

Fusion of Image and Symptoms for Skin Disease Prediction

Mekala Shashirekha Nayudu^{*1} and Juthuka Aruna Devi²

^{*1}M. Tech II Year Student, Department of CSE(DS), Anil Neerukonda Institute of Technology and Sciences, Bheemunipatnam, Visakhapatnam, 531162, Andhra Pradesh, India

²Associate professor, Department of CSE(DS), Anil Neerukonda Institute of Technology and Sciences, Bheemunipatnam, Visakhapatnam, 531162, Andhra Pradesh, India

Date of Submission: 25-05-2026

Date of Acceptance: 05-06-2026

ABSTRACT

This paper presents a multimodal skin disease prediction system that integrates deep learning-based image analysis with symptom-driven textual inputs to enhance diagnostic accuracy. The proposed system utilizes a Vision Transformer (ViT) model for extracting high-level visual features from skin lesion images and a BERT-based model for encoding patient-reported symptoms into contextual embeddings. These heterogeneous features are combined through a multimodal fusion mechanism to capture both visual and clinical correlations, enabling more robust classification. A fully connected layer with softmax activation is applied to the fused representation in order to forecast the most likely skin condition and provide a confidence score. In addition to prediction, the system provides supplementary outputs including disease description, severity level estimation, and precautionary recommendations to assist users in understanding their condition. A chatbot interface is also incorporated to facilitate interactive symptom input and basic guidance. The system is deployed as a web-based application, enabling real-time analysis and improving accessibility for the end users. Experimental evaluation demonstrates that the multimodal approach outperforms traditional unimodal method by leveraging complementary information from both image and text data. The proposed framework serves as an effective decision-support tool for early skin disease detection and awareness, while emphasizing that it is not a substitute for professional medical diagnosis.

KEYWORDS: Vision Transformer, Deep Learning, Image classification, Skin disease detection, Multimodal learning.

I. INTRODUCTION

Skin problems affect patients of all ages and backgrounds and represent a major worldwide public health issue. Many skin diseases cause discomfort for their sufferers, affect their mental

health, reduce their quality of life and place a heavy burden on public health spending. For skin cancer in particular, early diagnosis at the earliest stage can save a patient's life and transform what could be a long and arduous journey into a short one. Currently, analysis of a patient's eyes by a dermatologist might be aided by dermoscopy images, with biopsy and histological examination being the gold standard in situations of uncertainty. However, such methods are generally labour intensive, subjective and require the input of a qualified specialist who may not always be available. There are many regions of the world where access to dermatologists is restricted or limited, causing distress to their patients. The need for a scalable and robust diagnostic system to aid doctors in their analysis and to overcome restrictions to access to dermatology care is evident. While skin disease can be difficult to treat via image, recent advances in Artificial Intelligence (AI) have allowed machine learning and deep learning algorithms to analyse and categorise a variety of skin diseases with a high degree of accuracy. These systems are being increasingly utilized in a variety of settings for tasks including diagnosis, monitoring between inperson appointments, facilitating teleconsultations and supporting general practitioners. This paper explores recent advances in skin disease detection algorithms and models and details our current model's effectiveness. It then discusses current challenges and ethics related to deployment of such systems into healthcare settings. Many areas of the world have limited access to dermatologists for disease diagnosis. However, recent advances in deep learning and computer vision have achieved state-of-the-art results in various medical image classification tasks, such as skin disease diagnosis. Specifically, transformer models, including Vision Transformers (ViT) and other variants, have demonstrated exceptional performance in various medical image classification tasks.

Objectives of the Proposed System:

The main purpose of this project is to develop a smart and reliable system that can predict early skin disease based on picture and symptom analysis. The following are the specific objectives of this project.

- Develop a multimodal model that accurately predicts illness from both unstructured symptom freeform text and structured multimodal skin images.
- Extracting features of text and image data using advanced models such as BERT and Vision Transformer (ViT) and state-of-the-art deep learning techniques.
- Develop a mechanism to fuse the two modalities to improve predictive accuracy.
- Create disease categorization system (including predictions, relevant information and confidence score).
- Development of an interactive web interface chatting with a web based artificial intelligence.

