EAST WEST UNIVERSITY

Department of Computer Science and Engineering

Dermalytics: Skin Disease and Skin Cancer Detection using Machine Learning (ML)

Course Code: CSE400A

Submitted to:

Dr. Md Mostofa Kamal Rasel

Associate Professor

Submitted By:

Name	ID
Samia Jahan Mariam	2022-1-60-376
Khandhakar Shatu Moni	2021-3-60-119
Md. Sazzad Hossain	2021-3-60-112

Date of Submission: 02.07.2025

Dermalytics: Skin Disease And Skin Cancer Detection Using Machine Learning(ML)

Abstract

Skin-related disorders represent a significant global health challenge, with conditions like skin cancer posing serious risks due to their potential for metastasis and mortality if not detected early. Other common skin diseases—such as eczema, psoriasis, rosacea, and fungal infections—also lead to discomfort and frequent misdiagnosis due to overlapping visual symptoms. This project proposes a machine learning-based diagnostic system that uses image processing and deep learning, specifically Convolutional Neural Networks (CNNs), to automatically detect and classify various skin conditions. A curated dataset of annotated dermoscopic and clinical images undergoes preprocessing steps such as resizing, normalization, and augmentation to improve model performance. The trained model is evaluated using standard metrics including accuracy, precision, recall, and F1-score. To ensure practical usability, the system is deployed as both a local offline app for privacy and a remote server-based app for advanced processing. This work aims to deliver an accurate, scalable, and accessible diagnostic tool that supports early detection, aids dermatologists, and expands access to dermatological care through AI-powered medical imaging.

Introduction

Skin diseases, including skin cancer, represent a major global health burden, affecting millions of people annually. These conditions range from minor infections and chronic inflammatory disorders to serious and potentially life-threatening diseases like melanoma. Among them, skin cancer—particularly melanoma, basal cell carcinoma, and squamous cell carcinoma—poses a significant threat due to its potential for rapid progression and high mortality rates if not detected early. Traditional diagnostic methods, such as visual inspections and biopsies, are often time-consuming, subjective, and highly dependent on the availability and expertise of trained dermatologists. This creates a critical challenge, especially in remote or underserved areas, where access to specialized medical care is limited or nonexistent. As the incidence of skin disorders continues to rise due to factors such as ultraviolet (UV) radiation exposure, environmental pollution, and genetic predisposition, there is a growing need for faster, more accurate, and widely accessible diagnostic solutions. In this context, machine learning (ML) and image processing technologies offer promising alternatives. ML algorithms can be trained on large datasets of dermoscopic and clinical skin images to learn patterns associated with various conditions such as melanoma, actinic keratosis, vascular lesions, dermatofibroma, nevus, and other common disorders like eczema, psoriasis, vitiligo, and fungal infections. These tools can help identify visual cues—such as discoloration, texture, and shape—that may not be easily noticeable to the human eye. This project aims to develop an intelligent diagnostic system that utilizes machine learning to automatically analyze skin lesion images and classify them into different skin disease categories. By integrating the trained model into a mobile or web-based application, the system will offer real-time diagnostic support for both healthcare professionals and individuals. The ultimate goal is to create a reliable, accessible, and efficient skin disease detection tool that enhances clinical decision-making and lays the foundation for teledermatology solutions, particularly in resource-limited environments

Related Work:

Here briefly discusses what various researches have discussed in the past about skin cancer and skin diseases, the aims of various research papers and the applied methodologies to achieve those aims, and the results obtained through the research.

The paper [1] addresses the problem of accurately detecting and classifying skin diseases using image processing and machine learning, particularly in areas with poor access to dermatologists. It achieves this by carrying out preprocessing, segmentation (via Otsu's method), and feature extraction (GLCM and LBP) on skin images, with disease classification done using the Naive Bayes algorithm. Among the tested classifiers—SVM, KNN, and Naive Bayes—the Naive Bayes model achieved the highest accuracy of 97.5%. The final result is found to be highly effective for multi-class skin disease classification with Naive Bayes, along with LBP and GLCM features. Future work entails the creation of this model as an Android application to promote ease of access for remote and disadvantaged populations.

