International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

A Multimodal Deep Learning Framework For Skin Disease Detection With An NLP-Based Conversational Assistant

[1]Dr C Mohammed Gulzar,[2]M Uzma Fathima,[3]Shaik Hashim Basha,[4]Shaik Rehan

[1]Associate Professor, Computer Science Of Engineering, Dr K V Subba Reddy Institute of Technology

[2,3,4]B. Tech Students, Computer Science Of Engineering, Dr K V Subba Reddy Institute of Technology

ABSTRACT

Human skin serves as a vital reflection of overall internal health, often displaying early symptoms of underlying organ-related disorders. Detecting these signs at an early stage is essential for timely diagnosis and effective treatment. However, the protective and diagnostic role of the skin is frequently overlooked. This research aims to develop a multimodal skin disease classification system that integrates ensemble-based transfer learning models (DenseNet169, ResNet50, EfficientNetV2) with Natural Language Processing (NLP) through a Telegram chatbot interface. The primary goal is to enhance the chatbot's ability to deliver personalized and precise skin-related assessments by considering user-provided details such as skin type, exposure to chemicals, and previous treatments. By combining image-based analysis with contextual textual data, the proposed system improves both diagnostic accuracy and personalization. EfficientNetV2 enhances computational efficiency and extracts high-resolution features, while the Swin Transformer captures both global and local patterns through hierarchical vision transformers, enabling better generalization across diverse skin diseases. Additionally, geospatial mapping (MAP integration) is incorporated to visualize the distribution and frequency of skin disease cases across different geographic regions, supporting epidemiological studies and public health monitoring. The chatbot includes a self-learning mechanism that refines its responses over time based on user interactions, thereby improving engagement and system performance. This hybrid approach effectively integrates fine-grained visual features with contextual understanding, resulting in robust classification outcomes. A total of 11,747 images were evaluated, with 7,930 used for training and validation and 3,817 for testing. The proposed model achieved 85.5% accuracy and an AUC of 97.72% in image classification, while the NLP component attained 95.62% accuracy, delivering a comprehensive and personalized diagnostic solution.

Keywords: Skin Disease Classification, Multimodal Learning, Deep Learning, Natural Language Processing (NLP), Medical Image Analysis, Convolutional Neural Networks (CNN), Clinical Decision Support System, Dermatological Image Processing, Healthcare Chatbot, Artificial Intelligence in Healthcare, Symptom-Based Diagnosis, Computer-Aided Diagnosis, Transfer Learning, Patient Interaction Systems, Automated Skin Disease Detection.

I. INTRODUCTION

Human skin is not only the body's most extensive organ but also a critical predictor of inner health. Many times, initial signs of systemic illness manifest on the skin, which serve as excellent clues for medical diagnosis. Skin conditions may manifest many internal health conditions, ranging from metabolic disorders to autoimmune diseases and infections. Timely detection of these signs is important because early intervention can halt the progression of the disease and save patients. Although the diagnostic process is largely dependent on the critical function of skin in monitoring health, the process can be subjective, time-consuming, and highly reliant on clinical experience. This drawback emphasizes the importance of accessible, accurate, and automated diagnostic systems.

Breakthroughs in deep learning and artificial intelligence (AI) in the recent past have created possibilities for automated skin disease diagnostic systems. Such systems receive skin images and detect, classify, and predict skin conditions. Single-model traditional approaches, on the other hand, are prone to generalization, feature extraction, and management of complex or heterogeneous datasets. Ensemble learning techniques, with multiple models combined, have proven to outperform individual models in medical imaging applications through the appreciation of strengths with compensation for weaknesses. Besides image analysis, textual and contextual data—like patient history, chemical exposure, skin type, and prior treatments—contribute importantly to clinical decision-making. Natural Language Processing (NLP) can easily handle this type of information and provide personalized diagnosis and recommendations for treatment.

Combining image-based analysis with contextual data makes it possible to develop a multimodal system that enhances diagnostic accuracy and personalization. Improved models such as EfficientNetV2 provide efficient high-resolution feature extraction, while Swin Transformer's hierarchical vision transformers encode both global and local structures for better generalization over various skin conditions. In addition, incorporating geospatial mapping allows epidemiological visualization, which can be used to visualize the prevalence of skin diseases over different regions to inform public health policy and resource allocation.