II. RELATED WORK

[1] Sarala et al.'s 2025 This concept presents an image-based approach for the diagnosis of various common skin diseases. The System will incorporate an AI powered solution using a Convolutional Neural Network (CNN) with an interactive Chatbot. The system was trained and tested using images from the ISIC 2020 dataset, a total of 25331 tagged images of the skin. The model uses Transfer Learning by utilizing pre-trained models such as ResNet50 and Efficient Net to increase the accuracy of classification. Preprocessing of images was done to enhance the images. This was achieved by segmenting images and applying different techniques to contrast stretch the images. Images were also smoothed using a Gaussian filter to reduce noise. The images were also augmented to increase the dataset of images. With appropriate training and testing, the system reached an accuracy of 74.4% further proving the efficiency of a CNN DL model in skin disease classification. The system will also incorporate a chatbot which will direct users to the proper treatment and provide an interpretation of the results as well as explain the given symptoms. [2] In 2025, Rachel et al. Skin cancer is one of the most serious diseases in human body, and early diagnosis is crucial for its treatment. In this study, we propose ResNetVision, a DL approach for early skin cancer detection using Convolutional Neural Networks (CNNs) and transfer learning. To extract features

from pre-processed images of skin lesions, we largely employ the ResNet architectures. For the final classification, we employed conventional machine learning methods such as Support Vector Machine (SVM), random forest learning (RF), neural network algorithms (NN), Logistic Regression (LR), K-Nearest Neighbours (KNN), and AdaBoost. Additionally, we utilized several image processing techniques to enhance the lesions' visibility in images, such as cropping, pixel resizing, and Dull Sharp hair elimination. Results of the experimental studies showed that ResNet-18 with SVM achieved the best results (accuracy: 80.28%) and outperformed deeper versions of ResNet. This study also compared the proposed approach with several other CNNs, namely VGG16, VGG19, and Inception-V3, in which ResNet-18 achieved the highest accuracy. This demonstrated that residual learning improves the performance of deep networks and alleviates the vanishing gradient problem in skin cancer detection, and can distinguish between different types of skin tumours for accurate diagnosis. [3] In 2023, Mridha et al. introduced an interpretable skin cancer classification system based on improved (CNN). This improved system addresses some of the challenges faced by current deep learning systems in healthcare such as class imbalanced and lack of explainability. In this paper, we explored the application of improved CNN in classifying seven types of skin cancer lesions from the HAM10000 dataset. We employed several activation functions (ReLU, Swish, and Tanh) in order to improve the performance of our proposed system. We fine-tuned the pre-trained CNN models using Adam and RMSprop as optimizers. Moreover, we have incorporated Explainable AI (XAI) techniques called Grad-CAM and Grad-CAM++ to graphically highlight the regions of the skin lesions that contribute the most to the classifier's decision which will lead to increased confidence and improve the overall model interpretability. Experimental results show promising accuracy results reaching around 82% improvement in classification accuracy with decreased loss values. Combining explanation power and deep learning models, the system is capable of making reliable and fair skin cancer diagnosis decisions. [4] Raghupathy and Nithyashri et al (2025) proposed a hybrid model which combined CNN and GNN for categorizing different types of skin diseases. The GNN was used to capture the spatial correlation between different regions of image. On the other hand, the CNN was used to learn the prominent features such as texture, shape, color, etc. that CNN lacked to learn in global

fashion. This model achieved higher accuracy in skin disease classification (up to 92.4%) as compared to stand-alone CNN. The approach to skin disease diagnosis was further enhanced by the model's ability to differentiate and accurately predict between closely related conditions such as psoriasis and eczema. [5] Skin predicting illnesses using imaging data is the core idea that Madhur Shalini et al. (2020) propagate in their paper. They utilized neural network to forecast illnesses from patients' symptom input and DCNN to predict from skin images. An ensemble method was incorporated to enhance accuracy in illness classification. The model was trained on many illnesses classification and achieved 87.71% detection rate. The team also created a web-based online application that makes access to healthcare service wider and more convenient. [6] Nikam et al. (2025) developed a skincare recommendation platform that recommends personalizing skincare products based on skin morphology. The system identified various skin types, tones and acne severity levels using deep learning models like EfficientNet and image processing algorithms, and then utilized clustering techniques like K-Means and content-based product filtering. The model incorporates parameters like age, gender and various skin concerns, outperforming existing solutions and providing a viable system for product recommendations that integrates deep learning. [7] In this paper, Chhajer and Gadicha (2025) presented a hybrid model consisting of both dividing and categorizing approaches for skin disease diagnosis. Features from images were learnt using convolutional neural network AlexNet, subsequently the infected region of interest was segmented and masked using ResUNet++. The features so learnt were further classified using a Random Forest approach to support decision making. The approach yielded superior results as accuracy of 91.3% was achieved on the HAM10000 dataset and improved as model was trained more. The paper showed potential of deep learning, segmentation, and hybrid approach (ensemble) in improving classification accuracy of skin diseases. [8] A machine learning model to predict complexion, skin type and disorders was developed by Kugaraj et al. in 2024. The model utilized a decision tree to predict the skin type of user based on their input and a CNN (ResNet-50) for image processing. The model integrated NLP and web scraping to acquire real-time dermatology updates and interpret them. The skin analysis model performed well and demonstrated accuracy in all tasks. Despite some issues, the model could be very useful for giving personalized and up-to-date

