DermAI.

3.7 Model Interpretability Using Grad-CAM

Ensuring interpretability is a fundamental requirement in medical AI systems, particularly when model predictions may influence clinical decision-making. For this purpose, Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated into the DermAI diagnostic workflow to provide visual explanations of how the model arrives at its predictions.

Grad-CAM operates as a post-hoc interpretability technique that does not modify the model's parameters. Instead, it computes the gradients of the predicted class score with respect to the final convolutional feature maps of the ResNet50 backbone. These gradients are then used to generate a class-specific heatmap that highlights the spatial regions most responsible for the prediction. A conceptual illustration of the Grad-CAM mechanism is shown in Figure 3.6, which provides a general explanation of how feature-map activations and gradients are combined to

produce class-specific heatmaps. This figure serves only as a visual aid to clarify the underlying process, rather than depicting outputs from the DermAI model itself. When overlaid on the original dermoscopic image, the Grad-CAM heatmap allows assessment of whether the model is focusing on clinically meaningful structures such as lesion borders, atypical pigmentation, asymmetry, or irregular texture patterns rather than irrelevant background artifacts. This visual confirmation is particularly important in dermatological imaging, where diagnostic conclusions often depend on subtle morphological cues.

Within the DermAI pipeline, Grad-CAM served two main roles:

- Clinical Transparency: It enables clinicians and users to visually verify the rationale behind the model’s decision, supporting trust and facilitating acceptance of AI-assisted diagnosis.
- Model Validation: During evaluation, Grad-CAM helped detect cases where the model mistakenly focused on noise (such as hair, shadows, or ruler marks), guiding improvements in preprocessing and model refinement.

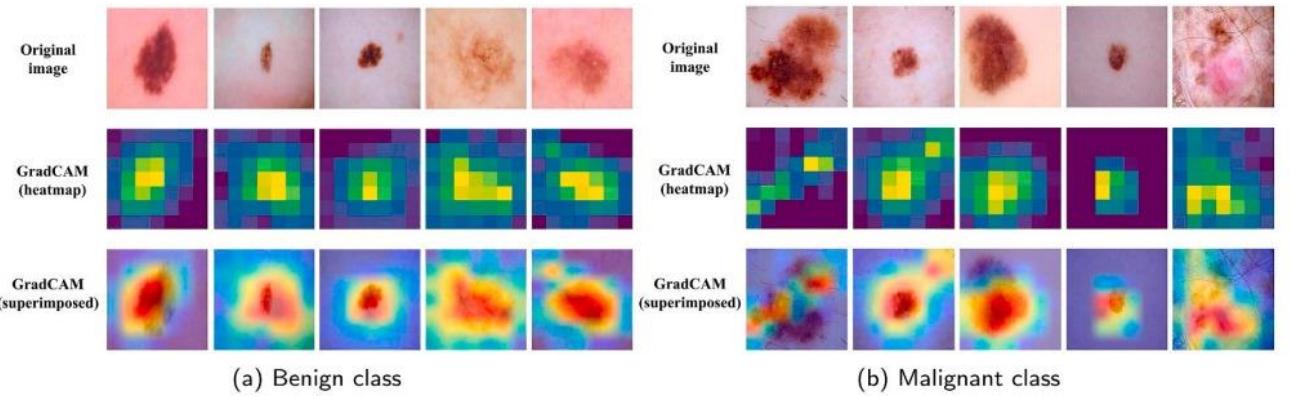


Figure 3.9: Grad-CAM visualizations shown for ten image samples.

3.8 Decision Threshold Strategy

In binary medical image classification, particularly when using sigmoid-based neural networks, deep learning models typically output a continuous probability score representing the likelihood of malignancy rather than a discrete class label. Converting this probabilistic output into a final diagnostic decision requires defining a decision threshold that separates benign from malignant predictions.

In the context of skin cancer detection, threshold selection plays a critical role due to the asymmetric cost of classification errors. False-negative predictions may delay the diagnosis of malignant lesions and pose significant clinical risk, whereas false-positive predictions primarily lead to additional medical examination.

Consequently, relying on a default threshold value is not always appropriate for medical decision-support systems.

In the DermAI system, the decision threshold was defined as a fixed component within the classification pipeline. Its role was examined during the experimental phase to assess its influence on the diagnostic behavior of the model, particularly in balancing sensitivity and specificity for skin lesion classification.

An analytical evaluation was conducted to study how variations in the decision threshold affected model predictions. This analysis focused on understanding performance trends and clinical trade-offs rather than aggressively optimizing numerical metrics, in order to preserve methodological transparency and avoid bias toward a specific evaluation measure.

Based on this assessment, a single threshold value was established prior to deployment and applied consistently during inference. This strategy ensures that probabilistic outputs are translated into clinically meaningful and reproducible diagnostic decisions. A quantitative evaluation of different threshold settings, along with the justification for the selected value, is presented in the Results chapter.

3.9 Cross-Validation Strategy for ResNet50

To ensure that the selected ResNet50 architecture demonstrated stable and generalizable behavior across the full DermAI dataset, a three-fold cross-validation procedure was implemented. This validation strategy was designed to evaluate the model's robustness under different data partitions and to reduce the risk of overfitting to any specific split of benign and malignant images.

3.9.1 Rationale for Using K-Fold Cross-Validation

Medical datasets especially dermoscopic image collections commonly exhibit heterogeneity in lesion appearance and noticeable class imbalance. Relying on a single train/validation split may cause the model's performance to depend heavily on the specific data partition.

Cross-validation mitigates this issue by:

- Providing multiple independent evaluations of the same model architecture.
- Assessing model stability under varying benign/malignant distributions.
- Ensuring that every sample is used for both training and validation at least once.

- Producing more reliable insights about the model's expected real-world behavior.

A value of $K = 3$ was chosen to ensure balanced evaluation while keeping training time feasible, given the computational cost of fine-tuning a deep CNN.

3.9.2 Cross-Validation Workflow

A stratified three-fold cross-validation procedure was implemented using Stratified Fold to ensure that each fold preserved the original benign–malignant class distribution. This process generated three balanced and non-overlapping splits. For each fold, the following workflow was executed:

1- Stratified Data Partitioning

- One fold was assigned as the validation set. The remaining two folds were combined to form the training set, ensuring that each image appeared once in validation and twice in training across the full cycle. The class-balanced organization of these splits is shown in Figure 3.7, which visualizes the distribution of benign and malignant samples within each fold.
- The remaining two folds were merged to form the training set.
- Every image participated exactly once in validation and twice in training across the full cycle.

2- Preprocessing Pipeline

- All images had already undergone standardization (resizing and cleaning) during the dataset-preparation stage.
- Within each fold, the training generator applied data augmentation (horizontal/vertical flips, rotations, zoom, brightness adjustments), while the validation generator applied only normalization.
- The preprocessing protocol remained identical across all folds to ensure methodological consistency.

3- Model Construction and Fine-Tuning

- Each fold instantiated a fresh ResNet50 backbone pre-trained on ImageNet, loaded without the final classification layers.
- Lower layers were kept frozen, while selective unfreezing of the upper convolutional block enabled controlled domain adaptation.
- An identical classification head was attached across all folds to maintain uniformity.

4- Training Configuration

- Class-weighted binary cross-entropy was used to address class imbalance.
- Training employed AdamW, early stopping, learning-rate scheduling, and model checkpointing.
- Each fold saved its own best-performing model parameters.

5- Evaluation Procedure (Qualitative)

- Each fold's model was evaluated on its corresponding validation subset using standardized metrics (accuracy, precision, recall, F1-score), along with confusion-matrix and ROC-curve analysis.
- These evaluations were used to assess stability and generalization, without reporting numerical scores at this stage.

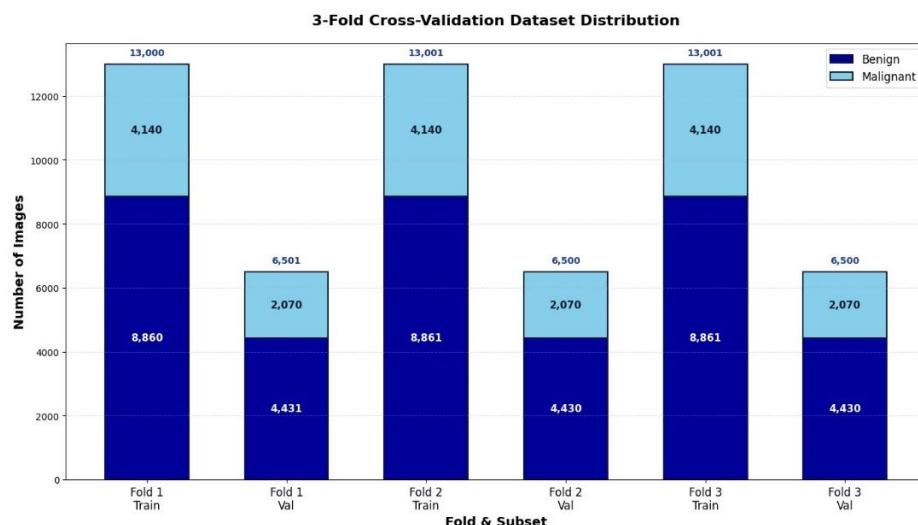


Figure 3.10: 3-Fold Cross-Validation Dataset.

Final Model Training Procedure

After completing the cross-validation phase and identifying the fold that demonstrated the most stable and optimal performance, the final training stage was executed to build the production-ready model that would be deployed within the DermAI system. In this stage, the entire dataset (19,505 images) was utilized to maximize visual variability and enhance the model's generalization capability.

Dataset Preparation and Splitting

The final training stage followed the same data-preparation pipeline used earlier, maintaining a stratified split to preserve the benign malignant class balance. An 80% training, 10% validation, and 10% test split was adopted. A wide range of augmentation operations including rotation, shifting, zooming, and brightness modulation was applied to increase robustness to lesion variability.

Additionally, class weights were automatically computed to mitigate the effects of class imbalance.

The stratified 80–10–10 split used for the final model is illustrated in Figure 3.8, which visualizes the balanced distribution of benign and malignant samples across the training, validation, and test subsets.

Model Initialization and Fine-Tuning Strategy

The final training stage was initialized using the best-performing ResNet50 model obtained from the cross-validation phase.

The fine-tuning strategy included:

- Freezing all ResNet50 base layers acting as the feature extractor.
- Unfreezing the last 40 trainable layers (excluding Batch Normalization) to enable domain-specific adaptation to dermoscopic images.
- Retaining the same classification head:
 - Global Average Pooling
 - Dense(512) + Dropout
 - Dense(1, sigmoid) for binary classification

This strategy preserved the representational strength of the pre-trained backbone while tailoring the upper layers to the medical-imaging domain.

Training Configuration

The final model was trained using the configuration proven effective during cross-validation, including:

- AdamW optimizer with a low learning rate (1e-5) and weight decay for improved stability.
- Binary Cross-Entropy loss.
- Tracking of core metrics: Accuracy, Precision, Recall, and AUC.
- Early Stopping based on validation loss to prevent overfitting.
- Automatic learning-rate reduction via ReduceLROnPlateau.
- ModelCheckpoint to store the best-performing version.
- Logging of all training details using CSVLogger.

Training Process

Training was performed using batches generated through ImageDataGenerator, with augmentation applied solely to the training set. The process included:

- Up to 50 training epochs.
- Continuous monitoring of validation performance.
- Saving:
 - The best model (final_model_best.keras)
 - The full final model (final_model_complete.keras)
 - Training logs
 - Loss and accuracy curves

This ensured that the final model was the most stable and generalizable version compared to any individual fold from the cross-validation stage.

Post-Training Evaluation and Outputs

After training, the best model was used to generate all evaluation outputs, including:

- Validation-set performance
- Independent evaluation on the unseen test set
- Visualization outputs:
 - Training curves (Loss/Accuracy)
 - Confusion matrices
 - ROC curves
 - Validation–test performance comparisons
 - Summary performance metrics

These outputs were later used for result interpretation and explainability analyses such as Grad-CAM.

Purpose of the Final Training Stage

This stage aimed to:

- Produce the official final version of the DermAI model.
- Improve overall stability using the full dataset.
- Deliver a model ready for real-world deployment.
- Support explainability through interpretability tools.
- Prepare the model for integration into the operational system.

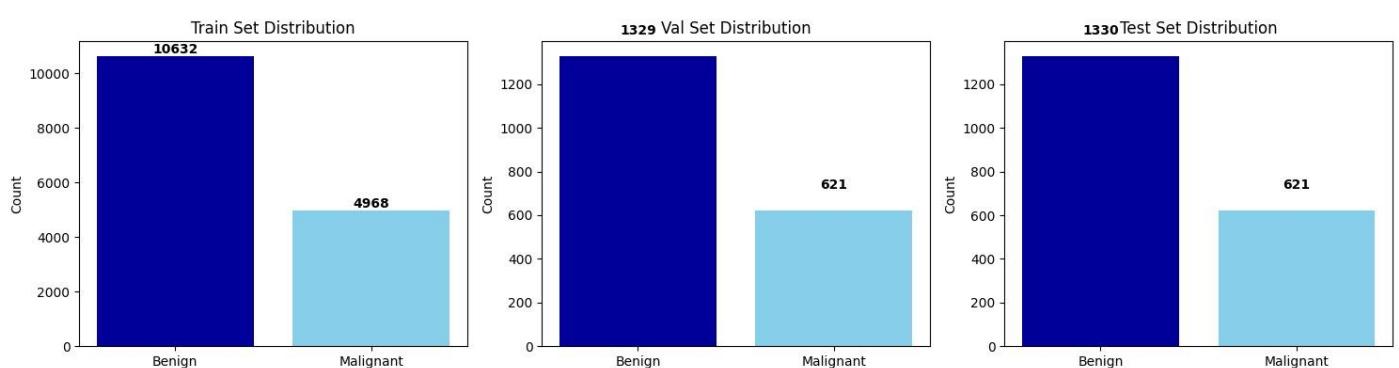


Figure 3.11: Final Training Dataset Distribution (Train/Val/Test).

3.10 Used Libraries and Tools

This section presents all libraries, frameworks, and tools utilized throughout the development of the DermAI system. The tools are organized according to the system's end-to-end workflow, covering dataset engineering, image validation, deep-learning development, model evaluation, and system deployment preparation. Each library is listed once and categorized based on its primary role in the project.

- **Deep Learning Libraries**

TensorFlow / Keras

Used to construct, train, and fine-tune the CNN classifier based on ResNet50, manage layers, define loss functions, metrics, and callbacks, and implement transfer-learning techniques.

TensorFlow Mixed Precision

Improves computational efficiency and reduces GPU memory usage during training.

Keras Functional API

Used to build modular models, define the classification head, and manage custom layers.

ImageDataGenerator

Provides image preprocessing and advanced data-augmentation operations for training and validation pipelines.

AdamW / Adam Optimizers

Used for stable model optimization, applying weight decay regularization to enhance convergence.

ModelCheckpoint, EarlyStopping, ReduceLROnPlateau, CSVLogger

Callback utilities used to prevent overfitting, dynamically adjust learning rate, store best model versions, and log training progress.

- **Data Processing and Preprocessing Libraries**

NumPy

Used for tensor manipulation, array operations, and numerical computations during preprocessing and metric calculations.

Pandas

Used to construct metadata tables, manage dataset partitions, and analyze data distribution.

Scikit-learn

Used for stratified data splitting, computing class weights, and generating evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

Pathlib

Provides object-oriented path management across dataset directories.

Shutil

Used for coping, organizing, and restructuring images when cleaning or splitting dataset folders.

Random

Ensures reproducibility in data shuffling and sampling.

JSON

Used to export structured validation outputs from the image-quality pipeline.

Math

Used for computations such as determining steps per epoch and scheduling parameters.

Collections.defaultdict

Used during duplicate-image detection to group images based on computed hashes.

- **Image Processing and Quality Validation Tools**

OpenCV (cv2)

Used extensively for image loading, resolution verification, blur detection using Laplacian variance, edge-density texture analysis, and color-space transformations (HSV, YCrCb) for skin-region extraction.

Pillow (PIL)

Used to validate file integrity, detect corrupted images, and convert formats during dataset cleaning.

Mediapipe

Used optionally for supplementary skin-region detection during quality screening.

Hashlib

Used for MD5 hashing to identify duplicated images in the dataset.

- **Visualization and Evaluation Tools**

- Matplotlib**

- Used to visualize training and validation curves, ROC graphs, and dataset statistics.

- Seaborn**

- Used for creating confusion matrices, heatmaps, and statistical visualizations.

- Plotly**

- Used to generate interactive graphs during exploratory analysis.

- **System Utilities and Support Libraries**

- OS**

- Used for directory inspection, file-system operations, and environment path handling.

- TQDM**

- Provides progress bars during long-running preprocessing, validation, and cleaning tasks.

- Warnings**

- Used to suppress unnecessary runtime warnings for clean output logs.

- Datetime**

- Used for logging timestamps, training durations, and tracking process metadata.

- GC (Garbage Collector)**

- Used to free memory between cross-validation folds and reduce GPU/CPU load.

- Sys / Traceback**

- Used to handle exceptions and produce detailed debug traces during image loading and processing.

- **Development Environment**

The DermAI model was developed and trained using a cloud-based computational environment provided by Google Colab. The training sessions utilized a dedicated GPU accelerator to enhance the efficiency of deep-learning operations.

Hardware Specifications (Google Colab GPU):

- GPU Model: NVIDIA Tesla T4
- GPU Memory: 16 GB GDDR6
- CUDA Version: 12.4
- Driver Version: 550.54.15

This hardware configuration provided sufficient computational capacity for training deep neural networks, optimizing transfer learning models, and running cross-validation efficiently.

Additionally, Jupyter Notebook was used locally for prototyping, debugging preprocessing functions, and testing individual modules prior to full pipeline execution.

3.11 Ethical Considerations

Given that DermAI operates in the medical domain, ethical responsibility is a core element of the system's development and deployment. Several ethical principles will be carefully considered to ensure fairness, transparency, and patient safety.

The main ethical considerations include:

Data Privacy and Anonymity: All images and metadata used for training and inference will be anonymized to remove personal identifiers. The system will comply with international data protection regulations such as the GDPR and HIPAA, and will adhere to the licensing policies of publicly available medical datasets to ensure ethical data usage.

Fairness and Bias Mitigation: The dataset will be reviewed to ensure representation across diverse skin tones, age groups, and genders. Class imbalance will be addressed through augmentation and weighting strategies to avoid biased predictions, particularly against underrepresented categories.

Clinical Responsibility: DermAI is intended to support—not replace—medical professionals. A clear disclaimer will notify users that the system is a diagnostic support tool, and final decisions must remain under clinical supervision. A “clinician override” feature will also be included to allow human experts to review and override AI-generated predictions when appropriate.

Transparency and Explainability: To enhance trust and interpretability, Grad-CAM visualizations will be provided alongside predictions to indicate which image regions contributed most to the model’s decision.

Continuous Monitoring and Feedback: After deployment, the system will be continuously monitored for bias, error trends, and false predictions. Feedback from clinicians and users will be collected to inform retraining and system updates.

By proactively addressing these ethical concerns, DermAI aims to deliver a responsible AI solution that aligns with medical standards and public trust.

CHAPTER 4

TECHNOLOGIES

4.1 User Interface Design

4.1.1 Figma

Figma is a collaborative web application for interface design, with additional offline features enabled by desktop applications for macOS and Windows. The feature set of Figma focuses on user interface and user experience design, with an emphasis on real - time collaboration, utilizing a variety of vector graphics editor and prototyping tools. The Figma mobile app for Android and iOS allows viewing and interacting with Figma prototypes in real - time on mobile and tablet devices.



Figure 4.1: Figma Logo.

4.1.2 Image Validation and Quality Assessment

Prior to forwarding user-submitted images to the deep learning model, the DermAI backend applies an image validation and quality assessment pipeline to ensure that only reliable and medically relevant inputs are processed. This step is critical in medical image analysis, as low-quality or irrelevant images may lead to unreliable predictions and compromise diagnostic safety.

The validation process is implemented as a server-side mechanism and operates immediately after image upload. It consists of a sequence of automated checks designed to assess image integrity, resolution adequacy, visual clarity, and consistency with expected skin characteristics.

First, the system verifies the existence and readability of the uploaded image file to detect corrupted or unsupported inputs. A minimum resolution requirement is then enforced to ensure compatibility with the convolutional neural network input dimensions and to preserve sufficient visual detail for feature extraction.

To evaluate image sharpness, a blur detection technique based on Laplacian variance is applied. Images exhibiting excessive blur are rejected, as blur significantly degrades the visibility of lesion boundaries and texture patterns essential for accurate skin lesion analysis.

In addition, the backend performs skin-content verification to ensure that the submitted image predominantly contains skin tissue. This verification combines color-space analysis using HSV and YCrCb representations with texture-based analysis, including edge density and local intensity variation. These complementary checks help distinguish true skin images from non-relevant patterns, background objects, or unrelated visual content.

Images that fail any of the validation criteria are excluded from further processing and returned to the user with an appropriate rejection message. For images that

satisfy all quality and relevance checks, a standardized resizing operation is applied to match the model's required input resolution. The validated image is then forwarded to the AI inference pipeline for classification.

By incorporating this image validation mechanism within the backend, the DermAI system enhances robustness, reduces the risk of unreliable predictions, and ensures that model inference is performed only on suitable and clinically meaningful image inputs.

4.2 Front-End Technologies

Front-end development, also known as client-side development, focuses on building user-friendly and interactive interfaces that allow users to interact effectively with web applications. This section describes the technologies and tools used to design and implement the front-end of the DermAI system, including React.js, CSS3, and Bootstrap. These technologies collectively enable the creation of a responsive, dynamic, and visually appealing user interface.

4.2.1 Front-End Framework

React.js

React.js is a popular, open-source front-end JavaScript library used for building user interfaces, particularly single-page applications (SPAs). It follows a component-based architecture, where the user interface is divided into independent, reusable components. This approach improves code maintainability, scalability, and development efficiency.

Why React.js?

React.js was selected over other front-end technologies such as Angular.js, Vue.js, and Vanilla (pure) JavaScript for several important reasons:

- **High performance using the Virtual DOM:**

React utilizes a Virtual DOM mechanism that efficiently updates only the parts of the interface that change, resulting in faster rendering and smoother user experience.

- **Single Page Application (SPA) support:**

React enables the development of SPAs, allowing seamless navigation between different views without requiring full page reloads, which enhances usability and performance.

- **Dynamic and interactive user interfaces:**

React easily integrates JavaScript logic for handling user input, conditional rendering, state management, and dynamic navigation, making the application highly interactive.

- **Modern JavaScript (ES6+) support:**

React is built on modern JavaScript standards, allowing the use of features such as asynchronous operations, modular code structure, and efficient state handling. This is particularly important in DermAI, where real-time interaction with backend APIs and AI results is required.

Overall, React.js provides a powerful and flexible framework that supports the development of modern, responsive, and scalable front-end applications, making it well-suited for the DermAI system.

4.2.2 React Ecosystem (React, React DOM, React Router DOM)

The React ecosystem represents the core technologies used to build the front-end of the DermAI system. It provides a structured, efficient, and scalable approach to developing modern web applications.

React is an open-source JavaScript library used for building interactive user interfaces. It follows a component-based architecture, allowing the user interface to be divided into reusable and independent components. This approach improves code organization, scalability, and maintainability.

React DOM acts as a bridge between React components and the browser's Document Object Model (DOM). It is responsible for rendering React components into the actual web page and efficiently updating the UI using the Virtual DOM, which enhances performance and minimizes unnecessary re-rendering.

React Router DOM is a routing library designed for Single Page Applications (SPA). It enables seamless navigation between different views without full page reloads and supports dynamic routes, protected routes, and authentication-related flows such as email verification and password reset.

Together, these technologies form the foundation of the front-end architecture, enabling dynamic navigation, high performance, and a smooth, responsive user experience.

4.2.3 HTTP Communication Layer

Axios

Axios is a JavaScript library used to handle HTTP requests between the front-end and back-end.

