SMART DERMATOLOGY: SKIN DISEASE DETECTION AND RECOMMENDATION SYSTEM USING VISION TRANSFER

1 Mr.V. UDHAYAKUMAR, 2 DEEPA SHIVANI

1 Assistant Professor, Department of Computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605008, India
udhayakumar.mca@smvec.ac.in

2 Post Graduate student, Department of Computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605008, India
deepashivaniaroumougam@gmail.com

ABSTRACT:

Skin disorders, including acne, eczema, psoriasis and melanoma, are prevalent and can result in serious health concerns. They often represent some of the first signs of serious illnesses. Timely and accurate identification of skin disorders is paramount, however the traditional diagnostic approach is primarily visualized; therefore it is tedious, subjective and reliant on expertise. These approaches tend to miss the subtle differences between diseases that have a similar visual presentation, in addition, they lack personalized value propositions for caring for patients. In this project we are proposing a next generation diagnostic solution in which the classifications utilize a hybrid of Vision Transformers (ViT) and Convolutional Neural Networks (CNNs). The ViT model partitions the input images into patches and then some self-attention module is used to learn the global context as well the subtle features of the lesion. The CNN modules will use the ViT-processed feature to provide overall classification of skin diseases including melanoma, basal cell carcinoma, and benign keratosis in addition to being able to provide care options. The proposed system will suggest treatments, check physician availability to dermatologists or specialized hospitals nearby, and suggest a remedy to take responsibly at home based on the diagnosis. The proposed solution will employ a recommendation system to connect the deep learning to the skin disease function with the system that is expanding the diagnostic accuracy, provide results significantly faster and egregiously improve the process accessibility of the progressively serious illness. It is a potent and scalable model that gives us a chance to overhaul the way skin disease diagnostics is done, while also empowering patients by better, more timely and higher-quality medical intervention possibly resulting in better patient outcomes.

Keywords: *Skin Disease Detection, Vision Transformer (ViT), Convolutional Neural Network (CNN), Deep Learning, Self-Attention Mechanism, Image-Based Diagnosis, Basal Cell Carcinoma.*

1. INTRODUCTION

The overall significance of the proposed project is to improve the diagnosis and treatment of skin diseases through advanced deep learning strategies. Skin diseases like acne, eczema, psoriasis, and melanoma are among the most prevalent type of clinical conditions, and they can also represent underlying health issues. Diagnostic work takes significant amounts of time for investigators since the majority of current diagnosticized rely on inspecting an image to provide a diagnosis. However, human inspection can fail due simply to the limited range and appreciation that humans have to identify changes in image abnormalities even on high-resolution skin abnormalities in less than 30 seconds. Additionally, current systems are unable to address the differentiation of various diseases as well as provide personalized treatments options. Our project proposed an intelligent, automated system composed of Vision Transformers (ViT) and Convolutional Neural Networks (CNN) to accurately identify and classify skin conditions.

The proposed system uses an Intelligent ViT as the core of the system, which will provide high-resolution representation from skin lesion images obtained from public skin lesion datasets by assuming patch representation structure in an image to account for the ViT architecture. Thus, the model will take a single high-resolution input image. This image will be cut into patches and the self-attention from ViT would be able to capture local image features and global image features allowing accurate identification of any abnormality down to the fine-grained image representation.

2. LITERATURE SURVEY

Adegun (2020) [1]: Adekanmi A. Adegun proposed a deep learning framework combining Fully Convolutional Networks (FCNs) with DenseNet to automate detection and classification of skin lesions in dermoscopy images. The study addressed challenges such as fuzzy lesion boundaries, artifacts, and limited datasets. By integrating segmentation and classification, the model improved accuracy in detecting complex lesions while minimizing computational load, enhancing the robustness of CAD systems.

Andreasen and Crandall (2021) [2]: Andreasen and Crandall explored the use of skin electrical resistance as a non-invasive biomarker for diagnosing and monitoring breast cancer. Focusing on lymphatic regions, their research demonstrated that electrical resistance measurements could effectively differentiate between malignant and benign breast lesions. This method presents a promising supplement to conventional diagnostic tools, offering accessibility and continuous therapy assessment.

