Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 12, No. 4, December 2024, pp. 1040∼1062
ISSN: 2089-3272, DOI: 10.52549/ijeei.v12i4.5806

DermAI: An Innovative AI-Driven Chatbot for Enhanced Dermatological Diagnosis and Patient Interaction

Pradeep Rajeshkumar[1], Shubhangi Kharche[1], Prithvi Poojari[1], Sachet Utekar[1], Sahil Saini[1], Samrridhi Bdwai[2]
[1]Department of Electronics & Computer Science, SIES Graduate School of Technology, Nerul, Navi Mumbai
[2]People-oriented Data Scientist, Consultant, Mumbai

Article Info

Article history:

Received Aug 6, 2024
Revised Dec 8, 2024
Accepted Dec 16, 2024

Keywords:

Deep Learning (DL)
Transfer Learning (TL)
Natural Language Processing (NLP)
Chatbot
Dermatology
Retrieval Augmented Generation (RAG)
Large Language Model (LLM)
Telemedicine platform

ABSTRACT

Skin disorders constitute a noteworthy public health concern globally, with earnest impacts on both physical and mental well-being. However, effective dermatological care faces challenges in resource-limited regions due to poor infrastructure, limited access to medical facilities & expertise, and inadequate advanced diagnostic tools. The existing research work majorly focuses on cancer and uncommon skin diseases with models trying to achieve a higher training accuracy with no regards to misclassification rate. The products currently available in the market provide a limited initial diagnosis and suggest consulting a doctor to get an accurate diagnosis or offer a list of other possible skin disorders. To address these challenges, we propose DermAI, an innovative AI-based Chatbot made entirely of open-source technologies, which integrates the ResNet-50 model and LLM via Chainlit, with Retrieval Augmented Generation(RAG), utilising AstraDB vector database and OpenAI embedding model for personalised responses. enabling accurate classification of common skin diseases. The proposed DermAI ensures minimal misclassification and comprehensive coverage of diseases, leveraging Retrieval-Augmented Generation and comparative model analysis. The metrics indicate that the model has a high true positive rate, with a misclassification rate of 2.17%, mean sensitivity, specificity & AUC of 92.6%, 99.8% & 99.9% respectively. This is demonstrated in the situations of melanoma, chickenpox, shingles, impetigo, and nail fungus, where it obtained 100% validation accuracy, a feat not attained by previous studies. Additionally, the model is highly capable of correctly identifying negative cases. The hallucination metric suggests the model may have a minimal tendency to hallucinate as the average hallucination score of 7% which falls far within the manually set threshold value of 50%. By setting the threshold value to 50%, the model generates grounded answers that are pertaining to the knowledge base and also allows it to be flexible with its responses. Overall, DermAI outperforms all solutions proposed in research literature.

Corresponding Author:

Shubhangi Pravin Kharche.
Department of Electronics & Computer Science, SIES Graduate School of Technology.
Nerul, Navi Mumbai, India.
Email: shubhangik@sies.edu.in

1. INTRODUCTION

Skin health is an essential aspect of our overall well-being, and dermatological issues can have a significant impact on our quality-of-life (QoL). However, accessing timely and accurate information about skin conditions can sometimes be challenging. Skin diseases should be tackled as soon as possible because it can result in increased risk of other health problems, such as heart disease, stroke, and diabetes [1]. To address these challenges and to empower individuals to take better care of their skin, the paper proposes an AI based chatbot termed DermAI. All Previous efforts in existing literature could only classify the severity of skin lesions but not the other skin-related diseases. If the segmentation of the diseased area is computed together with the classification of skin diseases, the treatment of patients will be more effective. Numerous challenges, including geographical barriers, scarcity of dermatologists in rural regions [2], scheduling constraints, and more, hinder individuals from accessing timely and expert guidance for their skin-related concerns. Furthermore, current research predominantly focuses on cancer and rare skin diseases, prioritising training accuracy over minimising misclassification rates. What's lacking are highly accurate OpenAI-based chatbot solutions for various skin disease detection, readily integrable into telemedicine platforms. The intent of our research implementation is to develop & provide a system that can be utilized to diagnose skin diseases autonomously utilising the images captured through cell phone cameras. The system can help the users understand various skin conditions, symptoms, and treatment options. The system also empowers the user with self-care tips, preventive measures, and lifestyle choices to maintain healthy skin. Hence creating a cost-efficient system that can help diagnose skin disease in rural areas with fewer facilities, thereby optimising healthcare resources. To facilitate disease classification, DermAI integrates ResNet-50 [3], a DL model trained with a diverse dataset of common skin disease images. The ResNet-50 model is integrated into the system using Chainlit [4], an open-source Python package that enables seamless integration of machine learning models into production-ready conversational AI applications. This integration ensures efficient and effective disease diagnosis directly from user-captured images. Trying to tackle the limitations of existing work, a comparative study of five different models (ResNet-50, VGG-19, Inception-V3, MobileNet-V2, and EfficientNet-B4) for the common diseases' dataset was performed and recorded. DermAl utilizes sophisticated technologies like Retrieval Augmented Generation (RAG), OpenAI's embedding model, and AstraDB [5], which acts as the foundation for knowledge storage and retrieval within DermAl, to enhance conversation between the system and user. RAG is integrated into DermAI to provide contextual understanding and improve the accuracy of disease classification. OpenAI's embedding model is being used to represent textual information in a rich semantic space. The embedding model encodes text inputs into high-dimensional vectors, capturing semantic relationships and contextual nuances. The organized flow of the paper is depicted in Figure1.

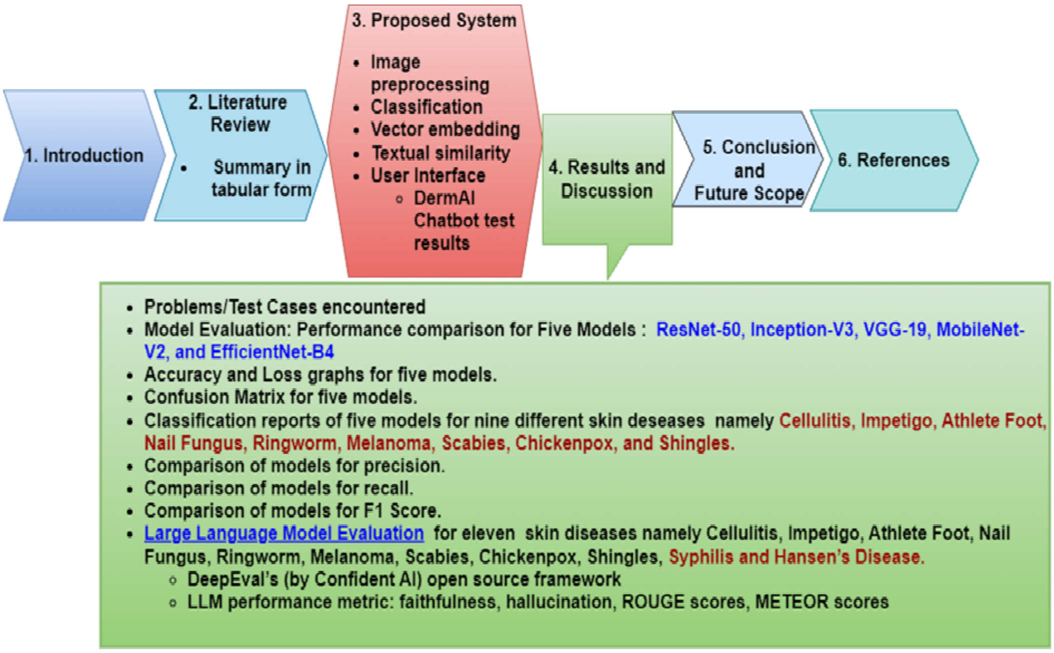

1. Introduction
2. Literature Review
• Summary in tabular form
3. Proposed System
• Image preprocessing
• Classification
• Vector embedding
• Textual similarity
• User Interface
∘ DermAI Chatbot test results
4. Results and Discussion
5. Conclusion and Future Scope
6. References
• Problems/Test Cases encountered
• Model Evaluation: Performance comparison for Five Models : ResNet-50, Inception-V3, VGG-19, MobileNet-V2, and EfficientNet-B4
• Accuracy and Loss graphs for five models.
• Confusion Matrix for five models.
• Classification reports of five models for nine different skin deseases namely Cellulitis, Impetigo, Athlete Foot, Nail Fungus, Ringworm, Melanoma, Scabies, Chickenpox, and Shingles.
• Comparison of models for precision.
• Comparison of models for recall.
• Comparison of models for F1 Score.
• Large Language Model Evaluation for eleven skin diseases namely Cellulitis, Impetigo, Athlete Foot, Nail Fungus, Ringworm, Melanoma, Scabies, Chickenpox, Shingles, Syphilis and Hansen's Disease.
∘ DeepEval's (by Confident AI) open source framework
∘ LLM performance metric: faithfulness, hallucination, ROUGE scores, METEOR scores

Figure 1. Organisation of Paper

2. LITERATURE REVIEW

Skin diseases represent a significant burden on public health worldwide, ranging from common conditions like psoriasis and acne to more severe and rare disorders. Traditional diagnostic approaches often rely on manual examination by dermatologists, leading to time-intensive processes, subjectivity, and limited accessibility, especially in remote or underserved areas. Recent advancements in ML and AI have inspired the development of programmed skin disease detection systems, offering the prospective for cost-effective, accurate, and rapid diagnosis. This literature review delves into several neoteric research efforts in this sphere, fixating on the utilisation of ML/DL techniques/algorithms for skin disease detection and classification. The literature review encompasses a wide range of studies spanning from 2017 to 2024, examining diverse methodologies and approaches to skin disease detection. Notable contributions include studies utilizing deep convolutional neural networks (CNNs) for dermatologist/doctor level classification of skin cancer, such as the work by Esteva et al. [22], which demonstrated the comparable performance of CNNs to board-certified dermatologists in diagnosing melanomas. With just pixels and disease labels as inputs, he showed how to classify skin lesions using a single CNN that was trained end-to-end (E2E) from images to differentiate between different skin lesions. His results were encouraging and point to the possibility of computer-aided skin disease diagnosis [6]. Another research, by Hu et al. [13], explores the integration of CNNs with chatbot technology for personalised feedback and guidance in skin cancer detection. Furthermore, studies by Verma et al. [20] and Huang et al. [15] highlight the potential of ensemble data mining techniques and image processing-enabled chatbots, respectively, in improving diagnostic accuracy and accessibility. However, these studies primarily focus on cancerous or rare skin diseases, leaving a gap in addressing common dermatological concerns. Additionally, challenges such as dataset bias, interpretability issues, and ethical considerations pose significant hurdles to the widespread adoption of AI-driven diagnostic tools. C. Barata et al. [23] employed two distinct methods for the identification of melanomas in dermoscopic pictures. While the second system makes use of the bag-of-features classifier along with local features, the first system classifies skin lesions using global approaches. To improve the melanoma detection accuracy which is essential for early diagnosis and treatment, the authors combine texture and colour information. Sensitivity & Specificity of 96% & 80% respectively for global methods against Sensitivity & Specificity of 100% & 75% respectively for local methods indicate that both approaches produce extremely good outcomes [23]. In 2018, the primary aim or challenge was to promote research in the automated analysis of skin lesions [21], with a specific focus on early detection of melanoma, a lethal type of skin cancer. The aim was to evaluate and compare various computer vision and machine learning techniques for the automated diagnosis of skin lesions. This included tasks like lesion classification, segmen-

Table 1. Literature Review for performance measures

TA: % Training accuracy, VA: % validation accuracy, MCR: Miss classification rate,
AUC: % area under ROC, Se: % sensitivity, Sp: % specificity, A: % Accuracy, P: % Precision.