Why Axios was used:

- Communicates with the backend REST API.
- Sends and receives data efficiently (GET, POST, etc.).
- Provides better error handling compared to native fetch.
- Simplifies API request management.

4.2.4 Form Handling & Client-Side Validation

4.2.4.1 Formik

Formik is a library designed to simplify form handling in React applications. It manages form state, validation, and submission efficiently.

Why Formik was used:

- Simplifies form state management.
- Reduces repetitive boilerplate code.
- Integrates easily with validation libraries.
- Improves reliability of form handling.

4.2.4.2 Yup

Yup is a JavaScript schema validation library, commonly used with Formik to validate form inputs.

Why Yup was used:

- Validates user inputs such as email and password.
- Enforces strong password rules.
- Displays clear validation error messages.

- Enhances data integrity before submission.

4.2.5 User Feedback & Notifications

React Toastify

React Toastify is a notification library used to display success and error messages in a user-friendly way.

Why React Toastify was used:

- Provides real-time feedback to users.
- Displays non-intrusive notifications.
- Improves user experience.
- Supports customizable and auto-closing alerts.

4.2.6 UI Frameworks & Component Libraries

4.2.6.1 React Bootstrap

React Bootstrap is a library that reimplements Bootstrap components as React components without relying heavily on JavaScript-based Bootstrap plugins.

Why React Bootstrap was used:

- Integrates Bootstrap components seamlessly with React.
- Provides cleaner and more readable JSX-based UI code.
- Eliminates direct DOM manipulation.
- Enhances component reusability and maintainability.

4.2.6.2 Bootstrap

Bootstrap is a front-end framework that provides pre-designed UI components and a responsive grid system. It simplifies the process of building consistent and mobile-friendly user interfaces.

Why Bootstrap?

Bootstrap was chosen for the following reasons:

- **Rapid development:** Pre-built components such as forms, buttons, modals, and grids speed up the development process.
- **Responsive design:** Bootstrap ensures consistent behavior across different screen sizes without extensive custom styling.

- **Consistency and compatibility:** The framework enforces a uniform design language throughout the application.

In the DermAI system, Bootstrap is combined with custom CSS3 styles to achieve both rapid development and a customized visual identity that aligns with the project's requirements.

4.2.7 Mapping & Geolocation Visualization

4.2.7.1 Leaflet

Leaflet is an open-source JavaScript library for creating interactive maps.

Why Leaflet was used:

- Displays geographic maps within the application.
- Enables visualization of clinic locations.
- Lightweight and efficient mapping solution.
- Supports user interaction with map elements.

4.2.7.2 React Leaflet

React Leaflet is a React wrapper for the Leaflet library, allowing maps to be used as React components.

Why React Leaflet was used:

- Integrates Leaflet maps into React architecture.
- Allows map control using React state and props.
- Improves code organization and maintainability.
- Simplifies map rendering in React applications.

4.2.8 Styling & Layout Technologies

CSS3

CSS3 (Cascading Style Sheets Level 3) is a styling language used to control the appearance and layout of web content. It plays a critical role in defining the visual design of the DermAI interface, including colors, typography, spacing, and component alignment.

Why CSS3?

CSS3 introduces advanced styling features that significantly enhance user experience:

- **Visual effects:** Features such as shadows, gradients, rounded corners, and custom fonts improve the overall aesthetic quality of the interface.
- **Animations and transitions:** CSS3 allows smooth animations and transitions without relying on JavaScript, improving performance and visual consistency.
- **Flexible layout management:** Layout systems such as Flexbox and Grid simplify the creation of complex, well-structured layouts.
- **Responsive design:** CSS3 enables the application to adapt seamlessly to different screen sizes, ensuring usability across desktops, tablets, and mobile devices.

In DermAI, CSS3 is used extensively to ensure a clean, professional, and user-friendly interface.

4.3 Back-End Runtime & Framework

4.3.1 Node.js

Node.js is a server-side JavaScript runtime environment built on Google's V8 JavaScript engine. It enables the execution of JavaScript code outside the browser while adopting an asynchronous, event-driven, and non-blocking I/O architecture. This model allows Node.js to efficiently handle a large number of concurrent connections, making it suitable for scalable, high-performance backend systems.

Why Node.js?

Node.js was selected for the DermAI project because it provides:

- **High performance** through non-blocking I/O operations, ensuring fast processing of user requests.
- **Scalability**, making it suitable for an AI-based platform expected to manage many simultaneous users.

- **Unified development language**, as both backend and frontend use JavaScript, improving consistency and productivity.
- Strong compatibility with RESTful architectures used in modern AI systems.

In summary, Node.js offers a robust and efficient backend foundation that aligns perfectly with the technical needs of DermAI.

4.3.2 Express.js Framework

Express.js is a lightweight and flexible web application framework built on top of Node.js. It simplifies backend development by providing a structured approach to routing, middleware integration, and HTTP request-response handling. Its minimalist design allows developers to construct clean, modular, and maintainable server-side architectures.

Why Express.js?

- Simplifies the development of RESTful APIs through intuitive routing mechanisms.
- Seamlessly supports middleware-based architectures for authentication, file handling, and request validation.
- Provides the flexibility required to integrate AI services, secure authentication flows, and image-processing endpoints.
- Benefits from a mature ecosystem of third-party packages that accelerate backend development.

Consequently, Express.js serves as the core framework responsible for managing DermAI's backend API operations.

4.3.3 RESTful API Architecture

Representational State Transfer (REST) is an architectural style for designing stateless web services that communicate over the HTTP protocol. RESTful APIs expose system functionality through clearly defined endpoints (e.g., /api/auth/login, /api/analyze) using standard HTTP methods.

Justification for REST Architecture in DermAI

- Provides a consistent and lightweight communication model suitable for AI-based medical platforms.
- Ensures reliable and efficient data exchange between the React frontend and the Node.js backend.
- Utilizes JSON as the primary data format, which is well-suited for transferring prediction results, metadata, and medical image analysis outputs.
- Supports scalability and extensibility, allowing the backend services to evolve without impacting existing clients.

The adoption of RESTful architecture ensures that DermAI remains modular, maintainable, and easily integrable with future clinical or third-party systems.

4.3.4 Authentication & Security Modules

4.3.4.1 JSON Web Token (JWT)

JSON Web Token (JWT) is a compact and secure mechanism for transmitting authenticated user information between system components. It enables stateless authentication by embedding signed claims within tokens, eliminating the need for server-side session storage.

Security Benefits

- Enables stateless authentication, reducing server overhead and improving scalability.
- Ensures token integrity through cryptographic signing.
- Restricts access to sensitive endpoints, such as medical analysis history and report downloads.
- Supports multi-client environments, including web and mobile platforms.

4.3.4.2 bcrypt.js

bcrypt.js is a cryptographic hashing library used to securely store user passwords. It employs salting and adaptive hashing techniques to protect against brute-force and rainbow-table attacks.

Contribution to System Security

- Protects user credentials even in the event of database compromise.
- Applies automatic salting and configurable hashing complexity.
- Follows widely accepted security standards in healthcare-related applications.
- Integrates seamlessly with the DermAI authentication workflow.

4.3.4.3 Crypto Module

The Crypto module is a native Node.js library that provides encryption, hashing, and secure random value generation. It supports essential security operations without external dependencies.

Its Role in DermAI

- Generates secure verification codes for email confirmations.
- Ensures cryptographically strong tokens for password reset workflows.
- Enhances overall data protection by supporting built-in encryption utilities.
- Contributes to maintaining the confidentiality of personal health information.

4.3.5 File Handling & AI Communication Libraries

4.3.5.1 Multer

Multer is a middleware for handling multipart/form-data in Express applications, primarily used for file uploads.

Usage in DermAI

- Handles secure uploading of skin lesion images.
- Validates file types and size limits.
- Prepares images for transmission to the AI analysis service.
- Enhances overall user experience with timely automated communication.

4.3.5.2 Axios

Axios is a promise-based HTTP client used for performing asynchronous API requests.

Usage in DermAI

- Sends uploaded images to the AI prediction server.
- Communicates with external APIs when required.
- Retrieves image resources for inclusion in PDF medical reports.

4.3.5.3 Form-Data

Form-Data is a library used to construct multipart/form-data payloads programmatically.

Usage in DermAI

- Packages uploaded images into multipart requests.
- Enables seamless transmission of image data to the AI model endpoint.

4.3.6 Report Generation Module

PDF-Lib

PDF-Lib is a JavaScript library used for creating and modifying PDF documents.

Usage in DermAI

- Generates structured medical analysis reports.
- Embeds prediction results, confidence levels, and image references.
- Provides users with downloadable, well-formatted diagnostic reports.

4.3.7 Email Service Integration

Sib API v3 SDK (Brevo)

The Sib API v3 SDK is an official Node.js client library for the Brevo (formerly Sendinblue) email service.

Usage in DermAI

- Sends account verification emails during user registration.
- Delivers secure password reset links.
- Enables reliable and automated system communication with users.

4.3.8 Database Layer

4.3.8.1 MongoDB Atlas

MongoDB Atlas is a fully managed cloud database service offering scalable and secure MongoDB clusters. It supports flexible document-based storage suitable for dynamic and evolving datasets.

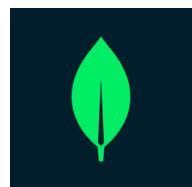


Figure 4.2: MongoDB Logo.

Advantages for DermAI

- Provides high availability and automatic scaling for growing user traffic.
- Offers built-in encryption and access control suitable for healthcare systems.
- Accommodates diverse data formats, including user metadata and prediction logs.

- Eliminates the overhead of server maintenance through its managed infrastructure.

4.3.8.2 Mongoose ODM

Mongoose is an Object Data Modeling library that provides schema-based structures for MongoDB collections. It enforces data validation and simplifies interaction with the database.



Figure 4.3: Mongoose Logo.

Contribution to DermAI

- Ensures consistent user, token, and prediction data structures.
- Supports middleware hooks for validation and sanitization.
- Simplifies CRUD operations, improving backend reliability.
- Enhances maintainability of DermAI's database layer.

4.3.9 Integrated Development Environment

4.3.9.1 Visual Studio Code

Visual Studio Code (VS Code) is a lightweight and extensible editor that offers integrated debugging, terminal access, and Git support, making it well-suited for backend development. Its smooth edit-build-debug workflow and intelligent code assistance streamline the implementation of DermAI's RESTful APIs, authentication logic, and data handling operations. Additionally, the availability of Node.js and API-testing extensions enhances productivity and ensures a consistent and efficient development environment throughout the project.



Figure 4.4: VS Code Logo.

4.3.9.2 Thunder Client (API Testing Extension)

Thunder Client is a VS Code extension designed for testing RESTful APIs directly within the development environment. It provides an intuitive interface for sending HTTP requests, managing authentication tokens, and validating backend responses.

Role in DermAI Backend Development

- Enables efficient testing of authentication endpoints such as login, registration, and token verification.
- Facilitates debugging of image upload requests and prediction-related APIs.
- Eliminates the need for external tools, allowing the developer to test endpoints without leaving VS Code.
- Supports repeated testing sequences during backend development workflows.

4.3.10 Version Control

Git & GitHub

Git is a widely adopted distributed version control system originally developed by Linus Torvalds in 2005. Its distributed architecture allows every developer to maintain a complete local copy of the project history, ensuring performance, flexibility, and secure version tracking across different environments and IDEs.

GitHub, built on top of Git, is a cloud-based collaboration platform that provides hosted repositories and tools for managing code contributions, reviewing changes, and coordinating team workflows. Within the DermAI project, Git and GitHub enabled structured collaboration between backend, frontend, and AI development teams, maintained an organized and accessible commit history, and ensured reliable versioning and safe storage of all system components throughout the development cycle.

CHAPTER 5

SOFTWARE DESIGN

5.1 Use Case Diagram

User Use Case Diagram

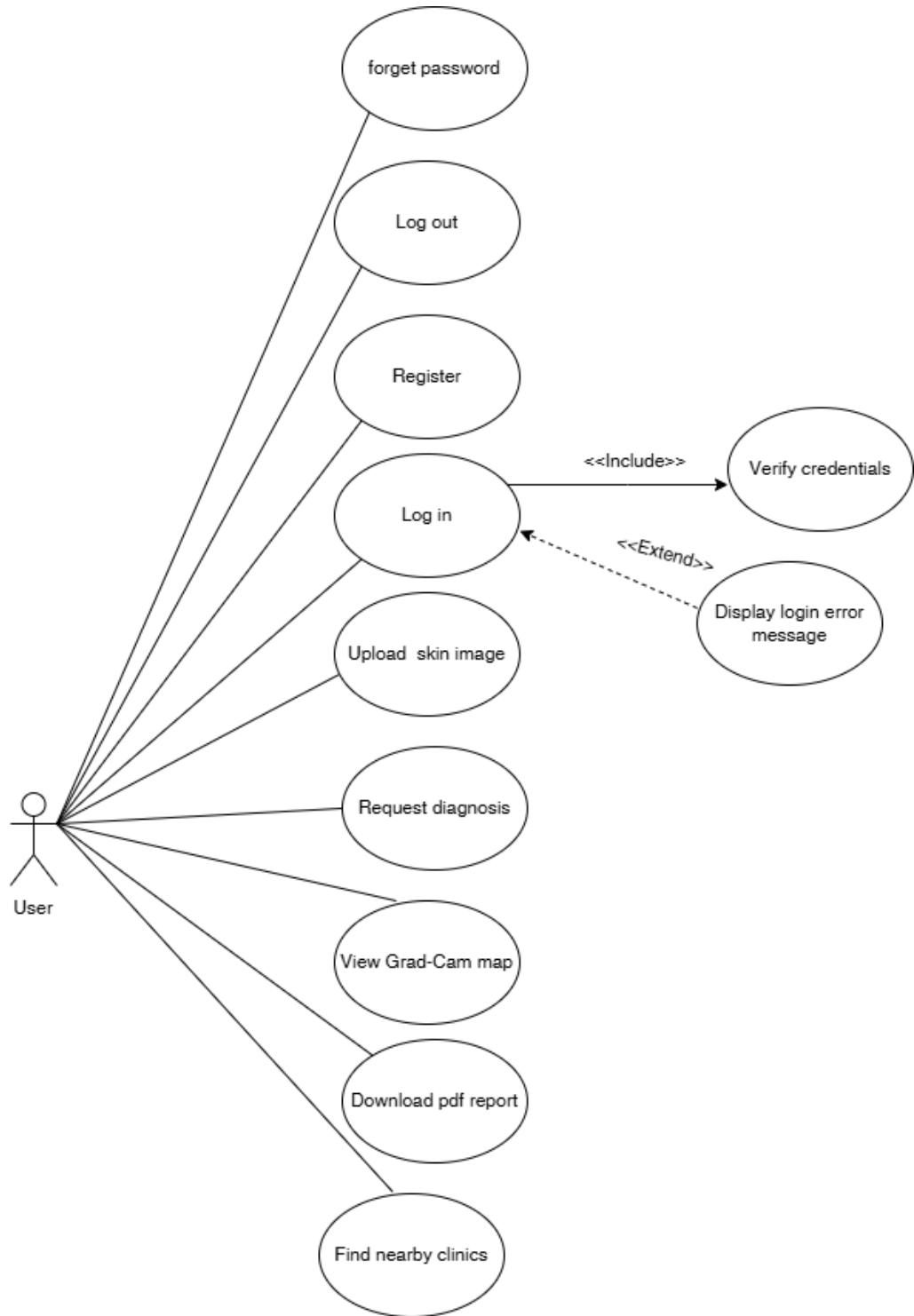


Figure 5.1: User Use Case Diagram.

5.2 Activity Diagrams

5.2.1 Signup Activity Diagram

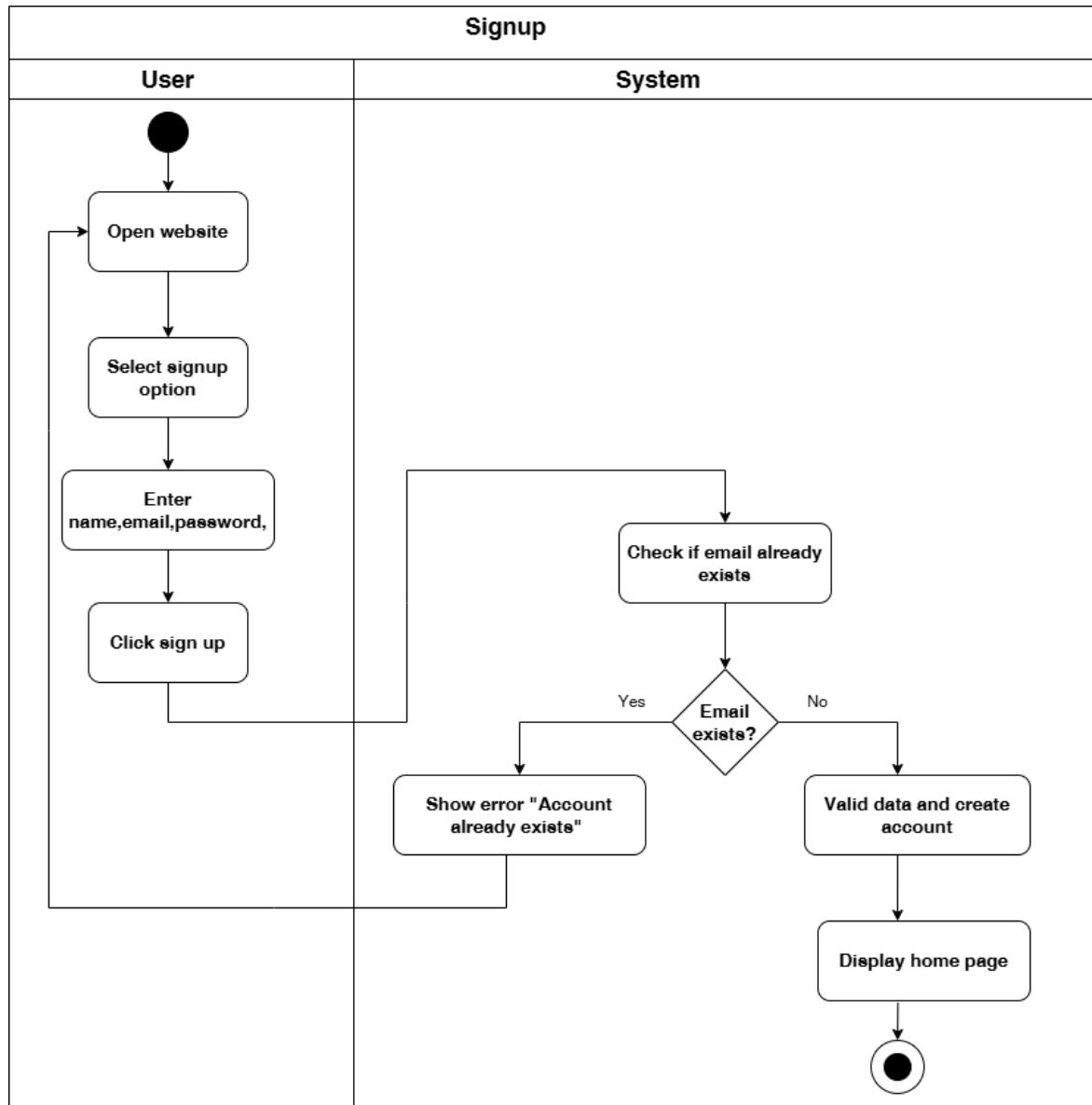


Figure 5.2: Signup Activity Diagram.

5.2.2 Login Activity Diagram

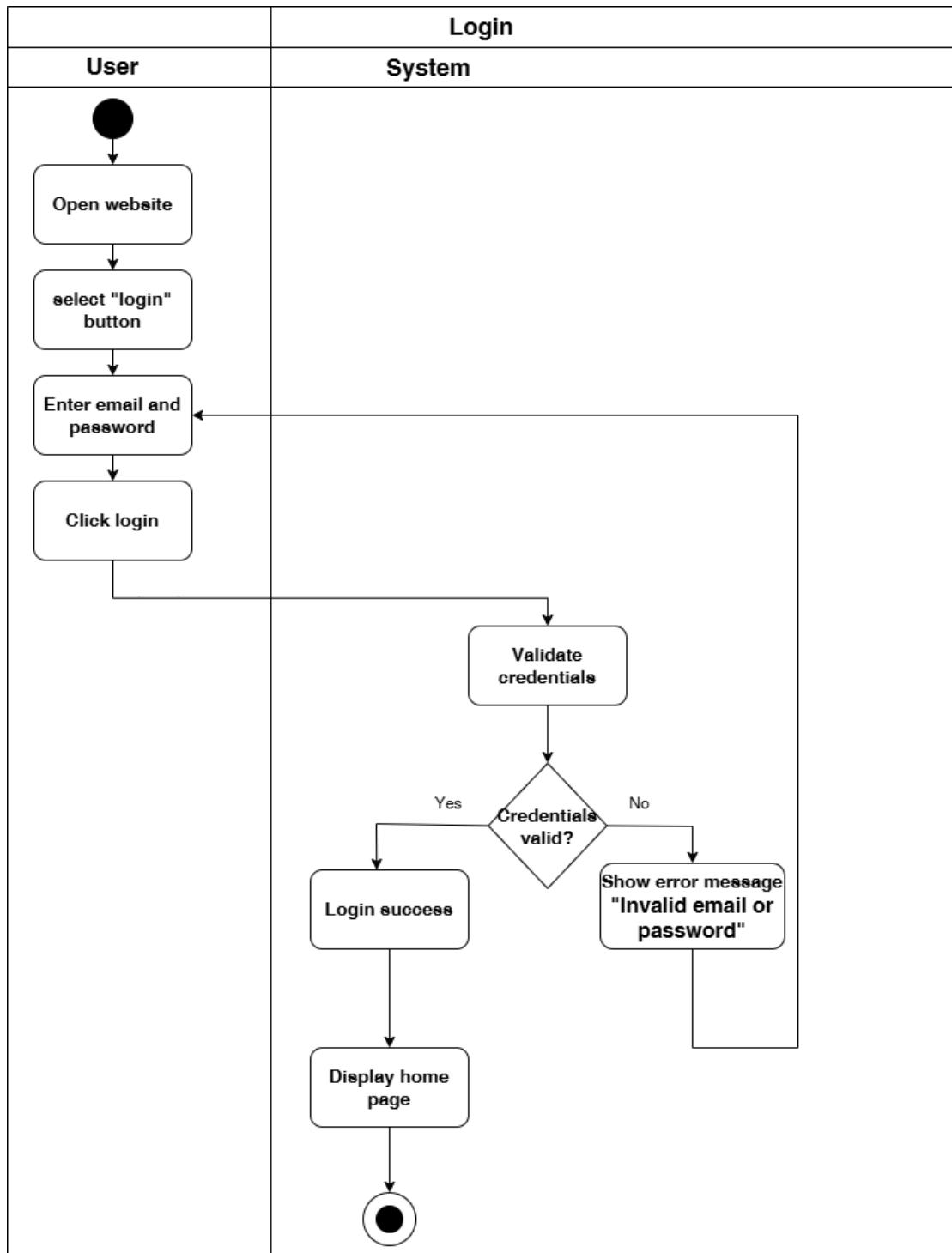


Figure 5.3: Login Activity Diagram.