skincare recommendations. [9] Fernandes et al. (2025) investigated current literature on the use of AI for the identification and diagnosis of various skin conditions. The main limitation of these models is the lack of external validation. While the models achieve excellent results while being optimized to diagnose skin diseases, their performance decreases drastically when tested on new patients. Many of the current works fail to establish a valid validation strategy, lack diversity in datasets and sample sizes. Most works fail to assess bias across institutions and instead rely on retrospective data. A more thorough validation protocol is necessary to create models that can be consistently used in clinical practice. [10] Narendra et al. in their work proposed a multimodal approach for skin disease detection using both text and images through a chatbot interface. To enhance the classification performance for the disease, they employed a combination of hybrid transfer learning techniques along with deep learning techniques. Telegram chat was integrated in order to collect the user symptoms and provide them with real-time interactive diagnosis. Experimental results on a large dataset of various skin diseases indicate very good results for both text and image processing. Results indicate that infection of the skin prediction systems (SKPs) can be improved by fusing textual and visual data. [11] Akshay et al. (2023) proposed a novel mobile based vision model called Skin-Vision. The model utilizes deep learning architecture to analyse skin problems through images. The model uses CNN and visual transformer, which can instantly predict potential risk of different skin diseases as soon as image of skin is uploaded. This prediction is provided through a mobile app. The model was able to achieve high accuracy in few days when it was trained on huge dataset. In general, the model is helpful in making early diagnosis and can increase accessibility to detect different skin diseases. [12] To detect skin diseases Balasundaram et al. (2024) proposed deep learning technique, namely Genetic Algorithm (GA) based skin disease image classification using Stacking Method to improve the performance of skin disease image classification by optimising the ensemble of skin disease classification models like CNN, DenseNet and ResNet etc. with an enhanced learning capacity. The technique demonstrated promising results with improvement in terms of accuracy and Top-5 accuracy by training and testing the models using dermatology reference information (DermNet) and large dataset (HAM10000). The average accuracy obtained by the proposed technique was about 91.7% and outperformed many existing approaches

for classification of various types of skin diseases. The results achieved using improved ensemble techniques further supports the potential of enhanced ensembling techniques for skin disease classification. [13] In 2022, Imran et al. have presented skin cancer identification technique that incorporates many models based on deep learning techniques. Instead of using single model they have fused VGG, CapsNet and ResNet models to generate features and employed ensemble technique to classify predictions from individual models. Results from the technique showed a lift in accuracy compared to individual models that have been trained on ISIC dataset, with an average accuracy of around 93.5%. This paper has shown that incorporating many models can result in more accurate medical predictions. [14] In the paper published at MICCAI 2019, the work Ganesan et al. 2025 compared and analysed state-of-the-art algorithms used for early skin disease diagnosis using ML and Computer Vision. We observed that deep learning architectures such as hybrid-based models and convolutional neural network (CNN) offered superior performance over traditional algorithms. Importantly, we highlighted existing opportunities offered by Explainable AI, attention mechanisms and ensemble methods that can boost the confidence of dermatologists in adopting AI-based diagnostic solutions for skin diseases. However, performance of current models is limited by factors such as class imbalance, limited data diversity and a lack of validation on clinical datasets. Importantly, future algorithms should focus on achieving scalability, robustness and clinical usability to be truly useful in healthcare. [15] Gomes et al. (2023) in their recent work have proposed an intelligent framework for identifying skin diseases and selecting appropriate treatment options by combining machine learning and traditional medical

system, Ayurveda practiced in Sri Lanka. The system employs deep learning neural networks (DNN) namely InceptionV3 and VGG16 for identification and severity level diagnosis of the identified disease. Experimental results demonstrate promising results for both disease recognition and classification. Furthermore, machine learning approach namely Random Forest algorithm is employed to select appropriate treatment options for each disease. The paper explores the effectiveness of merged approach of traditional medical system and intelligent system for improving accuracy of skin disease diagnosis and subsequently proposing optimal treatments.

III. METHODOLOGY

A novel skin disease prediction system is developed in this paper based on a multimodal learning framework, which jointly handles both visual and textual modalities. The system is organized into three steps: input preprocessing, feature extraction with fusion, and classification.