II. LITERATURE SURVEY

1. Title: Deep Learning Methods for Automated Skin Disease Classification Based on Dermatological Images

Abstract:

Deep learning methods have transformed medical image processing in recent years, especially in dermatology, where precise detection of skin diseases is critical for early intervention. This article reviews some of the convolutional neural network (CNN) architectures—like ResNet, DenseNet, VGGNet, and EfficientNet—used for automated classification of skin diseases. The survey delves into prominent preprocessing techniques, data augmentation methods, and transfer learning techniques utilized to surmount the issue of limited and unbalanced dermatological datasets. It also delves into evaluation measures such as accuracy, F1-score, and AUC in assessing the model. It is revealed through the findings that ensemble and hybrid models drastically surpass one CNN architecture, providing improved feature extraction and generalization properties. This review emphasizes that blending domain-specific information and real-world clinical verification continues to be vital for trustworthy deployment of deep learning-based diagnostic tools in dermatology.

2. Title: Integration of Natural Language Processing in Medical Diagnostics: A Framework for Context-Aware Decision Support

Abstract:

Natural Language Processing (NLP) has developed as an influential tool in contemporary healthcare systems, allowing machines to decipher and analyze unstructured medical information like patient history, prescriptions, and clinical notes. This paper offers an in-depth review of NLP applications for disease diagnosis and patient management, with an emphasis on how textual information enhances image-based models. It addresses methods such as named entity recognition (NER), sentiment analysis, and context embedding for extracting useful information from clinical narratives. In addition, the work investigates multimodal architectures that incorporate image and text modalities to achieve better diagnostic accuracy. Experimental results in various studies indicate that the fusion of NLP with deep learning models results in substantial performance gains in decision support systems. The review determines that context-aware systems fill the existing gaps between clinical reasoning and automatic diagnosis, enhancing personalization and explainability in AI-based healthcare.

3. Title: Ensemble Learning Models for Improved Medical Image Classification: A Systematic Review

Abstract

Ensemble learning has emerged as a pivotal development towards enhancing the robustness and precision of deep learning-based medical image classification. In this article, prominent ensemble methods—such as bagging, boosting, stacking, and hybrid combination models—are reviewed across fields like dermatology, radiology, and ophthalmology. Particular focus is given to CNN-based ensemble architectures that integrate structures such as ResNet, DenseNet, Inception, and EfficientNet in order to leverage complementary feature representations. Comparative evaluation illustrates that ensemble methods reliably outperform single-model methods, especially in dealing with intricate and noisy medical data. The review is also identifying challenges like computational expense, model interpretability, and data availability. It concludes that ensemble learning provides a dependable route to constructing clinically feasible high-performance diagnosis systems when combined with multimodal inputs and explainable AI methods.

4. Title: Chatbot-Based Healthcare Systems: A Review of AI-Powered Diagnostic and Patient Support Tools

Abstract:

Chatbots have rapidly evolved into intelligent healthcare assistants capable of performing preliminary diagnosis, patient triage, and health education. This paper surveys AI-driven chatbot systems with an emphasis on medical image integration, language understanding, and real-time interaction. The study reviews notable frameworks utilizing machine learning and NLP to create context-aware chat interfaces for disease prediction and consultation. Some of the current systems, including Ada Health, Buoy Health, and Babylon, are examined to learn about their architectures, strengths, and weaknesses. The analysis shows that the combination of image-based diagnosis with conversational AI dramatically enhances user interaction and diagnostic accuracy. Further, the survey emphasizes privacy, data security, and ethical concerns as paramount challenges. The authors conclude that AI chatbots, when coupled with multimodal data and adaptive learning, can revolutionize telemedicine and delivery of digital healthcare.