5.2.3 Logout Activity Diagram

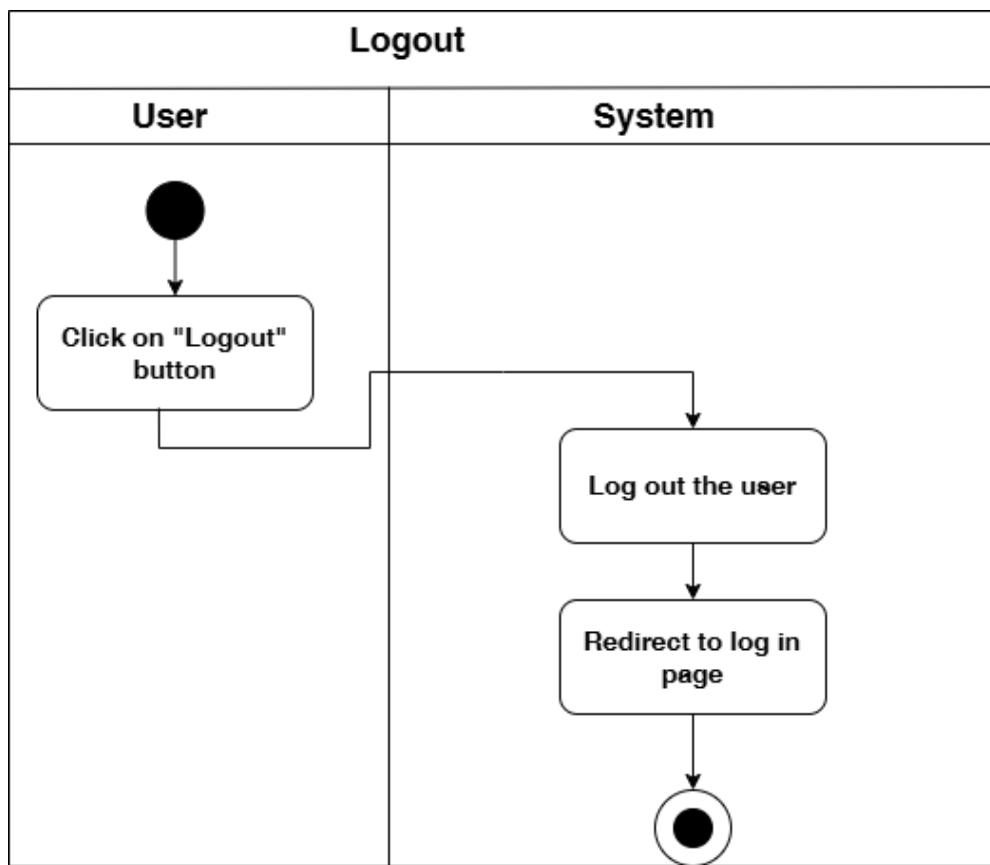


Figure 5.4: Logout Activity Diagram.

5.2.4 Forgot Password Activity Diagram.

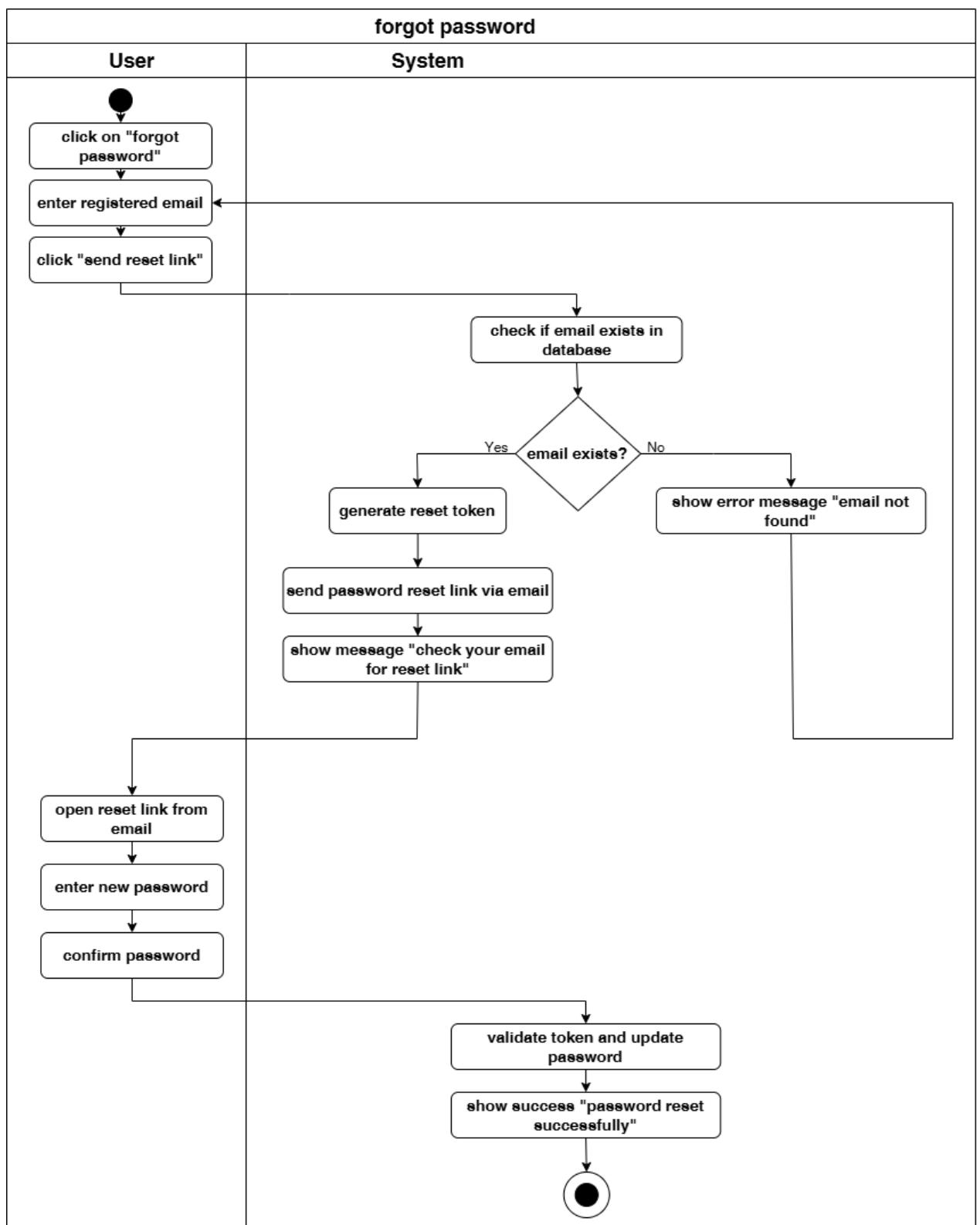


Figure 5.5: Forgot Password Activity Diagram.

5.2.5 Upload Image Activity Diagram

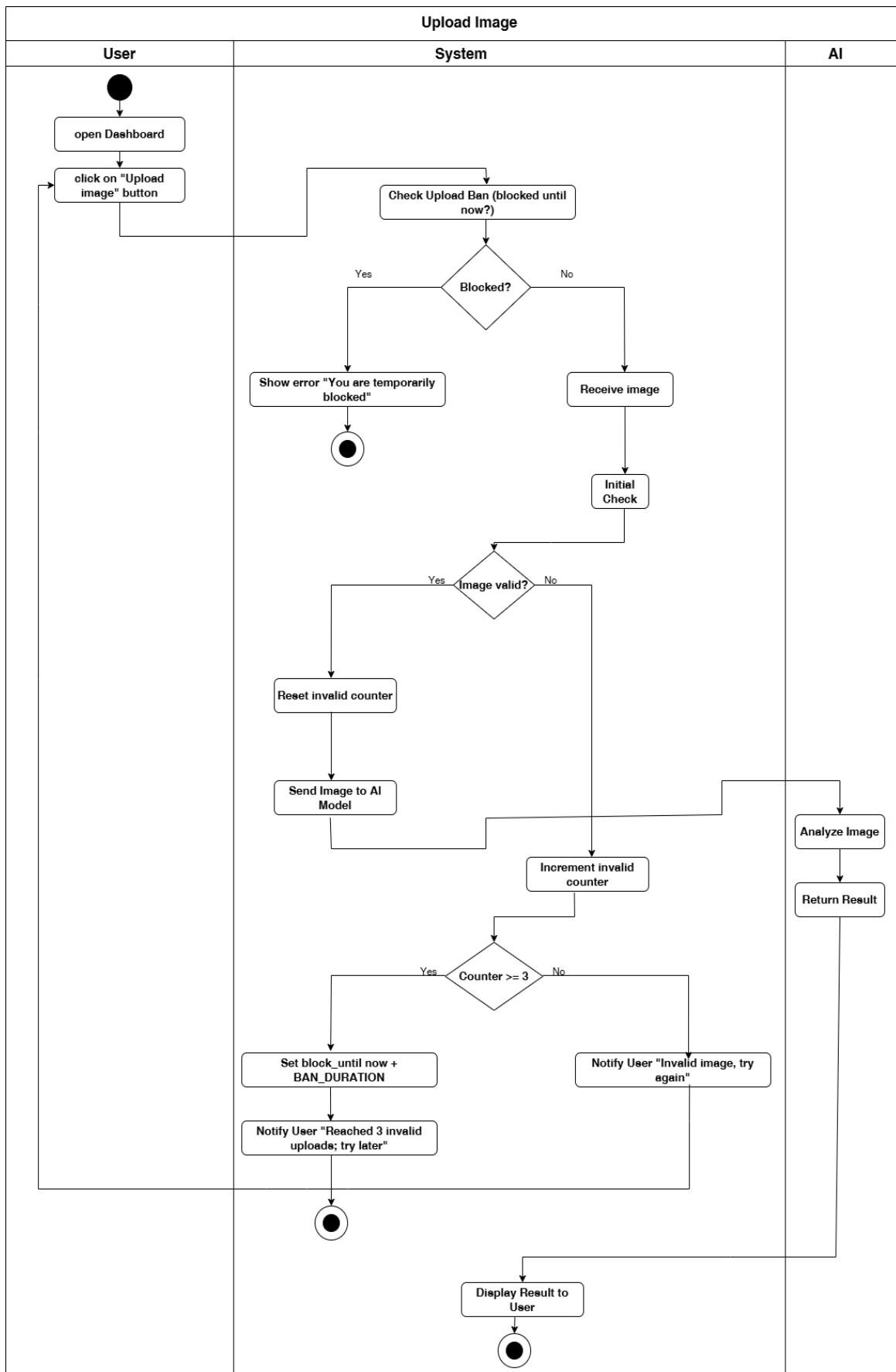


Figure 5.6: Upload Image Activity Diagram.

5.2.6 “View Nearby Clinics” button Activity Diagram

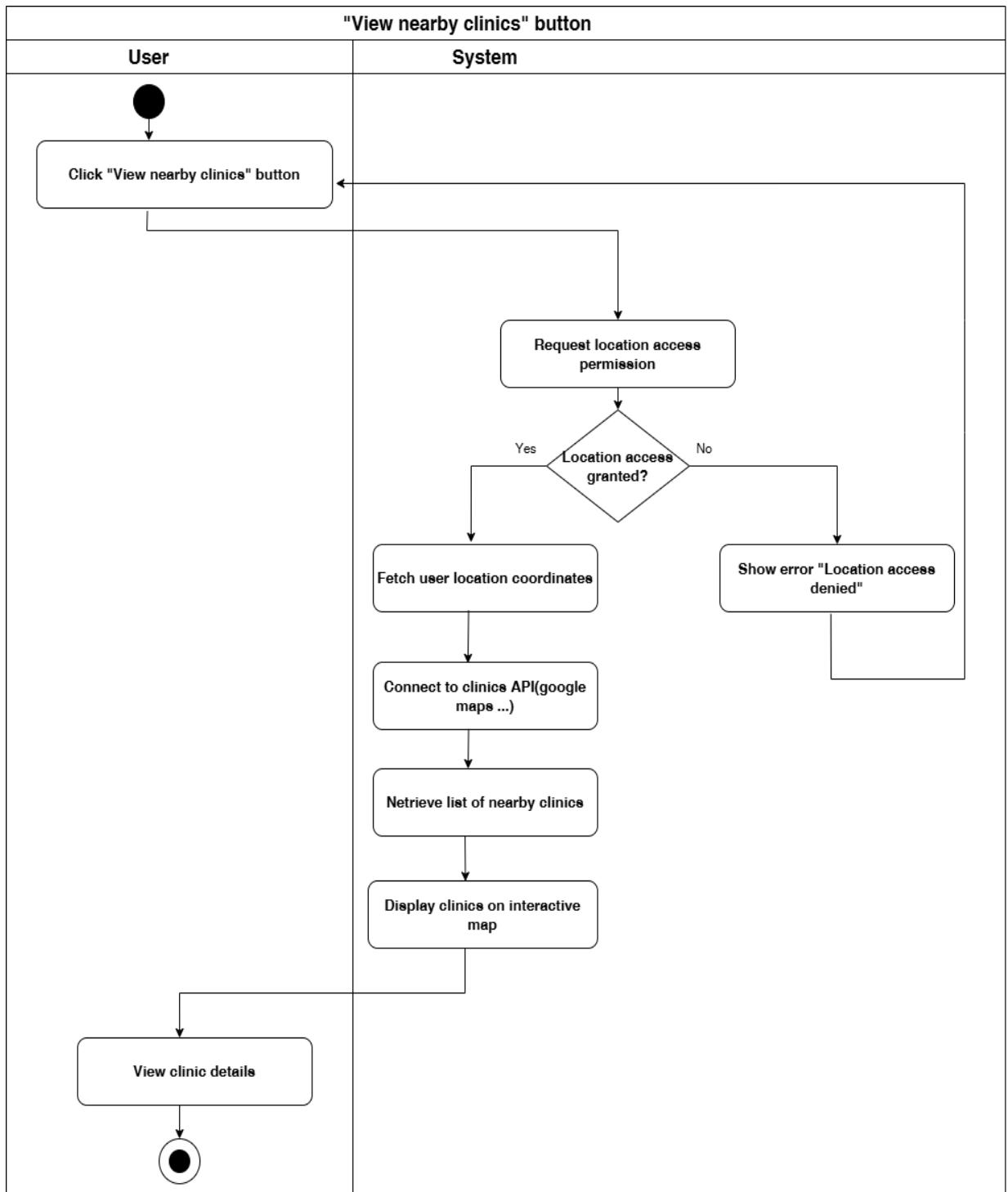


Figure 5.7: “View Nearby Clinics” button Activity Diagram.

5.2.7 “View Grad-Cam” button Activity Diagram

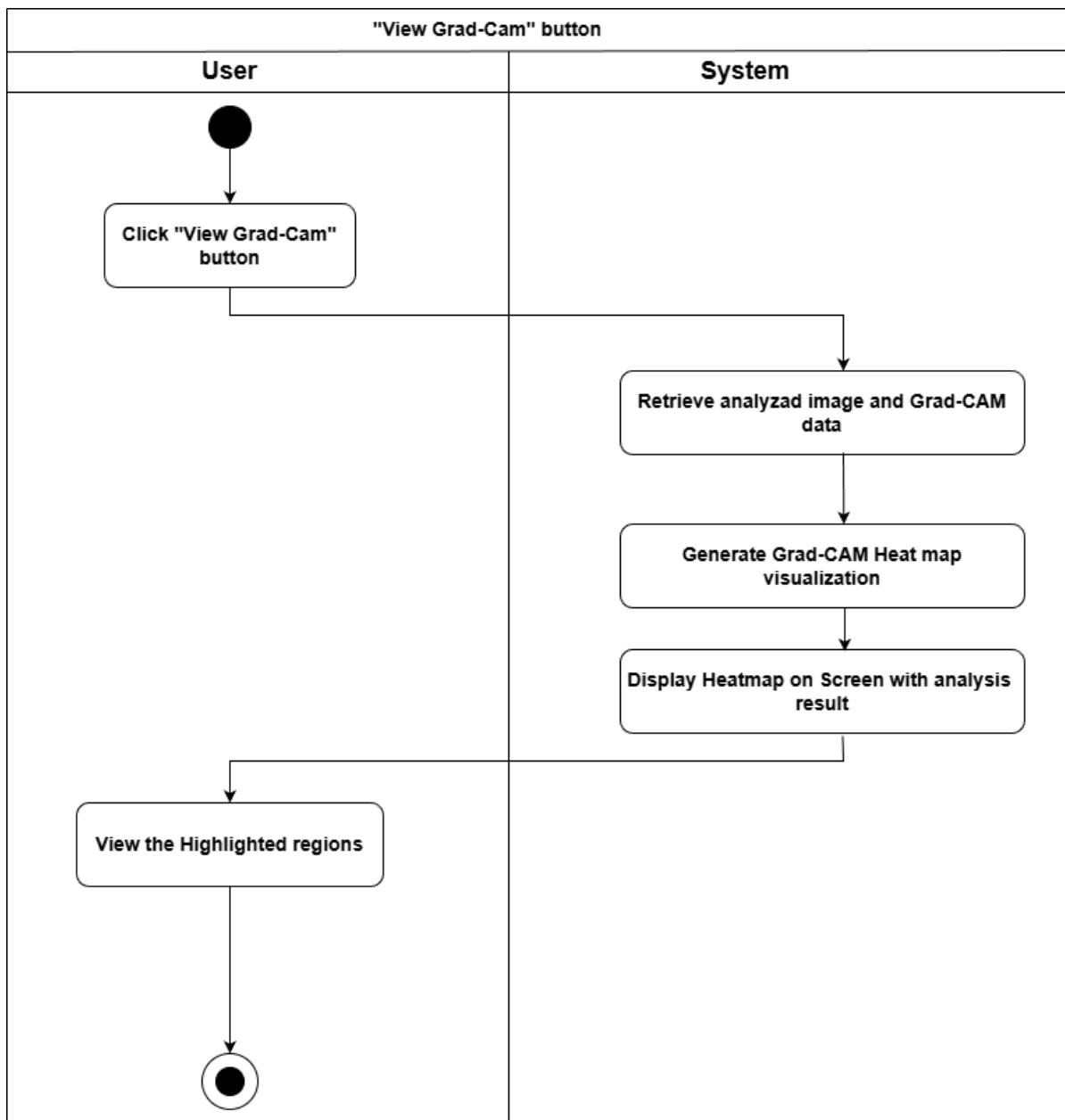


Figure 5.8: “View Grad-Cam” button Activity Diagram.

5.2.8 “My Reports” button Activity Diagram

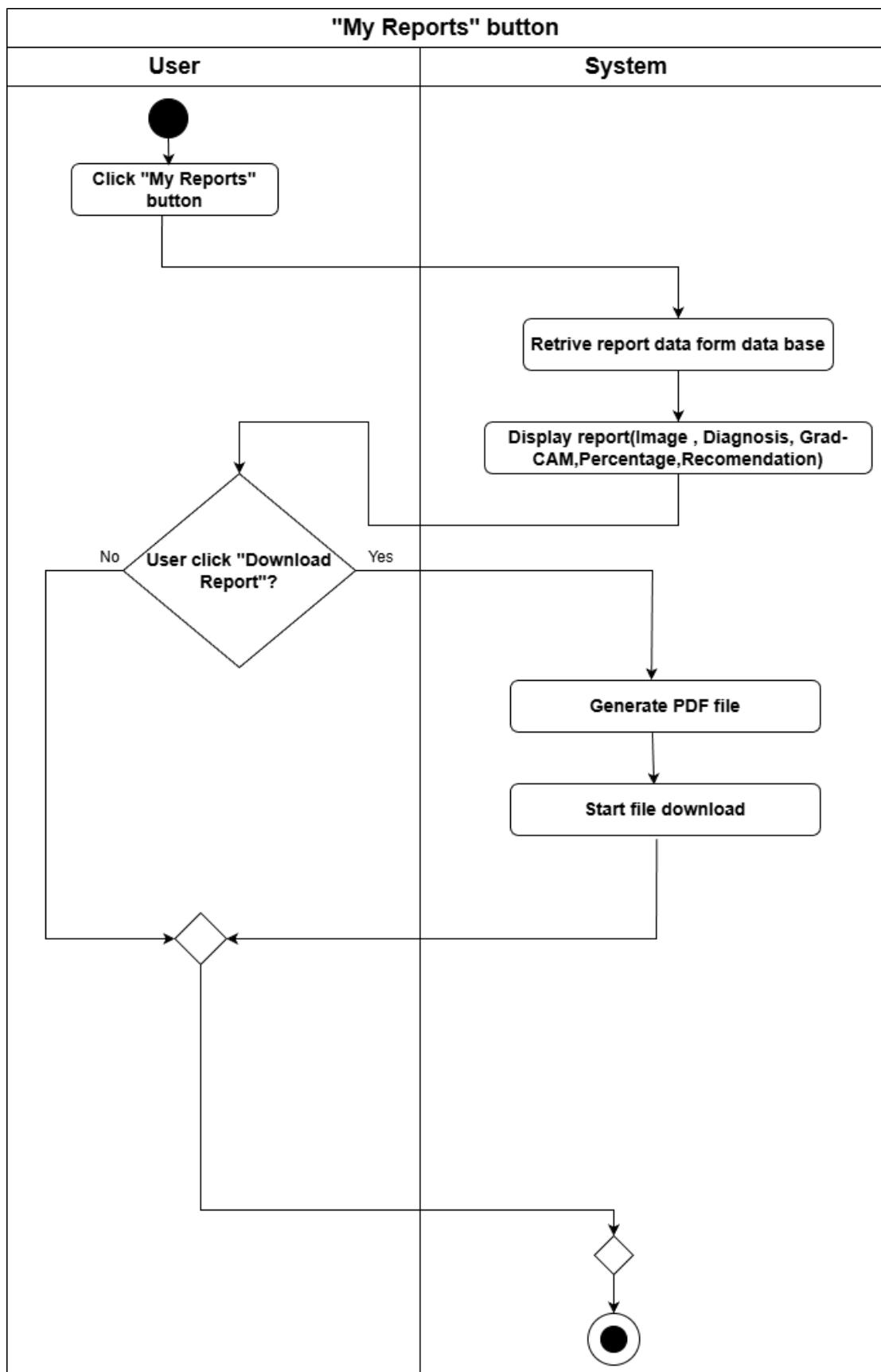


Figure 5.9: “My Reports” button Activity Diagram.

5.2.9 Imag-Valdition Activity Diagram

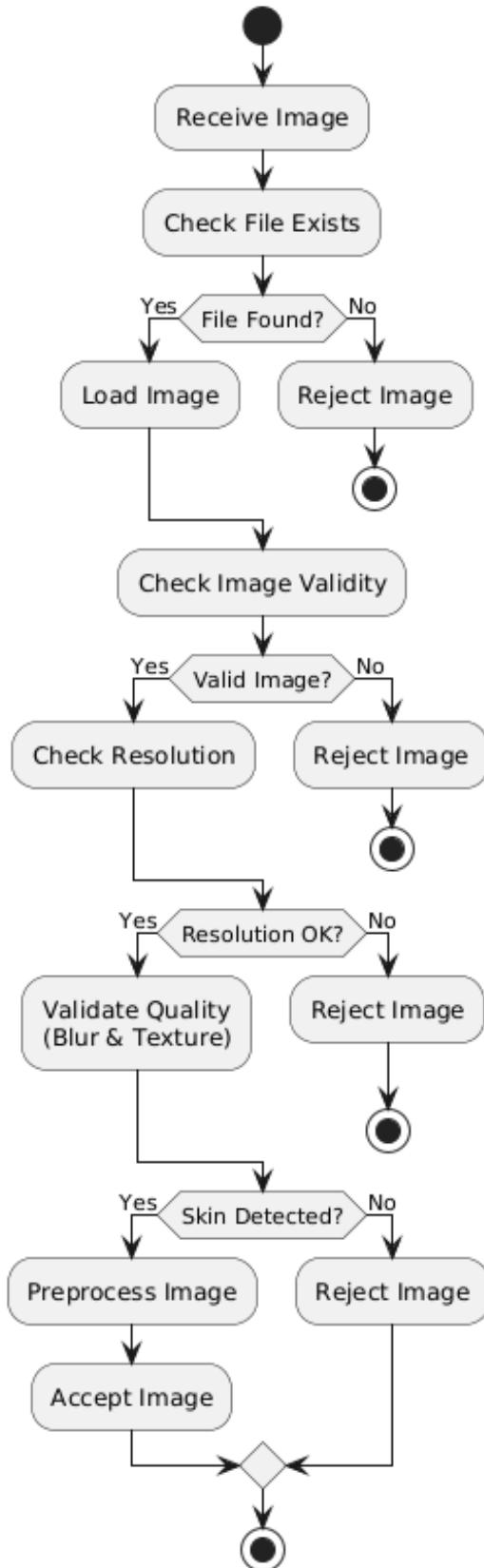


Figure 5.10: Imag-Valdition Activity Diagram.