Imran et al. (2022) [3]: Imran et al. developed an ensemble deep learning model combining VGG, CapsNet, and ResNet to improve skin cancer detection. Using the ISIC dataset, their integrated model outperformed individual classifiers in accuracy, sensitivity, and specificity. This ensemble strategy enhances diagnostic reliability, addressing the weaknesses of traditional machine learning methods in medical imaging.

Schiavoni and Maietta (2022) [4]: Schiavoni and Maietta introduced a microwave reflectometry system for in-vivo, low-cost skin cancer diagnostics. This approach utilizes dielectric contrast at microwave frequencies to detect cancerous tissues, offering high sensitivity, reduced invasiveness, and compact design. The study emphasized its potential as a fast, accessible alternative to conventional diagnosis methods.

Mridha and Uddin (2023) [5]: Mridha and Uddin proposed a Region-of-Interest (ROI) based transfer learning framework for skin cancer detection. By applying an improved k-means algorithm to isolate lesion-focused patches, their method enhances feature extraction in visually complex datasets. The CNN-based model demonstrated high accuracy in distinguishing melanoma from nevi, addressing overfitting and data limitation issues in CAD systems.

Riaz and Qadir (2023) [6]: Riaz and Qadir introduced a joint learning system combining CNNs and Local Binary Patterns (LBPs) to detect skin cancer from dermoscopy images. Their framework improves feature extraction and classification accuracy, particularly in large-scale datasets. The model supports early diagnosis and efficient clinical decision-making, contributing to improved patient outcomes.

3. PROBLEM STATEMENT

Skin diseases are among the most prevalent health issues globally and affect people of all ages. Therefore, early recognition and accurate diagnosis of skin diseases are essential under control complications. However, the traditional methods of diagnosis depend on dermatologists or other trained professionals. This can lead to subjectivity in interpretation and also can make access to dermatologists difficult in certain regional areas. Therefore, a patient can wait weeks or months to get a specialist consultation, which can lead to flare-ups or exacerbations of skin diseases. Another major hurdle is differentiating between skin diseases that look similar. Skin diseases such as eczema, psoriasis, and fungus all have similar traits. Therefore, traditional image processing techniques may lead to high accuracy, but medical professionals find it difficult because of visual overlap and similarities. Automated systems often struggle to recognize fine distinctions and may not generalize well across skin tones, lesion types, and lighting conditions. Traditional Machine Learning Models do not deal well with large, varied datasets, which also leads to issues in terms of scaling and trustworthiness. Most systems do not include rigorous feature extraction, meaning that they are not as helpful in real-world situations which would often be much more complex. Furthermore, current automated and/or AI solutions in dermatology mostly deal with detection, offering no treatment recommendations or information about what patients should do next after having a condition diagnosed. Therefore, there is a need for an AI-based approach in dermatology which utilizes Vision Transformers (ViT) and Convolutional Neural Network (CNN) methods.

4. PROPOSED SYSTEM METHODOLOGY

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04466
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

The proposed system utilizes a multi-tier client-server architecture, which provides for efficient, scalable and maintainable operations. Users will interact with the application using a responsive and easy-to-use frontend built with Bootstrap, allowing it to work seamlessly across platforms and devices with varying screen sizes. The frontend will communicate with the backend that is built with Flask using RESTful APIs to enable data flow and interaction in real-time. The backend will handle the image uploads to the system, apply business logic, and integrate with Artificial Intelligence (AI) models. Specifically, we use a Vision Transformer (ViT) for extracting the features, and a Convolutional Neural Network (CNN) for classifying skin diseases. Our AI pipeline ensures highly accurate collection and analysis of medical images. The backend will also connect to a MySQL database, which will store user information, diagnosis history, and personalizations recommendations of treatment in a secure manner.

The architecture will provide modularity and separation of concerns for our work, meaning we can send updates to the frontend, the backend, AI modules, or the database independently of the rest of the system. This layered architecture will improve performance and better secure the system, as well as reduce debugging time and provide for potential future expansions of the AI pipeline, thus making it more applicable for healthcare projects.

