Multimodal Dermatological Disorder Assessment via Conversational Bot Integration for Textual and Visual Diagnostic Interpretation

[1]Dr.B.Sravan Kumar, [2]SABA, [3]SHAFIA SUMEEN, [4]RAPARTHI RANJITH KUMAR

[1]Associate Professor, [2,3,4] UG STUDENT

[1,2,3,4,]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI & ML)

[1,2,3,4] VAAGDEVI COLLEGE OF ENGINEERING Autonomous

Bollikunta, Khila Warangal (Mandal), Warangal Urban-506 005 (T.S), www.vaagdevi.edu.in

Abstract

The skin of a person is a good sign of their health because it often shows the first signs of problems with their internal organs. It is very important to notice these signs early on so that you can get a diagnosis and treatment. People sometimes forget how important the skin is as a defence mechanism for the body. The goal of this research is to make a multimodal skin disease classification system that uses a Telegram chatbot to combine ensemble transfer learning models (DenseNet169, ResNet50, EfficientNetV2, and Swing Transformer) with Natural Language Processing (NLP). The main goal is to make the chatbot better at giving personalized and correct skin-related diagnoses based on information from users, such as their skin type, chemical exposure, and past treatments.

The system makes diagnoses more accurate and personalised by combining both image-based analysis and textual contextual information. EfficientNetV2 helps speed up computations and extract high-resolution features, while Swing Transformer uses hierarchical vision transformers to learn both global and local features so that it can work with different skin conditions. In addition, geospatial mapping (MAP Integration) shows the number and location of skin disease cases so that epidemiological analysis and public health information can be gathered.

The chatbot can learn on its own, which means it can change how it responds to users based on how they interact with it. This keeps users interested over time. The hybrid strategy makes the most of fine-grained features and understanding the context, which leads to great classification performance. We tested 11,747 images, of which 7,930 were used for training and validation and 3,817 for testing. The proposed model achieved 77.07% accuracy and an AUC of 96.72% for image classification, whereas the NLP model attained 93.62% accuracy, delivering a comprehensive and tailored diagnostic experience.

Keywords: Skin disease diagnosis, multimodal learning, ensemble CNNs, NLP chatbot, geospatial health analytics.

I.INTRODUCTION

The skin is the body's largest organ and a very important sign of how healthy the inside is. The skin is often the first place where signs of systemic illness show up, which is very helpful for doctors trying to figure out what's wrong. Skin problems can be signs of many other health problems, such as metabolic disorders, autoimmune diseases, and infections. It is important to find these signs early because early treatment can stop the disease from getting worse and save lives. The diagnostic process is mostly based on the important role that skin plays in keeping an eye on health. However, the process can be subjective, take a long time, and depend a lot on clinical experience. This issue shows how important it is to have diagnostic systems that are easy to use, accurate, and automated.

Recent advances in deep learning and artificial intelligence (AI) have made it possible to create automated systems for diagnosing skin diseases. These kinds of systems get pictures of skin and figure out what kind of skin problems they are and how likely they are to happen. On the other hand, single-model traditional methods tend to generalise, extract features, and handle complex or mixed datasets. Ensemble learning techniques, which combine several models, have been shown to work better than single models in medical imaging applications by using each model's strengths and making up for its weaknesses. In addition to analysing images, textual and contextual information, such as the patient's history, chemical exposure, skin type, and previous treatments, are also very important for making clinical decisions. Natural Language Processing (NLP) can easily deal with this kind of information and give personalized diagnoses and treatment suggestions.

By combining image-based analysis with contextual data, it is possible to create a multimodal system that improves the accuracy of diagnoses and personalisation. EfficientNetV2 is an improved model that extracts features from high-resolution images quickly. Swin Transformer's hierarchical vision transformers encode both global and local structures, which helps the model generalise better to different skin conditions. Adding geospatial mapping also makes it possible to see the spread of skin diseases in different areas, which can help public health policy and resource allocation.

II.LITERATURE REVIEW

Recent research in dermatological diagnostics has shown that deep learning-based image classification models like DenseNet, ResNet, and EfficientNet work well for automatically finding skin diseases. Transfer learning makes these convolutional neural networks better at extracting features and making diagnoses. However, single-model approaches often have trouble generalising and being robust when dealing with a wide range of skin conditions.