This paper [2] meets the challenge of accurate diagnosis of skin illnesses, generally difficult due to subjective visual estimation and limited access to professionals. It proposes new machine learning techniques—namely CNN, SVM, and KNN—coupled with methodologies like attention mechanisms and data augmentation to enhance diagnostic performance. Results put CNN-based methods at the center stage, specifically based on ensemble learning like SkinGenius, with a best accuracy of 98.2%. The study points to the possibility of AI improving dermatological care, especially in rural locations. Future studies involve increasing dataset diversity, explaining AI decisions, and developing hybrid systems for broader, more generalizable diagnosis.

This paper [3] provides a rigorous review of deep learning approaches to early skin cancer detection from datasets like ISIC, PH2, DermIS, and DermQuest. It outperforms the weaknesses of traditional biopsy by cross-comparing models like ANN, CNN, KNN, and GAN. The maximum accuracy attained is 98.3% by a better KNN model, while the worst overall accuracy is 70.1% by a GAN model. The study adopts the power of deep learning to enhance the precision of diagnosis and remove human error. Future directions include perfecting data quality, dealing with class imbalance, and developing real-time diagnostic systems.

The study [4] addresses the problem of detection of skin disease using precise image processing and machine learning techniques. It fixes this by applying a CNN model to classify skin diseases from dermoscopic images. The outcome indicates that the model achieved an accuracy rate of 97% in various skin conditions, proving its competence. As future work, the paper suggests adding more to the system with a bigger, more diverse dataset, and even implementing the model into a mobile or web app for real-time usage.

The research [5] addresses the challenge of early and low-cost diagnosis of skin conditions, which are often misdiagnosed due to medical unaffordability or expensive diagnostic machinery. The authors solved this by developing an automated system that utilizes image processing with a pre-trained convolutional neural network (AlexNet) for feature extraction and Support Vector Machine (SVM) for classification. The final result proved a 100% detection rate of three skin conditions: eczema, melanoma, and psoriasis. The system requires merely a proofreader's camera and a computer, which makes it budget-friendly and realistic. Future work includes developing a mobile app, classifying lesions in deeper layers of skin, and expanding the model to detect all skin conditions globally. The work suggests an image processing and machine learning-based skin disease detection system, which is developed in MATLAB, but there is no public website or mobile application available to users.

Research Questions/Problem statements

Early and correct identification of skin conditions, particularly skin cancer, is extremely crucial to improving the patients' outcome. However, there are some severe shortcomings of AI diagnostic systems. Their most critical flaws include limited access to large, heterogeneous, and well-labeled dermatological image databases due to privacy constraints; low-quality images; difficulties in extracting prominent features; limited robustness across different skin types and conditions; and high computational cost of deep learning models. Moreover, existing practices tend not to distinguish between potentially malignant disorders, cancer, and malignant tumors, increasing the danger of overdiagnosis or missing early warning signs. These deficiencies challenge the credibility and usability of AI when it comes to dermatological diagnosis. Decide if the disease is cancer or has the potential to create it.

Main Question:

How can we develop a mobile-friendly AI system to detect skin diseases and assess the potential for skin cancer with high accuracy and interpretability?

Sub-questions:

1. Can transfer learning models like AlexNet effectively classify multiple skin diseases?
2. How can robustness be tested across different skin tones, disease types, and image variations to ensure model reliability?
3. What's the best architecture for real-time mobile deployment?
4. How can the model be made explainable to clinicians and users?
5. Can the system be trained to distinguish between benign conditions, malignant skin cancer, and pre-cancerous lesions that may evolve into cancer over time?

Objectives:

This study work will tackle skin disease and skin cancer detection issues by using artificial intelligence. Its prime aim is to improve diagnostic accuracy and early identification by eliminating bias from existing practices. The study aims to overcome the constraints of limited availability of diverse and tagged skin images due to privacy and data unavailability, small and low-quality data sets, and feature extraction issues. It also covers the lack of robustness testing and computational complexity of deep neural networks. In addition, the study aims to minimize risks of overdiagnosis and false positives, and maximize distinguishability between benign conditions, skin diseases with cancerous potential, and genuine skin cancer. Through these objectives, the research seeks to develop a reliable and efficient machine learning-based diagnostic model.