5.3 Sequence Diagrams

5.3.1 Signup Sequence Diagram

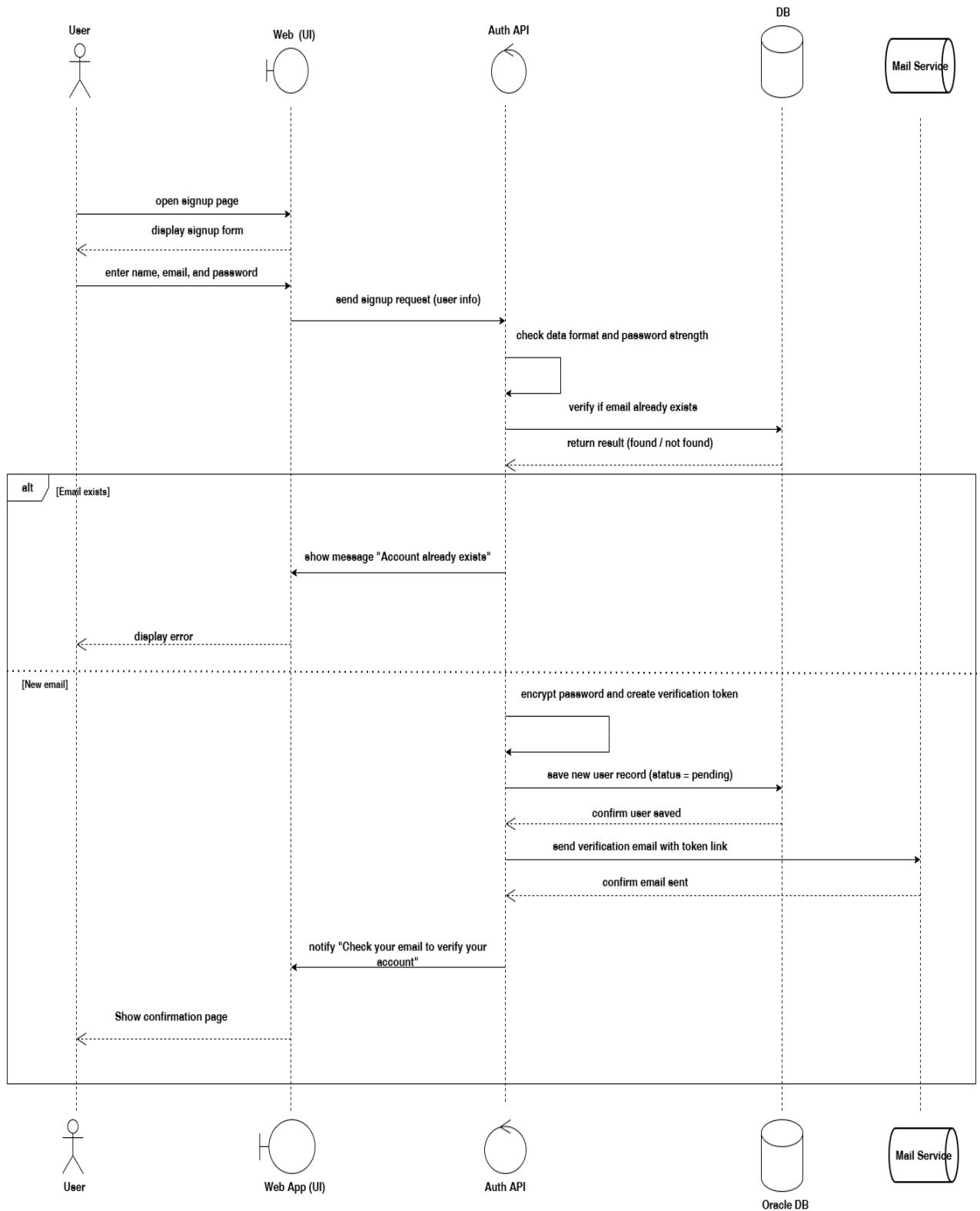


Figure 5.11: Signup Sequence Diagram.

5.3.2 Successfully Activate Account Sequence Diagram

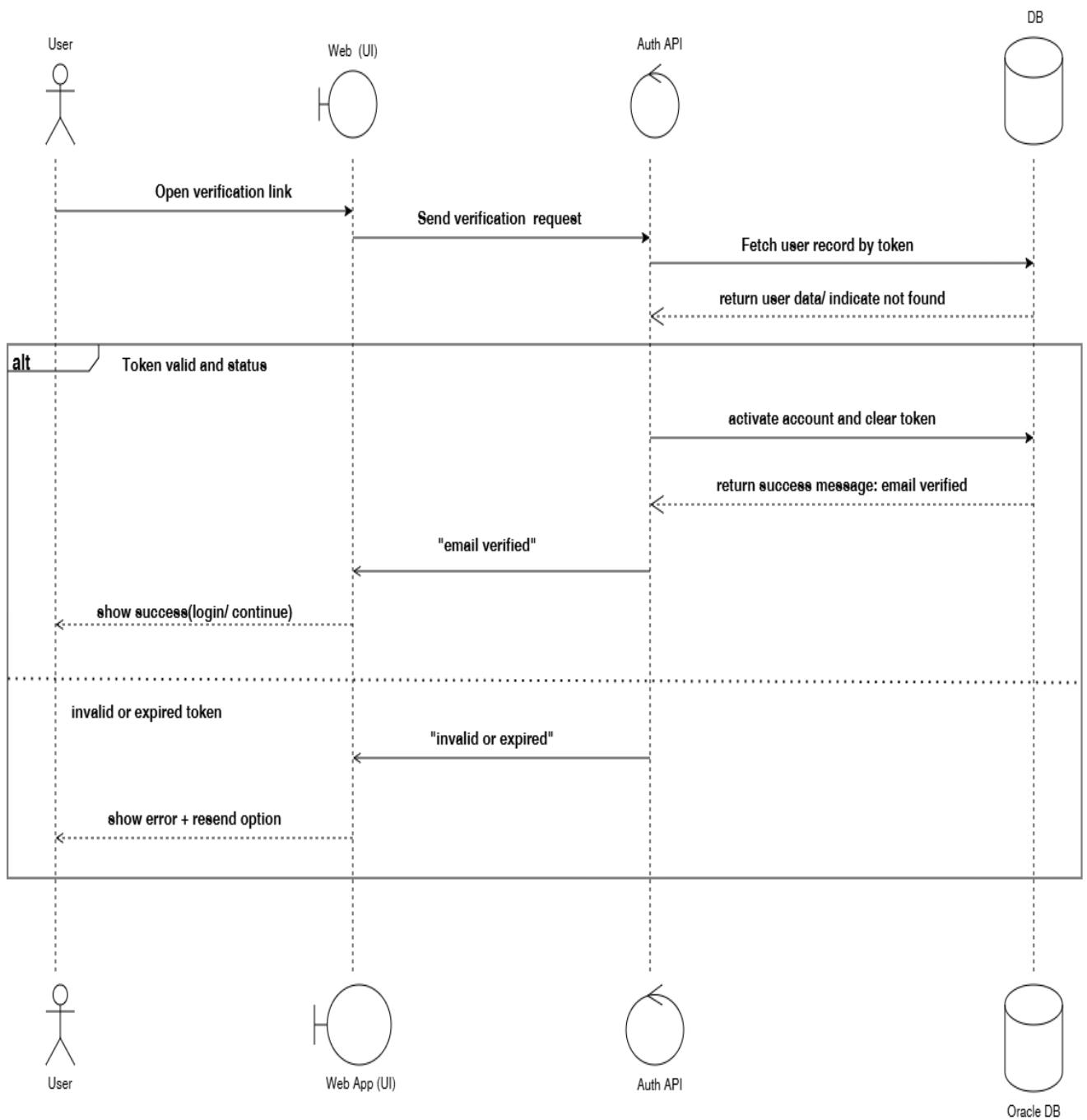


Figure 5.12: Successfully Activate Account Sequence Diagram.

5.3.3 Login Sequence Diagram

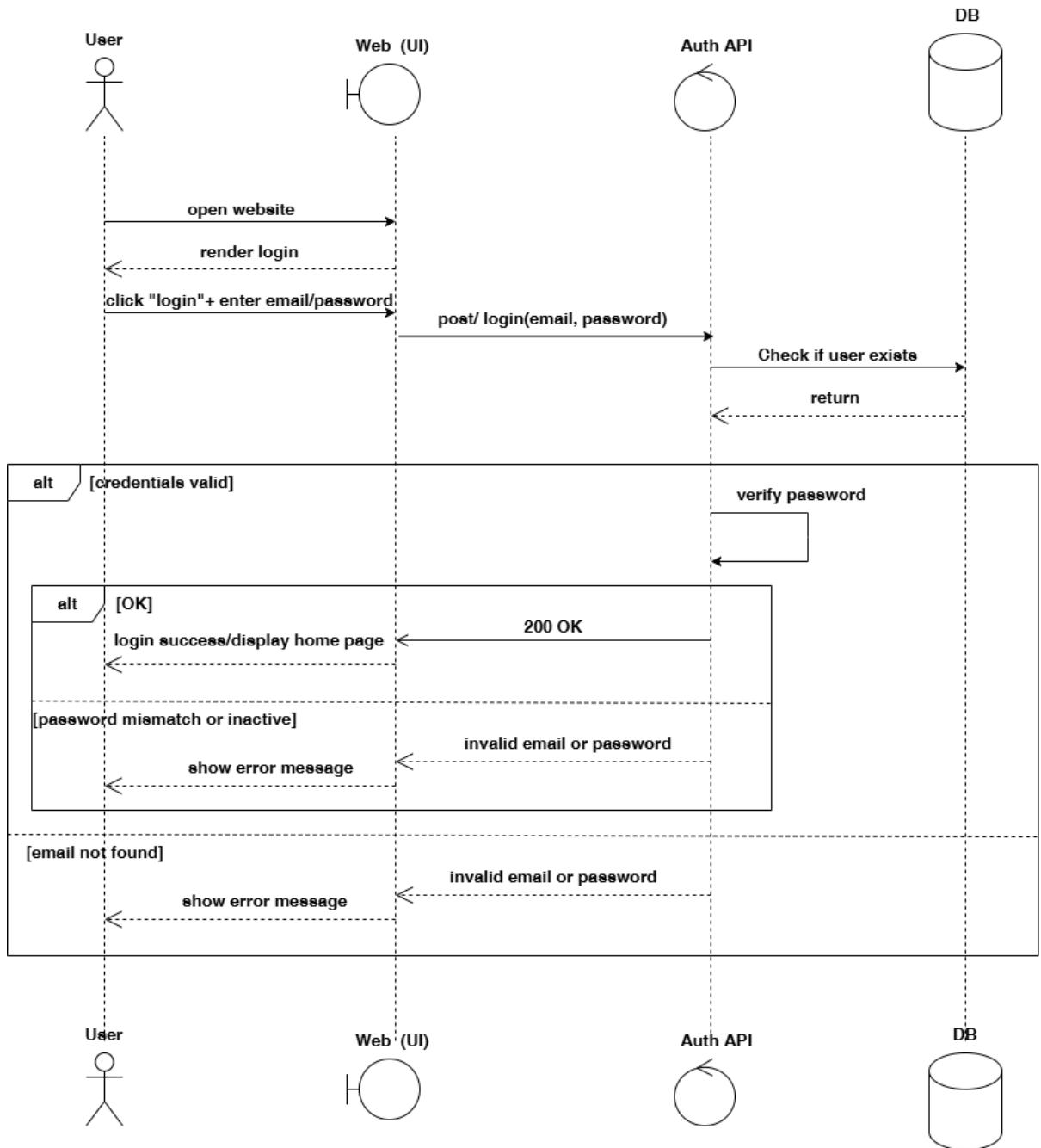


Figure 5.13: Login Sequence Diagram.

5.3.4 Password Reset Workflow Sequence Diagram

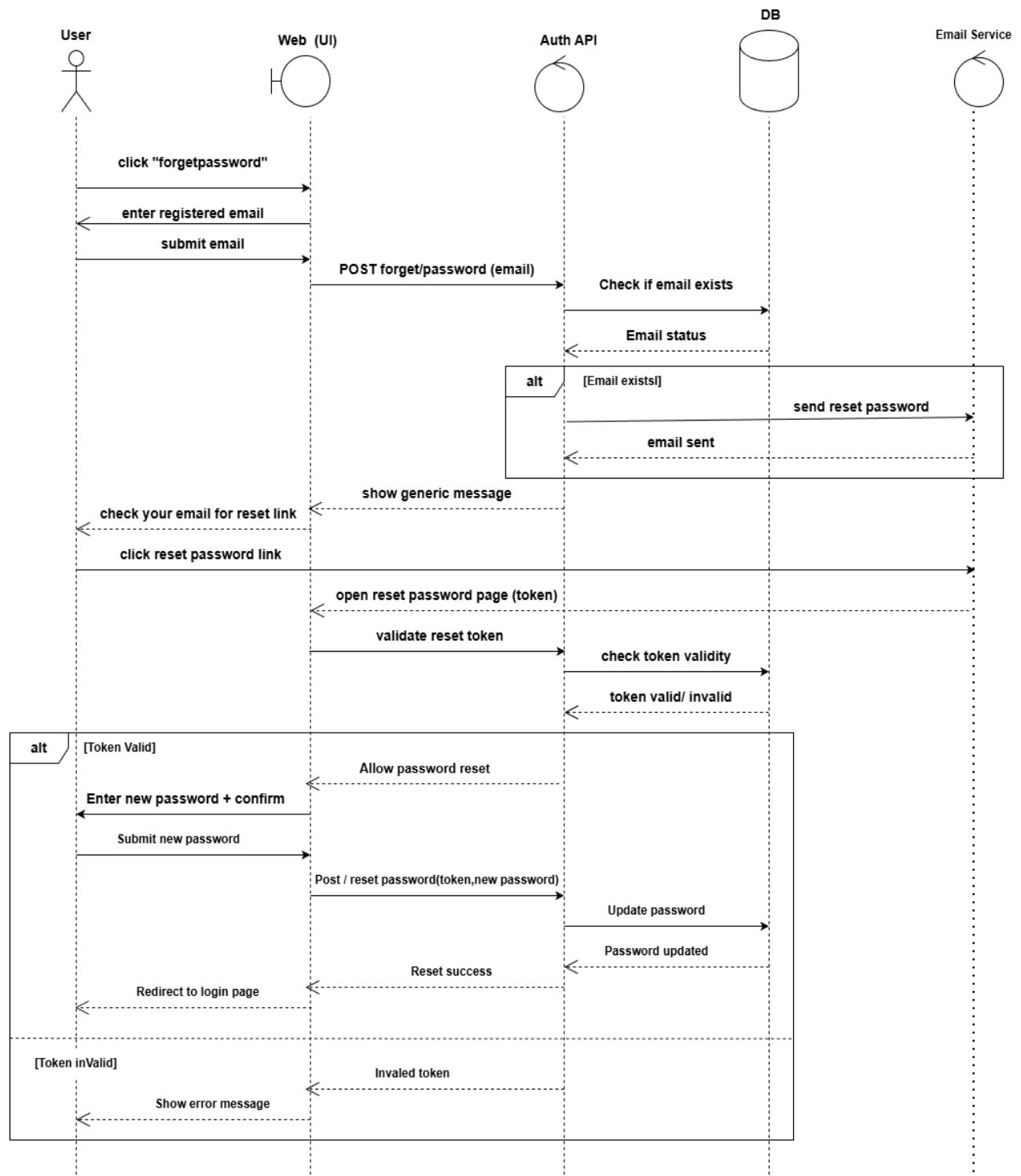


Figure 5.14: Password Reset Workflow Sequence Diagram.

5.3.6 Upload Image Sequence Diagram

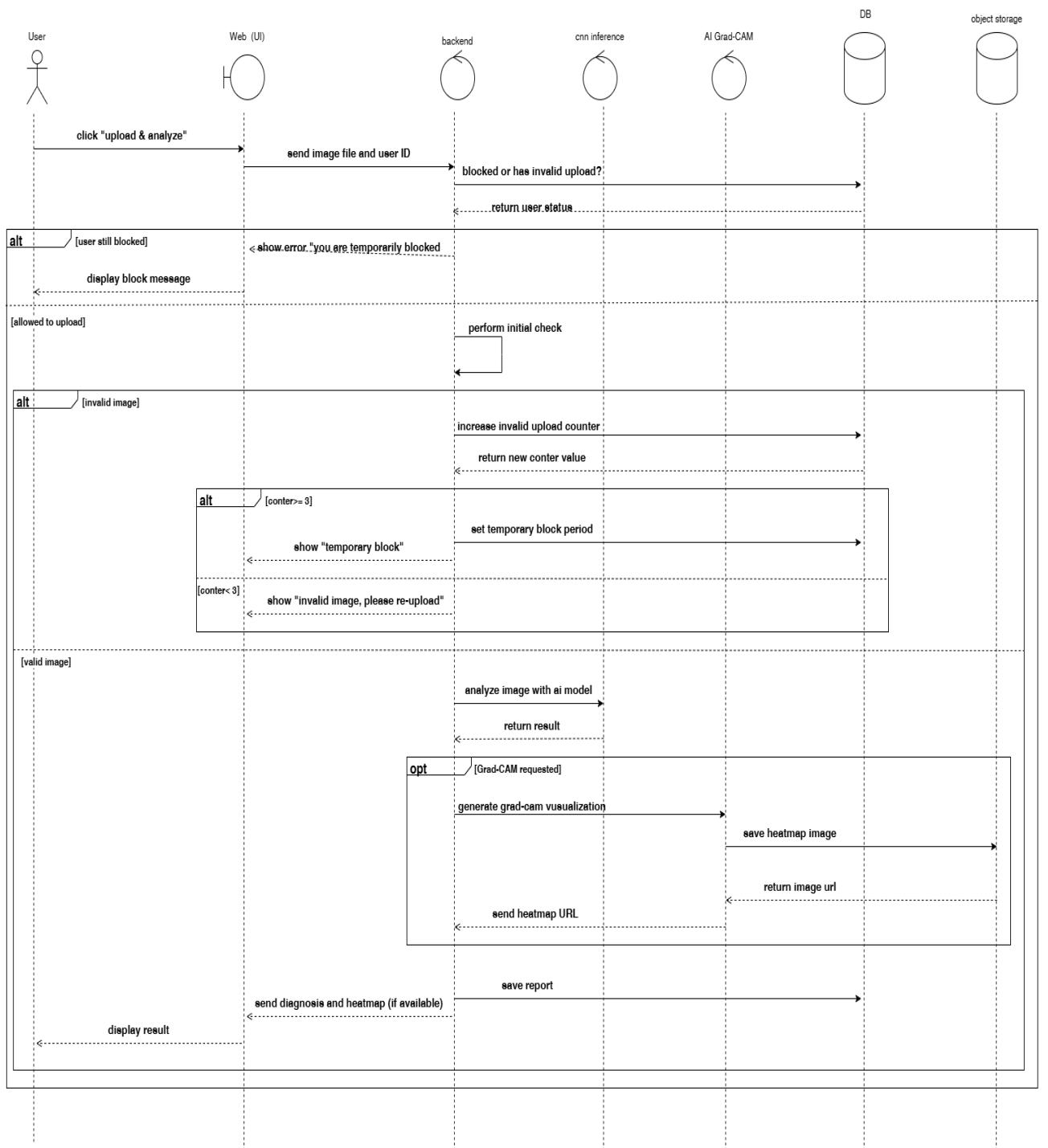


Figure 5.15: Upload Image Sequence Diagram.

5.3.7 “View Nearby Clinics” button Sequence Diagram

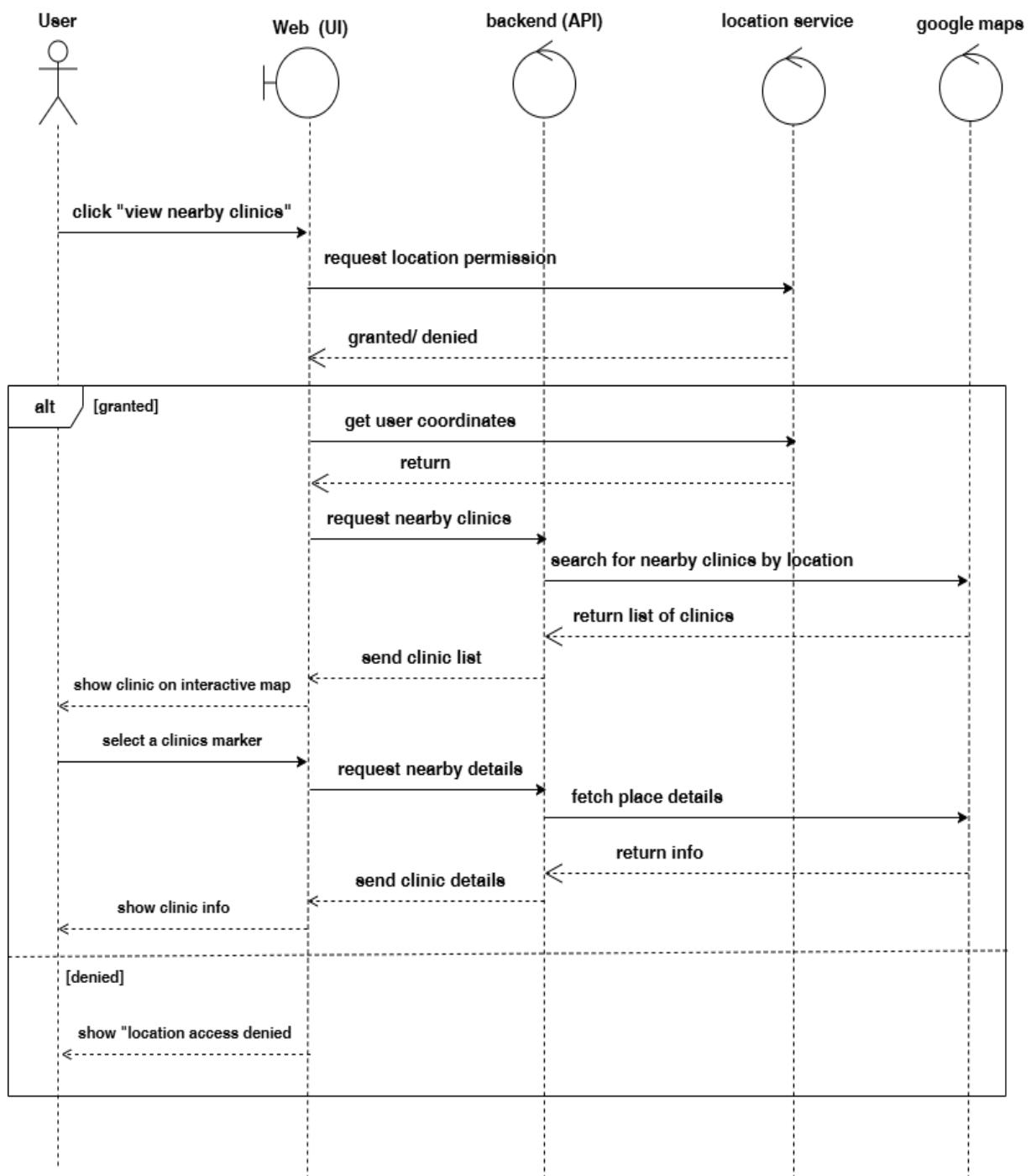


Figure 5.16: “View Nearby Clinics” button Sequence Diagram.