Ensemble learning techniques have been developed to improve performance by combining different architectures to take advantage of their strengths and make predictions that are more stable and accurate. Natural Language Processing (NLP) is also being used to help make more personalised and clinically relevant decisions by allowing systems to look at patient histories, symptoms, and other information. Studies show that multimodal systems that combine text and image data work better than unimodal diagnostic models.

Conversational AI and healthcare chatbots like Ada Health, Buoy Health, and Babylon Health make healthcare even more accessible by giving people real-time medical advice and the ability to talk to a doctor from anywhere. Still, many current solutions don't have built-in image analysis and adaptive learning. These limitations underscore the necessity for a cohesive multimodal framework that integrates ensemble deep learning, natural language processing, and chatbot-based interaction to facilitate precise and accessible skin disease diagnosis.

III.METHODOLOGY

A. Collecting Data

- I gathered a huge set of dermatological images with 11,747 pictures of skin diseases.
- Collected information about the user, like their skin type, symptoms, chemical exposure, and past treatments.
- To get a good idea of how well something works, split the data into training/validation and testing sets.

B. Preparing the Data

- Resized images so that they all had the same resolution for model compatibility.
- Used normalisation to change the size of pixel values.
- Used techniques like rotation, flipping, and zooming to make the model more general and less likely to overfit.

- For NLP processing, cleaned and tokenised text inputs.

C. Extracting Features from Images

Used transfer learning models to get deep visual features:

- DenseNet169 for reusing features and making gradient flow better
- ResNet50 to fix problems with vanishing gradients
- EfficientNetV2 for speed and efficiency
- Swing Transformer for learning global and local features in a hierarchical way

D. Learning in Groups

- Used weighted averaging or voting to combine predictions from all models.
- Used the best parts of each architecture to their full potential.
- Better accuracy, reliability, and robustness of classification.

E. Contextual Analysis Based on NLP

- Used Natural Language Processing techniques to process written inputs.
- Got medical entities, symptoms, and history information.
- Made personalised suggestions based on data about each user.

F. Fusion of Multiple Modes

- Combined image classification results with NLP insights.
- Made diagnostic decisions that were aware of the context and tailored to the person.
- Improved the accuracy of the whole system compared to methods that only use one modality.

G. Putting the Chatbot to Use

- Used a Telegram chatbot interface to set up the system.
- Let users upload pictures and talk to each other.
- Gave real-time health advice and diagnostic feedback.
- Added a way for the system to learn on its own so it can keep getting better.

H. Mapping in Geospatial

- Recorded location data of detected cases without names.

- Used maps to show how diseases spread in different areas.
- Helped with public health monitoring and epidemiological analysis.

I. Evaluating Performance

- Used accuracy and AUC metrics to evaluate the system.
- Looked at how well each model and the group of models worked.
- Test data that wasn't seen before showed that it worked.