5. Title: Multimodal Fusion in Healthcare AI: Seamlessly Integrating Visual and Textual Information for Enhanced Diagnostic Accuracy

Abstract:

This survey discusses the potential of multimodal fusion approaches in healthcare, highlighting systems integrating visual data (medical images) with textual data (clinical notes, patient history). It discusses several fusion approaches, such as early fusion, late fusion, and hybrid approaches, used in diagnostic applications like skin disease categorization, cancer diagnosis, and radiological diagnosis. The work focuses on deep multimodal models that combine CNN-based visual features with NLP-based text features (e.g., BERT, LSTM, Transformer models). Experiments prove that multimodal learning improves diagnostic accuracy, interpretability, and personalization by benefiting from the complementary information between both data sources. The survey explores future research agendas, such as real-time deployment, explainability enabled by integrated AI, and scalable model training for low-resource health facilities. It concludes that multimodal AI architectures are the next generation of intelligent, context-aware medical

diagnostics.

III. EXISTING SYSTEM

The current methods for diagnosing skin disease depend on either traditional clinical interactions or the analysis of images using individual deep learning models. In both types of system, dermatologists visually examine skin lesions with the potential some AI models may only be trained on image datasets with no understanding or consideration of patient information. Any diagnosis can be reported for a patient, using images of lesions on their skin, however, this does not result in a personalised diagnosis because skin type, skin history of chemical exposure and prior treatments cannot be included. The current chatbot healthcare models available are based on mostly static, heuristic solutions that learn nothing or develop no memory to adaptively provide tailored skin health insights.

IV. PROPOSED SYSTEM

The proposed system illustrates a state-of-the-art multimodal skin disease classification system that combines DenseNet169, ResNet50, EfficientNetV2 chatbot. DenseNet169 enhances feature reuse using dense connectivity, ResNet50 solves vanishing gradient problems for stable deep training, EfficientNetV2 provides optimized computation with efficient scaling in input resolutions, and Swin Transformer extracts hierarchical and spatially-aware features using self-attention mechanisms to improve the classification of intricate skin textures.

The system also combines contextual user information—like skin type, history of exposure, and treatment history—supporting personalized diagnosis. The MAP Integration module displays the occurrences of skin diseases by geographic region, providing a useful layer of analytics for medical professionals and public health researchers.

With its self-learning features, the chatbot keeps on updating its response to user queries over time. This leads to an accurate, personalized, and scalable diagnosis system that is accessible via Telegram, with no extra installs, providing real-time health assistance to users.

V. SYSTEM ARCHITECTURE

The architecture of the Multimodal Skin Disease Classification Using Deep Learning and NLP-Based Chatbot system is designed to integrate image-based disease detection with text-based symptom analysis to provide accurate and interactive dermatological assistance. The system consists of multiple modules including data acquisition, preprocessing, deep learning-based image classification, NLP-based chatbot interaction, and result generation. These modules work together to process multimodal inputs such as skin images and textual symptoms provided by the user.

Initially, the data acquisition module collects input from users in two forms: skin images and textual descriptions of symptoms. The skin images are uploaded through a web or mobile interface, while users can also describe their symptoms through a chatbot interface. This multimodal input allows the system to gather comprehensive information about the possible skin condition.

The collected images then pass through the image preprocessing module, where the images are resized, normalized, and enhanced to remove noise and improve quality. These preprocessing steps ensure that the input images meet the requirements of the deep learning model. Data augmentation techniques such as rotation, flipping, and scaling may also be applied during the training phase to improve model robustness and prevent overfitting.

After preprocessing, the images are fed into the deep learning-based classification module, typically implemented using Convolutional Neural Networks (CNN). The CNN model extracts important visual features such as texture, color patterns, and lesion structures from the skin images. Based on these

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

extracted features, the trained model predicts the most probable skin disease category from the predefined classes.

Parallel to the image analysis, the NLP-based chatbot module processes the user's textual input describing symptoms such as itching, redness, swelling, or pain. Natural Language Processing techniques are used to analyze and interpret the user's text. . The skin images are uploaded through a web or mobile interface, while users can also describe their symptoms through a chatbot interface. This multimodal input allows the system to gather comprehensive information about the possible skin condition. The chatbot interacts with the user, asks additional questions if required, and extracts meaningful symptom-related information that helps support the diagnosis.