Planned Methodology:

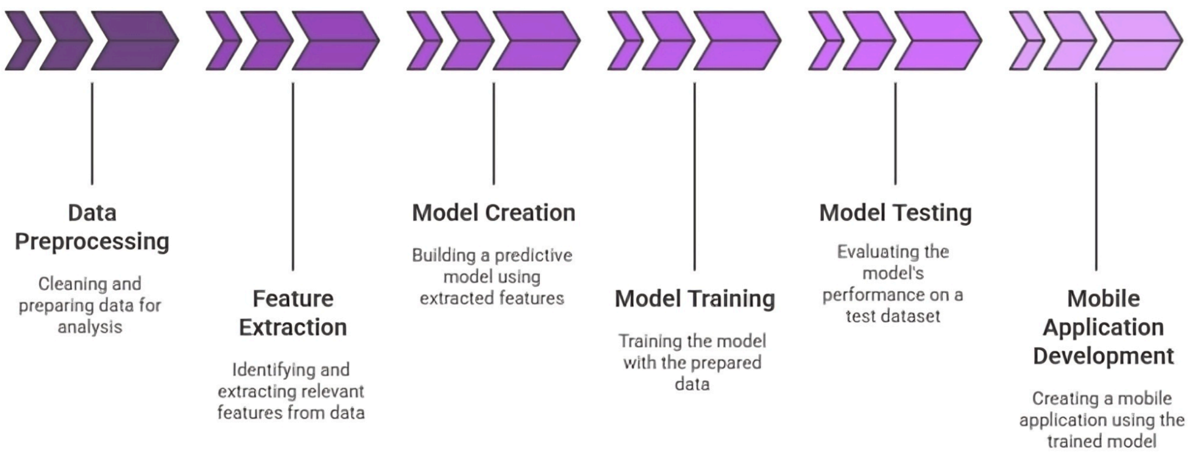

Figure 1: System Overview

Skin disease detection and classification, especially skin cancer and precancerous conditions, is a challenging task that requires robustness, explainability, and real-time usage on mobile devices in addition to excellent diagnostic accuracy. Several other approaches were taken into consideration, each with their own advantages and disadvantages.

<u>Method Choices: Strengths and Weaknesses</u>

1. **Traditional Machine Learning (e.g., SVM, Decision Trees)**
   Strengths: Works well on structured features

Weaknesses: Requires manual feature extraction, poor performance on raw image data, not scalable for multi-class skin disease detection

2. **Convolutional Neural Networks (CNNs)**
   Strengths: Automatic feature extraction from images, state-of-the-art performance in image classification
   Weaknesses: Requires large datasets and high computational power, may lack interpretability
3. **Transfer Learning with Pre-trained Models (e.g., AlexNet, ResNet, MobileNet)**
   Strengths: Reduces data requirements, enables faster training, supports mobile deployment (especially MobileNet)
   Weaknesses: Requires careful fine-tuning, may underperform without domain-specific adaptation

Justification for Chosen Method:

Convolutional neural networks (CNNs) and transfer learning, utilizing models such as MobileNet and AlexNet, have been selected as the research methodology. This method supports mobile deployment, performs well with little data, and is perfect for processing medical photos. Real-time use on smartphones is made possible via MobileNet, and transfer learning increases accuracy while cutting down on training time. Data augmentation improves robustness, while techniques such as Grad-CAM offer model explainability. Most significantly, this approach enables the system to distinguish between benign, malignant, and precancerous lesions, which is in line with the main objective of the research, which is to accurately, understandably, and easily assess the risk of skin cancer.

Table 1: Summary of Justification for Chosen Method

Requirement	How Transfer Learning With CNNs Helps
Small dataset	Uses knowledge from large pre-trained models
Image-based classification	CNNs are specialized for image tasks
Mobile deployment	MobileNet is optimized for mobile devices
Robustness across skin tones/types	Transfer learning + augmentation helps generalization

Model interpretability	Explainable AI tools (e.g., Grad-CAM) available
Distinguishing cancer/pre-cancer/benign	Multi-class and risk-level classification supported

Proposed Model and Framework

A. Dataset Collection and Preprocessing

B. Model Design

C. Model Robustness Testing

D. Model Explainability

E. Cancer Risk Classification

Train the system with labels for:

- **Benign** (e.g., eczema, psoriasis)
- **Pre-cancerous** (e.g., actinic keratosis)
- **Malignant** (e.g., melanoma, basal cell carcinoma)

F. Deployment Plan

Data Analysis Plans

The primary goal of this project is to develop an intelligent system capable of diagnosing various skin diseases, including skin cancer, using image processing and machine learning techniques. This system will analyze clinical or dermoscopic images of skin lesions and classify them into different disease categories, assisting in early detection and diagnosis.