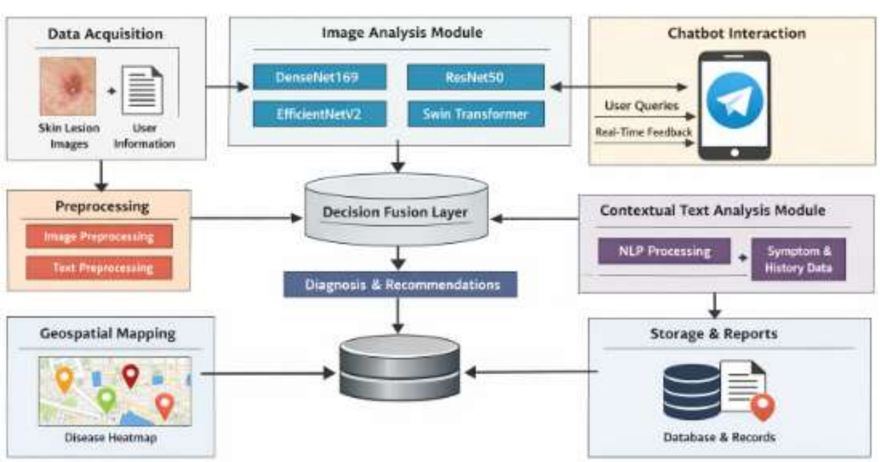
IV.SYSTEM ARCHITECTURE

The suggested Skin Disease Diagnosis System is a modular, scalable, and multimodal architecture that can process dermatological images and user information in real time. The framework combines image capture, text-based symptom analysis, ensemble deep learning models, and an interactive chatbot interface to give you accurate and personalised help with your diagnosis. There are input pipelines, processing modules for each type of data, a central fusion engine, and a layer for delivering results. The overall design makes sure that both individual users and healthcare monitoring apps can use it with high accuracy, reliability, and ease of use.

A. Overview

The system is made up of three main parts: an Image Analysis Module, a Contextual Text Analysis Module, and an Interaction & Analytics Module. Using ensemble transfer learning models like DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer, the Image Analysis Module sorts skin diseases. The Contextual Text Analysis Module uses Natural Language Processing to understand a user's symptoms and medical history so that it can make personalized suggestions. The Interaction & Analytics Module provides services through a Telegram chatbot, which lets people talk to each other in real time, make reports, and see data on a map. A central fusion layer connects all the modules and combines outputs from different modes to make the final diagnosis.

B. Architecture Diagram



V.EXPERIMENTAL SETUP

A. Getting the dataset ready

- Used a dermatological image dataset with 11,747 labelled images of skin lesions.
- Gather more information about the situation, such as symptoms, skin type, exposure to chemicals, and past treatments.
- To get an unbiased evaluation, divide the dataset into training, validation, and testing sets.

B. Preparing the Image

- Changed the size of the images so that they would work with the model.
- Used normalisation to make pixel values the same.
- Used techniques like rotation, flipping, and zooming to add more data and reduce overfitting while making the model work better on new data.

C. Processing Text Data

- Organised and cleaned up text that users gave us.
- Used natural language processing to find symptoms and medical information in context.

D. Setting up the model

- Used deep learning architectures based on transfer learning:
- DenseNet169
- ResNet50
- EfficientNetV2
- Transformer for Swing
- Fine-tuned pretrained models using the prepared dataset to classify skin diseases.

E. Strategy for the Ensemble

- Predictions from all of the individual models put together.
- Used weighted averaging and voting systems.
- Better strength, stability, and accuracy of diagnostics.

F. Setting up the system

- Put together the whole framework with a Telegram chatbot.

- Let users upload pictures and give health information.
- Made diagnostic outputs and suggestions in real time.

G. Evaluation of Performance

- Used accuracy and Area Under the Curve (AUC) to measure how well the system worked.
- Looked at how well each model did compared to the ensemble.
- Results that were checked against new test data

VI.RESULT ANALYSIS

A. Performance of Classification

The suggested multimodal framework showed great results in automatically classifying skin diseases by using transfer learning and ensemble learning methods. Individual deep learning models, such as DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer, made reliable predictions by picking out unique visual features from dermatological images. However, there were small differences in accuracy between models because of how they represented features and how well they learned.

B. What Ensemble Learning Does

The ensemble strategy made diagnostic performance much better than using models on their own. The system used weighted averaging and voting to combine predictions, which made it less likely to make mistakes and more stable. The ensemble method gave more stable and consistent results, showing that it could better generalise to test samples that had never been seen before.

C. How Multimodal Fusion Affects Things

The system worked even better when it combined contextual text information with predictions based on images. The Natural Language Processing module included symptoms, skin type, and medical history, which made it possible to make decisions that were tailored to each person and took into account their situation. This combination of different types of data made it easier to tell the difference between skin conditions that looked similar and made the overall accuracy of the diagnosis better than just looking at pictures.