Refe-rence	**Lesion Detection**	**Performance Measures**	**Chat-bot solution**	**LLM-/RAG?**	**Tele-medicine Platform**
Pro-posed work	Yes	TA 100, VA 97.35, MCR 2.214, AUC 99.90, Se 92.60, Sp 99.80	Yes	Yes	Yes
[6]	Yes	P 89.55, Se 90.12, Sp 90.24	No	No	No
[7]	No	A 93	Yes	Yes	No
[8]	Yes	A 90.8	No	No	No
[9]	Yes	A 75.25, P 82.77	No	No	No
[10]	Yes	A 99.8	No	No	No
[11]	Yes	AUC 82.2	No	No	No
[12]	Yes	A 90	Yes	Yes	No
[13]	Yes	TA 96, VA 93	No	No	No
[14]	No	A 92.23	Yes	No	No
[15]	No	A 93	No	No	No
[16]	Yes	A 89.90	No	No	No
[17]	No	A 95.20, Se 81.70, Sp 97.10	No	No	No
[18]	Yes	A 88	No	No	No
[19]	Yes	A 50-100	No	No	No
[20]	Yes	A 97.27	No	No	No
[21]	Yes	P 91	No	No	No
[22]	Yes	P 94.50	No	No	No
[23]	Yes	Se 96, Sp 80	No	No	No

Table 2. Literature Review for Models Used and Skin Diseases Detected

Ref.	Publication Year	Model	Skin Diseases Detected
Proposed work	Proposed work	The DermAi Chatbot, powered by chainlit for seamless ResNet-50 model integration, Retrieval Augmented Generation (RAG)	Eleven diseases namely Cellulitis, Impetigo, Athlete Foot, Nail Fungus, Ringworm, Melanoma, Scabies, Chickenpox, Shingles, Syphilis, and Hansen's Disease for LLM evaluation.
[24]	2024	ML, DL, Convolutional deep (CD), spiking neural networks (SNN)	Skin cancer
[25]	2024	SVM and AlexNet CNN models	Skin lesions
[26]	2024	Multiheaded convolutional neural network (CNN)	Skin lesions, Melanoma; Cepstrum
[27]	2024	Conditional generative adversarial networks (conditional GANs)	Melasma diagnosis
[28]	2024	Proposed HierAttn neural network	Skin lesions
[29]	2024	MobileNetV2	Monkeypox diagnosis
[30]	2024	Generative adversarial networks	Monkeypox detection
[31]	2023	CNN, LBP	Skin cancer
[32]	2021	Targeted Ensemble Machine Classification Approach (TEMCM)	Skin disorders
[33]	2021	Mobile deep neural network, MobileNetV3-Small	Herpes Zoster
[34]	2020	Proposed self-paced balance learning (SPBL) algorithm	Eczema
[35]	2019	CNN	Six common skin diseases [seborrheic keratosis (SK), squamous cell carcinoma (SCC), actinic keratosis (AK), rosacea (ROS), basal cell carcinoma (BCC), and lupus erythematosus (LE)]
[36]	2018	Computer vision	Commonly seen skin diseases
[21]	2017	Deep learning ensembles	Melanoma recognition

tation, melanoma detection and lay stress on improving the accuracy & efficiency of melanoma detection via computational methods. In 2021, K. Glock et al. [17] discussed the use of advanced ML techniques, principally deep CNNs & transfer learning (TL), which abolish the necessity for deep nets to be trained from the beginning when it comes to the singling out of measles rashes. With a sensitivity, accuracy & specificity of 81.7%, 95.2%, & 97.1% respectively the authors [17] successfully spotted the measles rashes. In the work [17], the MobileNet model was employed, using TL for seven skin disorders, to produce a lightweight model for setting a skin disease classification system moulded for an Android application. The work resulted in an accuracy of 84.28% with the default preprocessing of the input data combined with an under-sampling technique. However, a 93.6% accuracy was achieved with an unbalanced dataset & default input data preprocessing. After that, the researchers experimented with an oversampled dataset such that the model obtained an accuracy of 91.8%. Finally, accuracy of 94.4% is achieved by preprocessing the input data with the oversampling technique & data augmentation. Vayadande et al. [6] introduced a work deploying EfficientNet-B2 to provide an accuracy & precision of 89.55% & 90.12% respectively, exhibiting propitious results in lesion detection. Panagoulias et al. [7] accorded a multimodal evaluation aiming on LLM and GPT-4-Vision-Preview, suggesting advancements in AI-based skin disease diagnosis. The authors [8] proposed a multi-class SVM model that obtained an accuracy of 90.8%, with inherent limitations on scalability. Ansari et al. [9] developed a hybrid CNN-SVM model with an accuracy and precision of 75.25% & 82.77% respectively, showcasing the need for further research to address limitations & validate real-implementation fulfillment. The authors [10] introduced Dermatec, an AI-based platform achieving a magnificent accuracy (99.8%). Chiu et al. [11] used their built datasets and various CNN architectures, spotlighting the importance of colour contrast analysis for enhanced interpretability. Recent studies [11, 24, 25] illustrated advancements in ML/DL techniques, showcasing superior performance in skin disease diagnosis. These include Convolutional-Deep Spiking-Neural-Networks (CD-SNN), SVM, AlexNet CNN, and multi-headed CNN models, spotlighting accuracy, dependability, and efficiency. Thus, Skin disease classification/ detection have seen significant advancements in recent years, driven by the integration of ML-DL techniques. The literature review synthesises findings from various studies, focusing on model architectures, performance measures, and limitations. However, the reported research work lacks provision of chatbot solutions, utilisation of large language models (LLMs) or retrieval augmented generation (RAG), and support for

Table 3. Market Research

Name	Core Offerings	Target Condition	Market Position
Google Lens [37]	Image recognition technology	Broad application but not specifically tailored to skin diseases	A general-purpose tool with capabilities that can be extended to dermatology but lacks a specialized focus
Nirmai [38]	Breast cancer detection	Breast cancer only	Highly specialized and effective in its niche but not applicable to other skin conditions
Qure.ai [39]	Medical imaging analysis	Not focused on skin diseases; primarily targets brain, chest, and musculoskeletal conditions	Strong presence in radiology but not relevant to dermatology
SigTuple [40]	Blood and urine analysis	Blood and urine abnormalities	Relevant in clinical diagnostics but not in dermatological care
Mfine [41]	Telemedicine platform with chatbot integration	Acne and other common dermatological issues	A versatile platform providing direct-to-consumer medical consultations, including dermatology
SkinVision [42]	Melanoma and mole analysis	Melanoma and other problematic moles	Highly specialized in skin cancer detection, with strong clinical validation
CureSkin [43]	Dermatological consultation and skincare solutions	Acne, pigmentation, open pores, scars, and dandruff	A comprehensive platform providing AI diagnostics followed by professional advice
FirstDerm [44]	Dermatological consultation via mobile app	General skin conditions like rashes and bites, not AI-based	Simple, easy-to-use platform but lacks AI integration for diagnostics
Model Dermatol [45]	AI-based skin disease classification	Shingles, melanoma, and other skin diseases	Strong AI integration for dermatological diagnostics with a wide range of conditions covered
Mama Earth [46]	Skincare products	General skincare needs	Focuses on natural skincare products, without medical diagnostic capabilities
Cipla Limited [47]	Pharmaceutical products	General medical and skincare products	Established pharmaceutical company with a broad product range
iDoc24 [48]	Image-based dermatology consultations	General skin conditions but not comprehensive for all diseases	Convenient for users but lacks comprehensive AI support
Himalaya Wellness Company [49]	Herbal and natural products	General wellness and skincare	Well-known for natural products, without diagnostic capabilities
Aysa by Visualdx [50]	AI-driven dermatological diagnosis	Wide range of dermatological conditions	Robust AI platform with a broad diagnostic scope
mCaffeine [51]	Skincare products	General skincare needs	Focuses on caffeine-based skincare products without diagnostic tools

integration with telemedicine platforms (Table1). The literature encompasses diverse approaches, ranging from CNN-based models to ensemble methods and conditional generative adversarial networks (GANs), addressing a wide spectrum of skin diseases (Table2). Furthermore, it should be highlighted, nevertheless, that every study used a distinct sample of the complete dataset, leaving out instances with low-quality images. This makes a fair comparison of all the works reported impossible. Additionally, the proposed work showcases the integration of ResNet-50 with RAG, achieving high sensitivity, specificity, and AUC, indicating a robust model for skin disease identification. We used all the photos from the standard datasets in our research and replicated the results using other designs. Another challenge in comparing different methodologies is that different studies have reported their findings using different measures such as sensitivity, specificity, or accuracy, among others. Furthermore, the products currently available in the market (Table3) provide a limited initial diagnosis and suggest consulting a doctor to get an accurate diagnosis or offer a list of other possible skin disorders. To address these challenges, we propose an innovative AI-based solution made entirely of open-source technologies, a Chatbot DermAI integrated with image processing capabilities.

3. METHODOLOGY

3.1. Proposed System for Classification and Information Retrieval

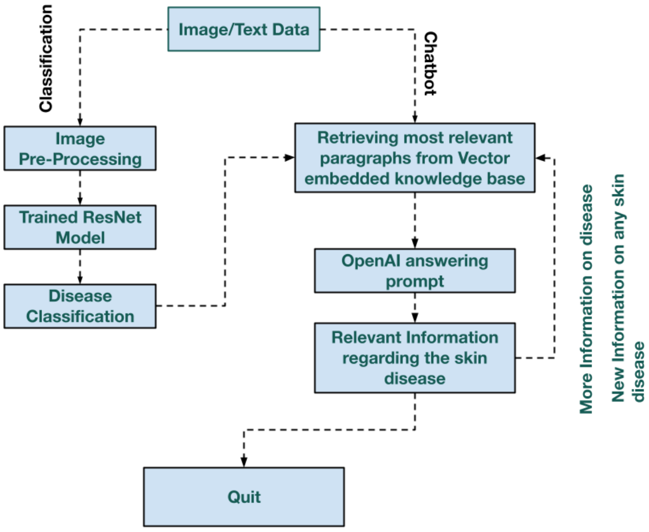

Image/Text Data
Classification
Chatbot
Image Pre-Processing
Trained ResNet Model
Disease Classification
Retrieving most relevant paragraphs from Vector embedded knowledge base
OpenAI answering prompt
Relevant Information regarding the skin disease
More Information on disease
New Information on any skin disease
Quit

Figure 2. Proposed system for classification and information retrieval

The workflow begins with a user interface, powered by Chainlit [4]. The user interface provides a welcome message followed by two options, which are information related to skin diseases and image analysis. Based on the user selection, an appropriate path is selected. In image analysis, the user provides image input, which undergoes preprocessing to attain the necessary size and scaling for subsequent processing. Following preprocessing, the image is forwarded to the trained ResNet50 model for prediction. The predicted output is presented to the user along with the information related to the predicted disease. This is performed with the help of LLM. In the information related to the diseases, the relevant information is retrieved from the vector databases, created from the self-curated knowledge base provided, based on certain keywords in the user queries with the help of Retrieval Augmented Generation (RAG). The retrieved information is then fed to OpenAI's embedding model, which simply acts as an interpreter or a translator. OpenAI makes sure that the retrieved information is meaningful and makes sense to the user by modifying the information. This information is then presented to the user through the user interface. Figure2 depicts the workflow schema of the system.

