DERMA AI: NEXT-GEN AI SKIN DIAGNOSIS

Abstract

Traditional skin disease detection relies primarily on image analysis, but incorporating patient symptoms can enhance diagnostic accuracy. This project develops a multi-modal AI system that combines CNN-based image classification with NLP-based symptom analysis. The system takes skin images and textual descriptions of symptoms as input and uses a Transformer-based model (BERT + Vision Transformer) to make predictions. The results are fused using an ensemble learning approach, improving accuracy compared to single-modality models. A chatbot interface allows users to describe their symptoms, making the system more interactive and user-friendly.

Introduction

Skin diseases affect millions of people worldwide, and accurate diagnosis is essential for effective treatment. Traditionally, skin disease detection has relied primarily on visual inspection and image analysis, where dermatologists or AI-based systems classify conditions based on photographs of affected areas. However, visual data alone may not always provide sufficient information for an accurate diagnosis, as many skin diseases share similar visual characteristics. Factors such as texture, color variation, and lesion shape can be ambiguous, leading to potential misclassification. To improve diagnostic precision, it is crucial to integrate additional patient information, such as symptoms and medical history.

This project aims to enhance skin disease detection by developing a multi-modal AI system that combines both image-based classification and text-based symptom analysis. Convolutional Neural Networks (CNNs) are widely used for medical image classification, but they do not consider textual descriptions of symptoms, which can provide crucial contextual information. By incorporating Natural Language Processing (NLP) techniques, the system can analyze patient-reported symptoms alongside images, leading to a more holistic approach to diagnosis. This integration of multiple data sources helps bridge the gap between visual analysis and patient experience, improving the overall accuracy of the model.

The proposed system leverages state-of-the-art deep learning architectures, combining a Vision Transformer (ViT) for image analysis with a Transformer-based language model (BERT) for text processing. Vision Transformers have demonstrated superior performance in image classification by capturing long-range dependencies in visual features, while BERT excels in understanding complex natural language inputs. By utilizing both models, the system can effectively process and interpret different modalities, making it more robust than traditional single-modality approaches. An ensemble learning strategy is employed to fuse the results from both modalities, ensuring that the final prediction is more reliable and accurate.

To enhance user accessibility and interaction, a chatbot interface is integrated into the system. This chatbot allows users to describe their symptoms in natural language, making it easier for non-experts to provide relevant information. The chatbot processes user inputs using NLP and feeds them into the symptom analysis module, ensuring that the system captures both explicit and nuanced descriptions. By facilitating a conversational interface, the system improves user engagement and usability, making advanced AI-driven skin disease detection more accessible to the general population.

The project introduces a novel approach to skin disease detection by integrating multi-modal deep learning techniques. By combining CNN-based image classification with Transformer-based text analysis, the system overcomes the limitations of traditional methods, enhancing diagnostic accuracy. The use of ensemble learning further strengthens prediction reliability, while the chatbot interface ensures ease of use and accessibility. This multi-modal AI system has the potential to revolutionize dermatological diagnostics, enabling more precise and patient-centric healthcare solutions.

Background

Skin diseases are among the most prevalent health concerns worldwide, affecting individuals of all ages and demographics. According to the World Health Organization (WHO), millions of people suffer from dermatological conditions such as eczema, psoriasis, fungal infections, and skin cancer. Early and accurate diagnosis is crucial for effective treatment and prevention of severe complications. Traditionally, dermatologists diagnose skin diseases through visual examination, patient history, and, in some cases, biopsy or lab tests. However, access to dermatologists is often limited, especially in rural and underdeveloped regions. This has led to the growing interest in AI-driven diagnostic tools that can assist in the early detection of skin conditions.

Deep learning, particularly Convolutional Neural Networks (CNNs), has significantly advanced the field of medical image analysis. CNNs have been widely used for classifying skin diseases from digital images with promising results. These models can learn complex patterns from large datasets and achieve performance comparable to dermatologists in some cases. However, CNN-based models rely solely on visual features, which may not always provide sufficient information for an accurate diagnosis. Many skin diseases exhibit similar visual characteristics, making it difficult to distinguish between conditions based only on images. This limitation highlights the need for a more comprehensive approach that incorporates additional patient data, such as symptoms and medical history.

Natural Language Processing (NLP) has emerged as a powerful tool in healthcare applications, enabling machines to understand and process human language. Recent advancements in Transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers), have significantly improved the accuracy of text-based analysis. In the context of dermatology, patient-reported symptoms provide valuable insights that complement image-based analysis. Symptoms like itching, burning, pain, or the duration of a skin lesion are often critical for an accurate diagnosis but are not captured in images alone. By integrating NLP with deep learning for image classification, a multi-modal AI system can leverage both textual and visual information for better diagnostic accuracy.