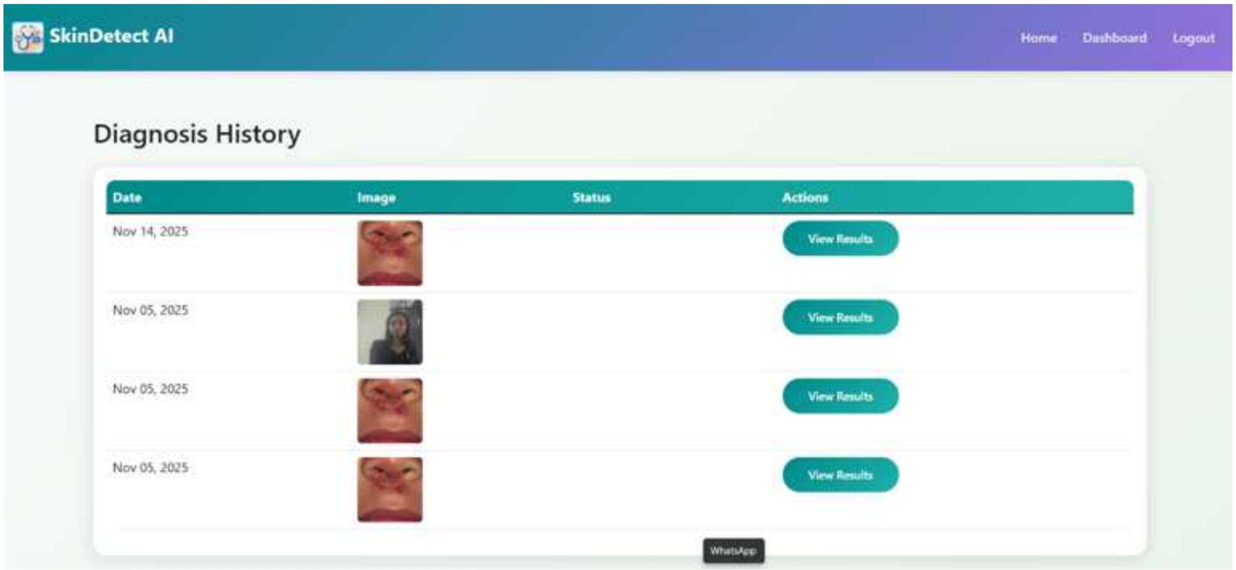
D. Evaluation of the System in Real Time

Deployment through the Telegram chatbot made it possible to interact and test in real time in real-life situations. Users could upload pictures and get predictions and suggestions right away. The system showed that it could be used for remote healthcare help by having a fast response time, clear communication, and easy access for users.

E. General Results

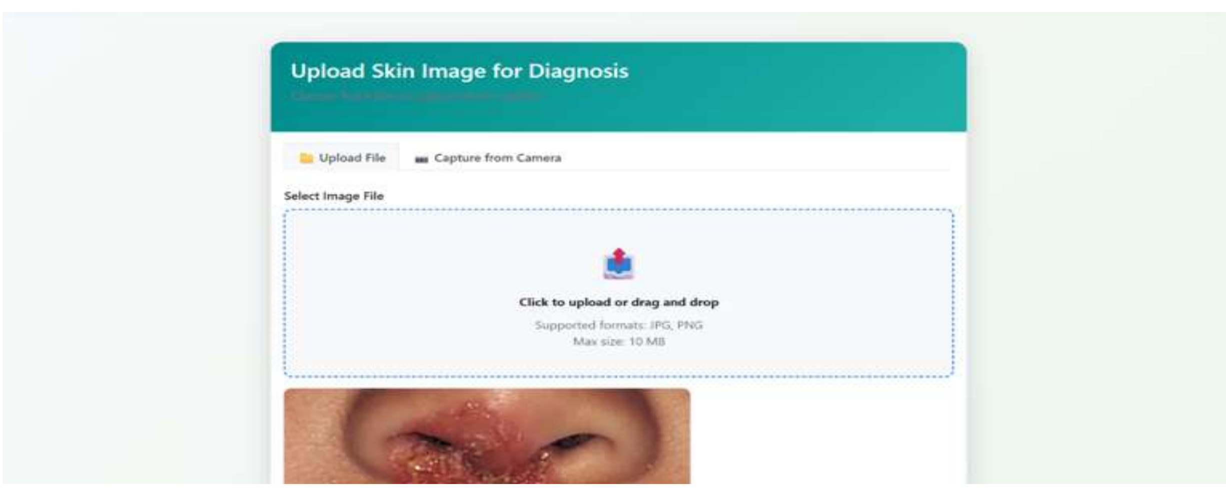
The results of the experiments show that using preprocessing, transfer learning, ensemble modelling, and multimodal fusion together makes the model more accurate, stable, and personalised. The proposed system is better than single-model and unimodal approaches, so it can be used for skin disease diagnosis that is both scalable and reliable.

