Data 606 Capstone in Data Science

AI-Driven NLP Chatbot for Skin Disease Diagnosis and Medication recommendations

Professor: Muhammed Ali Yousuf

Team:

Manisha Karnati (VZ42698)

Sai Sreeja Allepu (YD50976)

Yeswanth Reddy Chereddy (RD76099)

OBJECTIVE

The aim of this project is to design an AI-powered chatbot that will apply NLP and image recognition to help users diagnose dermatological conditions and provide them with relevant medication assistance. It will enable the user to upload pictures of skin conditions for preliminary diagnostic feedback and recommend over-the-counter medications or treatments along with dosages and guidelines.

By integrating the resources for diagnosis and medication in one place, the chatbot simplifies the usually complicated process of identifying appropriate treatments on one's own and avoids unnecessary medical consultations for minor conditions. This helps to reduce delays in diagnosis, improve access to medication, and also increase patient engagement through a user-friendly, fast response solution to skin health issues.

ABSTRACT

This project is focused on the development of an AI-powered chatbot based on NLP and image classification that helps in the diagnosis and medication recommendations for dermatology-related conditions. Utilizing CNNs trained on fine-quality skin disease datasets, such as DermNet, the model yielded considerable results in the classification of various skin conditions, hence proving its applicability in real-world practice. It enables the user to upload pictures and get immediate diagnostic feedback with suggestions for treatment. This enhances patient engagement and reduces unnecessary visits to the doctor. Future is to integrate generative AI and improve multilingual support to increase access.

INTRODUCTION

The prevalence of skin diseases worldwide requires novel solutions to enhance the diagnostic efficiency and accessibility of the service. Traditional diagnosis often involves expert dermatological consultations, which may not be always accessible or affordable. This project will bridge this gap by developing an AI-powered chatbot integrated with image classification and NLP for quick, accurate, and personalized dermatological care.

Convolutional Neural Networks form the core of this, which is quite efficient for analyzing medical images. Such datasets as DermNet, which have 20,000-plus images for over 200 conditions of skin, are useful in delivering reliable diagnostic feedback by the chatbot. NLP included in it can process user queries for medication recommendations, making the bot an all-inclusive tool for managing skin health. This project not only the explain us the potential of AI in healthcare industry but also on top of that makes us understand the need for continuous refinement to enhance accuracy and user experience.

LITERATURE REVIEW

- **Machine Learning and Deep Learning in Dermatology**: This paper by Zhang et al. (2023) showed us Machine Learning (ML) and Deep Learning (DL) have improved and showed us how accurately skin conditions can be diagnosed . The study highlights that Convolutional Neural Networks (CNNs) are particularly effective at analyzing and identifying skin conditions from images, outperforming older ML methods. When paired with techniques like transfer learning or advanced model designs, CNNs have achieved impressive accuracy rates between 91% and 99.7% across various datasets and architectures, setting the standard for diagnosing skin conditions. The review also mentions several well-known datasets, including DermNet, ISIC, and MED-NODE, which provide high-quality images crucial for training these powerful AI models.

- **Image Processing and Machine Learning for Skin Disease Detection**: ALEnezi (2020) introduces a method for skin disease detection using image processing and machine learning, focusing on essential preprocessing steps like resizing and normalizing images to enhance quality before model training. The study reveals that effective data which is preprocessing is crucial for enhancing performance of the model, which later aligns with the proposed methodology of preprocessing image data to optimize AI model accuracy for identifying skin conditions.
- **Chatbot-Based Disease Prediction and Treatment Recommendation**: This paper by Pathak and Ansari (2023) examines the use of AI chatbots in disease prediction and also treatment recommendation, highlighting the effectiveness as an initial point of contact in healthcare industry. The study shows that how chatbots can efficiently gather patient information, analyze symptoms, and suggest treatments accordingly. This aligns with the proposed chatbot's goal of using image recognition and natural language processing to provide personalized medication assistance and diagnostic support.
- **Dermatologist-level Classification of Skin Cancer with Deep Neural Networks:** Esteva et al.'s 2017 study which demonstrated the potential of deep neural networks (DNNs) in skin cancer classification, outperforming traditional methods helped in understanding the CNN integrations with chatbots. They demonstrated that DNNs can detect skin cancer from images with high diagnostic accuracy.
- **Human–computer collaboration for skin cancer recognition:** Tschandl et al. (2020) highlighted the significance of human-computer collaboration in skin cancer detection,

demonstrating that combining dermatologists' strengths with AI systems like CNNs can lead to better diagnostic outcomes. This study emphasized the need for AI tools to support healthcare professionals and provide reliable diagnostic feedback, helping us to understand the potential of AI in augmenting human expertise in dermatology.

- **Deep neural networks are superior to dermatologists in melanoma image classification:** Deep neural networks (DNNs) have shown they can outperform dermatologists in identifying melanoma from images, as demonstrated in a 2019 study by Brinker et al. In this research, a convolutional neural network (CNN) was trained using 4,204 biopsy-confirmed images of melanomas and nevi. The model was then tested against 95 dermatologists from nine German university hospitals using an additional 804 biopsy-confirmed dermoscopic images. The CNN achieved a sensitivity (ability to correctly detect melanoma) of 82.3% and a specificity (ability to correctly identify non-melanoma cases) of 77.9%, significantly better than the dermatologists, who achieved sensitivities of 67.2% and specificities of 62.2%. A statistical test (McNemar's test) confirmed the CNN's superior performance with strong significance ($p < 0.001$). This study highlights the ability of deep learning models to detect complex skin conditions like melanoma more accurately than human experts, reinforcing their potential in dermatology and their suitability for tools like our AI-driven chatbot.
- **Nutrient content, uptake and NUE of oats:** Kumar et al. (2022) study oat nutrient content and uptake efficiency, offering insights into agricultural practices and nutrient use efficiency. The paper emphasizes data-driven approaches for improving resource use and optimizing outcomes. Although not directly related to dermatology, it can inform chatbot data management strategies for optimal decision-making in diagnostic systems.