Dataset Description

For this task we used a publicly accessible skin disease dataset (around 6000 images) for training and testing. We selected 8 of the many available skin disease classes for this project. All images together were saved in a folder. Additionally, we saved relevant symptoms for each image in a CSV file. Each row in the CSV includes information about an image (classes, symptom, index). The fields image filename, x coordinate of image, y coordinate of image, filename, x coordinate of cropped image, y coordinate of cropped image, cropped filename, labels and corresponding symptoms were imported for the multimodal deep learning approach.

System Architecture

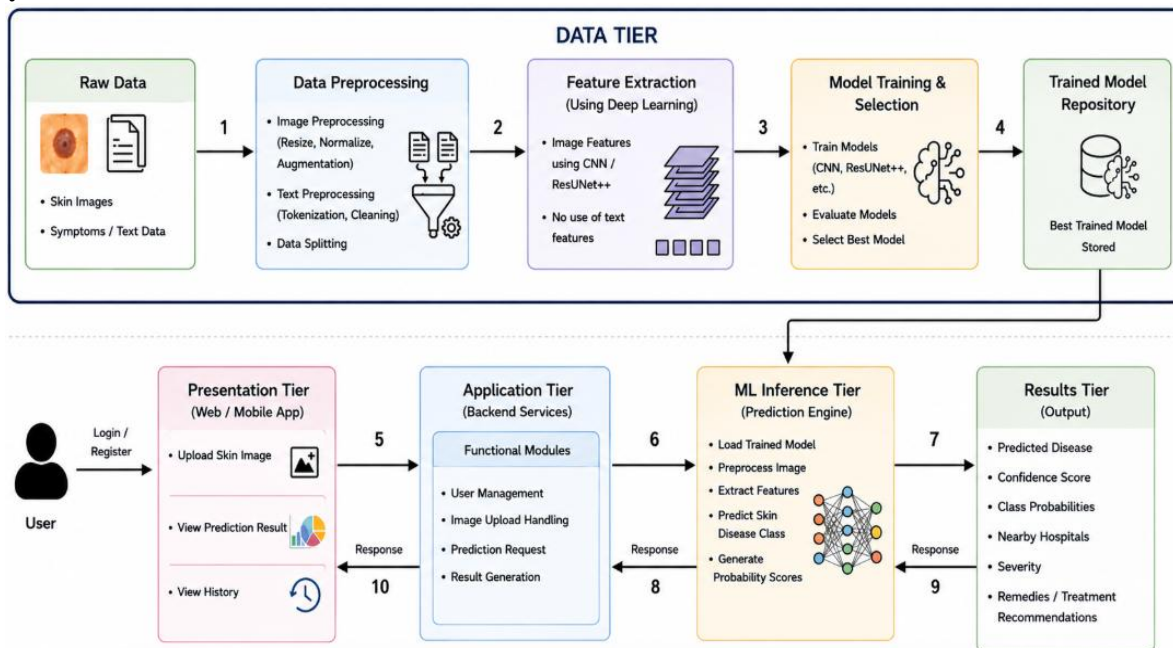


Fig 1: system architecture

Phase 1: Input Collection and Preprocessing

The system takes two forms of input, including images of skin diseases and text that describes symptoms experienced by the user.

Image Input and Preprocessing

This dataset is composed by clinical images from different skin diseases. It has been up-scaled to a uniform size to facilitate the analysis.

Image Rescaling:

For all images, we resize them to a uniform height of 224 pixels and then padding to image width of 224 pixels to facilitate input into the Vision Transformer model, which expects images of this specific size.

Pixel Standardization:

The pixel values of images are scaled to a fixed range, or are normalized statistically to improve training stability and efficiency.

These techniques help the model learn better by forcing it to focus on relevant visual patterns and ignoring unnecessary variation in the training set.

Symptom Text Processing

Data for symptoms, as seen in the following file, is stored in the form of a list of

images with corresponding descriptions, and is saved as a CSV file.

The text preparation process includes:

Breaking Text into Tokens:

The symptom sentences are tokenized into meaningful parts with the help of the BERT tokenizer.

Numerical Conversion:

The generated tokens are then converted to IDs which can be passed to the model.

Length Adjustment:

For fixed sequence size, inputs that are too short are padded with zeros (space padding or fill padding) and inputs that are too long are truncated.

This step converts symptoms to machine understandable input by parsing human readable text.

Phase 2: Feature Learning and Fusion

Once preprocessing has been done, both image and text inputs are independently learned for their features.