The introduction of Vision Transformers (ViTs) has further enhanced image classification capabilities by capturing global dependencies in images, unlike traditional CNNs that focus on localized patterns. ViTs have demonstrated superior performance in various computer vision tasks, making them well-suited for skin disease detection. When combined with BERT for text analysis, these models can process and understand different types of input data, leading to a more robust diagnostic system. Moreover, ensemble learning techniques, which combine predictions from multiple models, have been shown to improve the reliability and accuracy of AI-driven diagnosis. This approach ensures that the final decision is based on the strengths of both image-based and text-based analysis.

To enhance accessibility and usability, AI-powered chatbots are increasingly being integrated into healthcare applications. A chatbot interface allows users to describe their symptoms in natural language, making the system more interactive and user-friendly. This is particularly beneficial for individuals who lack medical knowledge or those in remote areas with limited access to dermatologists. By processing user inputs through NLP and combining them with image analysis, the system can provide preliminary diagnostic support, guiding users toward appropriate medical advice or professional consultation. The combination of deep learning, NLP, and chatbot technology represents a significant step toward democratizing dermatological care and improving early skin disease detection.

Literature Review

Author & Year	**Title**	**Methodology**	**Advantages**	**Limitations**
Ananthakrishnan Balasundaram et al & 2024	Genetic Algorithm Optimized Stacking Approach to Skin Disease Detection	Deep learning-based genetic algorithm	Improve adaptability and performance	Potential overfitting due to ensembling, and reduced generalizability to unseen or rare skin conditions
Nader Shafi & 2023	A Portable Non-Invasive Electromagnetic Lesion-Optimized Sensing Device for the Diagnosis of Skin Cancer (SkanMD)	Support Vector Machine (SVM)-based classification model	The results differentiate between cancerous and non-cancerous skin lesions.	Include the small sample size, potential variability in sensor performance across different skin types
Jufeng Yang & 2019	Self-Paced Balance Learning for	Self-paced balance learning	The iterative learning process enhances	The reliance on initial complexity

	Clinical Skin Disease Recognition	(SPBL) algorithm	discriminative feature representation by balancing complexity at each stage	estimation, potential sensitivity to hyperparameter tuning, and the need for extensive computational resources due to iterative training
Yasmeen George & 2019	Automatic Scale Severity Assessment Method in Psoriasis Skin Images Using Local Descriptors	Three-class machine learning classifiers	Outperforms in vocabulary building regarding accuracy and computation time	Small dataset size, potential overfitting due to handcrafted feature selection, and the need for further validation on diverse skin tones and lighting conditions

Sutra Verma & 2021	Digital Diagnosis of Hand, Foot, and Mouth Disease Using Hybrid Deep Neural Networks	Hybrid Deep Neural Networks	The combined features from both branches are merged for final classification, significantly improving diagnostic accuracy	The limitations, including potential overfitting due to the small dataset size
Adekanmi A. Adegun & 2020	FCN-Based DenseNet Framework for Automated Detection and	Fully Convolutional Network (FCN)	Hyperparameter optimization techniques are employed to reduce network	Limitations, including its reliance on a single dataset, which may limit

	Classification of Skin Lesions in Dermoscopy Images		complexity and improve computational efficiency, making the model effective even with limited data	generalizability across different populations and imaging conditions

Gaps Indentified

Lack of Multi-Modal Data Integration:

Current skin disease detection models primarily rely on CNN-based image classification, focusing solely on visual features while ignoring other crucial diagnostic factors such as patient history, symptoms, and environmental influences. Incorporating multi-modal data, including clinical metadata and dermoscopic imaging, could significantly improve diagnostic accuracy.

Limited Real-Time and Sequential Analysis:

Most existing systems analyze only static images, requiring manual interpretation by dermatologists. This approach overlooks disease progression and symptom variations over time. Developing AI models capable of processing sequential images or real-time video analysis could enhance early diagnosis and monitoring.