DATA SOURCE

For this project, DermNet NZ is a very valuable and perfect resource for training an AI model for recognizing skin conditions. It provides a rich dataset of high-quality dermatological images, which can be used in the chatbot's image recognition component. This data can improve precision in identifying dermatological issues and provide personalized medication recommendations, enhancing patient engagement and care outcomes.

DermNet. (2024, April 22). *Images A-Z | DermNet*. DermNet®.
https://dermnetnz.org/images

METHODOLOGY

The image classification data was sourced from the DermNet database, a publicly available and credible repository of dermatological images containing labeled data for 23 skin disease categories. Preprocessing steps included resizing images to 300x300 pixels, normalization of pixel values, and data augmentation techniques such as flipping, rotation, and zooming to address class imbalance and enhance model generalization. The preprocessed data was used to train a TensorFlow-based Convolutional Neural Network (CNN) for skin disease classification, with early stopping applied during training to prevent overfitting.

For the conversational component, the DialoGPT-medium model from Hugging Face was employed to generate context-aware responses to user queries. Medication suggestions were dynamically retrieved using the Cohere API, which generates treatment advice based on disease-specific prompts. The system integration was implemented using a Flask backend for managing API communication, while a Gradio interface provided a user-friendly platform for image uploads and chatbot interaction. Secure access to the application during testing was facilitated through ngrok tunneling, ensuring real-time accessibility and usability.

ARCHITECTURE

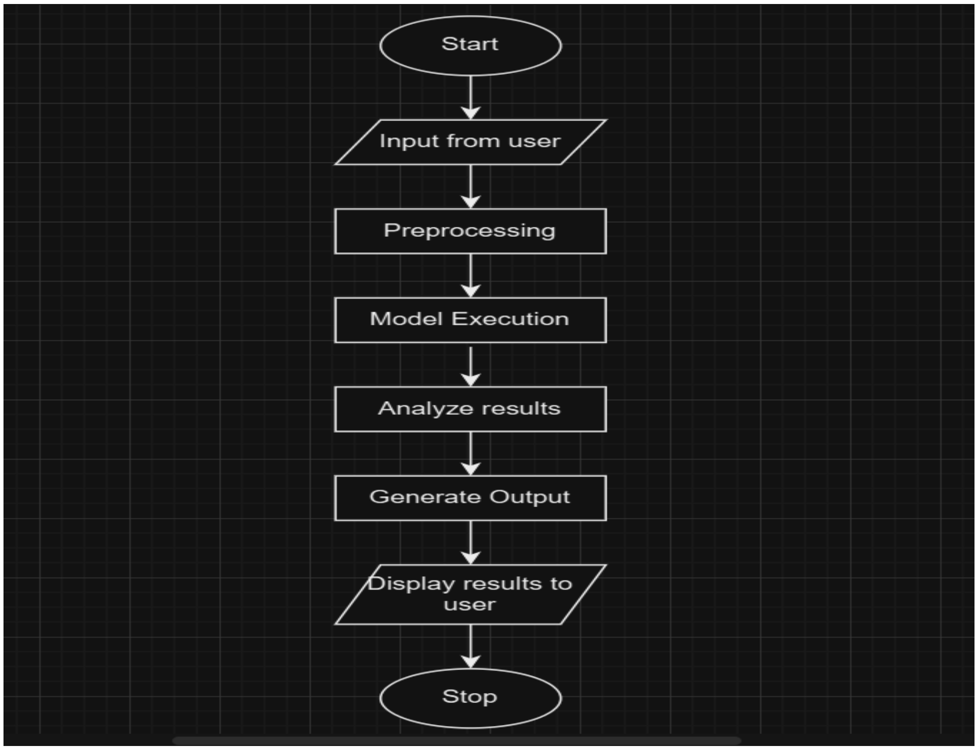

Figure (0)

Results and Analysis

This project is structured into five key phases:

Phase 01: Data Collection and Preprocessing: The skin disease dataset, referred to as SKIN (DermNet), was collected and preprocessed to ensure effective model training. The raw dataset was divided into an 80-20% train-test split, with randomized sampling used for this purpose. To enhance the model's generalization capability, different augmentation techniques like rotation, zooming, scaling of brightness, and flipping in the horizontal direction were employed. Images are resized to 300x300 pixels and normalized for consistency for the project. Medication-related queries were cleaned, tokenized, and preprocessed to support NLP model development. This pipeline ensured high-quality input data for training and evaluation. Two sample images:

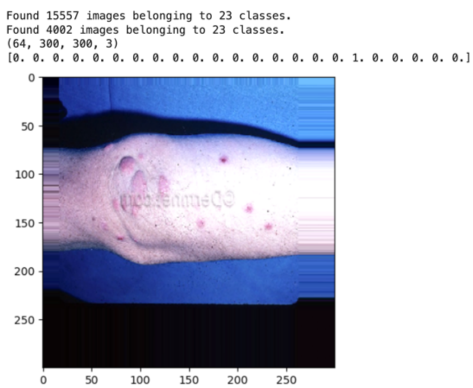

figure (01)