3.2. Dataset Distribution

Figure3 illustrates the distribution of samples between training and validation datasets for various skin conditions, highlighting the data split for each category. Most categories show an approximately 80-20 split between training and validation data, ensuring a standard ratio for model development and evaluation. The ”Non-diseased” category has significantly more samples, overshadowing other categories as one of the aims was to reduce misclassification rate. The images were scraped and consolidated from an archive, called DermNet [52], which is doctor-approved, HuggingFace dataset and some real-world images as well.

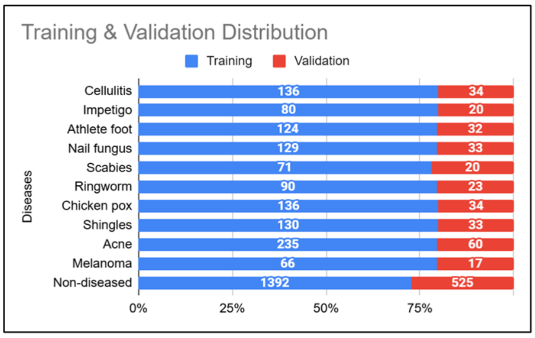


Figure 3. Overall Dataset Distribution

3.3. Image Pre-processing

In this work, the adoption of transfer learning in image processing is driven by the limitations of traditional techniques, which often require the explicit definition of rules and features for the extraction of relevant information from images. Transfer learning, on the other hand, adopts a data driven approach. The pre-trained model, in this case, ResNet-50 [3], has already learned rich hierarchical features from a diverse set of images during its training on ImageNet. This learned knowledge is then transferred and fine-tuned for the specific task of recognizing and classifying skin conditions. This approach enhances accuracy and generalization, making DermAI adapt at recognizing and classifying a broad spectrum of dermatological issues. We leverage the OpenCV library for efficient image loading and resizing, contributing to the overall robustness and effectiveness of our computational model. Furthermore, resizing the images is imperative to comply with the input requirements of our deep learning model, depicted in Figure4. The seamless integration of OpenCV facilitates a standardized preprocessing pipeline, ensuring that all images are consistently transformed before being fed into the deep learning model. This attention to preprocessing details is vital for maintaining the integrity of the model's training and validation processes.

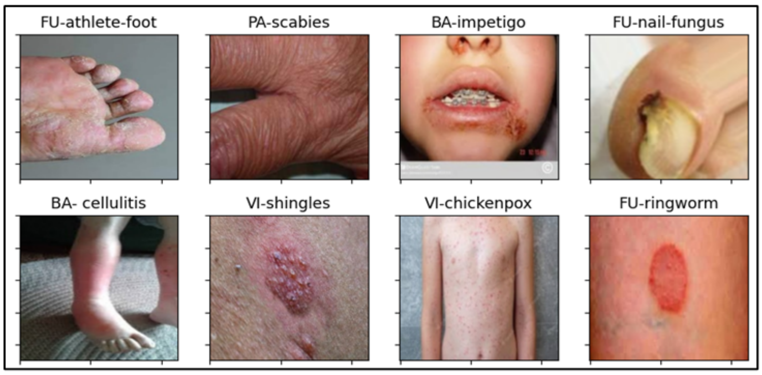


Figure 4. Pre-processed skin images [14]

3.4. Classification

This research employs a powerfully built classification process to validly identify and bracket skin diseases based on dermatological images. To excerpt features and catalogue data, the approach uses DL techniques amidst a trained ResNet-50 model. ResNet-50, a member of the residual network family, uses shortcut connections inside its 50-layer structure to solve the vanishing gradient issue in deep learning. These connections help to train deeper networks for complex feature extraction by facilitating gradient flow during backpropagation. Initially loaded with ImageNet weights, ResNet-50 shines in image classification, especially when fine-tuned with a Global Average Pooling (GAP) layer that condenses feature maps into a fixed-size vector, preparing it for deeper analysis of patterns [18]. To prepare the ResNet-50 model for a particular image classification task, the top layers of the model are altered. With the addition of the GAP layer, the spatial information is streamlined, resulting in a dense layer with ReLU activation that captures complex patterns in the data. The final section of the architecture is a SoftMax layer, designed for multiclass classification, where each unit represents a separate type of skin condition [18]. The weights of the ResNet50 layers are frozen to avoid additional changes during the early training phases, protecting the important information learned during pre-training. This arrangement makes sure the model can still identify common characteristics while enabling the upper layers to adjust to the particulars of our skin condition categorization task.

3.5. Vector Embedding

Vector embeddings convert textual data into vector spaces which is crucial for capturing contextual nuances and semantic relationships in machine learning applications. Our decision to use GPT as the language model stems from its sophisticated capacity to produce responses and interactions by keeping the provided context into consideration. AstraDB [5], a distributed NoSQL database intended for vector storage, was used to manage and store vector embeddings. This configuration allowed scalability and flexibility for analysis and retrieval of vectors by storing them in tables.This process is made smoother with the help of a wrapper that enables the chatbot to retrieve necessary and relevant vector representations in response to user queries. Figure5 illustrates the Retrieval-Augmented Generation (RAG) method, a cutting-edge natural language processing (NLP) framework that combines generative and retrieval-based model capabilities to accomplish complex performance goals and it retrieves relevant data from an existing knowledge base or corpus before generating the text response. On the other hand, the standard generative models that rely only on input prompts to generate text response. The user types in a prompt first. The application receives the user prompt and converts it to a vector representation. This model compares documents in a vector database to the user's prompt in order to find the most appropriate and suitable information. Following that, it uses a selected distance measure to order the results that are most suitable to the user's prompt. Finally, the user's prompt is combined with the most appropriate data from the vector database before being delivered to the model and the application sends the required response that the model generates back to the end user.

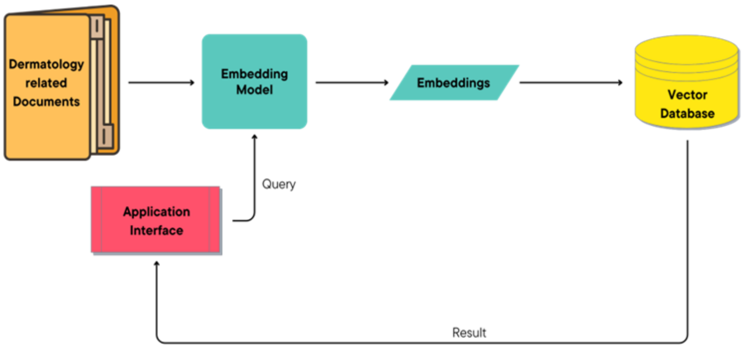


Figure 5. Retrieval-Augmented Generation flow

3.6. Textual Similarity

Textual similarity refers to the degree of similitude between two pieces of text. Measuring the cosine of the angle between two vectors that represent the text documents yields the cosine similarity measure, which is often used to assess textual similarity. The LLM employs cosine similarity for retrieving relevant context, in the form of vectors, from the knowledge base for generating responses, thereby identifying the most relevant information in the process. This is incorporated to match the symptoms existing in the knowledge base and suggest diagnosis for the user's queries. Here, we are comparing the similarity between two vectors using cosine similarity. Cosine does not require vectors to be normalised [36]. For two vectors A and B, the dot product of the vectors divided by the product of their magnitudes yields the cosine similarity :

$$\text{sim}_{\text{Astra, cosine}}(A, B) = \frac{1 + \frac{A \cdot B}{\|A\| \times \|B\|}}{2} \in [0, 1] \tag{1}$$

When returned by Astra DB, the result is a similarity score which is a number between 0 and 1:

A value of 0 indicates that the vectors are diametrically opposed. A value of 0.5 suggests the vectors are orthogonal (or perpendicular) and have no match. A value of 1 indicates that the vectors are identical in direction [5].

3.7. User Interface

The DermAI's interface, which was developed using Chainlit [4], provides users with a platform for help with skin issues. It has two types of interaction: examining skin diseases through image upload, followed by information related to the disease, or textual queries for information retrieval. The users can choose the method that best meets their needs. The user interface is capable of handling user inputs steadily, dynamically generating context-specific queries and responses. Figure6 shows the chatbot's Google Authentication Interface. Google's OAuth (Open Authorization) protocol is used in the Google authentication process to allow users to securely log in to applications using their Google account credentials. If a user opts to log in using Google, they are taken to a login page where they can verify their identity. Upon successful authentication, Google gives the application an access token that may be used to retrieve user data and redirects the user to the application's interface.

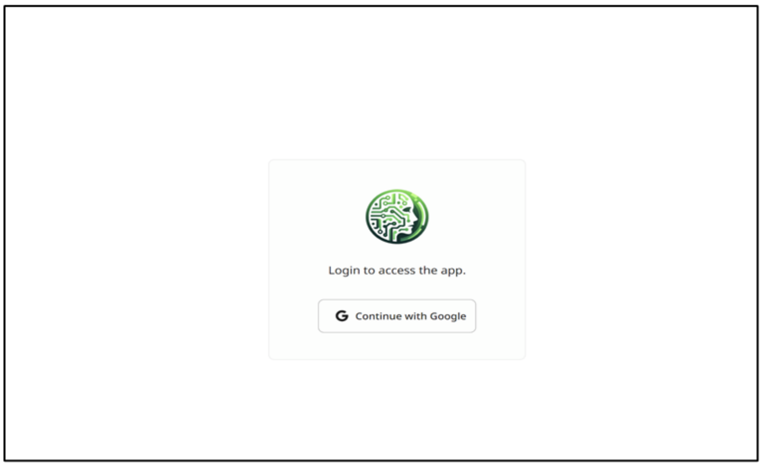
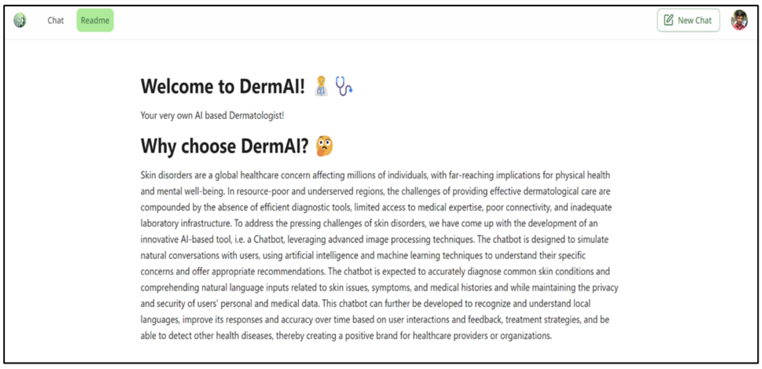

Login to access the app.

Figure 6. Login Interface



Welcome to DermAI! 🧑‍⚕️🩺

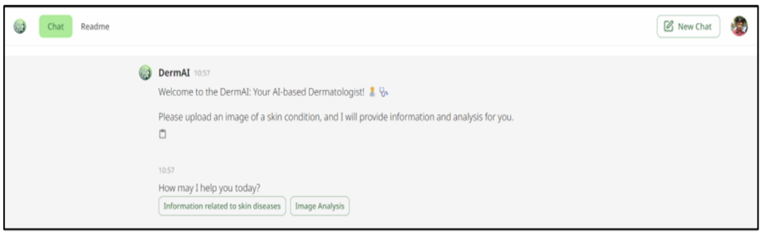
Your very own AI based Dermatologist!