Visual Feature Learning with ViT

Instead of traditional convolution methods, the Vision Transformer handles images by separating them into small patch units.

- The input image is partitioned into smaller blocks.
- Each block is transformed into an embedded representation.
- Position-related information is attached to preserve patch order.
- These representations are processed through transformer layers to learn important image relationships.

This allows the model to understand both local and overall disease patterns.

The extracted image representation can be written as:

$$F_{\text{image}} = \text{ViT}(I)$$

where I is the preprocessed image.

Text Feature Learning with BERT

BERT is applied to understand the symptom descriptions by analysing word relationships in both forward and backward directions.

The symptom representation is generated as:

$$F_{\text{text}} = \text{BERT}(T)$$

where T represents the symptom text.

The final contextual embedding is obtained from the special CLS token.

Fusion of Features

The learned features from both image and text are merged into one combined feature vector.

$$F_{\text{combined}} = F_{\text{image}} \parallel F_{\text{text}}$$

where \parallel represents feature joining (concatenation). This combined representation helps the system make decisions using both visual and symptom-related information.

Phase 3: Prediction and Classification

The fused feature vector is forwarded to dense neural network layers for final decision-making.

The classification output is computed as:

$$Z = W(F_{\text{combined}}) + b$$

where W and b are learnable parameters.

To convert output values into probabilities, the Softmax function is applied:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}}$$

where:

- $P(y_i)$ is the probability assigned to class i
- z_i is the output score of class i
- N is the total number of disease classes

The disease class with the maximum probability score is selected as the final result.

IV. RESULT AND ANALYSIS

In this study, we evaluated the performance of a multimodal approach for skin disease prediction using images and symptoms. We utilized standard performance metrics including accuracy, precision, recall, and F1-score to measure the performance of the deep learning-based models in predicting different skin diseases.

Model Type	Accuracy	Precision	Recall	F1-Score
Image-only(ViT)	78%	77%	76%	76.5%
Text-only(BERT)	72%	71%	70%	70.5%
Multimodel(Image+ Text)	89%	88%	87%	87.5%