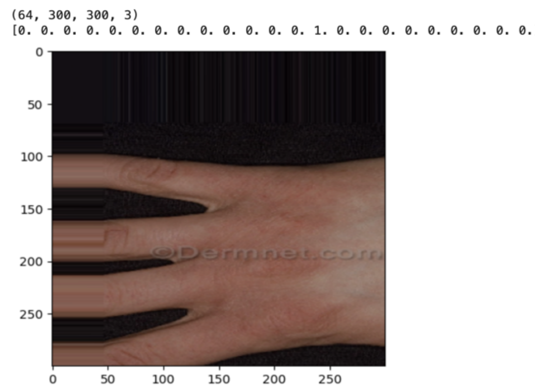

figure (02)

Phase 02: Model Development: A Convolutional Neural Network (CNN) is developed using the InceptionV3 model, pre-trained on ImageNet, as the base. This model is fine-tuned by freezing the first 249 layers and unfreezing the remaining layers for better adaptation to the skin disease classification task. Custom layers were added, including a global average pooling layer, a dense layer with 1024 units, a dropout layer, and a softmax output layer for multi-class classification. The model was compiled with the Adam optimizer, using a learning rate of 0.0001 and categorical cross-entropy loss.

```
Total params: 23,924,535 (91.26 MB)
Trainable params: 13,236,631 (50.49 MB)
Non-trainable params: 10,687,904 (40.77 MB)
```

figure (03)

The model was trained for 40 epochs [figure (04)] with early stopping and checkpoints in place to save the best-performing model based on validation loss. The training was conducted on the DermNet NZ dataset, where the CNN achieved a final training accuracy of 79.54% [figure (05)] and a final validation accuracy of 56.15%. The final training loss was 0.70, while the final validation loss was 1.87. Checkpoints were utilized to save the best model during training based on validation loss. The model's architecture was optimized to extract complex visual features such as texture and color patterns, which are essential for accurate classification of skin conditions like eczema, psoriasis, or melanoma. The final model, saved as skin_model_batch_final.h5.

```
Epoch 25/40
244/244 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.7917 - loss: 0.6696
Epoch 25: val_loss did not improve from 1.73585
244/244 ━━━━━━━━━━━━━━━━━━━━ 410s 2s/step - accuracy: 0.7916 - loss: 0.6697 - val_accuracy: 0.5615 - val_loss: 1.8680
Epoch 25: early stopping
```

figure (04)

```
WARNING:absl:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you train or evaluate the model.
123/123 ━━━━━━━━━━━━━━━━━━━━ 249s 2s/step
Test Accuracy: 79.54%
Classes in Test Generator: {'test': 0, 'train': 1, 'val': 2}
Updated Class Names: ['test', 'train', 'val']
```

figure (05)

Phase 03: Integration: The image classification and NLP models were seamlessly integrated into a chatbot system to provide a unified interface for diagnosing skin diseases and offering medication recommendations. The integration process utilized frameworks like Gradio for the user interface, TensorFlow for skin disease prediction, Transformers for the NLP-based chatbot, and additional APIs such as Cohere for treatment advice and Twilio for real-time notifications and DialoGPT-medium for conversational AI.

Framework Setup and Model Loading

To streamline the integration, all required libraries and dependencies, such as Gradio, TensorFlow, Transformers, OpenCV, and Twilio, were installed. These libraries enabled functionalities like image preprocessing, model inference, conversational responses, and real-time communication.

The image classification model, pre-trained using TensorFlow, was loaded to predict the type of skin condition based on uploaded images. The model's architecture accepts input images preprocessed to a standard size (300x300 pixels), performs classification, and outputs a label corresponding to one of the 23 predefined skin disease classes. The labels were then mapped to meaningful disease names using a disease dictionary, enhancing interpretability for end users.

Disease Severity and Treatment Suggestions

Each predicted skin condition was associated with a severity classification (e.g., Normal, Moderate, Severe) to alert users about the potential seriousness of the disease. A disease severity dictionary categorized conditions like Melanoma Skin Cancer and Bullous Disease as "Severe" and conditions like Acne and Rosacea as "Normal." For severe or risky conditions, users were prompted to consult healthcare providers immediately.

To provide treatment suggestions, the Cohere API was integrated. By generating responses using the API's large language model, relevant medications and treatments were dynamically recommended based on the diagnosed disease.

Communication via Twilio API

For critical skin conditions classified as "Severe" or "Risky," real-time alerts were sent to users through WhatsApp using the Twilio API. This feature bridges the gap between diagnosis and action by notifying users to seek immediate medical intervention. The Twilio client securely sends messages to the patient’s phone number, ensuring timely communication [figure (09)].