Why choose DermAI? 🤔

Skin disorders are a global healthcare concern affecting millions of individuals, with far-reaching implications for physical health and mental well-being. In resource-poor and underserved regions, the challenges of providing effective dermatological care are compounded by the absence of efficient diagnostic tools, limited access to medical expertise, poor connectivity, and inadequate laboratory infrastructure. To address the pressing challenges of skin disorders, we have come up with the development of an innovative AI-based tool, i.e. a Chatbot, leveraging advanced image processing techniques. The chatbot is designed to simulate natural conversations with users, using artificial intelligence and machine learning techniques to understand their specific concerns and offer appropriate recommendations. The chatbot is expected to accurately diagnose common skin conditions and comprehending natural language inputs related to skin issues, symptoms, and medical histories and while maintaining the privacy and security of users' personal and medical data. This chatbot can further be developed to recognize and understand local languages, improve its responses and accuracy over time based on user interactions and feedback, treatment strategies, and be able to detect other health diseases, thereby creating a positive brand for healthcare providers or organizations.

Figure 7. Description of the system

The README section of the user interface, as shown in Figure7, plays a critical part that gives users an in-depth overview of the application and its capabilities. This section tries to inform the users on the functionality and goal of the program, providing information on how it can help recognize and comprehend skin-related problems. The README may also contain instructions on how to use the interface, find sites for further skin health-related information, and navigate its features.



DermAI 10:57

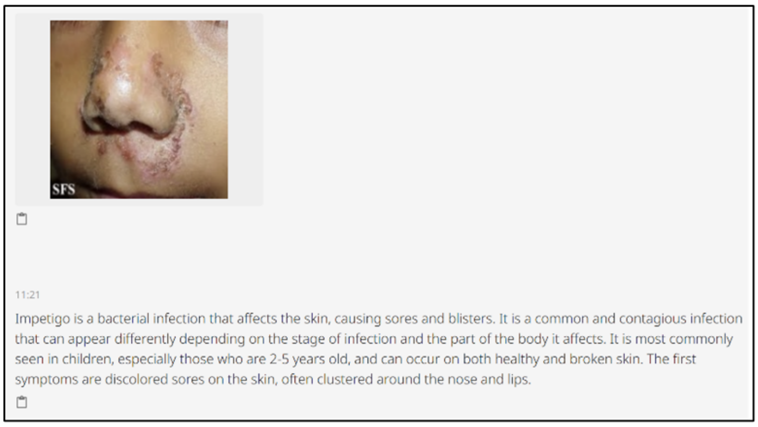
Welcome to the DermAI: Your AI-based Dermatologist!

Please upload an image of a skin condition, and I will provide information and analysis for you.

10:57

How may I help you today?

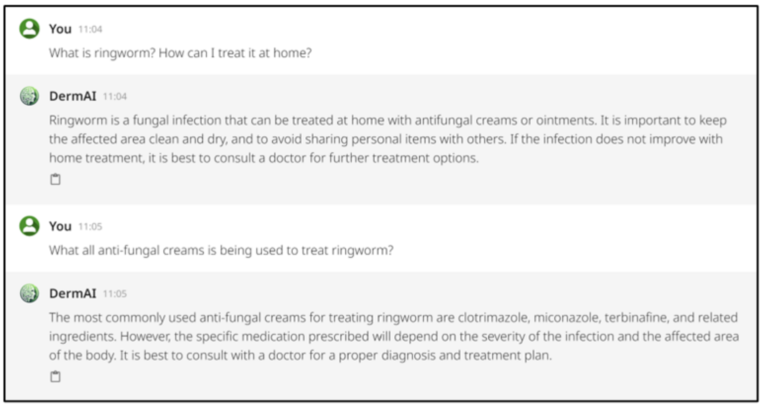
Figure 8. Prompt for user selection



11:21

Impetigo is a bacterial infection that affects the skin, causing sores and blisters. It is a common and contagious infection that can appear differently depending on the stage of infection and the part of the body it affects. It is most commonly seen in children, especially those who are 2-5 years old, and can occur on both healthy and broken skin. The first symptoms are discolored sores on the skin, often clustered around the nose and lips.

Figure 9. Image Analysis



You 11:04

What is ringworm? How can I treat it at home?

DermAI 11:04

Ringworm is a fungal infection that can be treated at home with antifungal creams or ointments. It is important to keep the affected area clean and dry, and to avoid sharing personal items with others. If the infection does not improve with home treatment, it is best to consult a doctor for further treatment options.

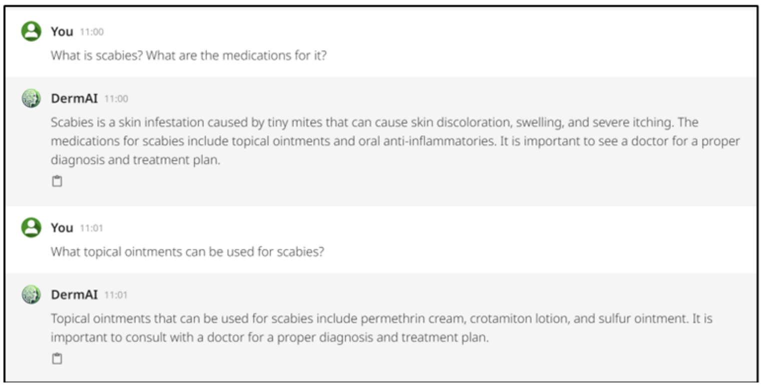
You 11:05

What all anti-fungal creams is being used to treat ringworm?

DermAI 11:05

The most commonly used anti-fungal creams for treating ringworm are clotrimazole, miconazole, terbinafine, and related ingredients. However, the specific medication prescribed will depend on the severity of the infection and the affected area of the body. It is best to consult with a doctor for a proper diagnosis and treatment plan.

Figure 10. Follow-up queries on prediction



You 11:00

What is scabies? What are the medications for it?

DermAI 11:00

Scabies is a skin infestation caused by tiny mites that can cause skin discoloration, swelling, and severe itching. The medications for scabies include topical ointments and oral anti-inflammatories. It is important to see a doctor for a proper diagnosis and treatment plan.

You 11:01

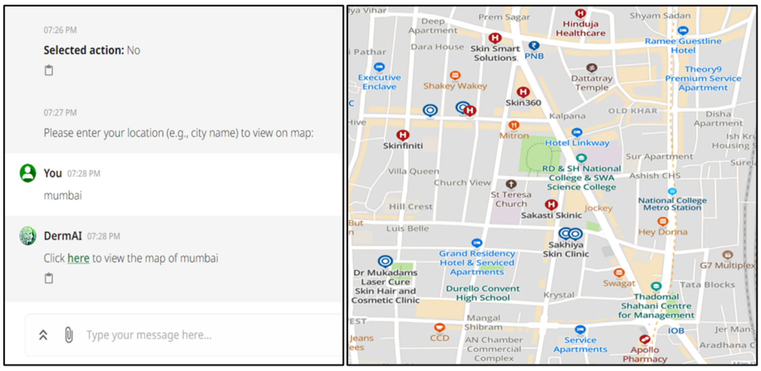
What topical ointments can be used for scabies?

DermAI 11:01

Topical ointments that can be used for scabies include permethrin cream, crotamiton lotion, and sulfur ointment. It is important to consult with a doctor for a proper diagnosis and treatment plan.

Figure 11. Information about skin disease

Figure8 shows the actual application interface where the user is asked to choose between two options: information related to skin diseases or image analysis for prediction. The user interface allows users to make selections based on their needs. Importantly, if a user selects an option by mistake or wishes to revisit a previous step, they have the flexibility to navigate back seamlessly within the interface. This ensures a smooth and intuitive user experience, accommodating user preferences and minimising potential errors in selection. Figure9 depicts the Image Analysis selection. When choosing the ”Image Analysis” option, the user is prompted to upload an image of the affected skin region. After uploading, the user is asked to confirm their choice to proceed with the uploaded image. Upon confirmation, the uploaded image undergoes processing, and the prediction results are provided to the user through the interface. Additionally, basic information related to the prediction, such as details about the predicted skin condition, is presented to the user via the user interface. This interactive process tries to ensure that users receive accurate and relevant information based on their uploaded images. After receiving the prediction and related information, users have the option to ask multiple follow-up questions pertaining to the prediction, as shown in Figure10, or inquire about other skin conditions. This feature allows for a more interactive and informative user experience, enabling users to seek additional clarification or explore various aspects of skin health beyond the initial prediction. The system is designed to accommodate diverse user queries and provide comprehensive responses to facilitate better understanding and engagement. If the user selects the option for information related to skin diseases, as shown in Figure11, they have the ability to ask queries about any skin diseases until they are satisfied with the responses. This functionality allows users to explore and gather detailed information about various skin conditions, symptoms, treatments, and preventive measures. Users can engage with the system to acquire comprehensive knowledge and address their specific concerns related to skin health through interactive dialogue and tailored information retrieval.



07:26 PM

Selected action: No

07:27 PM

Please enter your location (e.g., city name) to view on map:

You 07:28 PM

mumbai

DermAI 07:28 PM

Click here to view the map of mumbai

Figure 12. Map of dermatologists nearby

Figure12 depicts the display of map services to show the dermatologists close to the user’s location, achieved via the APIs provided by MapmyIndia [53], which is a leading digital mapping and location-based services provider based in India. Once the users have solved their queries, they are provided with the option to see the certified dermatologists close to their location using the latitude and longitude. These latitude and longitude values of major Indian cities are stored in the sqlite3 [54] database, which is a lightweight, self-contained, serverless, and zero-configuration relational database management system (RDBMS) and are retrieved based on the user’s location. The retrieved values are appended to generate map URLs, which redirect users to MapmyIndia’s map services to display location-specific maps.

4. RESULTS AND DISCUSSIONS

4.1. Problems/Test Cases Encountered

The central challenge we faced revolved around producing responses that were not only pertinent but also minimally prone to hallucination. To tackle this issue, we explored two potential solutions: fine-tuning and Retrieval Augmented Generation (RAG). After a thorough investigation, we determined that RAG represented the superior approach for erroneous or misleading information. Integration of Chainlit presented hurdles due to compatibility conflicts with certain Python modules since it being a recent thing in the market. The need to constantly update libraries for error prevention and to maintain compatibility with Chainlit’s own evolving versions adds complexity. Moreover, Chainlit applications can be computationally demanding, and its evolving syntax sometimes leads to deprecated code, requiring developers to refactor existing implementations. These factors contribute to a development process that necessitates ongoing maintenance and adaptation. The computational demands of DL applications, especially when dealing with exhaustive image datasets, contribute to potential performance bottlenecks. Furthermore, acquiring data, particularly medical related ones, are very difficult due to privacy concerns and doctor-patient confidentiality. Additionally, keeping sufficient images for both training and validation of the model to avoid underfitting and overfitting.

4.2. Model Evaluation

Table4 shows the performance comparison for five different models based on the parameters evaluated. The accuracy metric gauges the net correctness of the model by computing the ratio of correctly predicted instances (comprising both true positives and true negatives) to the total number of instances [55]. Precision quantifies the accuracy of the positive predictions made by the model [55].