5.3.8 “View Grad-Cam” button Sequence Diagram

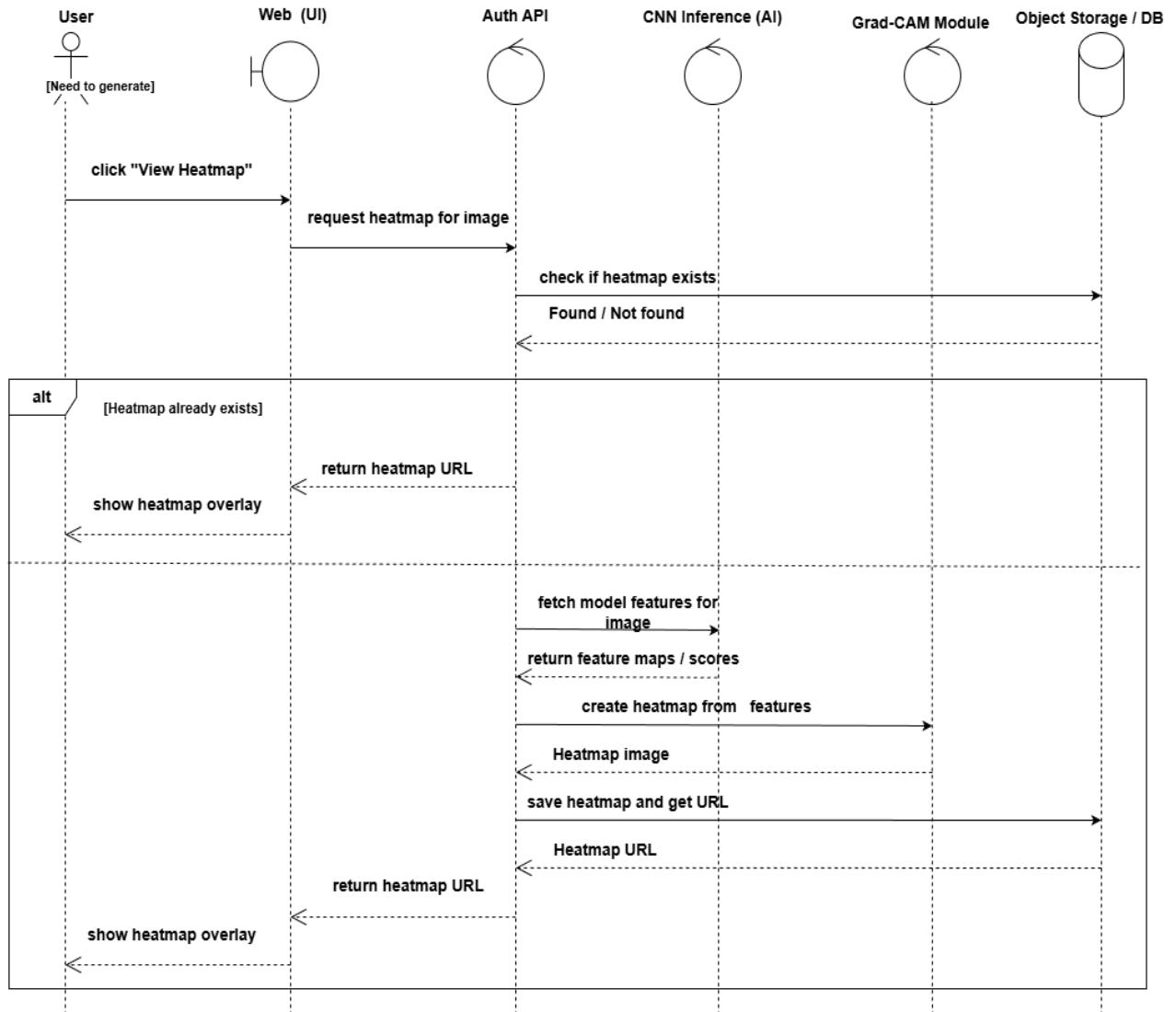


Figure 5.17: “View Grad-Cam” button Sequence Diagram.

5.3.9 “My Reports” button Sequence Diagram

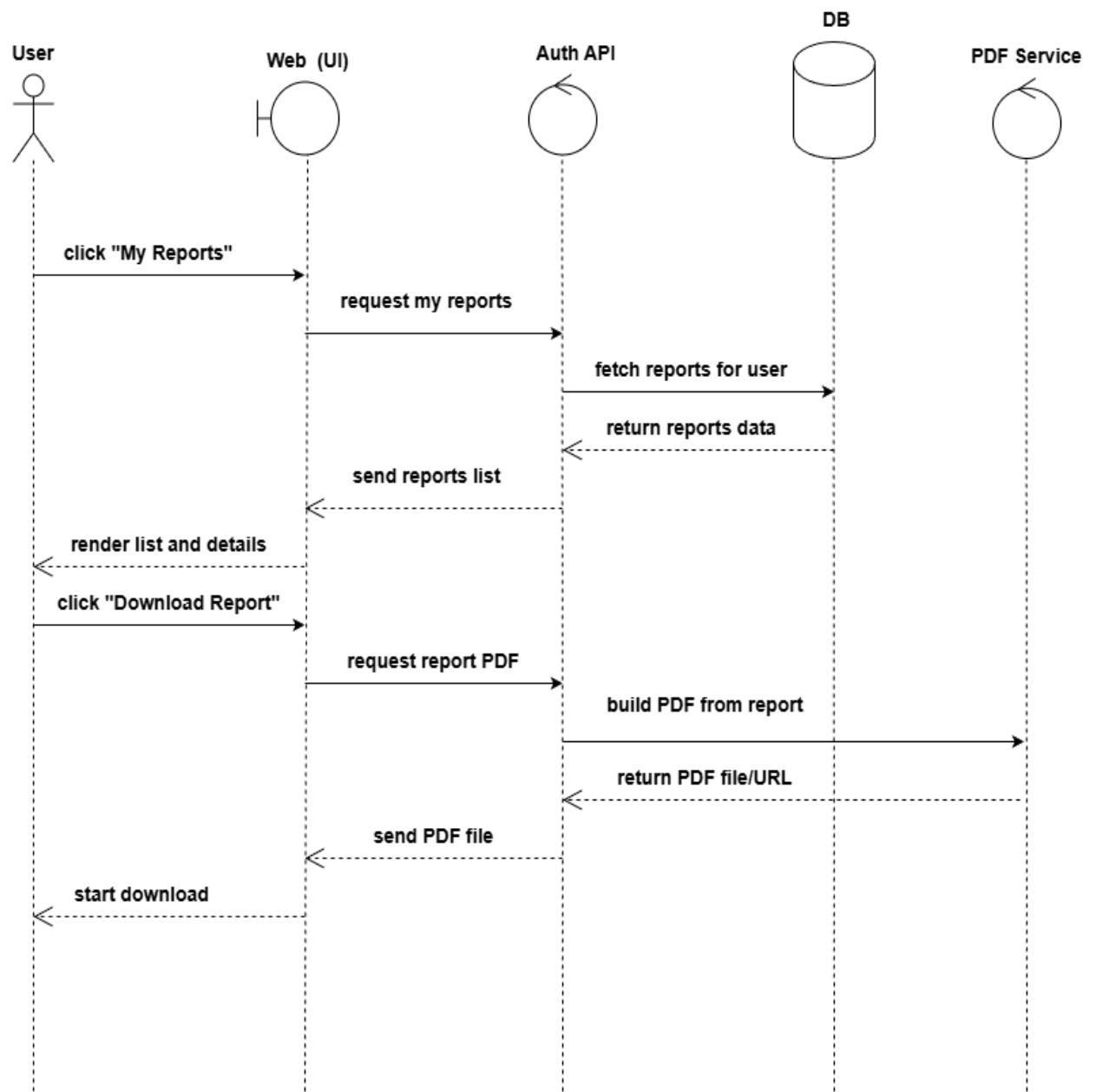


Figure 5.18: “My Reports” button Sequence Diagram.

5.3.10 Imag-Valdition Sequence Diagram

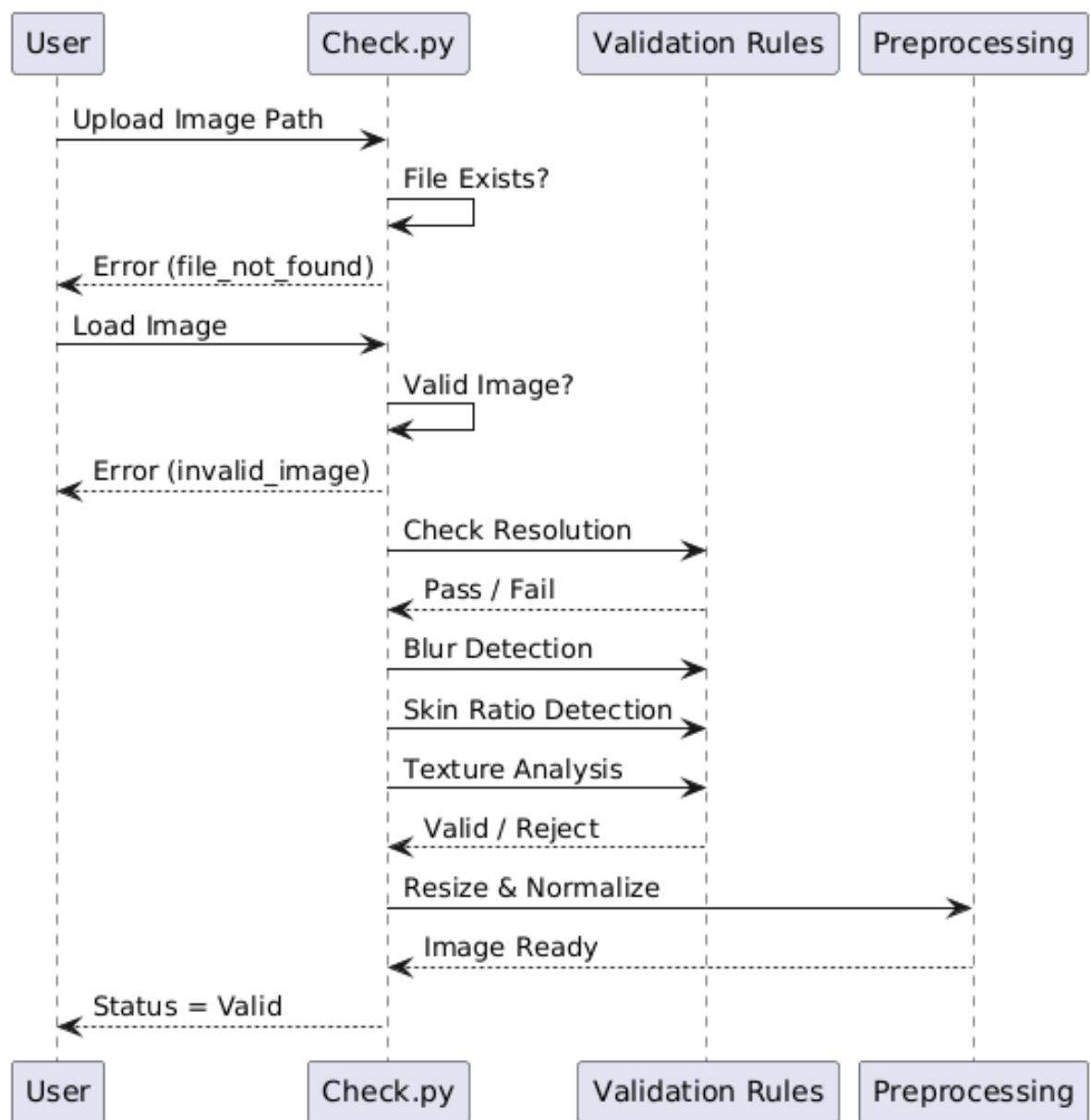


Figure 5.19: Imag-Valdition Sequence Diagram.

5.4 Relational Schema Diagram

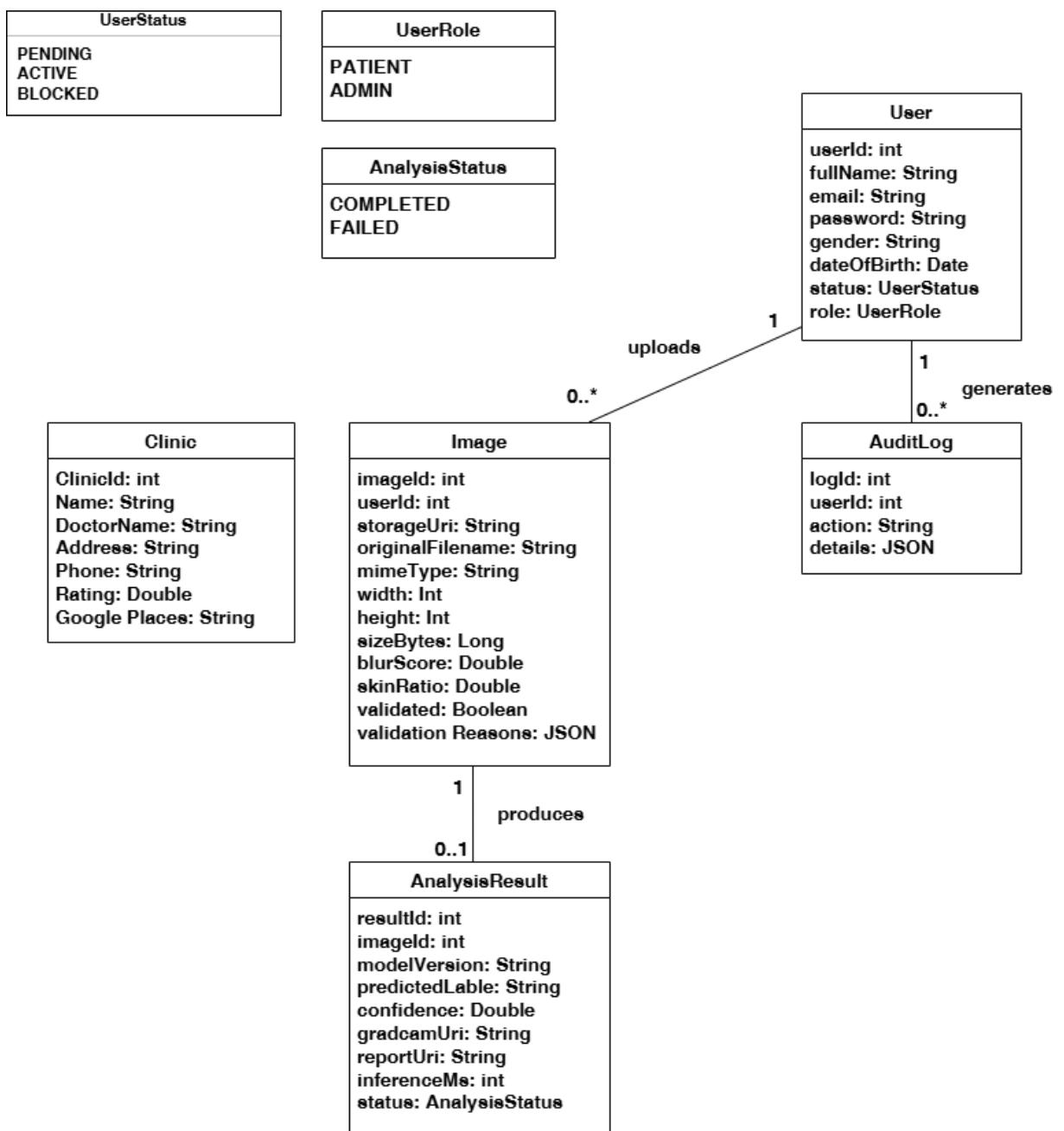


Figure 5.20: Relational Schema Diagram.

CHAPTER 6

RESULTS AND PERFORMANCE ANALYSIS

6.1 Introduction

This chapter presents the experimental results and performance analysis of the proposed DermAI system for skin cancer classification. The reported results are obtained through the experimental pipeline described in Chapter 3, including dataset preparation, comparative evaluation of multiple deep learning models, cross-validation, final model training, and the generation of Grad-CAM heatmaps for visual interpretation.

Special emphasis is placed on clinically relevant performance metrics, particularly the model's ability to accurately identify malignant lesions while minimizing diagnostically high-risk errors. Accordingly, the chapter analyzes the outcomes of the comparative model study, cross-validation experiments, and final training stage, together with detailed examination of confusion matrix patterns, decision threshold behavior, and Grad-CAM-based visual explanations. This structured analysis provides a coherent understanding of the system's diagnostic behavior across all evaluation stages.

6.2 Dataset Statistics After Preparation

After completing the dataset acquisition, cleaning, and preprocessing stages described in Chapter 3, the final dataset used for experimentation consisted of 19,500 dermoscopic images. All images were carefully reviewed to ensure data integrity, sufficient visual quality, and relevance to the task of skin lesion classification.

The final dataset included 13,290 benign images and 6,210 malignant images, resulting in an imbalanced class distribution. This imbalance is consistent with real-world clinical screening scenarios, where benign lesions are encountered more frequently than malignant ones. Consequently, the dataset provides a realistic and clinically relevant setting for evaluating the performance and reliability of the proposed classification model.

6.3 Evaluation Metrics Overview

In medical image classification tasks—particularly in cancer detection—model evaluation must extend beyond overall accuracy. While accuracy provides a general measure of correctness, it may be misleading in imbalanced datasets where one class dominates.

For this reason, the evaluation of the DermAI system prioritizes the following metrics:

Recall (Sensitivity)

Recall measures the proportion of malignant cases correctly identified by the model. In a clinical context, this metric is critical, as false negatives represent missed cancer cases, which can lead to delayed treatment and serious health consequences. Maximizing recall directly reduces clinical risk.

F1-Score

The F1-score represents the harmonic mean of precision and recall, providing a balanced evaluation of the model's diagnostic capability. This metric is particularly suitable when both false positives and false negatives carry significant implications, such as unnecessary anxiety versus missed diagnoses.

By emphasizing recall and F1-score, the evaluation framework aligns with clinical priorities, where patient safety and early detection outweigh purely statistical performance measures.

6.4 Results of the Comparative Model Study

6.4.1 Comparative Performance Analysis

This subsection presents a comparative analysis of the evaluated machine learning and deep learning models under a unified experimental setup. All models were trained and evaluated using the same dataset, preprocessing pipeline, augmentation strategy, optimizer configuration, learning rate, number of epochs, and decision threshold in order to ensure a fair and unbiased comparison.

Table 6.1 summarizes the performance of the evaluated models in terms of Accuracy, Precision, Recall, and F1-score. While accuracy provides a general indication of classification correctness, it is not sufficient on its own for assessing diagnostic reliability in medical applications. Therefore, particular attention is given to recall and F1-score, as these metrics better reflect the model's ability to detect malignant cases while balancing false positives and false negatives.

Model	Accuracy	Precision	Recall	F1-Score
DenseNet121	85.79%	77.40%	78.37%	77.88%
VGG16	85.29%	86.44%	63.95%	73.51%
InceptionV3	83.98%	76.24%	72.41%	74.28%
Xception	83.98%	79.34%	67.40%	72.88%
ResNet50	81.68%	76.56%	61.44%	68.17%
KNN	81.38%	81.52%	53.92%	64.91%
ANN	79.18%	77.07%	49.53%	60.31%
Custom CNN	73.57%	86.67%	20.38%	32.99%
EfficientNetB0	68.07%	0%	0%	0%

Table 6.1 presents the comparative performance of the evaluated models in terms of accuracy, precision, recall, and F1-score, highlighting differences in their ability to detect malignant skin lesions.

Among all evaluated architectures, DenseNet121 achieved the best overall performance, recording the highest accuracy (85.79%) and F1-score (77.88%), along with a strong recall of 78.37%. This indicates a robust capability to correctly identify malignant lesions while maintaining balanced classification behavior.

VGG16 demonstrated high precision (86.44%) but exhibited a substantially lower recall (63.95%), indicating a conservative prediction strategy that favors minimizing false positives at the cost of missing malignant cases. Such behavior, while producing fewer false alarms, is less suitable for clinical screening scenarios where false negatives carry a higher risk.

InceptionV3 and Xception achieved moderate and relatively balanced performance across all evaluation metrics, suggesting stable but less optimal diagnostic behavior compared to DenseNet121. ResNet50, while achieving acceptable accuracy, showed reduced recall (61.44%), indicating limitations in capturing malignant patterns under the standardized training configuration.

Traditional machine learning and shallow learning approaches, including KNN and ANN, exhibited inferior performance compared to convolutional neural network-based architectures. This outcome highlights the importance of spatial feature extraction when dealing with high-resolution dermoscopic images.

Notably, EfficientNetB0 achieved a reasonable accuracy (68.07%) but resulted in zero precision, recall, and F1-score. This behavior indicates that the model consistently favored the majority class (benign) and failed to detect malignant samples under the strictly unified experimental setup. Although EfficientNet is designed for parameter efficiency, this result demonstrates that architectural efficiency alone does not guarantee diagnostic robustness in imbalanced medical datasets.

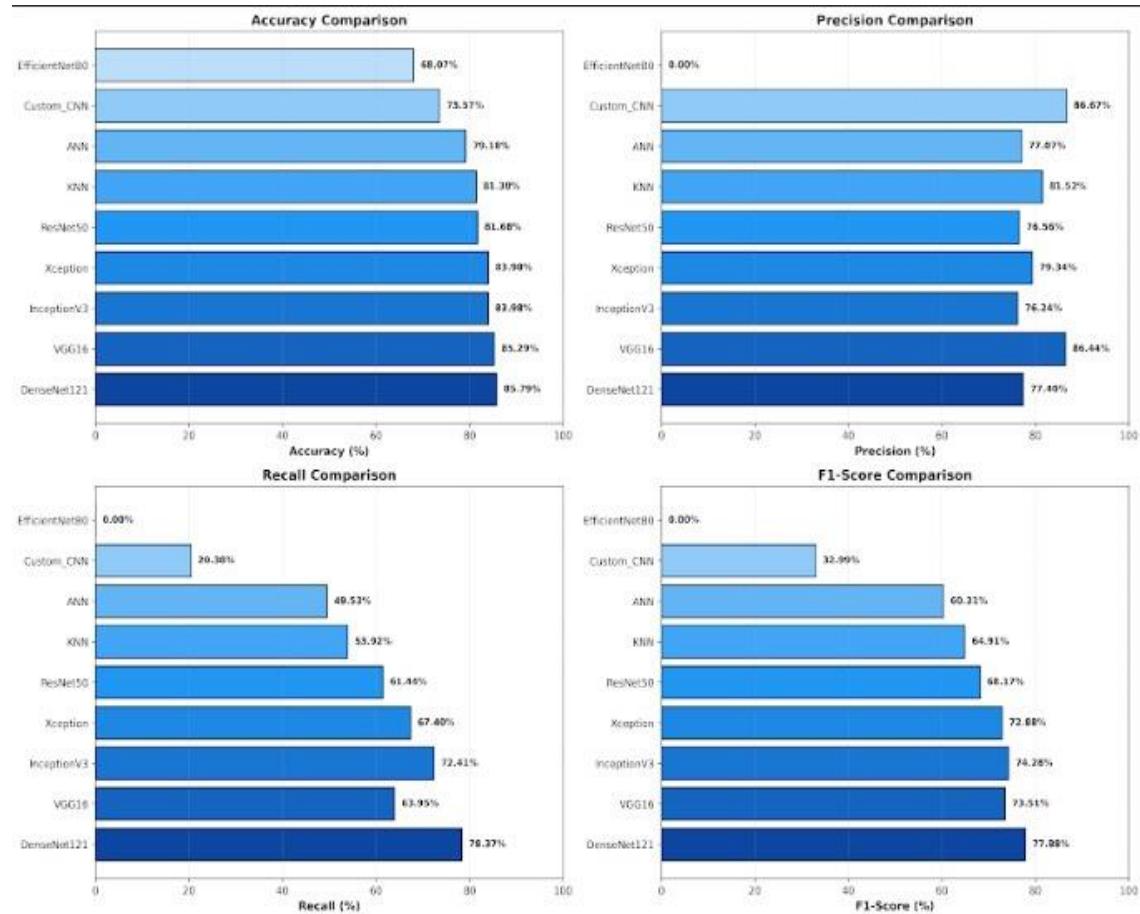


Figure 6.1: Comparative performance of the evaluated models on the DermAI sample dataset.

Overall, the comparative analysis confirms that deep convolutional architectures—particularly DenseNet121 and ResNet-based models—are more suitable for automated skin lesion classification, especially when recall and F1-score are prioritized for clinical relevance.