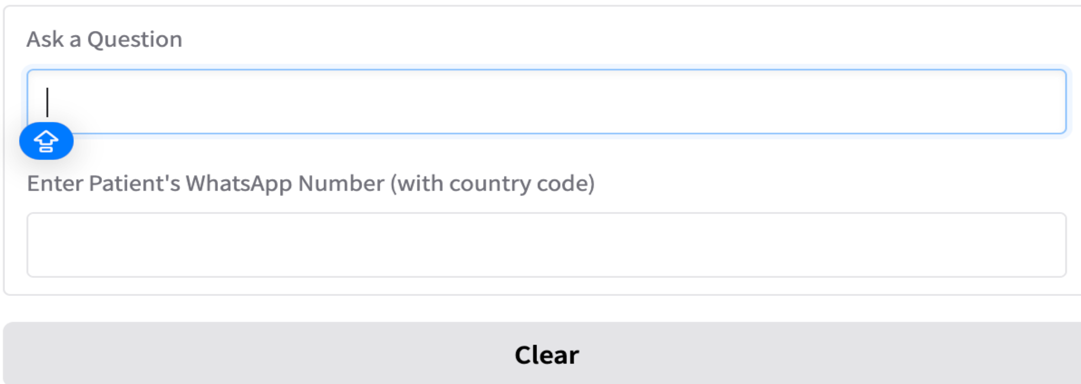

Figure (09)

NLP Chatbot Integration

To handle user queries and enhance interactivity, a pre-trained DialoGPT-medium model was integrated using the Hugging Face Transformers library. The chatbot could maintain context across conversations and generating coherent responses to user inputs. Tokenization and inference were handled efficiently using the tokenizer and model weights [figure (10)].

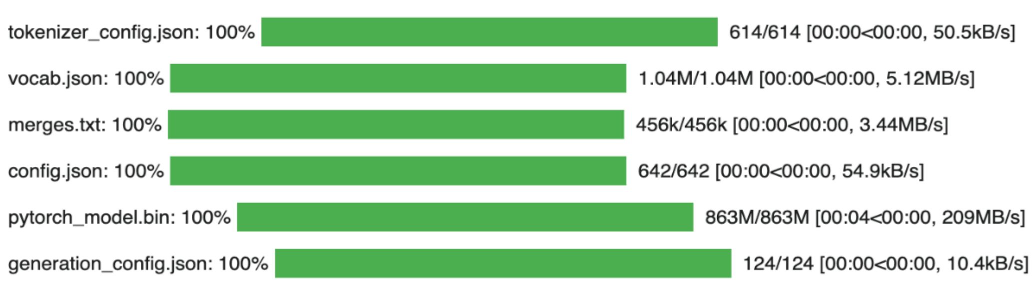

Figure (10)

Gradio User Interface

A user-friendly Gradio interface was developed to integrate all functionalities into a single platform. The interface accepts:

Image Input: Users upload skin images for disease classification.

Text Input: Queries for medication advice or explanations.

Phone Number Input: To send real-time notifications if required.

The Gradio interface combines all outputs, including predicted disease, severity, medication advice, and chatbot responses, into an organized layout for clarity. This interface was launched with options to share the application publicly.

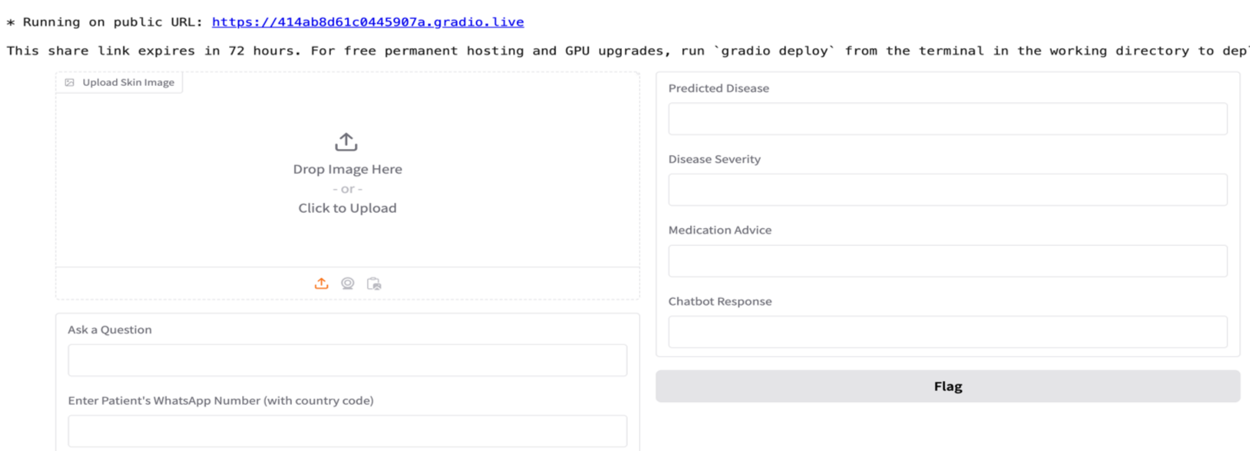

Figure (11)

Phase 04: Model Evaluation: The performance of the developed chatbot was systematically evaluated to ensure reliability in diagnosing skin conditions and providing accurate medication recommendations. The evaluation involved exploratory data analysis (EDA), visual inspections, and model performance metrics, including accuracy, precision, recall, and the F1-score. The results are presented below with detailed insights.

Exploratory Data Analysis

The class distribution of the dataset [figure (06)] revealed an evident imbalance across the training data. Some classes, such as Seborrheic Keratoses and Tinea Ringworm, contain more than 1300 images, while classes like Cellulitis Impetigo and Systemic Disease are underrepresented, with fewer than 300 samples. This imbalance can significantly impact the model's learning process, as it tends to favor majority classes, leading to poor predictions for underrepresented ones. Addressing such imbalance through data augmentation or class-weighted loss functions is critical for improving overall performance.