- **Dataset collection:** For this project, have to utilize publicly available skin disease datasets like Kaggle datasets related to skin lesion classification.
- **Data Cleaning & Preprocessing:** Have to remove corrupted or duplicate images and do data preprocessing to ensure consistency and improve model performance, some preprocessing steps will be applied such as: Image Resizing, Normalization, Data Augmentation, Label Encoding.

- **Exploratory Data Analysis (EDA):** EDA will be used to understand: Class balance (how many images per disease), Distribution of image resolution, contrast, brightness, Visual patterns across disease categories.
- **Data Splitting:** Have to Divide the dataset into Training Set, Validation Set and Test Set to ensure stratified sampling to maintain class balance across all sets.
- **Feature Extraction & Model Selection:** Feature Extraction will be needed for image-based classification. By feature extraction relevant features will be extracted using convolutional neural networks (CNNs). CNNs will help to automatically extract visual features from images. Optionally we can apply pre-trained models (e.g., VGG16, ResNet50, EfficientNet) for transfer learning to improve performance with fewer data.
- **Model Training & Evaluation:** Train different models and tune hyperparameters using the training set. We will try to Evaluate model performance using: Accuracy, Precision,Recall, F1-score, Confusion Matrix, ROC-AUC Curve.
- **Model Optimization for App:** There are two types of android application:

  Remote Server-Based App: A Remote Server-Based Android App is a mobile application that runs on an Android device but relies on a remote server (accessed over the internet) to perform data processing, such as running machine learning models or storing data. The app sends data to the server, which processes it and sends the results back to the app.

  Local Android App: A Local Android App is a mobile application installed and run entirely on an Android device. All processing, including any machine learning tasks, happens on the device itself without requiring an internet connection.

  Both have some pros and cons. We will test model inference time on the device or through the server.

- **Integration with App:** We will integrate the trained model into the app interface: The user uploads or takes a skin image initially, and the image is preprocessed to comply with the input demands of the model. The preprocessed image is then passed through the trained model, which processes it and generates a skin disease diagnosis. The system returns the model's confidence as a measure of confidence and the predicted outcome. To enable individuals to grasp the regions of the image that affected the model's choice, it also produces an explanation of the forecast, i.e., visual heatmaps.
- **Post-Analysis and Feedback:** Here we will collect user feedback and misclassification reports. We will analyze incorrect predictions and improve the model iteratively & we will use Grad-CAM or similar tools to visualize which image parts influenced the prediction.

Expected Results

The expected outcome of this project is a fully functional, user-friendly application (Android or web-based) capable of accurately detecting and classifying various skin diseases, including skin cancer, from images of skin lesions.

More specifically, the project is expected to achieve the following results:

Accurate Disease Classification:

The machine learning model should classify skin diseases (e.g., melanoma, eczema, psoriasis, etc.) with high accuracy, ideally above 85%, depending on the dataset quality

The model will output: The model's output will include the predicted name of the disease, along with an optional explanation to enhance interpretability, such as a heatmap or highlighted attention area indicating the regions of the image that contributed most to the prediction.

Real-Time Image Diagnosis:

The app will allow users (doctors or patients) to capture or upload an image of a skin lesion. After processing, the app will return a diagnosis in real-time or within a few seconds, depending on whether it's local or server-based.

Easy-to-Use App Interface: The application will feature a clean, intuitive interface for:

- Uploading or taking a skin image
- Viewing prediction results
- Accessing basic information about the detected disease

Privacy and Security:

In the case of a local app, all image data will remain on the user's device, ensuring user privacy. If remote servers are used, data will be transferred securely and processed with proper privacy measures in place.

Model Performance Metrics:

The final trained model is expected to achieve strong performance on the test dataset, with metrics like:

- Accuracy: >85%
- Precision and Recall: Balanced across all classes
- F1-score: High, especially for critical diseases like melanoma
- AUC-ROC: Indicative of good sensitivity and specificity

Potential Real-World Impact:

The app has the potential to serve as a valuable preliminary diagnostic tool, particularly in low-resource or remote areas where access to dermatologists is scarce or nonexistent. By analyzing images of skin conditions, it can offer initial assessments that may guide patients toward appropriate care or further medical evaluation. In more advanced healthcare settings, the app can also support medical professionals by functioning as a decision-support system. By offering second opinions based on image analysis, it can enhance diagnostic confidence, reduce the likelihood of oversight, and contribute to more accurate and timely treatment decisions.