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Positives (FP)}} \tag{2}$$

Recall/sensitivity, alternatively referred to as sensitivity or true positive rate (TPR) [55]. It measures the ability of the model to correctly identify positive cases.

$$\text{Sensitivity} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Negatives (FN)}} \tag{3}$$

Specificity quantifies the model's capability to accurately identify negative instances from all actual negative instances [54].

$$\text{Specificity} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN) + False Positives (FP)}} \tag{4}$$

The F1 score represents the harmonic mean of precision and recall, offering a balanced assessment between these two metrics. It's especially helpful in situations when there is an unequal distribution of classes or when recall and precision are both crucial [43].

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision + Recall}} \tag{5}$$

AUC represents the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity).

$$\text{Area under curve (AUC)} = \int_0^1 \text{Sensitivity } d(1 - \text{Specificity}) \tag{6}$$

Table 4. Performance comparison for five different models

Model Name	**Parameters**	**No. of layers**	**No. of class-es**	**% Training Accuracy**	**% Validation Accuracy**	**% Sensitivity**	**% Specificity**	**% AUC**
ResNet-50	24642443	50	11	100	97.35	92.6	99.8	99.90
Inception-V3	22857515	48	11	97.33	82.31	55.04	98.25	95.02
MobileNet-V2	2919499	53	11	100	93.02	79.35	99.34	99.10
VGG-16	20292683	19	11	100	97	92.35	99.75	99.67
EfficientNet-B4	18597482	19	11	98	92.78	80.18	92.78	92.52

4.3. Accuracy and Loss

Due to space limitations, results for all five models are discussed but figures and tables are included only for ResNet-50. Figure13 depicts the loss and accuracy of the ResNet-50 model over 20 epochs for both training and validation samples. During the training process, the model's ideal performance with respect to the peak validation accuracy was observed at epoch 8, whereas validation loss was observed at epoch 3, indicating a trade-off between maximising accuracy and minimising loss. After completing the initial training phase, the model displayed a good validation accuracy and minimal validation loss, indicating that the model generalises well to new untested samples without leading to overfitting. This stability represents the model's ability to learn meaningful patterns from the data and perform well on new samples.

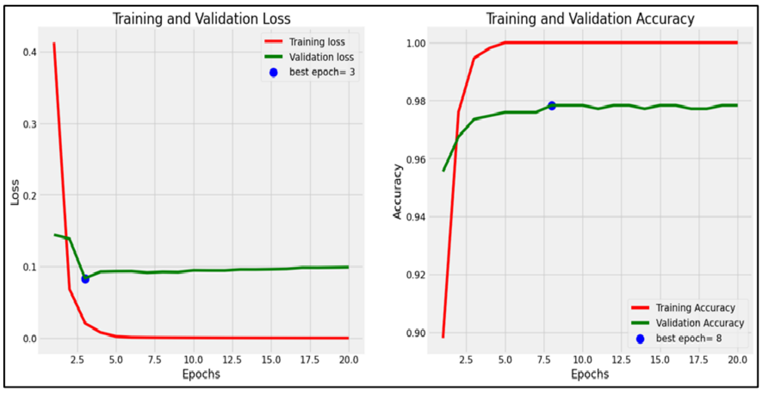
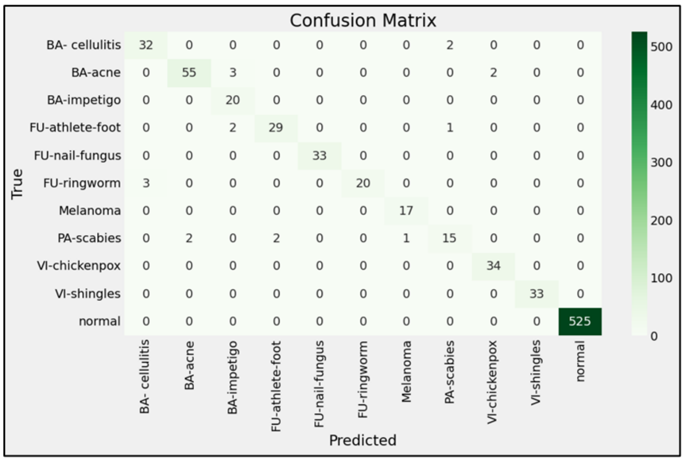


Figure 13. Accuracy & Loss Graph of ResNet-50

The MobileNet-V2 model's optimal performance in terms of validation loss was observed at epoch 10, thus making it almost similar to ResNet-50 and indicates a potential stopping point to avoid overfitting. But the peak validation accuracy was achieved at epoch 11, indicating a trade-off between maximising accuracy and minimising loss. The model's consistent low validation loss and high validation accuracy after the initial training phase, represents that the model generalises well to untested samples. At epoch 10, the Inception-V3 model exhibited optimal performance with respect to validation loss. indicating a possible moment of stopping to prevent overfitting. But the highest validation accuracy was obtained at epoch 20, emphasizing a trade-off between maximising accuracy and minimising loss. On the contrary, the VGG-19 model's optimal performance with respect to validation loss was at epoch 5. However, the peak validation accuracy was achieved at epoch 3, highlighting a trade-off between achieving maximising accuracy and minimal loss. Lastly, the EfficientNet-B4 model exhibited optimal performance with respect to both validation accuracy and validation loss, which coincided at epoch 4, highlighting a balance between achieving minimal loss and maximising accuracy.

4.4. Confusion Matrix

Figure14 depicts the confusion matrix of the ResNet-50 model, where the diagonal cells signify the count of correct predictions for each category. For instance, the model correctly predicted 'Cellulitis' 32 times, 'Acne' 55 times, 'Impetigo' 20 times, 'Athlete's Foot' 29. However, there is room for improvement in distinguishing between certain diseases, as indicated by a misclassification rate of 2.17%. In confusion Matrix of Inception-V3, the main diagonal cells showed the number of correct predictions for each category. For instance, 'Athlete's Foot' was mistakenly classified as 'Cellulitis' twice, 'Acne' once, and 'Impetigo' thrice. Notably, the 'Normal' category, likely representing healthy skin without disease, achieved the highest number of correct predictions with 515 instances. This matrix provides insights into which categories are frequently confused with one another. For example, 'Ringworm' was sometimes misidentified as 'Cellulitis' (1 time), 'Acne' (5 times), 'Impetigo' (1 time), 'Athlete's Foot' (2 times), 'Nail-Fungus' (3 times), 'Scabies' (2 times), 'Chickenpox' (2 times), and 'Shingles' (4 times). InceptionV3 encounters challenges in accurately identifying the 'Normal' skin category and shows limitations in effectively discerning specific skin conditions. Significant improvements are necessary to enhance disease discrimination, highlighted by a relatively high misclassification rate of 18.714%. In confusion Matrix of VGG-19, The main diagonal cells showed the number of correct predictions for each category. Notably, the model achieved correct predictions for 'Cellulitis' 30 times, 'Acne' 57 times, 'Impetigo' 20 times, 'Athlete's Foot' 30 times, and so forth. Conversely, the off-diagonal cells in the matrix signify instances of misclassifications. For example, 'Cellulitis' was mistakenly identified as 'Acne' once and 'Impetigo' three times. The 'Normal' category, likely denoting healthy skin, showed the highest number of correct predictions with 525 instances. This matrix is instrumental in identifying which disease categories are frequently confused with one another. For instance, 'Scabies' was sometimes misclassified as ’Cellulitis’ (4 times), ’Acne’ (1 time), ’Athlete’s Foot’ (2 times), and ’Melanoma’ (2 times). Despite VGG19’s commendable performance in accurately categorising normal skin and recognizing specific conditions, there remains scope for enhancement in disease discrimination, as indicated by a misclassification rate of 3.101%. This suggests a need for further refinement to improve the model’s ability to distinguish between certain skin diseases. In the confusion matrix of EfficientNet-B4, the main diagonal cells showed the number of correct predictions for each category. The model achieved accurate predictions for ’Cellulitis’ 28 times, ’Acne’ 49 times, ’Impetigo’ 18 times, ’Athlete’s Foot’ 27 times, and others. Conversely, the off-diagonal cells in the matrix represent instances of misclassifications, such as ’Cellulitis’ being misidentified as times, and so forth. On the other hand, the off-diagonal cells indicate instances of misclassification.



	BA- cellulitis	BA-acne	BA-impetigo	FU-athlete-foot	FU-nail-fungus	FU-ringworm	Melanoma	PA-scabies	VI-chickenpox	VI-shingles	normal
BA- cellulitis	32	0	0	0	0	0	0	2	0	0	0
BA-acne	0	55	3	0	0	0	0	0	2	0	0
BA-impetigo	0	0	20	0	0	0	0	0	0	0	0
FU-athlete-foot	0	0	2	29	0	0	0	1	0	0	0
FU-nail-fungus	0	0	0	0	33	0	0	0	0	0	0
FU-ringworm	3	0	0	0	0	20	0	0	0	0	0
Melanoma	0	0	0	0	0	0	17	0	0	0	0
PA-scabies	0	2	0	2	0	0	1	15	0	0	0
VI-chickenpox	0	0	0	0	0	0	0	0	34	0	0
VI-shingles	0	0	0	0	0	0	0	0	0	33	0
normal	0	0	0	0	0	0	0	0	0	0	525

Figure 14. Confusion Matrix of ResNet-50

For instance, ’Athlete’s Foot’ was incorrectly classified as ’Cellulitis’ twice and as ’Ringworm’ once. Notably, the ’Normal’ category, presumably representing healthy skin without disease, achieved the highest number of correct predictions with 525 instances. This confusion matrix enables us to pinpoint which categories are frequently confused with one another. For example, ’Scabies’ was sometimes mistaken for ’Acne’ (2 times), ’Athlete’s Foot’ (2 times), and ’Melanoma’ (1 time). ResNet-50 demonstrates robust performance in accurately identifying the ’Normal’ skin category and shows commendable ability in distinguishing several specific skin conditions. However, there is potential for improvement in differentiating between certain diseases, as evidenced by a misclassification rate of 2.17%. In confusion Matrix of MobileNet-V2, the main diagonal cells showed the number of accurate predictions for each category. For example, ’Cellulitis’ was predicted correctly 26 times, ’Acne’ 52 times, ’Impetigo’ 12 times, ’Athlete’s Foot’ 29 times, and so forth. Conversely, the off-diagonal cells indicate instances of misclassifications. Interestingly, the ’Normal’ category, presumably representing healthy skin without disease, achieved the highest number of correct predictions with 524 instances. This matrix serves as a tool to identify which categories are frequently confused with one another. MobileNetV2 demonstrates commendable proficiency in correctly classifying the ’Normal’ skin category and shows satisfactory performance in identifying various specific skin conditions. The ’Normal’ category, likely denoting healthy skin, displayed the highest number of correct predictions with 525 instances. Disease categories such as ‘Scabies’ are often mistaken for being ‘Cellulitis’, ‘Acne’, ‘Athlete’s Foot’ and ‘Melanoma’. This confusion matrix provides a better understanding of the disease categories that are frequently confused with one another. A commendable proficiency is noticed in accurately identifying the ‘Normal’ skin category using EfficientNetB4. It achieves a satisfactory level of performance in distinguishing between different skin conditions. Notable improvement in disease discrimination can be made. This is indicated by a misclassification rate of 7.782%. The model can be improved and refined to enhance its ability of differentiating between specific skin conditions and to reduce misclassifications.