In addition, sample images from the Light Diseases and Disorders of Pigmentation class were visualized as a grid [figure (05)] to assess labeling quality and dataset relevance. The images displayed high clarity, consistent labeling, and diversity of pigmentation-related conditions. This step validated that the dataset effectively represents real-world skin conditions, minimizing concerns about mislabeled or low-quality images that could undermine model training.

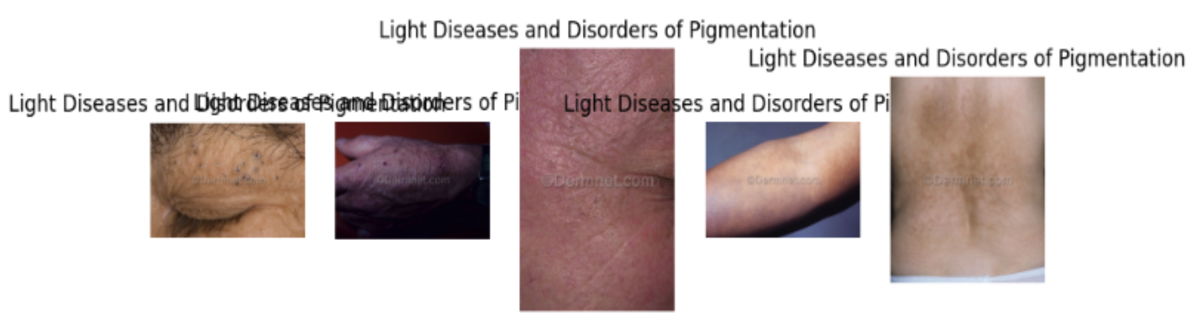

figure (05)

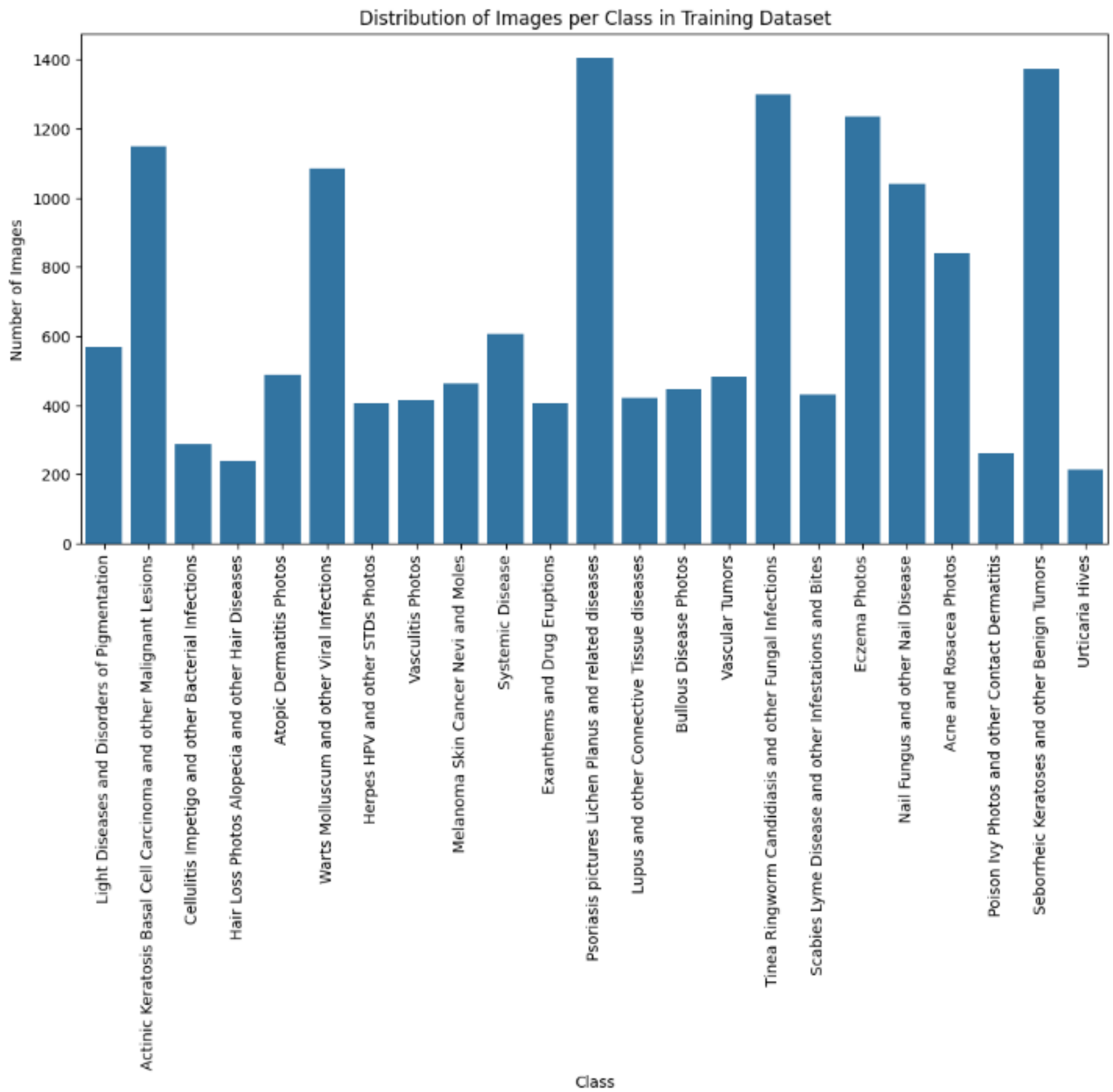

figure (06)

Classification Report

The classification report [figure (07)] provides a detailed breakdown of the model's performance for individual classes. Notably, the model performed well for majority classes such as Acne and Rosacea (F1-score: 0.84) and Nail Fungus (F1-score: 0.77), where sufficient training data was available. Conversely, classes like Systemic Disease and Cellulitis Impetigo recorded lower F1-

scores (0.37 and 0.33, respectively), likely due to insufficient samples and class overlap. These findings emphasize the need for more balanced datasets or advanced methods like data augmentation to improve generalization across all classes.