6.4.2 ResNet50 Performance Analysis

ResNet50 was selected for detailed analysis due to its widespread adoption in medical imaging tasks and its residual learning mechanism, which enables stable training of deep networks. Under the unified experimental conditions, ResNet50 achieved an accuracy of 81.68%, with a precision of 76.56%, recall of 61.44%, and an F1-score of 68.17%.



Figure 6.2: Confusion matrix of the ResNet50 model evaluated on the DermAI validation dataset.

The confusion matrix analysis (Figure X) reveals that ResNet50 correctly classified a substantial portion of benign samples, while a noticeable number of malignant cases were misclassified as benign. This behavior indicates a bias toward the majority class, which is consistent with the imbalanced nature of the dataset and the absence of class-specific weighting or threshold optimization.

From a clinical perspective, the recall value of 61.44% suggests that approximately four out of ten malignant cases may be missed by the model. Although this performance may be acceptable for general classification tasks, it is insufficient for standalone diagnostic use in medical screening applications, where sensitivity to malignant cases is critical.

Nevertheless, ResNet50 demonstrated stable convergence during training and consistent validation behavior, indicating good generalization under standardized conditions. The model's performance reflects a trade-off between stability and sensitivity, making it suitable as a baseline deep learning architecture rather than the final diagnostic model within the DermAI system.

These findings highlight the importance of prioritizing architectures that achieve higher recall and F1-score, particularly in clinically sensitive applications such as skin cancer screening.

6.5 Cross-Validation Results Analysis

This section presents the results obtained from a three-fold cross-validation experiment conducted on the fine-tuned ResNet50 model, following the validation strategy described in Section 3.8. The analysis aims to examine the model's performance stability and generalization capability when trained and evaluated on different stratified subsets of the dataset.

Each fold was trained independently under identical preprocessing, training, and optimization settings to ensure fair and consistent evaluation. By comparing fold-wise validation outcomes, this section highlights the extent to which the model's predictive behavior remains stable across varying data splits, particularly in terms of sensitivity to malignant lesions. The reported results reflect validation performance only and serve as an empirical assessment of the robustness of the proposed model under cross-validation.

6.5.1 Cross-Validation Performance Metrics

Table 6.4 summarizes the validation performance metrics obtained from the three-fold cross-validation experiment conducted on the fine-tuned ResNet50 model. The reported metrics include accuracy, precision, recall, and F1-score, with a particular focus on the malignant class due to its clinical importance.

The results indicate consistent model performance across different data splits, with moderate variations reflecting differences in validation subsets. Accuracy values ranged between 0.8148 and 0.8296, demonstrating stable overall classification capability. Precision values for malignant lesions remained relatively high across folds, indicating a low false-positive rate, while recall exhibited higher variability, highlighting differences in sensitivity to malignant cases among folds.

Notably, Fold 3 achieved the highest recall (0.7710), emphasizing its strong sensitivity to malignant lesions, whereas Fold 1 obtained the highest overall accuracy (0.8296). Fold 2 demonstrated a balanced trade-off between precision and recall, resulting in a competitive F1-score and stable overall behavior.

Fold	Accuracy	Precision (Malignant)	Recall (Malignant)	F1-Score
Fold 1	0.8296	0.7839	0.6415	0.7056
Fold 2	0.8182	0.7913	0.5826	0.6711
Fold 3	0.8148	0.6862	0.7710	0.7261

Table 6.2: Cross-validation performance metrics across three folds.

6.5.2 ROC Curve Analysis Across Folds

This subsection presents a detailed analysis of the Receiver Operating Characteristic (ROC) curves obtained from the three-fold cross-validation experiment. The ROC curves are used to evaluate the model's ability to discriminate between benign and malignant skin lesions, with particular emphasis on its sensitivity to malignant cases.

ROC Curve – Fold 1

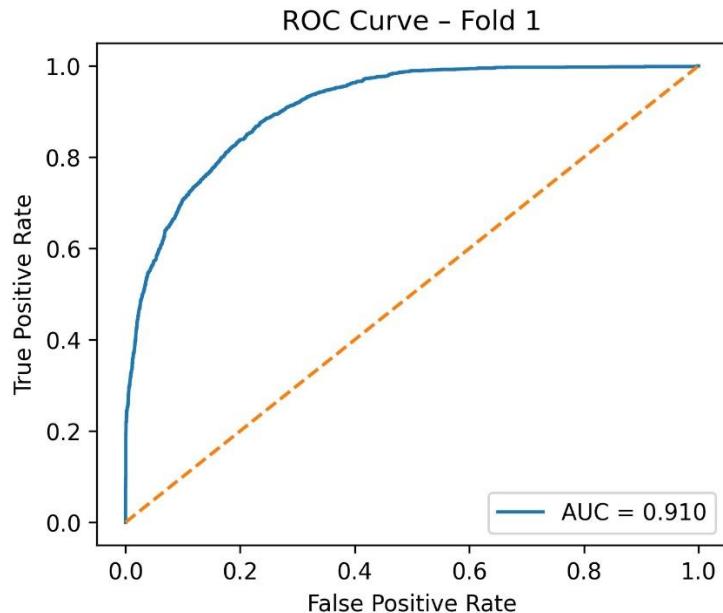


Figure 6.3: ROC curve and AUC value for Fold 1 obtained during cross-validation of the fine-tuned ResNet50 model.

Figure 6.3 illustrates the Receiver Operating Characteristic (ROC) curve obtained for Fold 1 during the cross-validation process. The model achieved an Area Under the Curve (AUC) value of 0.910, which represents the highest discriminative performance among the three folds. This result indicates a strong ability to distinguish between benign and malignant lesions across a wide range of classification thresholds.

The clear separation of the ROC curve from the diagonal baseline reflects high sensitivity to malignant cases while maintaining a relatively low false positive rate. This behavior demonstrates that the model effectively learned discriminative features within the validation subset of Fold 1.

ROC Curve – Fold 2

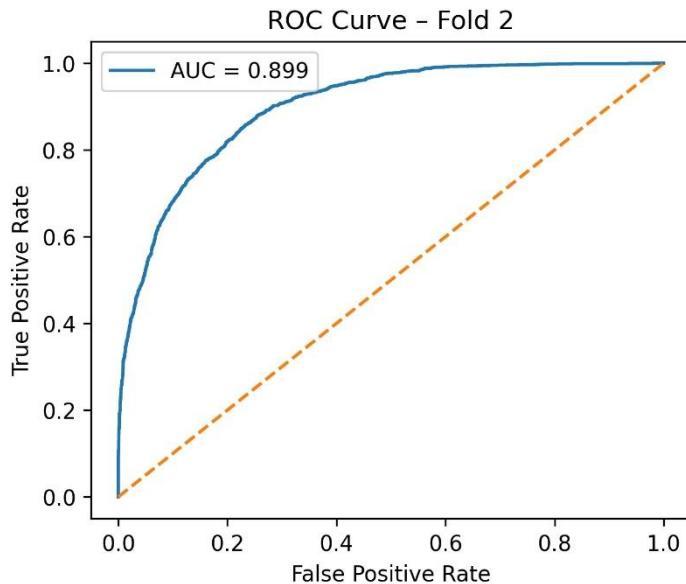


Figure 6.4: ROC curve and AUC value for Fold 2 obtained during cross-validation of the fine-tuned ResNet50 model.

Figure 6.4 presents the ROC curve corresponding to Fold 2, which achieved an AUC value of 0.899. Although slightly lower than Fold 1, this value still indicates strong classification performance and reliable discrimination between benign and malignant lesions.

Compared to the other folds, Fold 2 exhibits a balanced ROC profile, with a smooth increase in sensitivity and controlled false positive rate. This stable behavior contributed to the selection of Fold 2 as the representative fold for subsequent model training and deployment.

ROC Curve – Fold 3

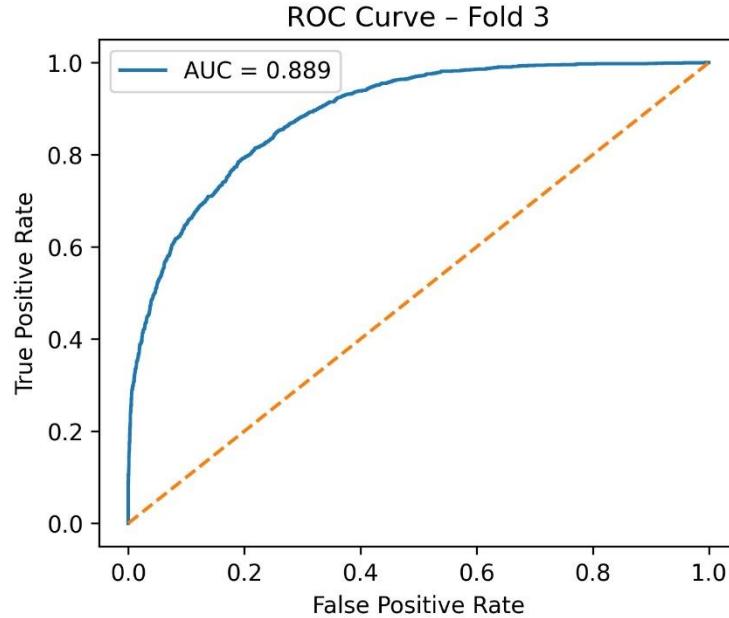


Figure 6.5: ROC curve and AUC value for Fold 3 obtained during cross-validation of the fine-tuned ResNet50 model.

Figure 6.5 shows the ROC curve for Fold 3, which achieved an AUC value of 0.889. While this is the lowest AUC among the three folds, it still reflects strong discriminative capability and confirms the robustness of the proposed model across different validation splits.

The ROC curve maintains a clear deviation from the diagonal reference line, indicating effective classification performance. The slightly reduced AUC can be attributed to differences in lesion characteristics or class distribution within the validation subset of Fold 3.

Overall, the ROC analysis across the three folds demonstrates consistent and reliable model performance under cross-validation. The relatively small variation in AUC values indicates that the fine-tuned ResNet50 model generalizes well across different data splits and is not overly sensitive to the specific composition of the validation subsets. These results further support the robustness of the proposed approach for skin lesion classification.

6.5.3 Accuracy and Loss Curves Analysis

This subsection analyzes the training and validation accuracy and loss curves obtained during the three-fold cross-validation process. These curves provide insight into the learning dynamics, convergence behavior, and generalization performance of the fine-tuned ResNet50 model across different data splits.

Fold 1 – Accuracy and Loss Curves

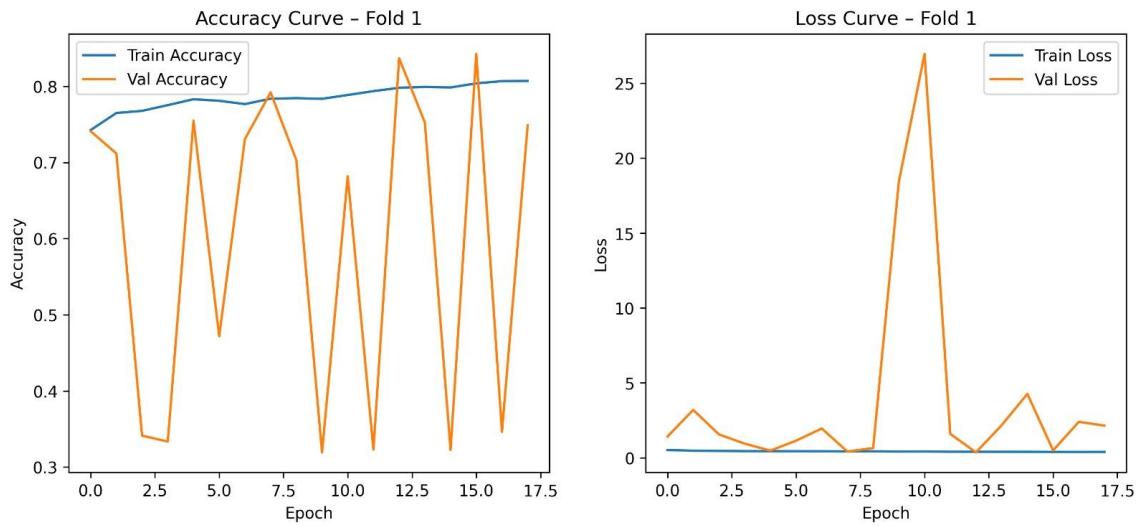


Figure 6.6: Training and validation accuracy and loss curves for Fold 1 obtained during cross-validation of the fine-tuned ResNet50 model.

Figure 6.6 illustrates the training and validation accuracy and loss curves for Fold 1. The training accuracy shows a steady increase over epochs, accompanied by a consistent decrease in training loss, indicating effective learning and convergence. In contrast, the validation accuracy and loss exhibit noticeable fluctuations, which can be attributed to the limited size and class imbalance of the validation subset. Despite these variations, no persistent divergence between training and validation curves is observed, suggesting controlled overfitting.

Fold 2 – Accuracy and Loss Curves

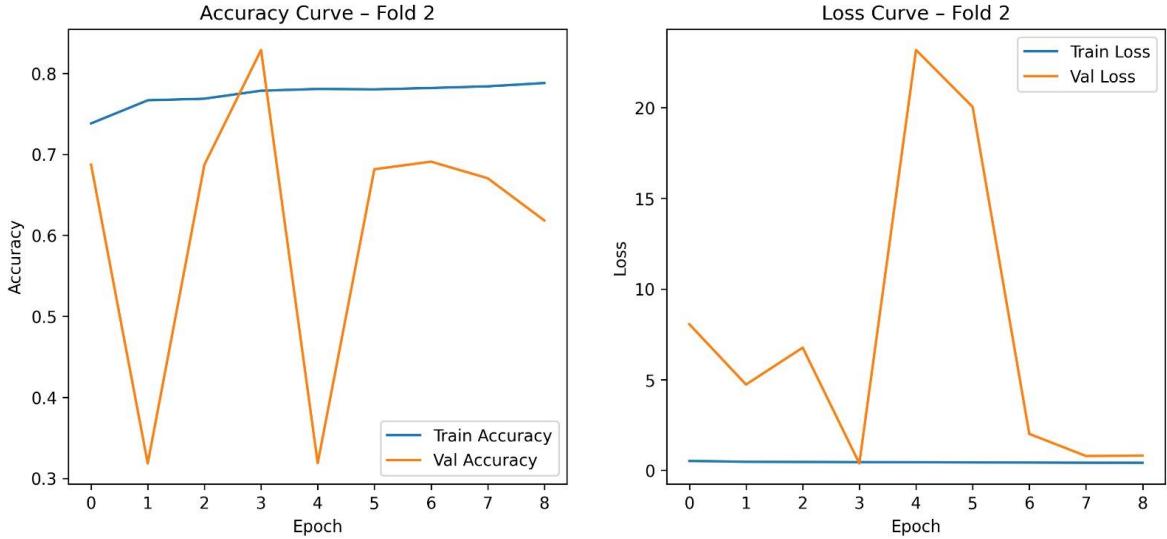


Figure 6.7: Training and validation accuracy and loss curves for Fold 2 obtained during cross-validation of the fine-tuned ResNet50 model.

Figure 6.7 presents the accuracy and loss curves for Fold 2. Compared to the other folds, Fold 2 demonstrates smoother validation behavior with reduced fluctuations. The validation accuracy remains relatively stable across epochs, and the validation loss shows fewer sharp spikes. This balanced convergence behavior indicates improved generalization performance and supports the selection of Fold 2 as the representative fold for subsequent model training.

Fold 3 – Accuracy and Loss Curves

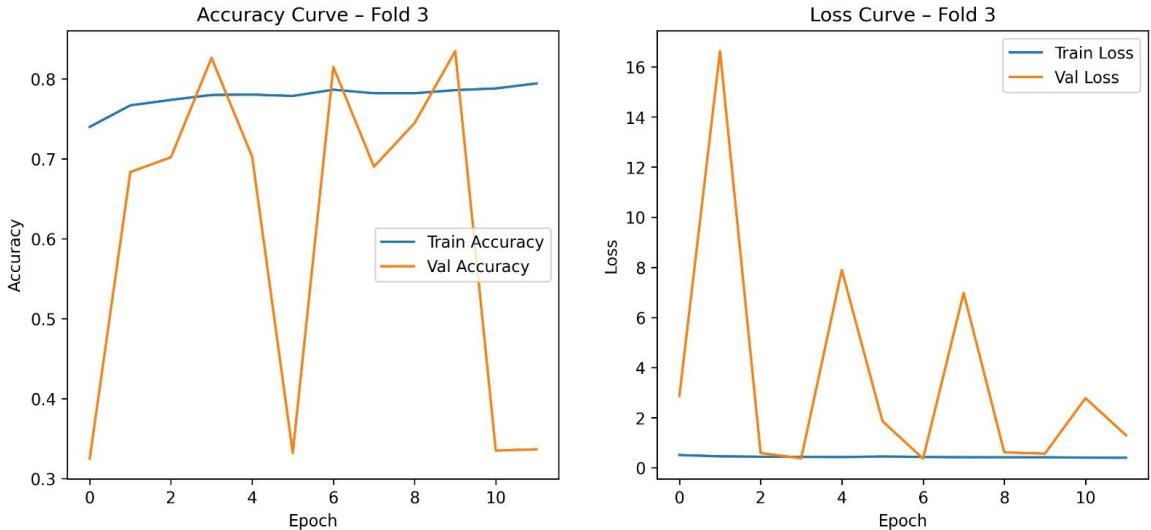


Figure 6.8: Training and validation accuracy and loss curves for Fold 3 obtained during cross-validation of the fine-tuned ResNet50 model.

Figure 6.8 shows the training and validation accuracy and loss curves for Fold 3. While the training curves follow stable trends similar to those observed in the other folds, the validation curves display higher variability. This behavior suggests increased sensitivity to the composition of the validation subset in Fold 3. Nevertheless, the overall convergence pattern remains acceptable, confirming the robustness of the training process.

Overall, the accuracy and loss curves across the three folds demonstrate consistent learning behavior and stable convergence of the fine-tuned ResNet50 model under cross-validation. Although variations in validation performance were observed, these fluctuations are expected given the characteristics of the dataset. Among the three folds, Fold 2 exhibited the most balanced and stable convergence patterns, further justifying its selection for final model training and evaluation.

6.5.4 Cross-Validation Findings and Fold Selection

This subsection consolidates the findings derived from the three-fold cross-validation analysis and explains the rationale behind selecting a representative fold for subsequent model training and evaluation. Overall, the fine-tuned ResNet50 model demonstrated stable and consistent performance across the three folds, as reflected by the evaluation metrics, ROC curve analysis, and accuracy–loss convergence behavior.

The cross-validation results revealed expected variations among folds. Fold 1 achieved the highest overall accuracy and AUC, indicating strong discriminative capability, while Fold 3 exhibited higher recall for malignant lesions, highlighting increased sensitivity to clinically critical cases. In contrast, Fold 2 presented a more balanced performance profile, achieving competitive accuracy, precision, recall, and F1-score, along with stable ROC characteristics.

Moreover, the training dynamics observed in Fold 2 showed smoother validation accuracy and loss curves with fewer fluctuations compared to the other folds, suggesting improved stability and generalization. Considering both quantitative performance and learning behavior, Fold 2 was selected as the representative fold. This choice reflects a balanced trade-off between overall classification performance and sensitivity to malignant lesions and serves as the basis for the final training and evaluation stages discussed in the following section.

6.6 Final Training Analysis

This section presents a comprehensive analysis of the final training stage of the proposed DermAI system. After completing the cross-validation process and selecting the most representative fold, the model was retrained using the full dataset to maximize learning from all available samples. The analysis focuses on quantitative performance evaluation, error distribution, decision threshold optimization, and visual interpretability using Grad-CAM. Together, these results provide a holistic assessment of the final model's reliability and clinical applicability.

6.6.1 Final Results Analysis

The final DermAI model was evaluated using both the validation and independent test sets to assess its classification performance under realistic operational conditions. The reported results, summarized in Table 6.3 and Figure 6.9, include accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), which are widely adopted metrics in medical image classification tasks.

Test Set Performance		Validation Set Performance	
Metric	Value	Metric	Value
Accuracy	83.70%	Accuracy	83.38%
Precision (Malignant)	74.80%	Precision (Malignant)	73.76%
Recall (Malignant)	73.59%	Recall (Malignant)	74.24%
F1-Score	74.19%	F1-Score	74.00%
AUC-ROC	91.44%	AUC-ROC	91.16%

Table 6.3: Final Performance Metrics of the DermAI Model on Validation and Test Sets.

This table summarizes the final classification performance of the DermAI model on the validation and independent test datasets, reported in terms of accuracy, precision, recall, F1-score, and AUC-ROC. The presented metrics reflect the model's discriminative capability and generalization performance under realistic evaluation conditions.

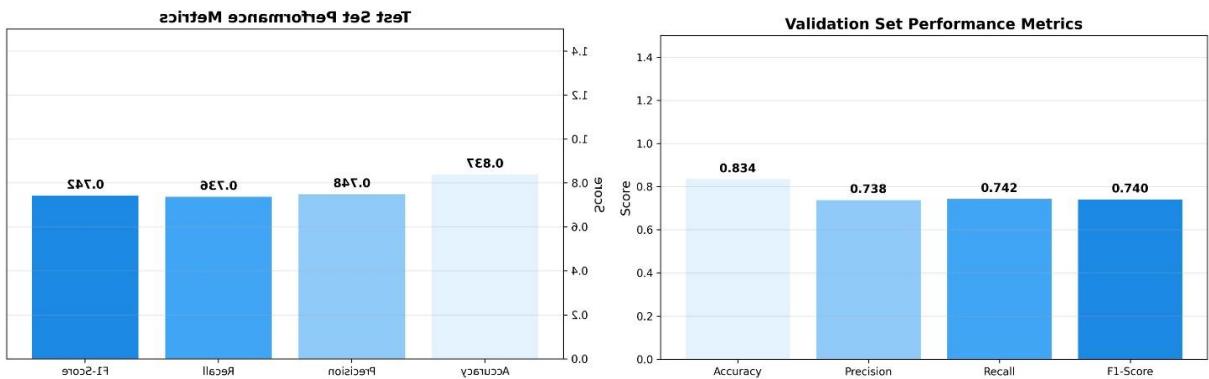


Figure 6.9: Comparison of final performance metrics (Accuracy, Precision, Recall, and F1-score) for the DermAI model on the validation and independent test sets.

On the validation set, the model achieved an accuracy of 0.834, with a precision of 0.738, recall of 0.742, and an F1-score of 0.740. These results indicate balanced classification behavior, reflecting the model’s ability to correctly identify malignant lesions while maintaining a controlled rate of false predictions. The corresponding AUC value of 0.912 demonstrates strong discriminative capability between benign and malignant skin lesions across varying decision thresholds.

Evaluation on the independent test set yielded comparable performance, with an accuracy of 0.837, precision of 0.748, recall of 0.736, and an F1-score of 0.742. The obtained AUC value of 0.914 further confirms the model’s robustness and its capacity to generalize effectively to previously unseen data.

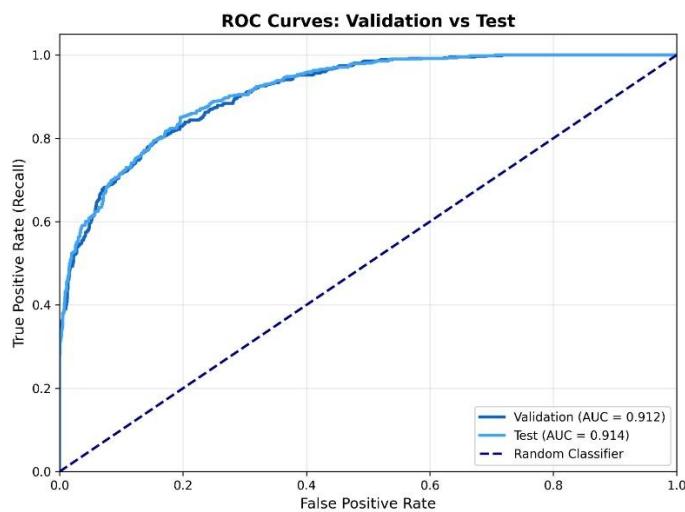


Figure 6.10: ROC curves of the final trained DermAI model evaluated on the validation and independent test sets.