A. Image showing diagnosed history



The image shows the Diagnosis History dashboard of the SkinDetect AI web application, which serves as the user interface for managing previously analyzed skin disease cases. The page presents a structured and user-friendly layout with a navigation header containing options such as Home, Dashboard, and Logout, indicating a secure and authenticated environment. Below the header, a tabular history of diagnostic records is displayed, where each entry includes the upload date, a thumbnail preview of the skin lesion image, the diagnosis status, and an action button labeled "View Results." The image previews allow quick visual identification of cases, while the results button enables users to access detailed predictions and recommendations generated by the deep learning models. This interface functions as a report management and tracking module, storing past diagnoses and providing easy retrieval of information. Overall, the dashboard enhances usability and supports real-time interaction, making the system practical for continuous monitoring and follow-up of skin health assessments.

B. Image uploaded for disease detection



The image shows the image upload interface of the SkinDetect AI application, designed for collecting skin lesion images for automated diagnosis. The page provides options to upload an image file or capture a photo directly using a camera, along with a drag-and-drop area that supports common formats such as JPG and PNG. A preview of the selected image is displayed to allow users to verify the input before submission.

In the system architecture, this component functions as the data acquisition layer, where images are gathered and forwarded to preprocessing and deep learning modules for analysis and classification. The simple and user-friendly design ensures accurate data collection and supports real-time diagnostic processing.

VII.CONCLUSION

This paper introduced a multimodal deep learning framework for the automated diagnosis of skin diseases by amalgamating dermatological image analysis, contextual text comprehension, and conversational support into a cohesive system. We used transfer learning-based architectures like DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer to get strong features. We also used an ensemble strategy to make the classification more accurate and stable than using just one model.

Adding Natural Language Processing to the system made it possible to look at a patient's symptoms and medical history and give personalized diagnostic suggestions based on the context. Using a Telegram chatbot to deploy the service made it possible for people to interact with it in real time, made it easy to get to, and made it easy for people to get help with their health. Geospatial mapping also made it easier to look at disease trends in different areas for public health monitoring.

The experimental results show that the suggested multimodal and ensemble-based method is more reliable, accurate, and scalable than systems that only use one mode. Overall, the framework is a smart, useful, and effective way to find skin diseases early and get help with diagnosis from a distance.

VIII.REFERENCES

1. Modigari, N., Harshini, T. S., & Anbarasi, L. J. (2024). Advancing skin disease diagnosis: A multimodal approach utilizing Telegram API token chatbot for text and image analysis in skin disease classification.
2. Alruwaili, M., & Mohamed, M. (2025). An integrated deep learning model with EfficientNet and ResNet for accurate multi-class skin disease classification. Diagnostics, 15(5), 551. https://doi.org/10.3390/diagnostics15050551
3. Sahoo, S. (2025). Deep learning-powered skin disease classification. Journal of Theoretical and Applied Information Technology, 103(15), 234–240. https://www.jatit.org/volumes/Vol103No15/27Vol103No15.pdf
4. Mohan, J., Sivasubramanian, A., Vishvanathan, S., & Ravi, V. (2024). Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable AI. Computers in Biology and Medicine, 190, 110007. https://doi.org/10.1016/j.compbiomed.2025.110007
5. Zhou, J., et al. (2024). Pre-trained multimodal large language model enhances skin disease diagnosis. National Institutes of Health.
6. Banerjee, A., Das, N., & Nasipuri, M. (2020). Skin diseases detection using LBP and WLD: An ensembling approach. arXiv. https://arxiv.org/abs/2004.04122
7. Putatunda, S. (2019). A hybrid deep learning approach for diagnosis of the erythemato-squamous disease. arXiv. https://arxiv.org/abs/1909.07587
8. Velasco, J. S., et al. (2023). Classification of skin disease using transfer learning in convolutional neural networks. arXiv. https://arxiv.org/abs/2304.02852