Classification Report:

	precision	recall	f1-score	support
Acne and Rosacea Photos	0.80	0.87	0.84	312
Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	0.66	0.59	0.62	288
Atopic Dermatitis Photos	0.49	0.57	0.53	123
Bullous Disease Photos	0.49	0.40	0.44	113
Cellulitis Impetigo and other Bacterial Infections	0.29	0.38	0.33	73
Eczema Photos	0.50	0.60	0.55	309
Exanthems and Drug Eruptions	0.50	0.44	0.47	101
Hair Loss Photos Alopecia and other Hair Diseases	0.49	0.72	0.59	60
Herpes HPV and other STDs Photos	0.49	0.41	0.45	102
Light Diseases and Disorders of Pigmentation	0.46	0.45	0.46	143
Lupus and other Connective Tissue diseases	0.60	0.31	0.41	105
Melanoma Skin Cancer Nevi and Moles	0.53	0.71	0.61	116
Nail Fungus and other Nail Disease	0.84	0.71	0.77	261
Poison Ivy Photos and other Contact Dermatitis	0.48	0.37	0.42	65
Psoriasis pictures Lichen Planus and related diseases	0.52	0.50	0.51	352
Scabies Lyme Disease and other Infestations and Bites	0.41	0.38	0.39	108
Seborrheic Keratoses and other Benign Tumors	0.68	0.53	0.59	343
Systemic Disease	0.59	0.27	0.37	152
Tinea Ringworm Candidiasis and other Fungal Infections	0.55	0.61	0.58	325
Urticaria Hives	0.40	0.47	0.43	53
Vascular Tumors	0.70	0.43	0.53	121
Vasculitis Photos	0.45	0.62	0.52	105
Warts Molluscum and other Viral Infections	0.45	0.66	0.53	272
accuracy			0.56	4002
macro avg	0.54	0.52	0.52	4002
weighted avg	0.58	0.56	0.56	4002

figure (07)

Confusion Matrix

To better understand misclassifications, the confusion matrix [figure (08)] was generated. The matrix highlights specific classes that the model struggles to distinguish. For instance, Actinic Keratosis was frequently misclassified as Seborrheic Keratoses, indicating potential visual similarity between these skin conditions. Such error patterns suggest that the model could benefit from enhanced feature extraction techniques or the application of transfer learning using pre-trained models like ResNet or EfficientNet.