Conclusion

The integration of machine learning (ML) and image processing into dermatological diagnostics represents a major advancement in modern healthcare, addressing many of the limitations posed by traditional diagnostic practices. In particular, Dermalytics, the solution developed through this project, demonstrates how cutting-edge technologies can offer a powerful and scalable approach to identifying and classifying a wide spectrum of skin diseases, including benign conditions, precancerous lesions, and malignant skin cancers such as melanoma and squamous cell carcinoma. Leveraging deep learning techniques—specifically Convolutional Neural Networks (CNNs) enhanced through transfer learning—Dermalytics achieves high classification accuracy across diverse skin disease categories. The model has been trained and validated on varied, real-world datasets encompassing different skin tones, lesion types, and image resolutions, ensuring both clinical robustness and algorithmic fairness. By incorporating visual explanation tools (such as Grad-CAM or heatmaps), the system enhances interpretability and builds trust among end-users, especially clinicians who require transparency in AI-assisted diagnostics. A key strength of Dermalytics lies in its flexible deployment architecture. It supports both on-device inference (ideal for mobile applications in offline environments) and remote server-based processing (allowing heavier models and centralized data analysis). This dual-mode capability enables the system to adapt to various use cases, ranging from personal self-screening by non-expert users to clinical decision support for healthcare professionals. Additionally, the application's ability to deliver real-time diagnostic results empowers users to seek timely medical intervention, which is particularly crucial in diseases like melanoma, where early detection significantly increases survival rates. In low-resource or rural settings, where dermatologists are scarce or entirely unavailable, Dermalytics serves as a practical teledermatology tool, significantly improving accessibility, affordability, and equity in dermatological care. Furthermore, the project illustrates the broader impact of AI in democratizing healthcare. By reducing dependency on in-person consultations, streamlining diagnostic workflows, and enabling mass screening at scale, this system supports public health goals related to early cancer detection, preventive care, and digital health innovation. In summary, Dermalytics is more than just a diagnostic tool—it is a scalable, intelligent health platform designed to close the gap between dermatological expertise and patient needs. With continued data expansion, user testing, and clinical integration, it holds the potential to become a vital asset in the global fight against skin diseases, transforming how dermatological care is delivered, especially in under-served communities.

References

1. Prasad, P. M. K., Jahnavi, S., Gayatri, N., & Sravani, T. (2023). Skin Disease Detection using Image Processing and Machine Learning.
2. Vayadande, K., Bhosle, A. A., Pawar, R. G., Joshi, D. J., Bailke, P. A., & Lohade, O. (2024). Innovative approaches for skin disease identification in machine learning: A comprehensive study. Oral Oncology Reports, 10, 100365.
3. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., ... & Mahnashi, M. H. (2021). Skin cancer detection
4. Aftab, A., Sattar, H., Aslam, M. F., Hussain, T., & Waseem, S. S. (2024). An IoT-Driven Real-Time Skin Disease Detection System Using Machine Learning. Sukkur IBA Journal of Computing and Mathematical Sciences, 8(2), 13-24.
5. ALEnezi, N. S. A. (2019). A method of skin disease detection using image processing and machine learning. Procedia Computer Science, 163, 85-92.

Appendix

CO	CO Descriptions	PO Descriptions	Learning Domains	Way to achieve	Rubrics Design
CO1	Integrate new and previously acquired knowledge for identifying a real-life complex computer science and engineering problem as the capstone project.	PO1: Engineering Knowledge	Cognitive	-Review previous research work - Brainstorm real-world problems	Able to understand the knowledge of mathematics, natural sciences, engineering fundamentals, and computer science and engineering.
					Able to apply new and previously acquired knowledge for solving problems.
CO2	Examine various problem domains (literature review), define the problems, and formulate the objectives for the capstone project.	PO4: Investigation	Cognitive, Psychomotor	-Literature survey - Problem Statement - Objective - Methodology	Able to analyse and/or compare and/or categorise investigation (experiment/modelling/survey/etc.) data.
					Able to critically evaluate and/or assess and/or interpret and/or making inferences and/or conclusions data from the investigation (experiment/modeling/survey/etc.).
					Able to design investigation methodology based on literature review and/or standards.
					Able to conduct investigation (experiment/modeling/survey/etc.) under some supervision.
					Able to perform investigation (experiment/modeling/survey/etc.) under minimum supervision.
					Able to adapt investigation (experiment/modelling/survey/etc.) to suit objectives.