Restricted Accessibility in Low-Resource Settings:

Many AI-based skin disease detection systems depend on cloud-based processing, requiring a stable internet connection, which limits their usability in rural and underserved areas. Developing lightweight, offline-capable models optimized for edge devices could improve accessibility and healthcare equity

Problem Statement

Traditional AI-based skin disease detection primarily relies on image analysis using deep learning models, but this approach has limitations due to the visual similarity of many skin conditions, leading to potential misclassifications. In real-world diagnostics, dermatologists consider both visual inspection and patient-reported symptoms, such as itching, pain, and duration of the condition, which existing AI models often overlook. To address this gap, this project develops a multi-modal AI system that integrates CNN-based image classification with NLP-based symptom analysis using Transformer models (BERT for text processing and Vision Transformer for image analysis). By employing an ensemble learning approach to fuse both modalities, the system enhances diagnostic accuracy compared to single-modality models. Additionally, a chatbot interface enables users to describe their symptoms in natural language, making the system more interactive, accessible, and user-friendly, ultimately improving early diagnosis and healthcare outcomes.

Objectives

Enhancing Diagnostic Accuracy with Multi-Modal AI

This project aims to improve skin disease classification accuracy by integrating image-based deep learning models with Natural Language Processing (NLP). By combining Vision Transformer (ViT) for image analysis and BERT for processing patient-reported symptoms, the system provides a comprehensive diagnosis, reducing false positives and negatives compared to traditional CNN-based models that rely solely on visual data.

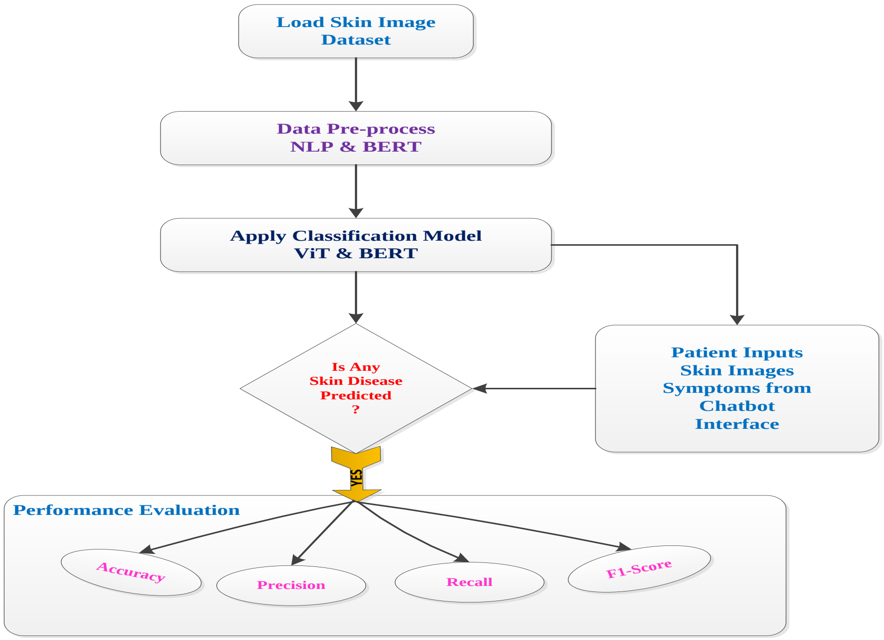
Real-Time and Explainable AI for User Trust

To enhance transparency and trust in AI-driven diagnoses, this project incorporates explainability techniques such as Grad-CAM for visualizing affected skin regions and attention mechanisms for highlighting key symptoms from patient inputs. This enables both patients and dermatologists to understand how the AI arrives at its conclusions, making the system more interpretable and reliable.

Interactive and Accessible AI with Chatbot Integration

This project focuses on making AI-driven skin disease diagnosis more accessible, particularly in remote areas with limited internet connectivity. By integrating a GPT-based chatbot for symptom collection and optimizing the model with TensorFlow Lite for on-device processing, the system reduces reliance on cloud computing, allowing users to receive real-time assistance without requiring constant internet access.

SYSTEM DESIGN


Load Skin Image Dataset
Data Pre-process NLP & BERT
Apply Classification Model ViT & BERT
Is Any Skin Disease Predicted ?
Patient Inputs Skin Images Symptoms from Chatbot Interface
YES
Performance Evaluation
Accuracy
Precision
Recall
F1-Score

The system design follows a structured pipeline for multi-modal skin disease detection, integrating both image and text-based analysis. It begins with loading a skin image dataset, followed by data preprocessing using Natural Language Processing (NLP) and BERT to process symptom descriptions. The classification model applies a combination of Vision Transformer (ViT) for image analysis and BERT for textual symptom interpretation. Patient inputs, including skin images and symptom descriptions, are received via a chatbot interface, enhancing user interaction. The model then determines whether a skin disease is detected, and performance evaluation is conducted using metrics such as accuracy, precision, recall, and F1-score. This approach ensures a robust, interactive, and accurate diagnostic system by leveraging both deep learning-based image classification and symptom-based analysis.