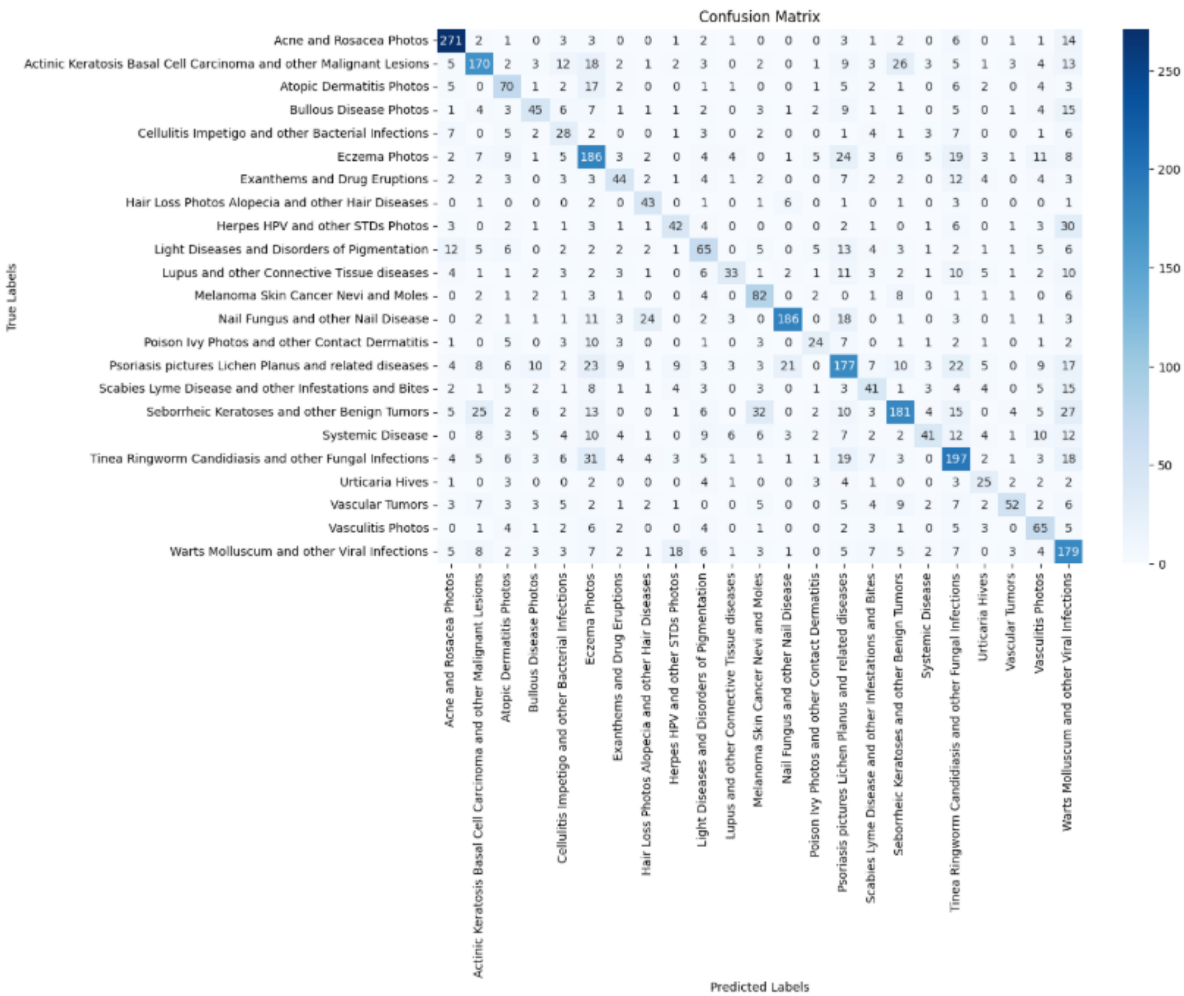

figure (08)

Phase 05: Deployment and Optimization: The chatbot system was successfully deployed and optimized to ensure accessibility, reliability, and scalability for end users. The deployment leveraged state-of-the-art APIs for model integration, secure web-based access, and text generation. This phase also outlines planned enhancements, including the incorporation of generative AI for diagnostic precision and multilingual support to expand usability across diverse user groups.

Deployment Using Hugging Face and Flask

The pre-trained models, including the TensorFlow-based image classification model for disease detection and the DialoGPT-medium model for conversational responses, were deployed using a Flask backend. Flask acted as the central processing hub for API requests, efficiently managing image uploads, model predictions, and chatbot responses. The deployment workflow included the following steps:

Image Preprocessing: Uploaded skin images were resized and normalized before being passed to the TensorFlow model for classification.

Model Inference: The skin condition was identified using the classification model, and a corresponding label was retrieved from a predefined disease dictionary.

Chatbot Interaction: User queries were tokenized using the DialoGPT-medium tokenizer, and coherent responses were generated through Hugging Face's conversational model:

This integration ensured seamless communication between the image classification and NLP components, enabling instant diagnostic feedback and conversational support.

Secure Tunneling with Ngrok

To provide secure and scalable access to the chatbot during deployment, ngrok was used for creating a secure public URL [Figure (12)]. The tool tunneled the Flask development server running locally to a publicly accessible endpoint, enabling real-time interaction with the chatbot.

The following live server link was created during the final test phase:

```
App running at: https://4564-35-243-140-153.ngrok-free.app
 * Serving Flask app '__main__'
 * Debug mode: off
INFO:werkzeug:WARNING: This is a development server. Do not
 * Running on http://127.0.0.1:5000
INFO:werkzeug:Press CTRL+C to quit
```

Figure (12)

Final Tested Output

The chatbot was rigorously tested for functionality and user experience. The final tested interface [Figure (13)] showcases:

Image Upload: Users uploaded a skin image, which the model classified accurately as "Warts Molluscum and other Viral."

Severity Classification: The predicted condition was classified as "Normal," providing reassurance to the user.

Medication Advice: The chatbot generated precise treatment recommendations, including options like Salicylic acid, Imiquimod cream, and Cryotherapy.

Chatbot Response: A follow-up user query resulted in an appropriate chatbot-generated response.

The tested output confirms the successful integration of all components, ensuring a smooth and reliable diagnostic experience.