4.5. Classification Reports

A summary of the classification model providing key metrics to evaluate its own effectiveness in predicting the class labels can be called a classification report. The classification report of the ResNet-50 model is shown in Table5. Exceptional performance metrics with an average precision (AP) of 94%, average recall (AR) of 94% and average F1 score (AF1) of 94% were obtained across 11 classes as shown in the classification report. The 11 classes were distributed into 10 diseased samples and 1 non diseased sample. The results indicate a high level of accuracy and sensitivity in the model's predictions, effectiveness in classifying different classes, its precision being positive which is indicated by the obtained average precision of 94%, average recall being 94% suggesting that it can rightly identify around 94% of positive instances, and an F1 score of 94% suggesting a strong balance between precision & recall metrics. Therefore, the findings based on the results obtained indicate the model's robust performance and reliability when it comes to handling diverse classification tasks. MobileNet-V2 scores an average precision of 79%, average recall of 79%, F1 score of 78% that reflects a balanced performance between recall & precision metrics. Inception-V3 scored an average precision of 62%, average recall of 59% and an average score F1 score of 58% for 10 diseased and 1 non diseased sample (overall the different classes are 11). These findings suggest that Inception-V3 is less accurate and effective when compared to other models also indicating that there could be opportunities for model improvement to enhance performance, especially when it comes to boosting precision, recall and overall F1 score. VGG-19 performed well with average precision, recall and F1 score values being around 91% which indicates high accuracy and effectiveness in all the classification tasks when compared to Inception-V3 and MobileNet-V2 but slightly lower when compared to ResNet-50. Results of EfficientNet-B4 demonstrates good performance, although slightly lower than ResNet-50 and VGG-19 but still effective for classification tasks. The x-axis shows different diseases, while the y-axis shows the precision and recall values. Precision is represented by the blue line and recall by the orange line. Generally, the higher the line, the better the model's performance for that disease.

Table 5. Disease Classification Results

Diseases	Precision (%)	Recall (%)	F1 Score (%)
Cellulitis	91	94	93
Acne	96	92	94
Impetigo	80	100	89
Athlete Foot	94	91	92
Nail Fungus	100	100	100
Ringworm	100	87	93
Melanoma	94	100	97
Scabies	83	75	79
Chickenpox	94	100	97
Shingles	100	100	100
Normal	100	100	100

4.6. Precision, Recall and F1 Score VS Models

Figure15 shows the precision of five different image classification models: ResNet-50, Inception-V3, VGG-19, MobileNet-V2, and EfficientNet-B4. Precision is a metric exploited in ML to measure the proportion of positive predictions that were correct.

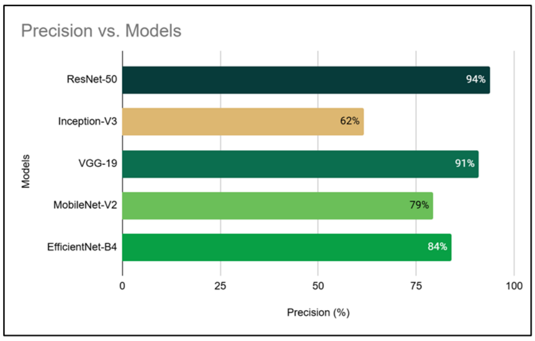


Figure 15. Precision vs Models

As depicted (Figure15), ResNet-50 achieved the highest precision (94%), while Inception-V3 achieved the lowest precision (62%). This suggests that ResNet-50 outperformed the other models in this experiment when it comes to accurately classifying images. Figure16 shows the recall of five different image classification models: ResNet-50, Inception-V3, VGG-19, MobileNet-V2, and EfficientNet-B4.

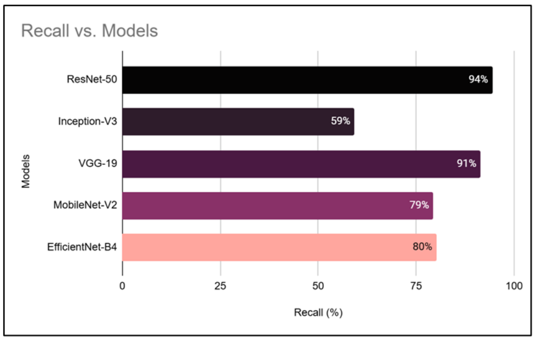


Figure 16. Recall vs Models

Recall is a metric exploited in ML to measure the proportion of actual positive cases that were identified correctly by the model. As depicted (Figure16), ResNet-50 and VGG-19 achieved the highest recall (over 90%), while Inception-V3 achieved the lowest recall (around 60%). This suggests that ResNet-50 and VGG-19 were more successful at finding most of the relevant images in this experiment.

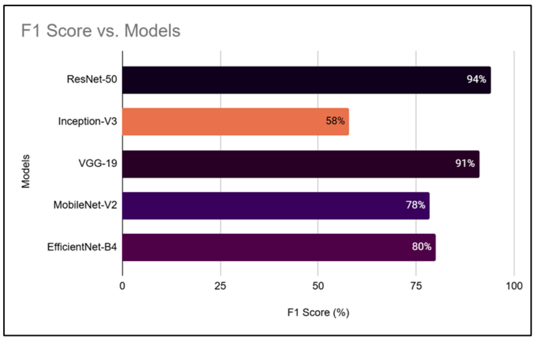


Figure 17. F1 Score vs Models

F1 score is a metric used in machine learning to measure the accuracy of a model on a test dataset. It combines precision and recall, which are two other common metrics used in machine learning. As seen (Figure17), ResNet-50 achieved the highest F1 score (94%), while Inception-V3 achieved the lowest F1 score (58%). This suggests that ResNet-50 outperformed the other models in this experiment when it comes to accurately classifying images.

4.7. Large Language Model Evaluation

In Table6, we have taken eleven diseases namely Cellulitis, Impetigo, Athlete Foot, Nail Fungus, Ringworm, Melanoma, Scabies, Chickenpox, Shingles, Syphilis and Hansen's Disease for LLM evaluation. Our LLM model made use of OpenAI's GPT 3.5-turbo model for generating answers and GPT-4 for evaluating the generated answers with the provided context. The key thing to note is that we've used GPT 3.5-turbo for the generation of answers because of its ability to stick to the provided context whereas GPT 4 tends to add content that is not mentioned in the provided context. GPT 4 being more improved and advanced than GPT 3.5-turbo makes it a reliable option for being used as an evaluation model. It can understand the nuances of the engendered information with more accuracy. DeepEval's (by Confident AI) open-source framework is used that employs GPT 4's API key for evaluating our LLM model. DeepEval quantifies the performance of the LLM based on aspects such as faithfulness, answer relevancy, contextual recall etc. Since we did not have access to the RAG's retrieval components, we went ahead with Faithfulness and Hallucination metrics provided by DeepEval [56] for the evaluation. For our LLM evaluation, we had asked our LLM model 10 specific questions to maintain the consistency, integrity and accuracy of the different metrics instituted in use. It's noteworthy that despite all the metrics used to quantify the evaluation of our LLM model, we made sure to cross verify every generated information with the knowledgebase by human intervention. The questions asked were as follows:

1. What is "name of disease"?
2. What are the symptoms of "name of disease"?
3. How does "name of disease" spread?
4. What are the types of "name of disease"?
5. How is "name of disease" diagnosed?
6. How is "name of disease" treated?
7. What are the complications caused by "name of disease"?
8. How to prevent "name of disease"?
9. What's the long-term outlook of "name of disease"?
10. How to know if you have "name of disease" from home?

Table 6. LLM Performance Metric

Diseases	**Faith-fulness %**	**Hallucination %**	**Rouge 1**			**Rouge 2**			**Rouge L**			**Meteor Score %**
			P %	**R %**	**F1 %**	**P %**	**R %**	**F1 %**	**P %**	**R %**	**F1 %**	
Cellulitis	100	10	92	89	90	89	86	87	92	88	89	94
Impetigo	92	0	53	47	47	31	28	27	50	45	45	48
Athlete's Foot	100	0	51	45	41	31	27	26	47	41	41	47
Nail Fungus	61	11	63	19	51	15	28	32	59	16	17	55
Ringworm	83	30	71	35	44	47	22	29	68	34	43	54
Melanoma	100	0	37	21	25	13	5	7	34	20	23	30
Scabies	100	10	80	53	61	60	44	48	77	52	59	70
Chickenpox	90	0	44	45	44	25	25	24	43	43	42	41
Shingles	74	20	81	80	81	71	71	71	79	78	79	83
Syphilis	90	40	45	44	42	22	21	20	41	40	38	41
Hansen's Disease	81	10	45	44	43	24	24	23	43	41	41	46

These questions were chosen to see if the LLM Model has a thorough understanding of the contexts provided and has the basic foundations covered by learning from the knowledgebase. The LLM metrics used for evaluation are Faithfulness, Hallucination, ROUGE score and METEOR score. These metrics cover the most important facets of the LLM model for our specific application.

4.8. Faithfulness

The faithfulness metric is used to measure the quality of our LLM model's ability to generate texts that factually align with the actual output and the context provided to the model. The evaluation of the model is conducted by considering the actual output generated by the model and the context provided to the model [23]. It is calculated according to the following formula:

$$\text{Faithfulness} = \frac{\text{Number of Truthful Claims}}{\text{Total Number of Claims}} \tag{7}$$

The faithfulness metric provided by DeepEval makes use of LLM to extract all the claims made in the actual output provided by the LLM before using the same to classify whether each claim is truthful based on the facts presented in the context. By taking a look at the scores from Table6, we can conclude that the average Faithfulness score is 95.6% that exceeds the set threshold of 70%. Therefore, the score indicates that the LLM is greatly faithful to the context that has been provided and has minimal number of inaccuracies and inconsistencies.

4.9. Hallucination

The hallucination metric is used to determine whether the LLM model generates factually correct answers by comparing the actual output generated by the LLM and context provided to it to generate the answers. It is calculated according to the following formula:

$$\text{Hallucination} = \frac{\text{Number of Contradicted Contexts}}{\text{Total Number of Contexts}} \tag{8}$$

The hallucination metric provided by DeepEval is quite similar to the faithfulness metric but it is calculated differently as it uses contexts as the source of truth. The context remains the only source of truth as there is no other way for the LLM model to gain knowledge from other sources. Therefore, the degree of hallucination is measured by the degree of which the contexts are disagreed upon. It's crucial to acknowledge that the threshold value for the hallucination metric was set to 50%. By taking a close look at the scores from the graphs, we can conclude from Table6. that the average hallucination score is 7.3% which is much improved (way below the set threshold of 50%) designating that our LLM model generates answers/responses that deviate very faintly from the provided context and generate substantially correct information by staying faithful to the provided context.