Table 1: Comparison Table of existing systems

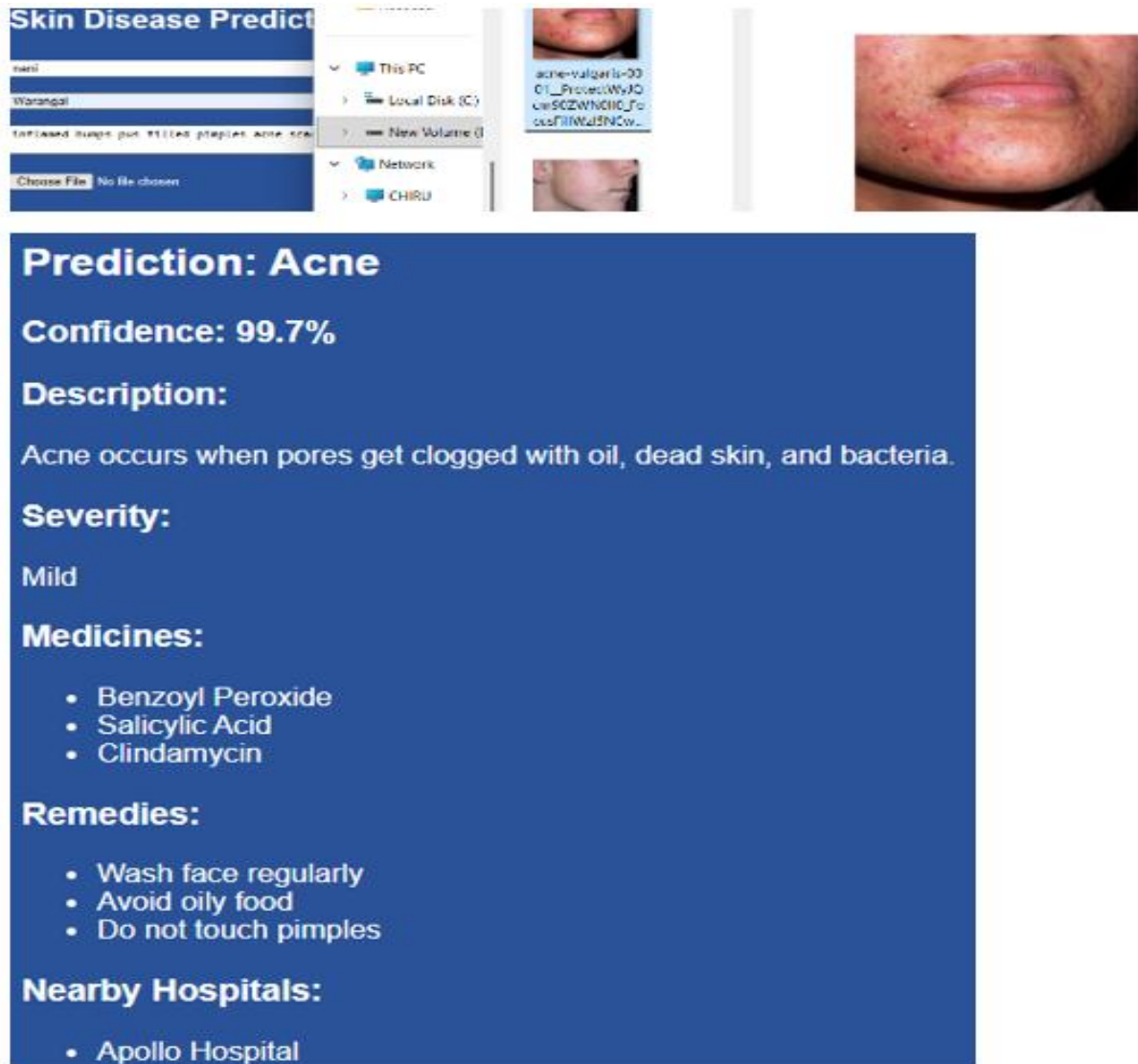


Fig 2: Input and Output

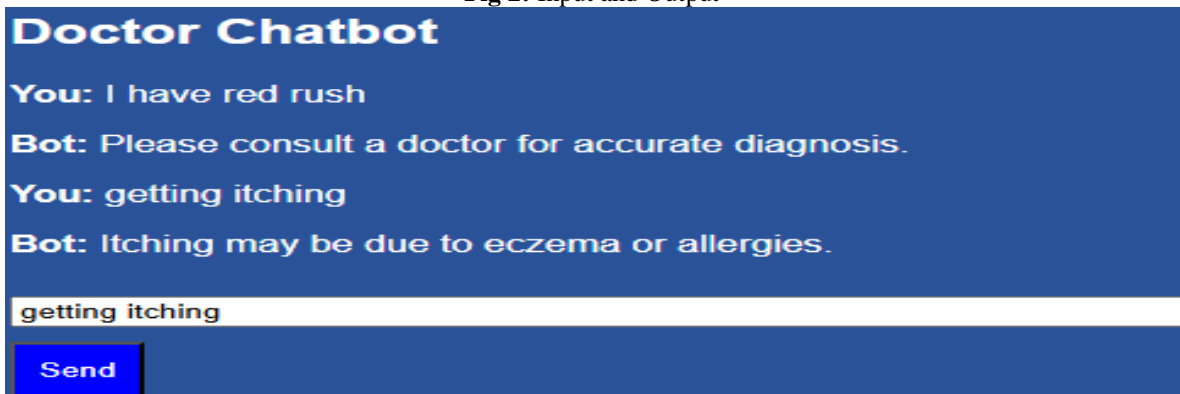


Fig3: Chatbot support

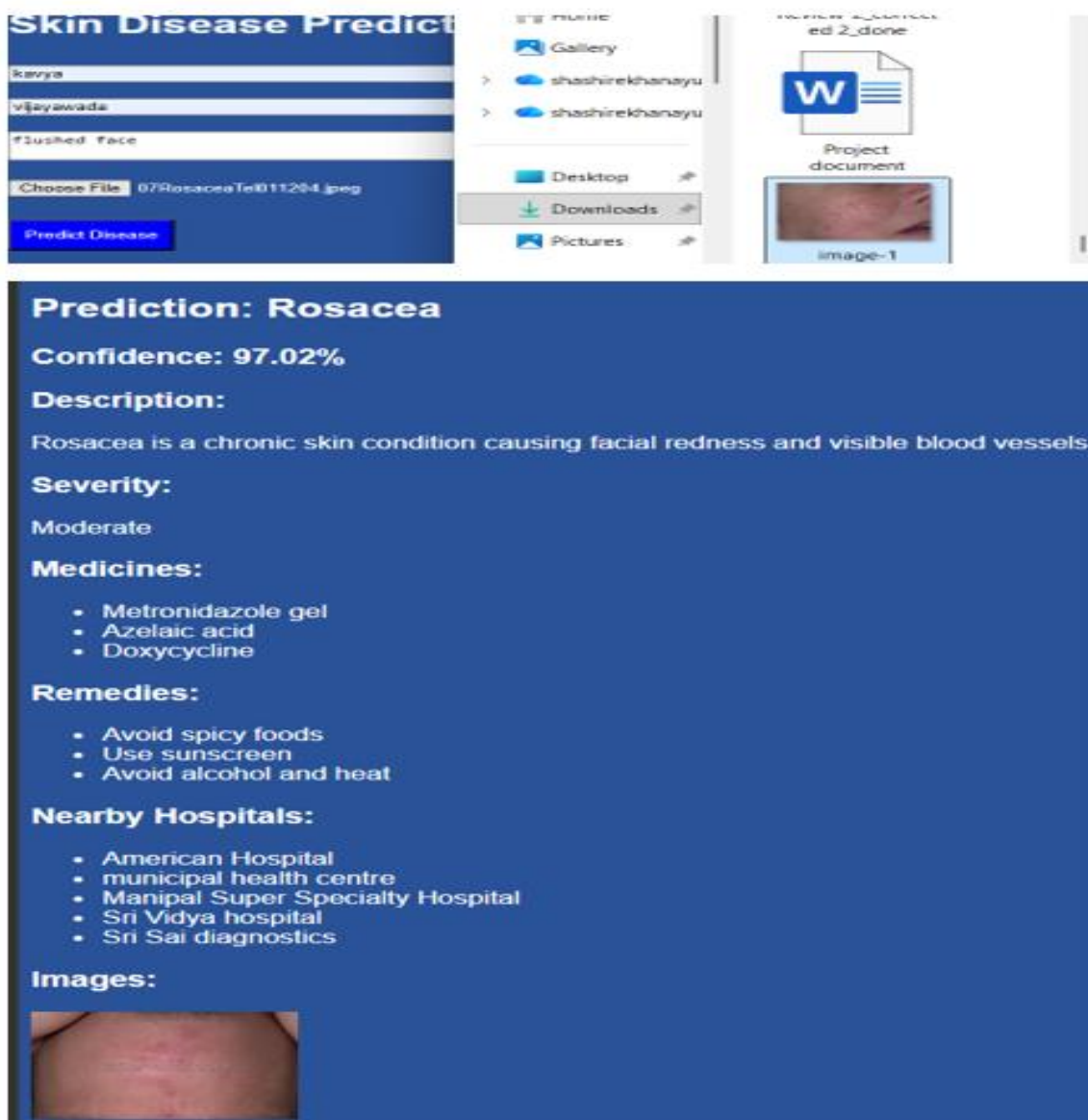


Fig 4: Sample input output

V. CONCLUSION

We present a multimodal system for skin disease prediction, which combines features extracted from images of the skin and symptoms reported by the patient using deep learning techniques. The image features are extracted using Vision Transformers (ViT) and enhanced by incorporating multimodal fusion in order to increase the robustness of the system especially for images that are visually ambiguous. Combining both image and text data enhances the system for realistic applications. The system is also available as an interactive web-based application. The architecture of the model is very scalable,

maintainable and enables easy integration of upcoming improvements. In summary, this innovative technique offers the potential of multimodal deep learning for identification of skin diseases at an early stage and improved access to dermatology services.

Future scope

Future work involves expanding the system to cover more diseases/medical conditions/patient groups, reducing bias and improving accuracy, using more advanced fusion strategies including attention, transformers, etc. The system can be ported to mobile/wearable

platforms with support for telemedicine, enabling real-time monitoring. Most importantly, ensuring data security, clinical compliance, multilingual chatbots are crucial to making this system usable in real-world scenarios.