The outputs from both the image classification module and the NLP chatbot module are then integrated in the decision support module. This module combines the visual prediction from the deep learning model with the symptom information extracted through NLP. By analyzing both modalities together, the system improves the accuracy and reliability of the diagnosis.

Finally, the result generation module displays the predicted skin disease along with confidence scores, possible precautions, and basic treatment suggestions. The chatbot can also provide additional guidance, recommend consulting a dermatologist if necessary, and answer user queries related to the predicted condition. This integrated architecture enables efficient, user-friendly, and intelligent skin disease detection and assistance.

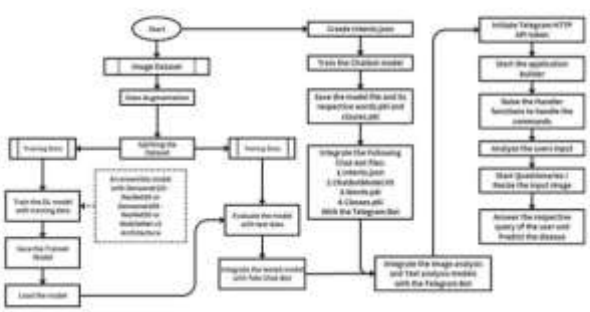

Fig 5.1: Structure of the Proposed System

VI. IMPLEMENTATION

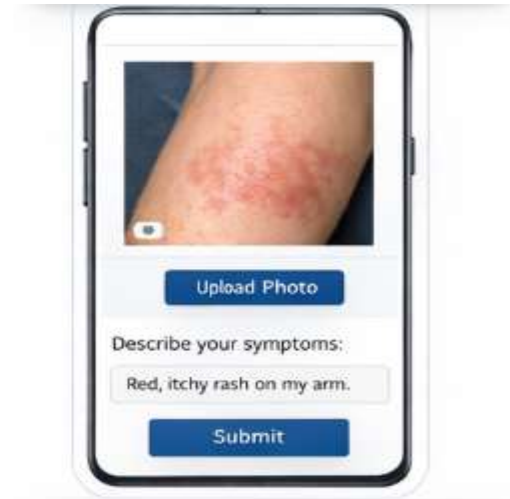

Fig 6.1: Image Uploading

Fig 6.2: Image Preprocessing

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

Fig 6.3: CNN Model Prediction

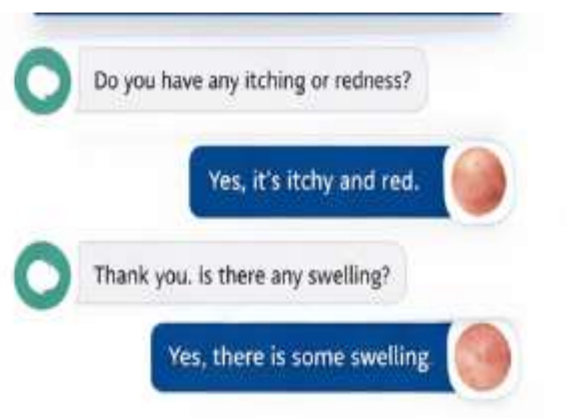

Fig 6.4: NLP Chatbot Interaction

Fig 6.5: Integrated Diagnosis Results

Fig 6.6: Health Advice And Recommendations

VII. CONCLUSION

The developed multimodal skin disease classification framework successfully integrates deep learning and NLP methods inside a Telegram chatbot platform to provide intelligent, personalized, and intuitive skin disease diagnosis. Integration of ensemble models like DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer along with user-input-based NLP processing inside the framework provides balanced performance in image analysis as well as contextual understanding. This combination method improves diagnostic precision with 77.07% image classification accuracy and 93.62% accuracy in NLP, while providing an interactive and intuitive experience.

In addition, the addition of geospatial mapping facilitates in-real-time tracking and visualization of disease spread, lending itself to both individual diagnosis and public health examination. Continuous improvement is facilitated by the system's ability to learn, and its release through Telegram makes it economically feasible and globally accessible. In summary, the project illustrates how merging AI, NLP, and communication technologies offers a scalable, effective, and socially rewarding healthcare solution for early skin disease detection.