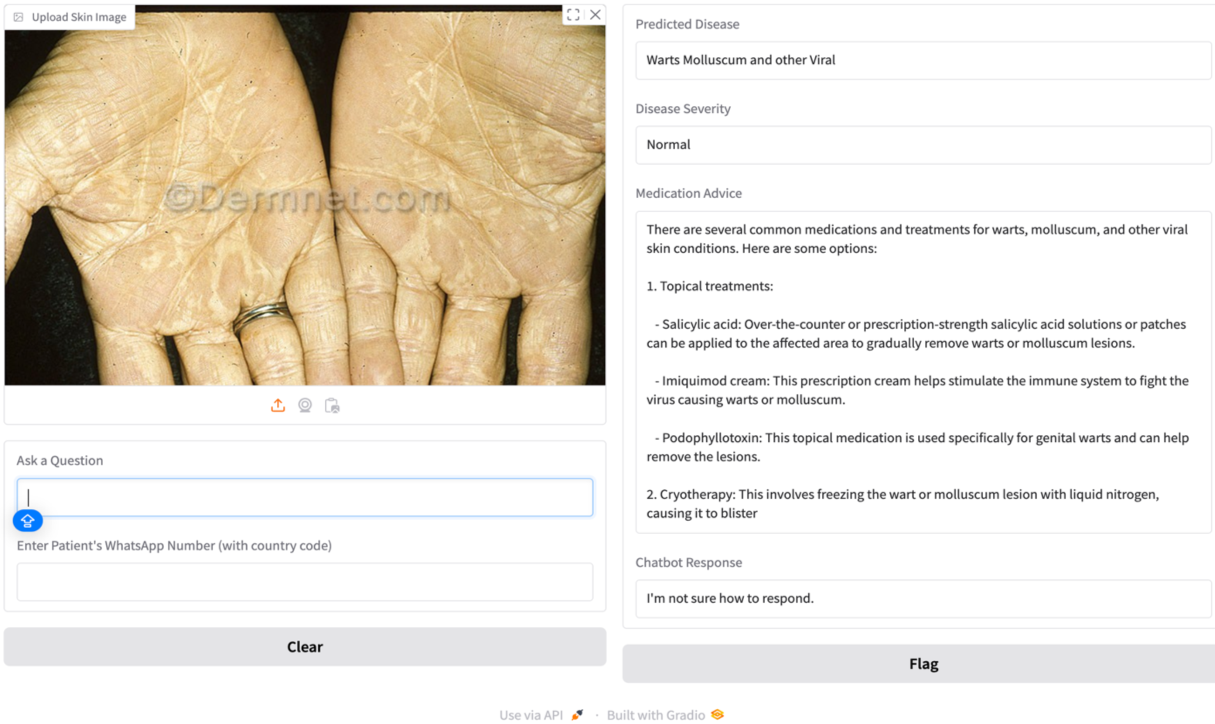

Figure (13)

CONCLUSION

Our project deals with the development of a chatbot using AI, integrated with NLP and image recognition, which will be able to provide fast and accurate diagnosis of dermatological conditions, recommending personalized medications. Using large, high-quality dermatological datasets and state-of-the-art machine learning techniques, the proposed chatbot simplifies the treatment process by improving medication accessibility and reducing doctor visits for minor conditions. Ultimately, this solution increases engagement by the patient, and simplifies diagnosis and treatment of skin health issues with ease in a non-complicated manner.

In our development phase, we created an image classification model using a comprehensive dataset of skin disease images. The model was trained over 25 epochs, achieving an accuracy of 79.54%

during training. This high level of accuracy indicates the model's strong potential for effectively classifying new data and highlights its capability for real-world skin disease recognition. With these promising results, we can further develop the AI-powered chatbot which provide reliable diagnostic feedback and personalized treatment recommendations, paving the way for more accessible and efficient dermatological care.

We also developed the chatbot's skin disease detection and NLP components. The system now allows users to upload images and ask queries for diagnostic feedback and personalized treatments. Currently, this chatbot uses the DialoGPT-medium model and Gradio interface for a user-friendly experience. Future improvements will focus on incorporating generative AI to enhance diagnostic precision and medication recommendations.

Finally, we obtained a model for skin disease classification and successful integration with conversational AI for medication advice. The chatbot's ability to combine image recognition and NLP into a single interface enhances the user experience, offering immediate diagnostic feedback and clear treatment suggestions. It successfully detects the images and give medication recommendation. The project utilized APIs like Hugging Face, OpenAI, Cohere, and ngrok for secure, scalable, and interactive application development.

The Future work will focus on refining the model's diagnostic accuracy, incorporating multilingual capabilities, and improving the chatbot's medical advice. The aim is to make dermatological care more accessible, efficient, and patient-centered, empowering individuals to take control of their skin health while minimizing unnecessary medical expenses and also saving time and energy.

REFERENCES

- Zhang, J., Zhong, F., He, K., Ji, M., Li, S., & Li, C. (2023). Recent Advancements and Perspectives in the diagnosis of skin diseases using machine learning and deep Learning: a review. *Diagnostics*, *13*(23), 3506. https://doi.org/10.3390/diagnostics13233506
- ALEnezi, N. S. A. (2019). A method of skin disease detection using image processing and machine learning. *Procedia Computer Science*, *163*, 85–92. https://doi.org/10.1016/j.procs.2019.12.090
- Pathak, C., & Ansari, N. (2023). *Chatbot-based Disease Prediction and Treatment Recommendation using AI*. K.J. Somaiya Institute of Engineering and IT.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level Classification of Skin Cancer with Deep Neural Networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056
- Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020). Human–computer collaboration for skin cancer recognition. *Nature Medicine*, *26*(8), 1229–1234. https://doi.org/10.1038/s41591-020-0942-0
- Brinker, T. J., Hekler, A., Enk, A. H., Berking, C., Haferkamp, S., Hauschild, A., Weichenthal, M., Klode, J., Schadendorf, D., Holland-Letz, T., von Kalle, C., Fröhling, S., Schilling, B., & Utikal, J. S. (2019). Deep neural networks are superior to dermatologists in melanoma image classification. *European Journal of Cancer*, *119*, 11–17. https://doi.org/10.1016/j.ejca.2019.05.023

- Kumar, D., Singh, M., & Manish Kushwaha. (2022). Kumar et al., 2021 @ IJA - Nutrient content, uptake and NUE of oats (2). *Indian Journal of Agronomy*, *66*((4)), 466–473 Dec (2021). https://www.researchgate.net/publication/359025271_Kumar_et_al_2021_IJA_-_Nutrient_content_uptake_and_NUE_of_oats_2

Git Hub: https://github.com/YeswanthReddychereddy/AI-Driven-NLP-Chatbot-for-Skin-Disease-Diagnosis-and-Medication-recommendations