4.10. ROUGE Scores

ROUGE scores, also referred to as Recall-Oriented Understudy for Gisting Evaluation [55], comprise an array of metrics utilized to assess the similarity between the quality of machine-generated texts or information with a set of references or human written contexts. Albeit there are several variants of ROUGE scores, we have put to use ROUGE-1, ROUGE-2 & ROUGE-L for the sake of evaluation. Other variants aren't relevant for our evaluation because their focus is on varied linguistic aspects and our priority is more on coherence and content fidelity or faithfulness. These metrics are very famously known and widely used in the field of Natural Language Processing and also show their usage in text summarization and machine translation tasks where the generated output has to be compared with the provided contexts. The higher the ROUGE scores, the higher the agreement between the generated texts and the provided contexts. It's worth highlighting that the ROUGE scores evaluate based on fundamental metrics namely Recall, Precision and F1 Score. Recall indicates how much of the provided context is covered, Precision indicates how much of the generated text is relevant and F1 score is the balance between Recall and Precision. Below are brief explanations of ROUGE-1, ROUGE-2 and ROUGE-L scores along with our generated scores. ROUGE-1 score is responsible for evaluating the overlap of unigrams or individual words between generated texts and provided contexts. Higher ROUGE-1 scores indicate a greater sense of similarity between the generated texts and the provided context in terms of the words being used indicating better content coverage and accuracy. The average scores generated are - precision value is 60.17%, recall score is 50.15% and the F1-Score is 51.98%. The precision score of approximately 60.17% means that about 60.17% of the words in the engendered answers are accounted for in the provided context. It is accurately selecting relevant words 60.17% of the time. The recall score of 50.15% means that our LLM model is capturing approximately 50.15% of the words from the provided context into its generated answers. It also indicates that sometimes captured is only half. The F1-score of 51.98% indicates that our LLM model strikes a balance between both precision and recall and improvements can be made for both the aspects. ROUGE-2 score is responsible for evaluating the overlap of bigrams or sequences of two adjacent words between the generated texts and the provided contexts. They measure the proportion of bigrams in the generated texts which also show up in the provided context. Higher ROUGE-2 scores indicate a greater sense of similarity in the sequence of word pairs between the generated texts and the provided contexts thereby indicating the improved coherence and content fidelity (faithfulness) in the generated texts. The average scores generated are - precision is 41.62%, recall is 34.59% and F1-Score is 35.82%. The precision score indicates that around 41.62% of the bigrams in the generated answer is also present in the reference context. This gives the measure of accuracy of our LLM model in selecting relevant and consecutive pairs of words. The recall score indicates that our LLM model is capturing 34.59% of the bigrams from the provided context in its generated answers, this score measures the ability of our LLM model to retrieve relevant and consecutive pairs of words from the provided context. The F1-Score of around 35.82% underscores the importance of striking a balance between precision recall. There is room for improvement for the recall and precision scores. ROUGE-L score is responsible for evaluating the Longest Common Subsequence or LCS between the generated texts and the provided context. It is responsible for measuring the longest continuous sequence of words in both the generated texts and the provided context giving a measure of summary coherence and content overlap. Higher

ROUGE-L scores mean a higher sense of similarity in structural and contextual organisation between generated texts and provided contexts thereby showing improved coherence and faithfulness in the generated texts. The average scores generated are - Precision is 57.50%, Recall is 48.01% and F1-Score is 49.74%. The precision score of approximately 57.50% means that around 57.50% of the words in the engendered answers are also part of the provided contexts. Recall score of approximately 48.01% indicates that 48.01% of the words from the provided context is in the generated answers. The F1-Score of around 49.74% indicates that our LLM model is performing moderately well when it comes to capturing relevant words from the provided contexts but there is still room for improvement for recall and precision.

4.11. METEOR Scores

The Metric for Evaluation-of-Translation with Explicit-Ordering (METEOR) is employed to assess the alignment between the quality of machine-generated translations. Unlike the other most commonly used metrics such as BLEU or TER scores, METEOR score makes use of synonyms and paraphrases into its evaluation allowing it to be more robust and aligning better with human judgement. It also takes into consideration the concept of stemming and synonymy to account for variations in several word forms and terminology. METEOR score makes use of alignment optimization which finds the best possible mapping between words in the machine generated text and provided context. The average METEOR score is 55.36%. This score indicates that our LLM model is generating answers that achieve a moderate sense of similarity and quantity when compared to the provided context.

5. CONCLUSION

The DermAI implementation emerges as a promising step toward addressing critical challenges in dermatological healthcare. The adoption of Retrieval Augmented Generation has substantially improved the accuracy and relevance of responses, marking a significant advancement in natural language processing for dermatological or any such queries. Existing research predominantly emphasises cancer and rare skin diseases, prioritising high training accuracy over misclassification rates. The existing platforms, mentioned in Table3, offer restricted initial diagnosis and necessitate professional consultation, except a few platforms such as the Model Dermatol and Aysa by Visualdx . These platforms consider almost all major types of skin diseases and list out all the possible skin diseases, more than one, without providing a justification. While the others simply offer product-based services or basic skin care routines. The ChatGPT-4o did manage to excel in certain conditions, such as Athlete's foot, Melanoma and Shingles, but struggled with others. To address these limitations, we did a comparative study of five models, namely, ResNet-50, MobileNet-V2, VGG-19, InceptionNet-V3 and EfficientNet-B4, which was conducted on a dataset of common skin diseases, aiming to evaluate and mitigate misclassification challenges. ResNet-50 emerged as a robust choice for image classification, guided by its well-established architecture, showcased better performance, laying a foundation for effective skin condition recognition. The Chainlit interface, along with the skin condition classifier, not only enhances user engagement but also positions the project for future advancements in diagnostic accuracy. While these achievements are notable, challenges in compatibility and continuous updates in Chainlit underscore the evolving nature of AI applications. Furthermore, the implementation emphasises the broader need for accessible and privacy-sensitive healthcare solutions. Looking ahead, DermAI holds promise for contributing to improved skin health awareness and diagnostics, albeit with the recognition of ongoing technical and ethical considerations. With a mean sensitivity of 92.6%, specificity of 99.8% and AUC of 99.9%, the metrics indicate that the model demonstrates a high true positive rate and is highly capable of correctly identifying negative cases. The excellent training accuracy of 100% and a high validation accuracy of 97.35% infer that the model has learned efficiently from the training data without any augmentation required and can generalise well to unseen examples. The low misclassification rate of 2.17% further supports the model's robustness and accuracy in classifying common skin diseases. Using strategies for data augmentation managed to enhance the model's dimensions to generalise to new data. The evaluation metrics for text summarization and machine translation outputs indicate varying levels of performance. For Rouge-1, the system achieved a high precision of 92%, suggesting good capture of relevant unigrams from reference summaries, but a lower recall of 89% implies some important words may be missing from system-generated summaries. In contrast, Rouge-2 shows moderate precision (42%) in predicting relevant bigrams related to skin diseases, but a lower recall (35%) indicates room for improvement in identifying all relevant bigrams. The F1 score (36%) for Rouge-2 reflects a trade-off between precision and recall, indicating potential for enhancement in overall performance. Similarly, Rouge-L exhibits moderate precision (57%) in matching predicted sequences with reference sequences, but a lower recall (48%) suggests gaps in capturing all relevant information. An F1 score of 50% for Rouge-L signifies a satisfactory equilibrium between precision and recall, highlighting areas for improvement in sequence prediction and capturing relevant content related to skin diseases. One significant enhancement lies in the integration of more extensive and diverse dermatological datasets. Acquiring datasets from various sources and populations can enrich the language model's understanding of dermatological conditions. Using cutting-edge computer vision techniques to enhance the sensitivity and precision of skin condition detection from uploaded photos. To broaden the accessibility of DermAI, multilingual support can be incorporated, expanding DermAI's language support to cater to a diverse global audience and address a myriad of skin-related concerns across different linguistic demographics, thereby supporting real-time interactions. Additionally, efforts will be dedicated to improving the cultural sensitivity of the system by accounting for regional nuances in expressing and comprehending skin-related issues. Our research is well capable of performing on far more complex diseases than the once being used. Improvements and adjustments will only be required if we take into consideration the scenario where a user uploads a blurry image of the affected area on their skin. This would require training the model on blurry images to give accurate diagnosis. This approach will refine the information retrieval process by narrowing the search scope. In addition to that, DermAI can have a strategic integration into telemedicine platforms, enabling seamless remote consultations with dermatologists. The chatbot can be tailored to a more holistic and personalised user experience by integrating with electronic health records (EHR) that considers individual medical backgrounds for more accurate and informed interactions.

ACKNOWLEDGEMENT

Authors are grateful to all reviewers and editors for their valuable time in successfully reviewing and publishing this paper.

CONFLICT OF INTEREST

The authors have no conflict of interests with any other work.

REFERENCES

[1] A. Dregan, J. Charlton, P. Chowienczyk, and M. C. Gulliford, "Chronic inflammatory disorders and risk of type 2 diabetes mellitus, coronary heart disease, and stroke: a population-based cohort study," *Circulation*, vol. 130, no. 10, pp. 837–844, 2014.

[2] H. Feng, J. Berk-Krauss, P. W. Feng, and J. A. Stein, "Comparison of dermatologist density between urban and rural counties in the united states," *JAMA Dermatology*, vol. 154, no. 11, pp. 1265–1271, 2018.

[3] "Resnet-50 guide," https://datagen.tech/guides/computer-vision/resnet-50/, accessed: 2023-08-31.

[4] "Chainlit: Get started overview," https://docs.chainlit.io/get-started/overview, accessed: 2024-01-16.

[5] "Astra db vector: Concepts and metrics," https://docs.datastax.com/en/astra/astra-db-vector/get-started/concepts.html#metrics, accessed: 2023-10-26.

[6] K. Vayadande, "Innovative approaches for skin disease identification in machine learning: A comprehensive study," *Oral Oncology Reports*, p. 100365, 2024.

[7] D. P. Panagoulias, E. Tsoureli-Nikita, M. Virvou, and G. A. Tsihrintzis, "Dermacen analytica: A novel methodology integrating multi-modal large language models with machine learning in tele-dermatology," *arXiv preprint arXiv:2403.14243*, 2024.

[8] N. H. Sany and P. C. Shill, "Image segmentation based approach for skin disease detection and classification using machine learning algorithms," in *2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2024, pp. 1–5.

[9] A. Ansari, A. Singh, M. Singh, and V. Kukreja, "Enhancing skin disease classification: A hybrid cnn-svm model approach," in *2024 International Conference on Automation and Computation (AUTOCOM)*, 2024, pp. 29–32.

[10] K. Sudharson, K. S. Essakki, V. R. Thanujaa, and S. Varsha, "Dermatec: Transformative ai-driven platform for comprehensive skin disease monitoring and dermatologist recommendations," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2024, pp. 1–5.

[11] M. C. Chiu, Y. Wang, Y. J. Kuo, and P. Y. Chen, "Ddi-coco: A dataset for understanding the effect of color contrast in machine-assisted skin disease detection," *arXiv preprint arXiv:2401.13280*, 2024.

[12] D. Xia, "Skin disease diagnosis using deep neural network and large language model," https://doi.org/10.32657/10356/172895, 2023.

[13] Q. Hu, H. Xia, and T. Zhang, "Chatbot combined with deep convolutional neural network for skin cancer detection," in *Methods*, vol. 2, 2023, p. 35.

[14] S. Kohli, U. Verma, V. V. Kirpalani, and R. Srinath, "Dermatobot: An image processing enabled chatbot for diagnosis and tele-remedy of skin diseases," in *2022 3rd International Conference for Emerging Technology (INCET)*, 2022, pp. 1–5.

[15] J. Huang, J. Li, Z. Li, Z. Zhu, C. Shen, G. Qi, and G. Yu, "Detection of diseases using machine learning image recognition technology in artificial intelligence," *Computational Intelligence and Neuroscience*, vol. 2022, p. 5658641, 2022, retraction in: Comput Intell Neurosci. 2023 Jul 19;2023:9845093.

[16] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm," *Sensors*, vol. 21, no. 8, p. 2852, 2021.

[17] K. Glock, C. Napier, T. Gary, V. Gupta, J. Gigante, W. Schaffner, and Q. Wang, "Measles rash identification using transfer learning and deep convolutional neural networks," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 3905–3910.