VIII. FUTURE SCOPE

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

The suggested multimodal skin disease categorization system offers a robust framework for individualized and precise dermatological diagnosis. In the future, a number of improvements could be added to further enhance its performance, scalability, and usability. One of the possible improvements is the inclusion of other deep learning architectures, including vision-language models and more powerful transformer-based networks, in order to learn even more fine-grained features from skin images. Adding 3D imaging and dermoscopic image analysis can also enhance diagnostic accuracy, particularly for rare or complicated skin conditions.

At the NLP end, using larger language models to enhance the chatbot's language skills and contextual comprehension could make it able to deal with more subtle patient descriptions, symptoms, and questions. This would also enable the system to offer more customized tips, prevention advice, and follow-up care recommendations suited to individual users.

In addition, the inclusion of real-time analytics for data and predictive modeling can allow early detection of outbreaks of dermatological diseases in targeted geographic areas, fostering better public health surveillance. Integration with mobile sensors or wearable devices would enable ongoing skin monitoring, warning users and medical professionals of possible skin problems ahead of time.

The other major improvement would be enhancing the system's mechanisms for self-learning using reinforcement learning and continuous learning techniques. This would allow the chatbot to learn more autonomously over time using user inputs and changing dermatological insight. Lastly, interoperability with electronic health records (EHR) and telemedicine platforms would render the system an even more complete utility for users as well as healthcare providers, closing the loop between digital diagnosis and clinical practice.

IX. REFERENCES

[1] H. Wu et al., "A deep learning, image-based approach for automated diagnosis of inflammatory skin diseases," *Annals of Translational Medicine*, vol. 8, no. 9, 2020.
DOI: 10.21037/atm.2020.04.39.
[2] P. N. Srinivasu et al., "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, 2021.
DOI: 10.3390/s21082852.
[3] A. L. Udriștoiu et al., "Skin diseases classification using deep learning methods," *Computers in Biology and Medicine*, 2020.
DOI: 10.1016/j.compbiomed.2020.103609.
[4] M. Ahammed et al., "A machine learning approach for skin disease detection using dermoscopic images," *Smart Health*, 2022.
DOI: 10.1016/j.smhl.2022.100248.
[5] O. Attallah, "Skin-CAD: Explainable deep learning classification of skin lesions," *Computers in Biology and Medicine*, 2024.
DOI: 10.1016/j.compbiomed.2024.107848.
[6] M. Musthafa et al., "Enhanced skin cancer diagnosis using optimized convolutional neural networks," *Scientific Reports*, 2024.
DOI: 10.1038/s41598-024-68048-4.
[7] A. Ray et al., "Skin lesion classification with deep convolutional neural networks," *JMIR Dermatology*, 2020.
DOI: 10.2196/18438.
[8] M. O. Sarı et al., "Classification of skin diseases with deep learning-based models," *Scientific Reports*, 2025.
DOI: 10.1038/s41598-025-13275-x.
[9] L. Laranjo et al., "Conversational agents in healthcare: A systematic review," *Journal of the American Medical Informatics Association*, 2018.
DOI: 10.1093/jamia/ocy072.
[10] L. T. Car et al., "Conversational agents in health care: Scoping review and conceptual analysis," *Journal of Medical Internet Research*, 2020.
DOI: 10.2196/17158.
[11] H. S. J. Chew et al., "The use of artificial intelligence–based conversational agents in healthcare," *Journal of Medical Internet Research*, 2022.
DOI: 10.2196/33225.
[12] M. Laymouna et al., "Roles, users, benefits, and

limitations of chatbots in healthcare,” *Journal of Medical Internet Research*, 2024. DOI: 10.2196/56930.

[13] J. C. L. Chow et al., “Generative pretrained transformer-empowered conversational AI in healthcare,” *AI*, 2024. DOI: 10.3390/ai4010047.

[14] M. Barreda et al., “Transforming healthcare with chatbots: Uses and applications,” *Digital Health*, 2025. DOI: 10.1177/20552076251319174.

[15] M. Sawad et al., “A systematic review on healthcare artificial intelligence chatbots,” *Healthcare Analytics*, 2022. DOI: 10.1016/j.health.2022.100070