REFERENCES

- [1]. R. Sarala, A. L, M. P, L. L and K. A. M.S, "Image-Based Skin Disease Detection using Convolutional Neural Networks," 2025 7th International Conference on Intelligent Sustainable Systems (ICISS), India, 2025, pp. 892-899, doi: 10.1109/ICISS63372.2025.11076343. <https://ieeexplore.ieee.org/document/11076343>
- [2]. S. R. S, R. K. S, V. J. S and R. R, "ResNetVision: Harnessing Deep Learning for Early Skin Cancer Detection," 2025 International Conference on Computing and Communication Technologies (ICCT), Chennai, India, 2025, pp. 1-5, doi: 10.1109/ICCT63501.2025.11019928. <https://ieeexplore.ieee.org/document/11019928>
- [3]. K. Mridha, M. M. Uddin, J. Shin, S. Khadka and M. F. Mridha, "An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System," in IEEE Access, vol. 11, pp. 41003-41018, 2023, doi: 10.1109/ACCESS.2023.3269694. <https://ieeexplore.ieee.org/document/1017401>
- [4]. B. L. Raghupathy and N. J, "Augmenting Convolution Neural Networks with Graph Models (GNN) for Skin Disease Classification," 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2025, pp. 1768-1774, doi: 10.1109/ICIRCA65293.2025.11089713. <https://ieeexplore.ieee.org/document/11089713>
- [5]. M. M, C. Nair and N. Goel, "Automatic identification of skin lesions using deep learning techniques," 2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G), Geneva, Switzerland, 2020, pp. 230-235, doi: 10.1109/AI4G50087.2020.9311068. <https://ieeexplore.ieee.org/document/9311068>
- [6]. S. Nikam, S. Chobe, O. Patil, R. Baviskar, G. Birla and S. Biradar, "Cosmetics and Skincare Recommendation based on Skin Characteristics using Machine Learning Techniques," 2025 9th International Conference on Computing, Communication, Control and Automation (ICCCBEA), Pune, India, 2025, pp. 1-6, doi: 10.1109/ICCUBEA65967.2025.11284176. <https://ieeexplore.ieee.org/document/11284176>
- [7]. K. k. Chhajed and A. B. Gadicha, "A Hybrid Approach for Detection of Skin Diseases Using Deep Learning," 2025 International Conference on Computing and Communications (COMPUTINGCON), Talegaon, India, 2025, pp. 1-6, doi: 10.1109/COMPUTINGCON64838.2025.11377415. <https://ieeexplore.ieee.org/document/11377415>
- [8]. L. Kugaraj, D. M. G. T. Dassanayake and K. Rajendran, "A Machine Learning Framework for Accurate Skin Tone, Type, and Disease Detection with Web Scraping and NLP for Dermatological Insights," 2024 9th International Conference on Information Technology Research (ICITR), Colombo, Sri Lanka, 2024, pp. 1-6, doi: 10.1109/ICITR64794.2024.10857787. <https://ieeexplore.ieee.org/document/10857787>
- [9]. T. Ribeiro Silva Fernandes et al., "External Validation of AI Models for Skin Diseases: A Systematic Review," in IEEE Access, vol. 13, pp. 114411-114427, 2025, doi: 10.1109/ACCESS.2025.3584904. <https://ieeexplore.ieee.org/document/11062591>
- [10]. M. Narendra, T. S. Harshini and L. Jani Anbarasi, "Advancing Skin Disease Diagnosis: A Multimodal Approach Utilizing Telegram Api Token Chatbot for Text and Image Analysis in Skin Disease Classification," in IEEE Access, vol. 12, pp. 189009-189023, 2024, doi: 10.1109/ACCESS.2024.3516884.

- <https://ieeexplore.ieee.org/document/10798433>
- [11]. G. Akshay, M. Irfan, S. KG and A. Singh, "Skin-Vision: An Innovative Mobile-Based Automated Skin Disease Detection Application," 2023 OITS International Conference on Information Technology (OCIT), Raipur, India, 2023, pp. 835-840, doi: 10.1109/OCIT59427.2023.10430941. <https://ieeexplore.ieee.org/document/10430941>
- [12]. A. Balasundaram, A. Shaik, B. R. Alroy, A. Singh and S. J. Shivaprakash, "Genetic Algorithm Optimized Stacking Approach to Skin Disease Detection," in IEEE Access, vol. 12, pp. 88950-88962, 2024, doi: 10.1109/ACCESS.2024.3412791. <https://ieeexplore.ieee.org/document/10552847>
- [13]. A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani and A. Almuhaimeed, "Skin Cancer Detection Using Combined Decision of Deep Learners," in IEEE Access, vol. 10, pp. 118198-118212, 2022, doi: 10.1109/ACCESS.2022.3220329. <https://ieeexplore.ieee.org/document/9940917>
- [14]. Naga, B & Ganesan, Subramanian. (2025). Early Skin Disease Detection Algorithms- Approaches, Challenges, and Future Directions. https://www.researchgate.net/publication/400694787_Early_Skin_Disease_Detection_Algorithms-Approaches_Challenges_and_Future_Directions
- [15]. M. P. O. M. Gomes, Y. N. Jayasekara, K. M. K. R. Kariyapperuma, H. P. M. N. Gunawardhna, N. H. P. R. S. Suwarnakantha and G. Wimalarathne, "Human Skin Diseases Identification and Treatment Suggestion by Sri Lankan Ayurveda Medicine Using Machine Learning," 2023 5th International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka, 2023, pp. 65-70, doi: 10.1109/ICAC60630.2023.10417632. <https://ieeexplore.ieee.org/document/10417632>