Image Data Loading

The Image Data Loading module is responsible for importing and storing the image data for the classification process. This includes loading skin images from datasets (e.g., SD-198 or custom datasets) into memory. Tools like OpenCV or PIL (Python Imaging Library) are used to read image files, ensuring the images are in a suitable format (e.g., PNG, JPEG). Additionally, this module will handle the organization of images into training, validation, and test sets, ensuring that the images are balanced across different skin disease categories. It also includes functionality for handling any missing data or corrupted files, allowing the system to handle real-world datasets effectively.

Data Pre-processing

The Data Pre-processing module processes both image and text data to make them ready for model training. For image data, this includes resizing images to a uniform dimension, normalizing pixel values, and applying augmentation techniques like rotation, flipping, and zooming to increase dataset diversity and improve model generalization. For the patient symptom data, NLP techniques are used, including tokenization, stopword removal, and stemming to process the textual descriptions of symptoms. Furthermore, the symptom data is tokenized using BERT’s tokenization scheme, which converts raw text into tokens that can be understood by the Transformer model. This module also handles combining the image and symptom data into a unified format suitable for multi-modal processing.

Train Model Based on Images and Patient Symptoms

The core of the system is the training of the multi-modal model that integrates image and textual data. The Image Classification Model, based on Vision Transformer (ViT), is trained on the preprocessed images to extract features from skin lesions and other visual cues. Simultaneously, the Text Classification Model, based on BERT, processes the patient’s symptom descriptions. Both models are independently trained on their respective data before their outputs are combined in the next step. An ensemble learning approach is employed to fuse the predictions of both models. The ensemble method improves accuracy by balancing the strengths of both models. This module involves optimizing the learning process, adjusting hyperparameters, and utilizing loss functions tailored for multi-modal learning.

Skin Disease Prediction

After training the individual models, the Skin Disease Prediction module is responsible for generating predictions based on the combined inputs of skin images and symptom descriptions. When a user inputs a skin image and describes their symptoms through the chatbot interface, the system first processes the image through the Vision Transformer model and the symptoms through the BERT model. The output predictions from both models are then fused using the ensemble approach to make a final diagnosis. This module returns the predicted disease type along with confidence scores, allowing both users and dermatologists to understand the likelihood of the prediction. It also incorporates attention mechanisms and Grad-CAM visualizations to highlight key regions in the image and key symptoms in the text that influenced the model's prediction.

Performance Evaluation

The Performance Evaluation module assesses the accuracy and effectiveness of the multi-modal AI system. It uses various performance metrics, including accuracy, precision, recall, F1 score, and confusion matrix, to evaluate the overall classification performance. The module evaluates how well the model performs on the test dataset, which includes unseen images and symptom descriptions. Additionally, it provides insights into the fusion process's impact, showing how combining image and symptom data improves the system's diagnostic capabilities compared to single-modality models. This module also handles error analysis, helping to identify patterns in misclassifications and guiding future model improvements.

REFERENCES

1.Balasundaram, A., Shaik, A., Alroy, B.R., Singh, A. and Shivaprakash, S.J., 2024. Genetic Algorithm Optimized Stacking Approach to Skin Disease Detection. IEEE Access.

2.Shafi, N., Costantine, J., Kanj, R., Tawk, Y., Ramadan, A.H., Kurban, M., Abou Rahal, J. and Eid, A.A., 2023. A Portable Non-Invasive Electromagnetic Lesion-Optimized Sensing Device for The Diagnosis of Skin Cancer (SkanMD). IEEE Transactions on Biomedical Circuits and Systems, 17(3), pp.558-573.

3.Yang, J., Wu, X., Liang, J., Sun, X., Cheng, M.M., Rosin, P.L. and Wang, L., 2019. Self-paced balance learning for clinical skin disease recognition. IEEE transactions on neural networks and learning systems, 31(8), pp.2832-2846.

4.George, Y., Aldeen, M. and Garnavi, R., 2019. Automatic scale severity assessment method in psoriasis skin images using local descriptors. IEEE Journal of Biomedical and Health Informatics, 24(2), pp.577-585.

5.Verma, S., Razzaque, M.A., Sangtongdee, U., Arpnikanondt, C., Tassaneetrithep, B. and Hossain, A., 2021. Digital diagnosis of hand, foot, and mouth disease using hybrid deep neural networks. IEEE Access, 9, pp.143481-143494.

6.Adegun, A.A. and Viriri, S., 2020. FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images. IEEE Access, 8, pp.150377-150396.