The figure illustrates the trade-off between the true positive rate and false positive rate across different decision thresholds.

The close alignment between validation and test metrics suggests stable generalization behavior and indicates that the model's predictive performance is not dependent on a specific data split. From a clinical perspective, the achieved recall values are particularly significant, as they reflect the model's effectiveness in identifying malignant lesions, which is a critical requirement in skin cancer screening systems where minimizing missed malignant cases is essential.

Overall, the reported results demonstrate that the final trained DermAI model delivers reliable, consistent, and well-balanced performance across all evaluation metrics, supporting its suitability for deployment as a decision-support tool for automated skin lesion classification.

6.6.2 Confusion Matrix Interpretation

Figure 6.11 presents the confusion matrices corresponding to the validation and independent test sets, providing a detailed breakdown of the classification outcomes achieved by the final DermAI model.

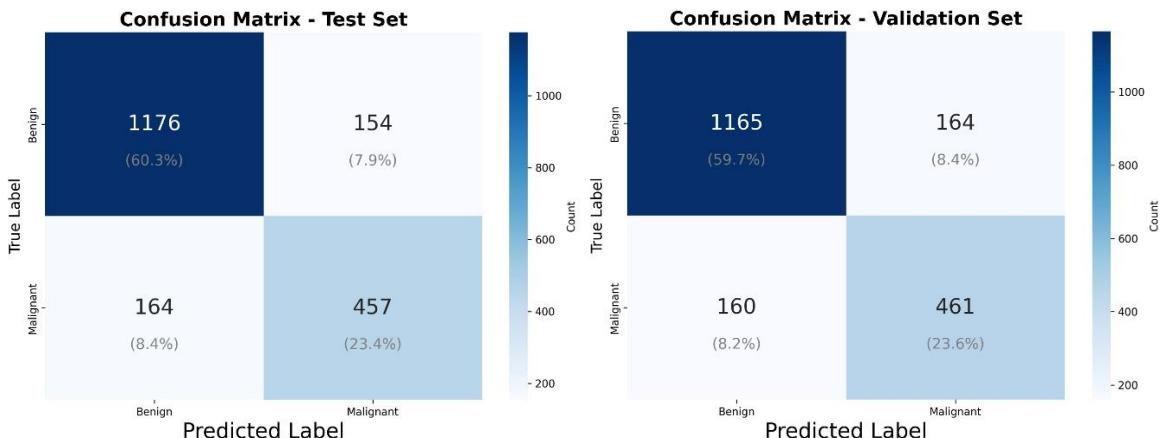


Figure 6.11: Confusion Matrices of the Final DermAI Model on Validation and Test Sets.

On the test set, the model correctly classified 1,176 benign and 457 malignant samples. Misclassifications included 154 benign lesions incorrectly predicted as malignant (false positives) and 164 malignant lesions incorrectly classified as benign (false negatives). These results indicate a strong capability to correctly identify malignant cases while maintaining a controlled false positive rate.

A similar distribution is observed on the validation set, where 1,165 benign and 461 malignant lesions were correctly identified, alongside 164 false positives and 160 false negatives. The close similarity between validation and test confusion matrices reflects consistent classification behavior and stable generalization performance.

From a clinical screening perspective, the achieved true positive rate for malignant lesions is particularly significant, as it demonstrates the model's effectiveness in detecting clinically relevant cases while preserving acceptable specificity. Overall, the confusion matrix analysis further supports the reliability and robustness of the final trained model.

6.6.3 Threshold Optimization for the Final Model

To further analyze the classification behavior of the proposed DermAI model, a threshold optimization study was conducted on the validation set. Instead of relying solely on the default decision threshold of 0.5, multiple thresholds ranging from 0.35 to 0.55 were evaluated to examine their impact on key performance metrics, including accuracy, precision, recall, and F1-score.

Table 6.4 presents a comparative summary of the evaluation results obtained at each tested threshold. As expected, lower thresholds (e.g., 0.35 and 0.40) resulted in substantially higher recall values, indicating improved sensitivity to malignant cases; however, this was achieved at the expense of precision and overall accuracy due to an increased number of false positive predictions. Conversely, higher thresholds (e.g., 0.55) improved precision but significantly reduced recall, increasing the risk of missing malignant lesions.

Threshold	Accuracy	Precision	Recall	F1
0.35	0.7779487179	0.6044444444	0.876006412	0.7153188692
0.4	0.7876932077	0.6260657734	0.8276972625	0.7128987517
0.45	0.8030796231	0.6569536424	0.7987117552	0.7209303236
0.5	0.8117948718	0.6862170088	0.7536231884	0.7536231884
0.55	0.8220512821	0.7245901639	0.7117552335	0.7117552335
0.55	0.8220512821	0.7245901639	0.7117552335	0.7181153534

Table 6.4: Performance comparison of the DermAI model on the validation set across different decision thresholds, illustrating the impact of threshold selection on accuracy, precision, recall, and F1-score.

The optimal threshold was selected based on the F1-score, as it provides a balanced measure between precision and recall. Among all tested values, a threshold of 0.5 achieved the highest F1-score on the validation set, while also maintaining strong accuracy and balanced class-wise performance. This confirms that the default threshold represents an effective trade-off between detecting malignant lesions and limiting false alarms.

Figure 6.12 illustrates the relationship between the decision threshold and the evaluated performance metrics. The plots clearly demonstrate the trade-offs involved in threshold selection, with the vertical dashed line highlighting the optimal threshold. Notably, while minor variations in accuracy and precision were observed across thresholds, the F1-score peaked at the selected value, supporting its suitability for final deployment.

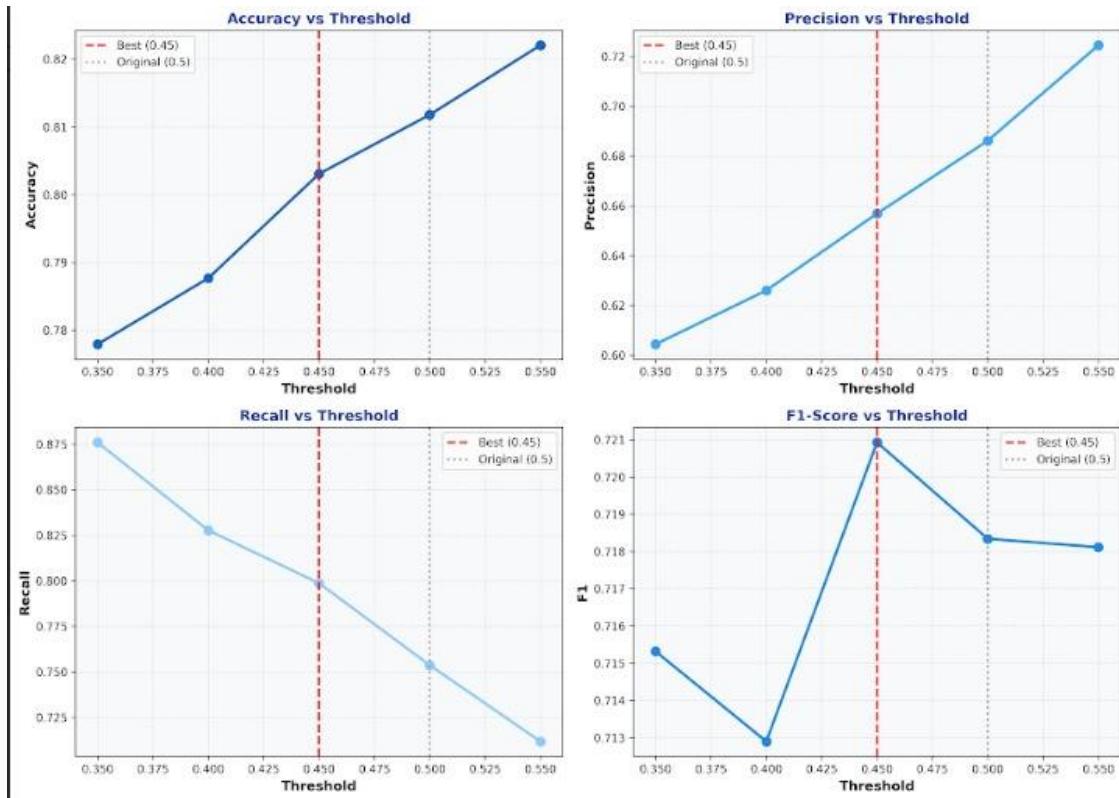


Figure 6.12: Effect of decision threshold variation on accuracy, precision, recall, and F1-score for the validation set. The dashed red line indicates the optimal threshold selected based on the maximum F1-score.

The optimized threshold was subsequently applied to the independent test set. The resulting performance metrics remained unchanged compared to the original configuration, indicating consistent generalization behavior and confirming that the chosen threshold does not introduce overfitting or instability. From a clinical perspective, maintaining balanced recall and precision is critical, as it ensures reliable

detection of malignant lesions while minimizing unnecessary false positive classifications.

Overall, the threshold optimization analysis validates the robustness of the final DermAI model and supports the use of a threshold of 0.5 as an optimal and clinically appropriate decision boundary for automated skin lesion classification.

6.6.4 Grad-CAM Visual Explanation Analysis

To provide visual insight into the final model's decision-making process, Grad-CAM visual explanations were generated for selected test samples.

Figure 6.13 presents a representative Grad-CAM visualization overlaid on a dermoscopic image, highlighting the regions that contributed most strongly to the model's prediction.

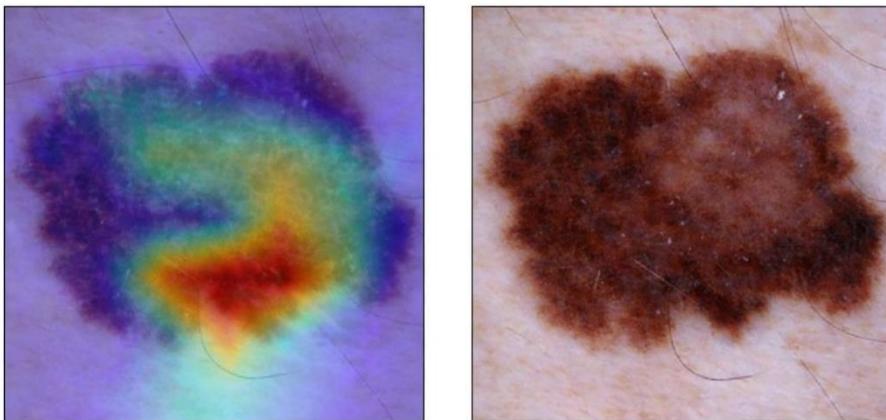


Figure 6.13: Example of Grad-CAM visualization showing the original dermoscopic image (right) and the corresponding heatmap (left), highlighting lesion regions that contributed most to the model's malignant classification decision.

As illustrated, the model primarily focuses on clinically relevant areas of the lesion, including regions of irregular pigmentation and internal structural patterns, rather than background artifacts. This behavior

suggests that the model's predictions are guided by meaningful lesion characteristics, supporting the reliability of the learned representations.

Although a limited number of Grad-CAM examples are presented, the observed activation patterns align with dermatological diagnostic cues, providing qualitative confirmation of the model's interpretability and reinforcing its suitability for clinical decision-support applications.

6.7 Discussion of Results

The results obtained in this study indicate that the final DermAI model achieved consistent and dependable performance within the adopted experimental methodology. The similarity between validation and test outcomes reflects good generalization ability and suggests that the model is not overly sensitive to data partitioning.

Clinically, the recall achieved for malignant lesion detection is a key strength of the proposed system, as reducing missed cancer cases is essential in skin cancer screening. At the same time, the obtained precision and F1-score demonstrate a balanced prediction behavior, avoiding excessive false alarms.

The ROC curve analysis further confirms the model's strong discriminative performance, as evidenced by high AUC values across both validation and test sets. Moreover, Grad-CAM visualizations qualitatively support that the model focuses on lesion-related regions rather than irrelevant background features, enhancing confidence in its decision-making process.

6.8 Chapter Summary

This chapter evaluated the final performance of the DermAI skin lesion classification system. The results showed stable accuracy, balanced classification metrics, and strong discriminative capability across different evaluation sets.

The inclusion of Grad-CAM provided interpretability by highlighting clinically relevant image regions, supporting transparency in model predictions. It is emphasized that DermAI is designed as a supportive screening tool and not as a replacement for professional medical diagnosis.

In summary, the presented results demonstrate the effectiveness of the proposed system and its potential applicability in assisting automated skin lesion assessment.

CHAPTER 7

SYSTEM USER INTERFACES AND FUNCTIONAL WORKFLOW

This chapter presents the user interfaces of the DermAI system and explains the functional workflow followed by the user while interacting with the platform. It highlights the main system screens and describes how each interface supports image submission, analysis, and result visualization within the overall system operation.

7.1 Home and User Entry Interface

This interface represents the Home Page of the DermAI system, allowing users to sign up or sign in, while providing a brief overview of the system's concept and its objective of early skin cancer detection using artificial intelligence techniques.

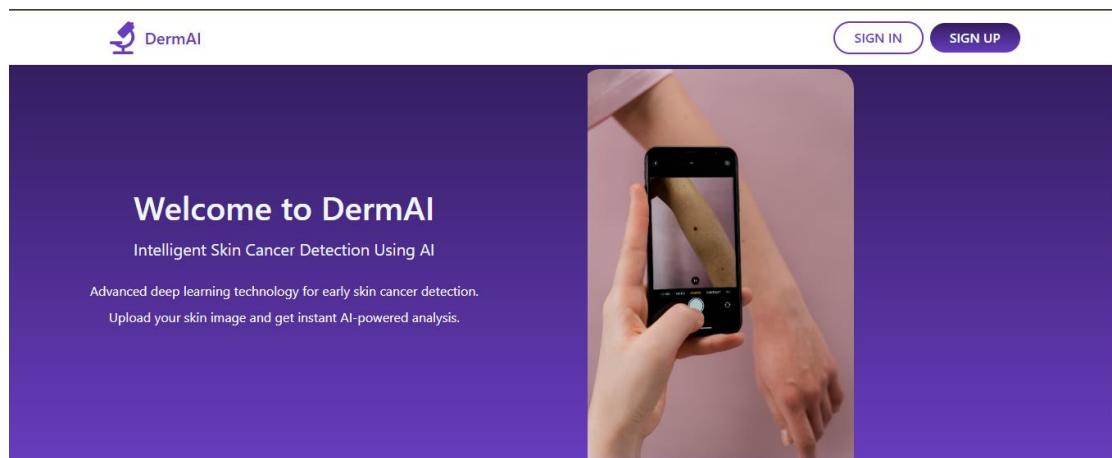


Figure 7.1: Home page DermAI.

7.2 User Authentication and Account Management

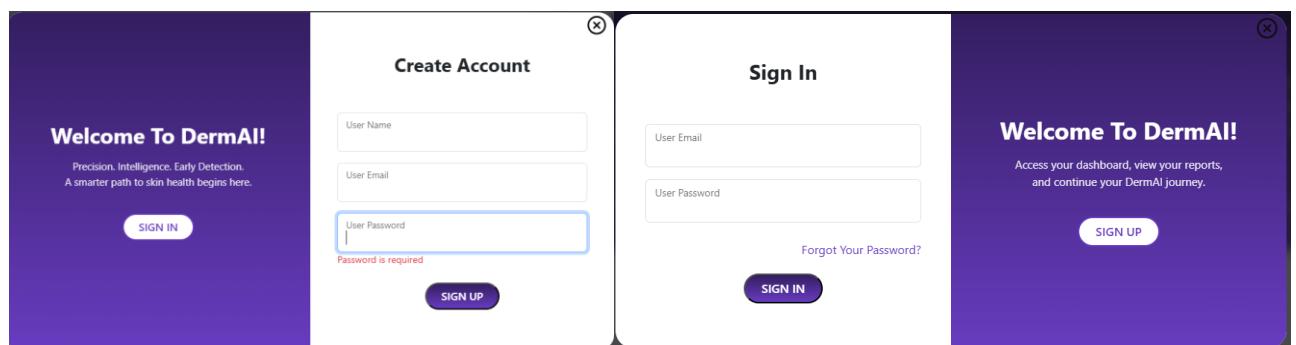


Figure 7.2: Sing-in and create account in DermAI.

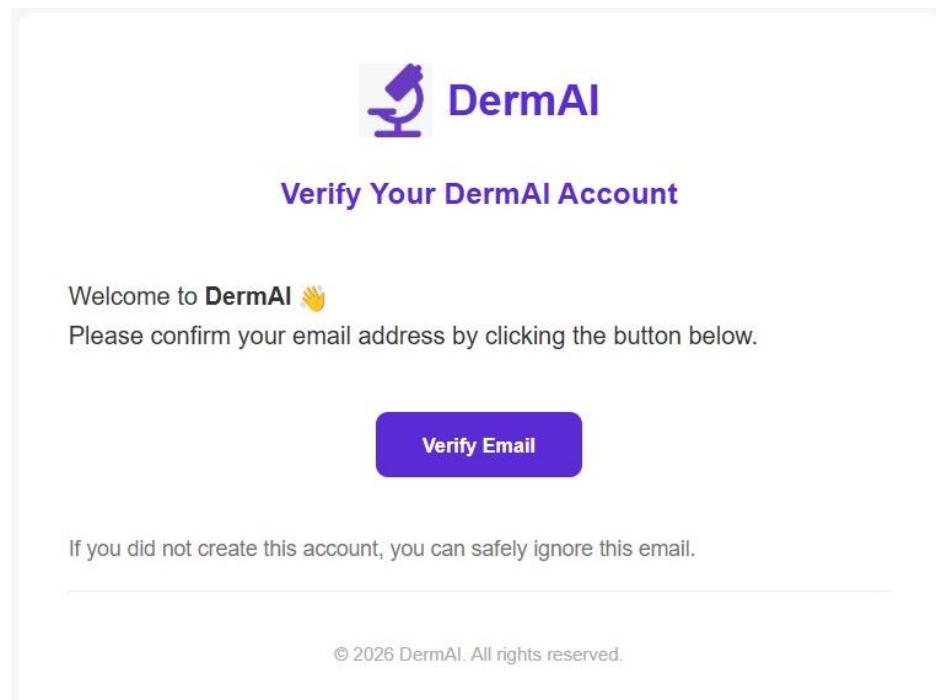


Figure 7.3: Once an account is created, this email will be sent to confirm the account..

This interface is displayed when the user selects the Forgot Password option. It allows the user to enter the same email address used during account registration, after which a password reset link is sent to the user's email to initiate the password recovery process.

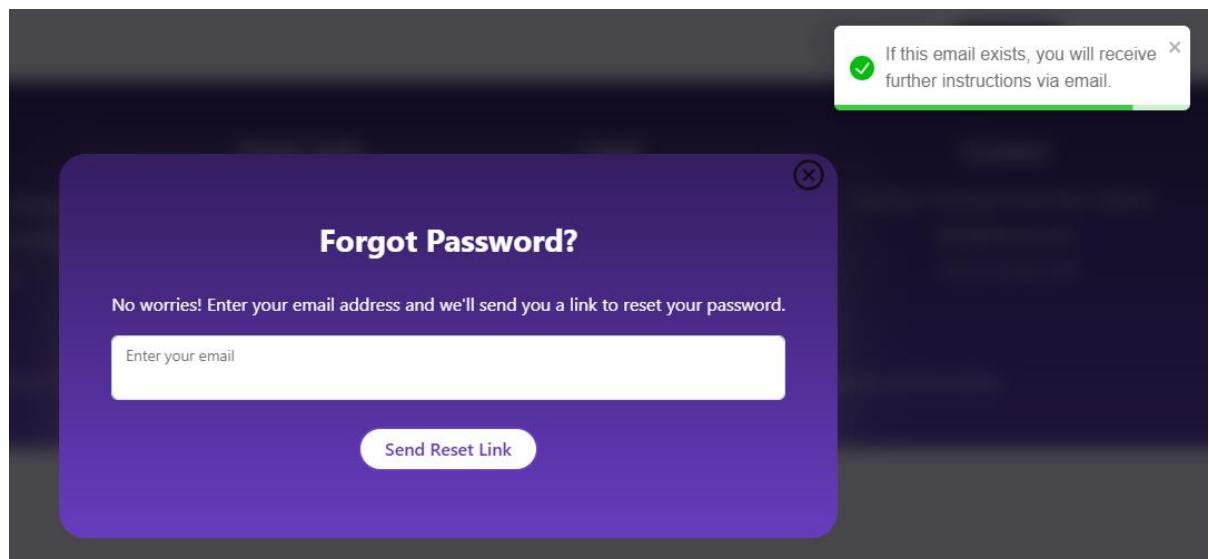


Figure 7.4: Forgot password in DermAI.

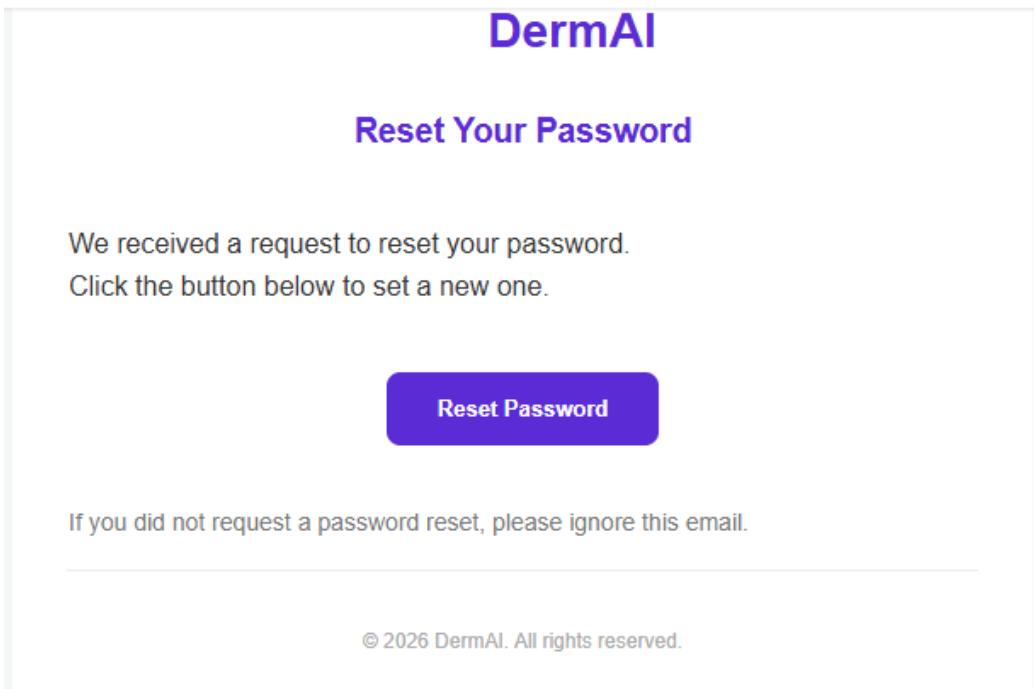


Figure 7.5: The message that the user receives when they log into their email to reset their password.

This interface appears after the user clicks the password reset link sent to their email. The user is required to enter the new password twice to confirm its correctness, and the system validates the inputs to ensure they match and meet the required password criteria before completing the reset process.

A screenshot of the "Reset password" page. It features a lock icon with a circular arrow above the title "Reset password". Below the title is the instruction "Please kindly set your new password." Two input fields are present: "New password" and "Re-enter password". The "New password" field contains "....." and has a red validation message: "Must contain at least one lowercase letter". The "Re-enter password" field contains ".....|" and has a red validation message: "Passwords do not match". A large purple "Reset Password" button is at the bottom.

Figure 7.6: Reset Password Page in DermAI.

7.3 Skin Image Upload Interface

This interface allows the user to upload a skin lesion image for analysis through the DermAI system. The user can capture an image directly using the camera or select an existing image from the device files on a computer or laptop. Additionally, the interface provides guidelines to ensure image quality, which contributes to more accurate analysis results.