[18] A. A. Elngar, R. Kumar, A. Hayat, and P. Churi, "Intelligent system for skin disease prediction using machine learning," in *3rd International Conference on Smart and Intelligent Learning for Information Optimization (CONSILIO 2021)*, vol. 1998, 2021, p. 012037.

[19] T. Goswami, V. K. Dabhi, and H. B. Prajapati, "Skin disease classification from image - a survey," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 599–605.

[20] A. K. Verma, S. Pal, and S. Kumar, "Classification of skin disease using ensemble data mining techniques," *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 6, pp. 1887–1894, 2019.

[21] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 168–172.

[22] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[23] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, 2014.

[24] R. Mittal, F. Jeribi, R. J. Martin, V. Malik, S. J. Menachery, and J. Singh, "Dermcdsm: Clinical decision support model for dermatosis using systematic approaches of machine learning and deep learning," *IEEE Access*, vol. 12, pp. 47 319–47 337, 2024.

[25] V. S. S. B. T. Sathvika, N. Anmisha, V. Thanmayi, M. Suchetha, E. D. Dhas, S. Sehastrajit, and S. N. Aakur, "Pipelined structure in the classification of skin lesions based on alexnet cnn and svm model with bi-sectional texture features," *IEEE Access*, vol. 12, pp. 57 366–57 380, 2024.

[26] A. Kumar, A. Vishwakarma, V. Bajaj, and S. Mishra, "Novel mixed domain hand-crafted features for skin disease recognition using multiheaded cnn," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024.

[27] C. Tsai, P. P. H. Huang, Z. C. Wu, and J. F. Wang, "Advanced pigmented facial skin analysis using conditional generative adversarial networks," *IEEE Access*, vol. 12, pp. 46 646–46 656, 2024.

[28] W. Dai, R. Liu, T. Wu, M. Wang, J. Yin, and J. Liu, "Deeply supervised skin lesions diagnosis with stage and branch attention," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 2, pp. 719–729, 2024.

[29] D. Raha, M. Gain, R. Debnath, A. Adhikary, Y. Qiao, M. M. Hassan, A. K. Bairagi, and S. M. S. Islam, "Attention to monkeypox: An interpretable monkeypox detection technique using attention mechanism," *IEEE Access*, vol. 12, pp. 51 942–51 965, 2024.

[30] D. Kundu, M. M. Rahman, A. Rahman, D. Das, U. R. Siddiqi, M. G. R. Alam, S. K. Dey, G. Muhammad, and Z. Ali, "Federated deep learning for monkeypox disease detection on gan-augmented dataset," *IEEE Access*, vol. 12, pp. 32 819–32 829, 2024.

[31] L. Riaz, H. M. Qadir, G. Ali, M. Ali, M. A. Raza, A. D. Jurcut, and J. Ali, "A comprehensive joint learning system to detect skin cancer," *IEEE Access*, vol. 11, pp. 79 434–79 444, 2023.

[32] H. Q. Yu and S. Reiff-Marganiec, "Targeted ensemble machine classification approach for supporting iot enabled skin disease detection," *IEEE Access*, vol. 9, pp. 50 244–50 252, 2021.

[33] S. Back, S. Lee, S. Shin, Y. Yu, T. Yuk, S. Jong, S. Ryu, and K. Lee, "Robust skin disease classification by distilling deep neural network ensemble for the mobile diagnosis of herpes zoster," *IEEE Access*, vol. 9, pp. 20 156–20 169, 2021.

[34] J. Yang, X. Wu, J. Liang, X. Sun, M. M. Cheng, P. L. Rosin, and L. Wang, "Self-paced balance learning for clinical skin disease recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2832–2846, 2020.

[35] Z. Wu, S. Zhao, Y. Peng, X. He, X. Zhao, K. Huang, X. Wu, W. Fan, F. Li, M. Chen, J. Li, W. Huang, X. Chen, and Y. Li, "Studies on different cnn algorithms for face skin disease classification based on clinical images," *IEEE Access*, vol. 7, pp. 66 505–66 511, 2019.

[36] Y. Xia, L. Zhang, L. Meng, Y. Yan, L. Nie, and X. Li, "Exploring web images to enhance skin disease analysis under a computer vision framework," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3080–3091, 2018.

[37] "Google lens," https://lens.google/, accessed: 2024-06-24.

[38] "Niramai," https://www.niramai.com/, accessed: 2024-06-24.

[39] "Qure.ai," https://www.qure.ai/, accessed: 2024-06-24.

[40] "Sigtuple," https://sigtuple.com/, accessed: 2024-06-24.

[41] "Mfine," https://www.mfine.co/, accessed: 2024-06-24.

[42] "Skinvision," https://www.skinvision.com/, accessed: 2024-06-24.

[43] "Cureskin," https://cureskin.com/, accessed: 2024-06-24.

[44] "First derm," https://www.firstderm.com/, accessed: 2024-06-25.

[45] "Model dermatol," https://modelderm.com/en.html, accessed: 2024-06-25.

[46] "Mamaearth," https://mamaearth.in/, accessed: 2024-06-25.

[47] "Cipla," https://www.cipla.com/, accessed: 2024-06-25.

[48] "idoc24," https://idoc24.com/, accessed: 2024-06-25.

[49] "Himalaya wellness," https://himalayawellness.in/, accessed: 2024-06-25.

[50] "Aysa," https://askaysa.com/, accessed: 2024-06-25.

[51] "mcaffeine," https://www.mcaffeine.com/, accessed: 2024-06-25.

[52] "Dermnet image library," https://dermnetnz.org/image-library, accessed: 2023-08-03.

[53] "Mapmyindia api addendums," https://github.com/MapmyIndia/mapmyindia-api-addendums/tree/main/mapmyindia-move-deeplinks/web#1-nearby-facilities, accessed: 2024-04-27.

[54] "Sqlite3 — python 3.13.1 documentation," https://docs.python.org/3/library/sqlite3.html, accessed: 2024-04-27.

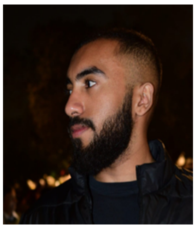
[55] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models," in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, 2020, pp. 79–91.

[56] R. Banjade, "Deepeval: An integrated framework for the evaluation of student responses in dialogue based intelligent tutoring systems," https://core.ac.uk/works/127173147, 2014.

BIOGRAPHIES OF AUTHORS

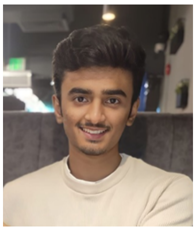
Pradeep Rajeshkumar has completed his bachelor's degree in Electronics and Computer Science Engineering from SIES Graduate School of Technology, Navi Mumbai, affiliated with the University of Mumbai. With a passion for Data Science, Artificial Intelligence, and Deep Learning and a strong foundation in Python programming, he is poised to explore advanced topics such as Advanced Analytics, Machine Learning, and Natural Language Processing. His expertise is further validated through hands-on experience and certifications. .He can be contacted at email: pradeeprecs120@gst.sies.edu.in

Shubhangi Kharche holds a PhD in Electronics and Communication and currently serves as an Associate Professor and Head of Electronics and Computer Science Department at SIES Graduate School of Technology, Nerul, Navi Mumbai. With expertise as an AWS certified Cloud Solutions Architect Associate and Cisco Certified Network Associate, she has adeptly trained numerous students and faculty in Linux, Computer Networking, and AWS Cloud skills. Accumulating over two decades of teaching, training, and research experience, Dr. Kharche stands as a beacon of knowledge in her field. Her contributions to academia and research have been duly recognized, notably with the receipt of the Women Researcher Award at the International Scientist Awards on Engineering, Science, and Medicine in Coimbatore, India, in September 2021. Her research interests encompass a wide array of cutting-edge domains, including Computer Communication Networks, Mobile Cellular Networks, Wireless Sensor Networks, IPv6 Over Low Power Wireless Personal Area Networks, Cloud Computing, Artificial Intelligence, and Data Analytics. With a portfolio boasting more than 35 research papers published in prestigious international conferences and journals, she has demonstrated her prowess as a scholar and innovator. She received the best paper award at the IEEE ANTS conference held at IISc Bangalore in 2016. Furthermore, her dedication to innovation is evident in her intellectual property achievements, having filed one patent and registered several copyrights. Her commitment to professional development is further underscored by her memberships in esteemed organisations such as ISTE, IETE, and Women in 6G. Dr. Shubhangi Kharche's multifaceted expertise and unwavering dedication to advancement make her a true trailblazer in the realms of academia and technology. She can be contacted at email: shubhangik@sies.edu.in, shubahngi.kharche@gmail.com



Prithvi Poojari has completed his bachelor's degree in Electronics and Computer Science Engineering from SIES Graduate School of Technology, Navi Mumbai, affiliated with the University of Mumbai. With a never-ending passion for technology and science and its ability to steer societal progress, Prithvi looks at the world with a learner's perspective, constantly seeking ways to expand his horizons. Driven by a desire to make a meaningful impact, Prithvi actively seeks out opportunities to contribute to projects aimed at advancing humanity. He can be contacted at email: prithvipecs120@gst.sies.edu.in

Sachet Utekar has completed his bachelor's degree in Electronics and Computer Science Engineering from SIES Graduate School of Technology, Navi Mumbai, affiliated with the University of Mumbai.He has a keen interest in Generative Artificial Intelligence (Gen AI), Machine Learning (ML), and Deep Learning (DL), coupled with proficient Python programming skills. He is eager to delve deeper into advanced topics within these fields, exploring cutting-edge concepts and applications. Sachet vigorously ventures out opportunities to engage in work that drive forward societal advancement and improvement. He can be contacted at email: sachetuecs120@gst.sies.edu.in



Sahil Saini has completed his bachelor's degree in Electronics and Computer Science Engineering from SIES Graduate School of Technology, Navi Mumbai, affiliated with the University of Mumbai. He excels in AI, and blockchain. His dedication to hands-on innovation has earned him recognition in prestigious competitions like TechXter 13.0 and the Smart India Hackathon. Sahil's diverse interests include leadership roles such as Technical Head for IEEE SIES GST and organising a blockchain conference. With his ambition and diverse skill set, Sahil is poised to make significant contributions in the engineering field and beyond. He can be contacted at email: sahilsecs120@gst.sies.edu.in

Samrridhi Bdwai a people-oriented Data Scientist, boasts extensive experience in data science and analytics, having worked with industry giants such as Unilever, Hindustan Coca-Cola Beverages, Mondelēz International, General Mills, Sutherland Global Services and Cartesian Consulting. She holds a Master's degree in Statistics from the University of Mumbai and has earned certifications in Tableau and advanced deep learning. Her passion for deep diving into data, business analytics, statistical analysis, and problem-solving has led to array of notable projects, from optimizing salaries for Ice Hockey players to Electrical Equipment Pricing Optimization. With her expertise in the FMCG and CPG domains, Samrridhi has collaborated across sales, marketing, HR, finance, quality, manufacturing, and supply chain sectors. Samrridhi has received multiple recommendations highlighting her teamwork, technical acumen, and leadership qualities. Her work has earned her the Trade Secret Recognition for reducing 90% of product trial time, the prestigious nomination for Olson Service Excellence Award from General Mills, and other accolades such as the Manager's Award and Certificate of Valuable Contribution for her impactful analytics work. She can be contacted at email: samrridhibdwai@gmail.com