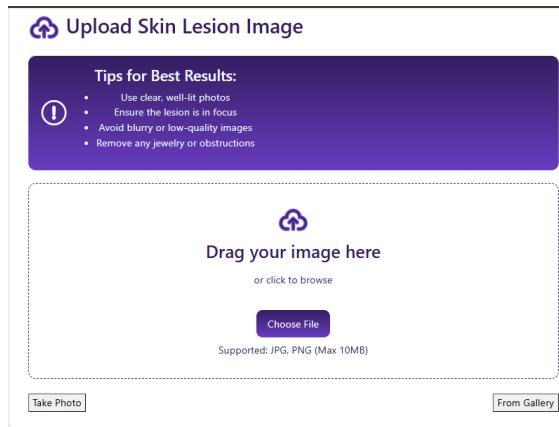


Figure 7.7: Upload image page in DermAI.

7.4 AI Analysis Result Interface

This interface presents the analysis result after the user uploads a skin lesion image. It displays the classification outcome as benign or malignant along with the confidence percentage generated by the AI model. The interface also includes three action buttons: the first to view the heatmap highlighting the regions of interest, the second to download a detailed report containing the uploaded image, analysis results, and heatmap visualization, and the third to display nearby dermatology clinics based on the user's location.

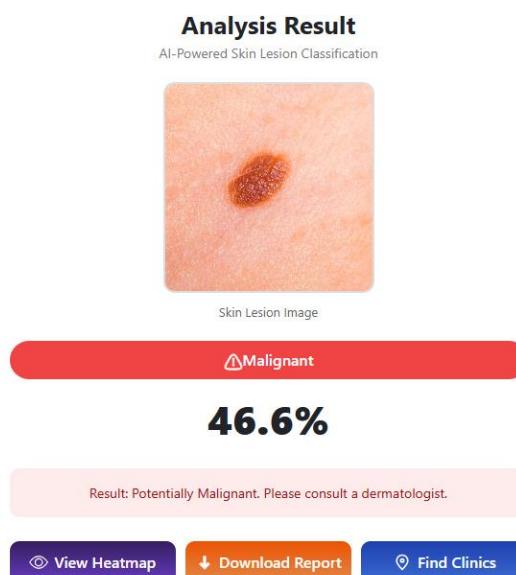


Figure 7.8: Analysis Result page in DermAI.

7.5 Model Explainability and Visualization (Grad-CAM)

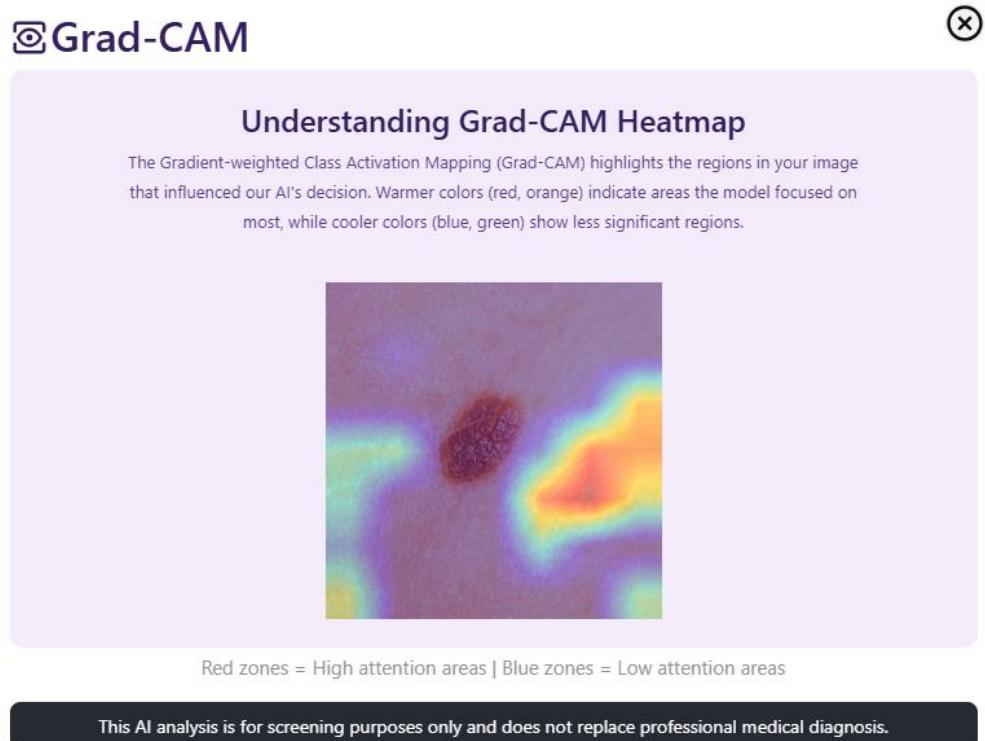


Figure 7.9: Grad-CAM Interface in DermAI.

7.6 Report Generation and Download

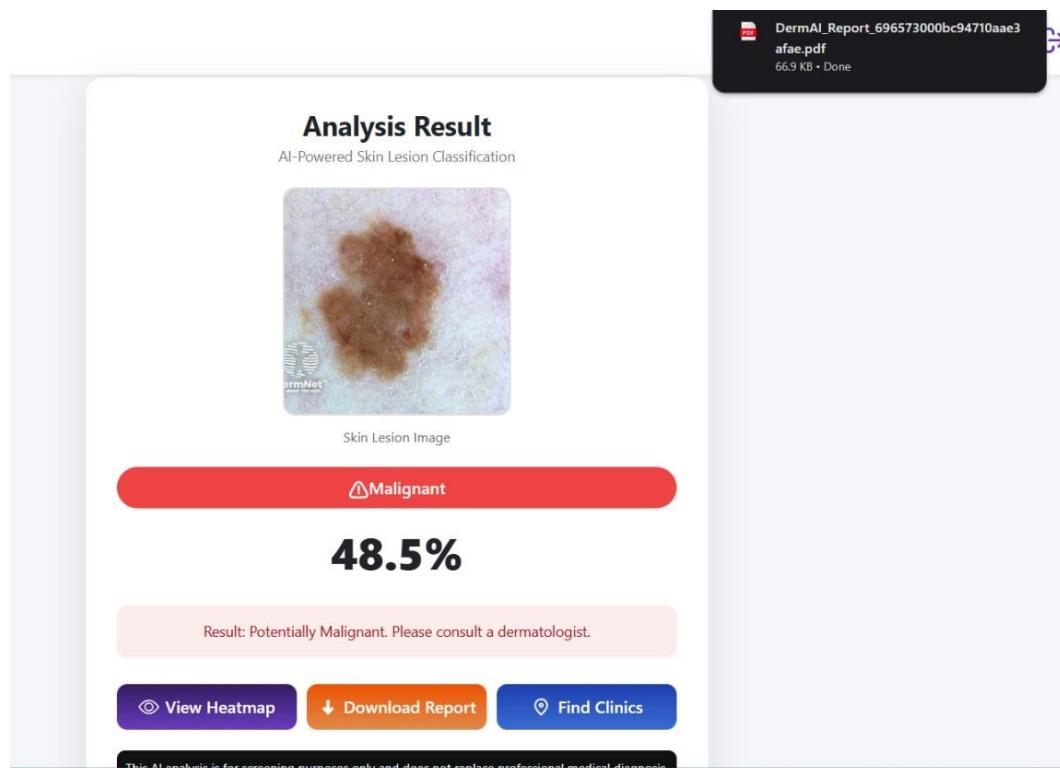


Figure 7.10: Clicking the "Download Report" button will directly download the PDF file to your device.

7.7 Nearby Clinics Locator Interface

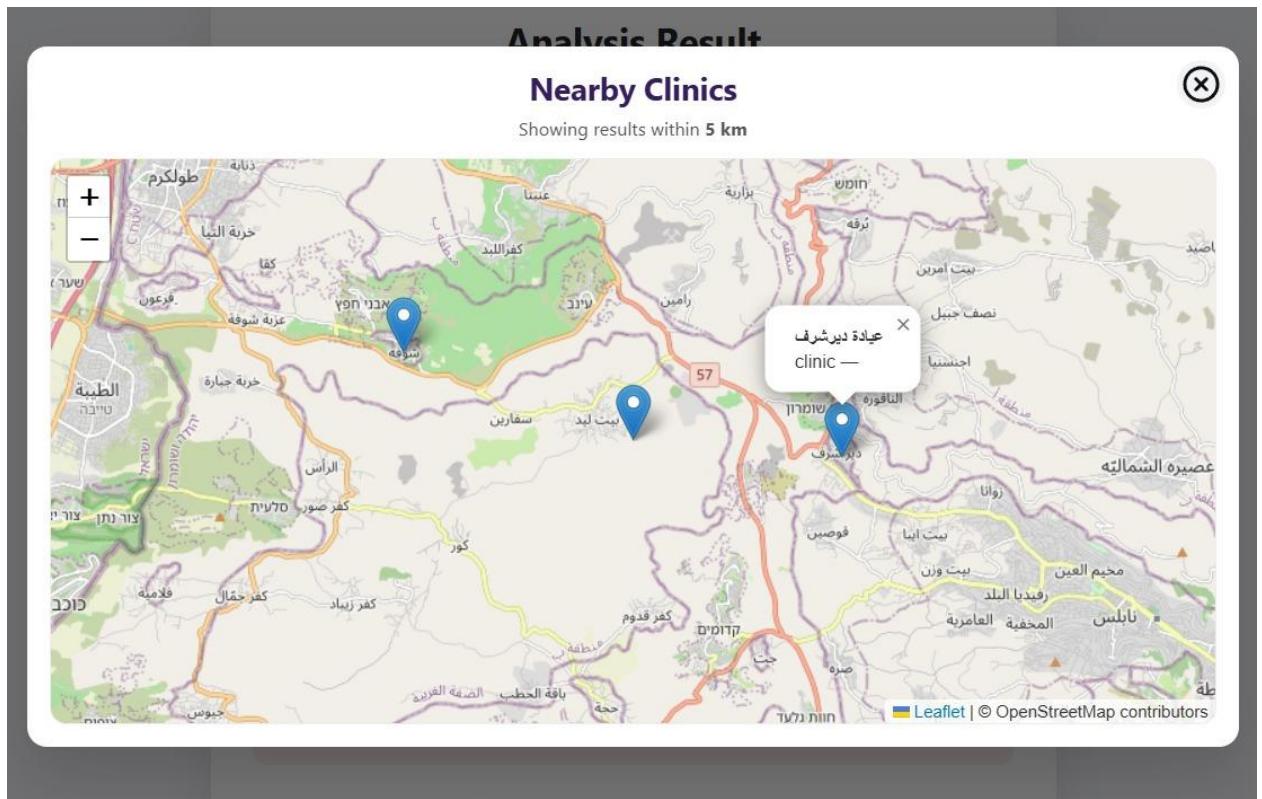


Figure 7.11: When you click the "Find Clinics" button, a map will appear showing the locations of clinics near the user.

CHAPTER 8

FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS

8.1 Introduction

This chapter presents the functional and non-functional requirements of the DermAI system as implemented in the final version of the project. The requirements were derived directly from the actual frontend and backend implementation, reflecting the real system behavior rather than preliminary design assumptions.

8.2 Functional Requirements

Functional requirements describe the system functionalities that were implemented and made available to end users.

8.2.1 User Authentication and Account Management

- The system allowed users to create a personal account using an email address and password.
- Passwords were validated using strong security rules, including minimum length, uppercase and lowercase letters, numeric characters, and special symbols.
- The system implemented a secure login mechanism using JSON Web Tokens (JWT).
- Protected routes were enforced to prevent unauthorized access to sensitive pages such as image upload and result viewing.
- A complete *Forgot Password* workflow was implemented, allowing users to request a password reset link via email and securely set a new password using a token-based validation mechanism.
- Email-based account verification was supported to ensure account authenticity.

8.2.2 Skin Lesion Image Upload

- The system enabled authenticated users to upload skin lesion images in standard formats such as JPG and PNG.
- Users were able to upload images from local storage or capture images directly using the device camera.
- Input validation was applied to ensure that only valid image files were accepted.
- The interface provided clear instructions to guide users in uploading high-quality images suitable for analysis.

8.2.3 AI-Based Skin Lesion Analysis

- The system applied a trained convolutional neural network (CNN) model to analyze the uploaded skin lesion images.
- Each image was classified into one of two categories:
 - Benign
 - Malignant
- The analysis result included a confidence score expressed as a percentage to indicate the certainty of the model's prediction.
- The analysis process was executed asynchronously to ensure a smooth user experience.

8.2.4 Grad-CAM Heatmap Visualization

- The system generated a Grad-CAM heatmap for each analyzed image.
- The heatmap visually highlighted the regions of the image that most influenced the model's decision.
- The Grad-CAM result was presented in a modal window to improve interpretability while maintaining a clean user interface.
- This feature enhanced transparency and supported user trust in the AI-generated results.

8.2.5 Analysis Result Presentation

- The system displayed a detailed analysis results page containing:
 - The uploaded skin lesion image
 - The predicted classification (Benign or Malignant)
 - The confidence score
 - A short explanatory message based on the classification outcome
- Clear visual indicators (color-coded labels and icons) were used to distinguish between benign and malignant cases.
- A medical disclaimer was included to emphasize that the result was for screening purposes only.

8.2.6 Diagnostic Report Generation

- The system generated a downloadable PDF diagnostic report for each analysis.
- The report included:
 - The uploaded image
 - The classification result
 - The confidence score
 - The Grad-CAM heatmap visualization
- The report was generated on demand and downloaded directly to the user's device.

8.2.7 Location-Based Clinic Recommendation

- The system provided a feature to locate nearby dermatology clinics based on the user's geographical location.
- The user's location was accessed only after explicit permission was granted through the browser.
- An interactive map was displayed showing the user's location and nearby clinics using map markers.
- This feature assisted users, especially in malignant cases, in identifying nearby medical facilities.

8.2.8 Error Handling and User Feedback

- The system provided real-time feedback for user actions such as login, registration, image upload, and analysis.
- Toast notifications and alert messages were used to communicate success or failure events.
- A custom 404 (Not Found) page was implemented to handle invalid routes gracefully.

8.3 Non-Functional Requirements

Non-functional requirements describe how the system performed in terms of security, usability, performance, and reliability.

8.3.1 Security and Privacy

- All communication between the client and server was secured using HTTPS.
- Authentication tokens were securely stored and validated for protected routes.
- User passwords were never stored in plain text and were handled securely by the backend.

- User location data was accessed only with explicit consent and used solely for clinic recommendation functionality.
- Legal pages such as Privacy Policy, Terms of Service, and Medical Disclaimer were implemented to ensure transparency and ethical compliance.

8.3.2 Usability

- The user interface was designed to be intuitive and user-friendly for non-technical users.
- Clear instructions and visual guidance were provided throughout the system.
- Responsive design principles were applied to ensure compatibility across desktops, tablets, and mobile devices.
- Loading indicators were implemented to enhance user experience during asynchronous operations.

8.3.3 Performance

- The system returned AI analysis results in near real-time, typically within a few seconds after image submission.
- Frontend performance was optimized using component-based architecture and state management.
- Asynchronous API calls were used to prevent UI blocking during analysis and data fetching.

8.3.4 Reliability and Availability

- The system handled errors gracefully without causing application crashes.
- Fallback mechanisms were applied for missing data, invalid states, or failed requests.
- Route protection ensured consistent system behavior even after page refreshes.

8.3.5 Maintainability

- The project followed a modular structure separating concerns between components, pages, contexts, and services.
- Reusable components were implemented to reduce redundancy and simplify future updates.
- The codebase was structured to support future feature expansion, such as enhanced analytics or administrative functionality.

8.4 Conclusion

This chapter presented the implemented functional and non-functional requirements of the DermAI system based on the final application code. The system successfully delivered a complete AI-powered skin lesion analysis workflow, including secure authentication, image analysis, explainable AI visualization, report generation, and clinic recommendations. The implemented requirements demonstrate that DermAI is a robust, secure, and user-centered platform that provides practical support for early skin cancer screening while maintaining ethical and usability standards.

CHAPTER 9

FUTURE WORK AND SYSTEM ENHANCEMENTS

Although the DermAI system was successfully implemented and demonstrated its ability to analyze skin lesion images and provide AI-assisted classification results, several enhancements and extensions can be considered for future development. These improvements aim to increase system scalability, usability, reliability, and real-world applicability by expanding both the frontend and backend functionalities, as well as strengthening the integration with advanced AI techniques.

Future Work for Front-End and Back-End Development

One of the most important future enhancements involves the introduction of a *dedicated administrative dashboard* within the web platform. While the current system focuses primarily on end-user interaction, an admin panel was originally planned but not implemented due to time constraints. This dashboard could provide administrators with the ability to monitor system usage, manage registered users, review uploaded images, and inspect analysis statistics. Additionally, administrators could track system performance metrics, error logs, and usage trends, which would contribute to improved system maintenance and decision-making.

Another significant enhancement is the development of a *personal user profile page*. In its current implementation, users can upload images, view results, and download reports; however, these reports are not persistently organized under a dedicated user space. A future version of DermAI could allow users to access a personal profile where all previously generated diagnostic reports are stored and categorized by date. This feature would enable users to review past analyses, compare results over time, and maintain a personal diagnostic history, thereby improving long-term usability and user engagement.

From a frontend perspective, future work may include further *user experience (UX) and accessibility improvements*. These enhancements could involve advanced form validation feedback, more interactive result visualizations, multilingual support, and improved accessibility features for users with disabilities. Additionally, enhanced UI customization and theming options could be introduced to improve overall usability across diverse user groups.

On the backend side, future improvements could focus on *scalability and performance optimization*. This includes introducing advanced caching mechanisms, background job processing for report generation, and more efficient file storage solutions for uploaded images and generated reports. Furthermore, the backend architecture could be extended to support role-based access control (RBAC), enabling differentiated permissions for administrators, medical professionals, and standard users.

Another potential enhancement is the integration of *structured logging and analytics services*, which would allow detailed monitoring of system behavior, API performance, and user interactions. Such data could be used to improve system reliability, detect anomalies, and support future research and system optimization.

Integration of Advanced AI Features and OOD Detection

In addition to web-based enhancements, an important direction for future work involves strengthening the AI component of the system through the integration of *Out-of-Distribution (OOD) detection mechanisms*. In real-world scenarios, the system may receive images that do not represent skin cancer cases, such as infections, inflammations, or other dermatological conditions not included in the training dataset. In the current implementation, the model is limited to binary classification (benign or malignant), which may lead to misleading predictions when unfamiliar inputs are provided.

By incorporating OOD detection techniques, the system could identify inputs that fall outside the model's learned distribution and avoid producing overconfident predictions. Instead of forcing a benign or malignant classification, the system could display a warning message indicating that the image is not suitable for automated analysis and recommend professional medical consultation. From a frontend perspective, this enhancement could be reflected through dedicated warning interfaces or result states, while the backend would handle confidence thresholding, uncertainty estimation, or auxiliary OOD models.

Integrating OOD detection would significantly improve the safety, reliability, and ethical alignment of the system, reinforcing the concept of DermAI as an *AI-assisted preliminary screening tool rather than a diagnostic replacement*. This enhancement would also make the system more robust and suitable for real-world deployment, where input variability is unavoidable.

Extended Clinical and System-Level Enhancements

Future versions of DermAI could also include deeper integration with healthcare workflows, such as allowing medical professionals to securely review shared reports or enabling optional consultation requests through verified clinics. Additionally, the system could support model versioning and performance tracking to ensure transparency when AI models are updated or retrained.

Overall, these future enhancements would transform DermAI from a standalone screening platform into a more comprehensive, scalable, and clinically supportive system. By combining advanced AI techniques with expanded frontend and backend functionalities, future iterations of DermAI could significantly enhance its impact, usability, and readiness for real-world medical environments.

CHAPTER 10

REFERENCES AND SOURCES

Abdel-Hamid, A., Soliman, A., & Elsayed, R. (2023). Artificial intelligence-driven enhanced skin cancer diagnosis: Leveraging convolutional neural networks with discrete wavelet transformation. *Journal of Imaging*, 10(12), 332. <https://doi.org/10.3390/jimaging10120332>

Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. *JAMA Dermatology*, 154(11), 1247–1248. <https://doi.org/10.1001/jamadermatol.2018.2348>

Ahmed, M., Alamri, A., & Alghamdi, R. (2023). Explainable AI-based skin cancer detection using CNN, particle swarm optimization and machine learning. *Journal of Imaging*, 10(12), 331. <https://doi.org/10.3390/jimaging10120331>

Al-Sarraf, Z., Al-Sarraf, H., & Jassim, A. H. (2024). Artificial intelligence in skin cancer diagnosis: A reality check. *Journal of Modern Human Genetics*, 5(1), Article 5. <https://doi.org/10.1186/s43042-024-00522-5>

American Cancer Society. (2023). Cancer Facts & Figures 2023. <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html>

Behara, K., Bhero, E., & Agee, J. T. (2024). AI in dermatology: A comprehensive review into skin cancer detection. *PeerJ Computer Science*, 10, e2530. <https://doi.org/10.7717/peerj-cs.2530>

Bonnett, M., Kennedy, M., Okala, O., & Garstka, T. (2024). Precision public health: Empowering communities with hyperlocal data for targeted interventions and improved outcomes. Semantic Scholar. <https://api.semanticscholar.org/CorpusID:270160142>

Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., ... von Kalle, C. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer*, 118, 91–96. <https://doi.org/10.1016/j.ejca.2019.01.013>

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT)*. <https://proceedings.mlr.press/v81/buolamwini18a.html>

Codella, N. C. F., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S. W., Gutman, D., ... Halpern, A. (2019). Skin lesion analysis toward melanoma detection 2018: A

challenge hosted by the International Skin Imaging Collaboration (ISIC). arXiv preprint.
<https://arxiv.org/abs/1902.03368>

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>

Garbe, C., & Leiter, U. (2019). Melanoma epidemiology and trends. *Clinics in Dermatology*, 35(1), 3–7. <https://doi.org/10.1016/j.clindermatol.2016.06.003>

Haenssle, H. A., et al. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition. *Annals of Oncology*, 29(8), 1836–1842.

<https://academic.oup.com/annonc/article/29/8/1836/5046152>

Han, S. S., et al. (2020). Assessment of deep neural networks for the diagnosis of skin cancer. *JAMA Network Open*, 3(5), e205324.

<https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2769429>

Jones, M., Bradley, J., & Sakimura, N. (2015). JSON Web Token (JWT) (RFC 7519).

<https://datatracker.ietf.org/doc/html/rfc7519>

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019).

Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17, 195. <https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1426-2>

Liu, Y., et al. (2020). A deep learning system for differential diagnosis of skin diseases. *Nature Medicine*, 26(6), 900–908. <https://www.nature.com/articles/s41591-020-0842-3>

MongoDB Inc. (2024). MongoDB Documentation.

<https://www.mongodb.com/docs/>

Mozilla. (2023). Cross-origin resource sharing (CORS). MDN Web Docs. <https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS>

Ramírez, S. (2023). FastAPI Documentation.

<https://fastapi.tiangolo.com/>

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. <https://arxiv.org/abs/1602.04938>

Selvaraju, R. R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV). <https://arxiv.org/abs/1610.02391>

Siegel, R. L., Miller, K. D., & Jemal, A. (2022). Cancer statistics, 2022. CA: A Cancer Journal for Clinicians, 72(1), 7–33. <https://doi.org/10.3322/caac.21708>

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (ICML). <https://arxiv.org/abs/1905.11946>

Tan, S.-T., Chong, Y., & Tan, J. (2023). Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. npj Digital Medicine, 6(1), 17. <https://doi.org/10.1038/s41746-023-00914-8>

Tschandl, P., Codella, N., & Winkler, J. K. (2020). Human–computer collaboration for skin cancer recognition. Nature Medicine, 26(8), 1229–1234. <https://doi.org/10.1038/s41591-020-0942-0>

DermAI project GitHub link:

https://github.com/DermAI-Project/DermAI_Graduation-Project