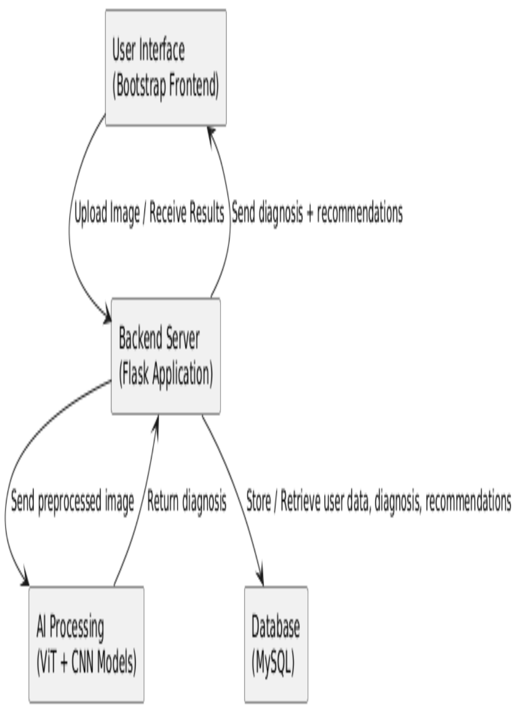

Fig 4.1

Proposed Architecture Design

A multi-tier architecture designed to detect skin diseases effectively using various deep learning techniques is shown in the system diagram. Components of the architecture include a Flask-developed backend server, a responsive user interface, AI processing components that use Python-developed Vision Transformer (ViT) and Convolutional Neural Networks (CNN), and MySQL as a data storage engine. Each part of the system is described in the paragraphs that follow, along with how they work together as a whole:

4.1 User Interface Layer

The system's user interface was developed in Bootstrap, an open-source, popular front-end framework that allows for responsive design and is mobile-ready across devices. This user interface is the primary place to interact with our application and has an intuitive and accessible design. Users can upload images of their skin lesions and analyze them with our application, regardless of whether or not they have technical know-how. The process is simple and straightforward. Once the user uploads an image, the front-end UI sends the image to the backend server through HTTP, using defined APIs. After processing the image, the backend server returns diagnostic results and medically relevant suggestions, which are easy to digest and comprehend in the UI. Our design helps the user experience by providing informative feedback quickly while allowing the user to effectively receive accurate diagnostic insights and recommendations. The Bootstrap-based front-end UI is a key aspect of the user experience for our application. In summary, the front-end UI allows us to effectively connect our users to the back-end server utilising our AI-driven diagnostic system while making use of rapid feedback through our real-time snapshots of the available output data.

4.2 Backend Server Layer

Flask-based back end serves as a control center, showing the communication, logic execution, and data flow between all parts of the software. When a user uploads a skin lesion image in the front end, the Flask server receives it, and begins the preprocessing steps of image resizing, normalization, and format conversion before using it to classify the image. Once preprocessing is completed, the image is then sent to the AI processing module, where features are extracted and classified with a Vision Transformer and Convolutional Neural Network. After the AI module returns the diagnostic result, the back end is responsible for the final storage of the diagnosis result and user data and personalized medical recommendation to the MySQL database, as well as retrieving relevant information from the MySQL when needed. The back end also consolidates the

ISSN: 2583-6129
DOI: 10.55041/ISJEM04466

diagnosis and recommendations and sends it to the user interface. The overall design goal of the Flask back end is to operate securely and efficiently to control overall communication and work reliably throughout the system.

4.3 AI Processing (ViT + CNN Models)

The combined deep learning architecture of the AI processing unit in the system is based on a hybrid of a Vision Transformer (ViT) and a Convolutional Neural Network (CNN) to improve the skin disease detection accuracy. In general terms, a Vision Transformer (ViT) proceeds by dividing the input image into small patches and applies self-attention mechanisms to characterise the global context and fine-grained visual features allowing the model to observe the subtle patterns and relations in spatial dependencies for the lesions. The features produced by ViT are passed to the CNN, which is tasked with the classification task. The structure of the CNN is hierarchical, where spatial patterns and texture of the lesion inform us if the skin disease is a melanoma, nevi, keratosis or carcinoma. Combining both the ViT and CNN is advantageous, as ViT has an explicit global representation and CNN has a strong spatial representation. Together they improve the models diagnostic accuracy ideally suited for distinguishing skin conditions with visual similarities.

4.4 Database (MySQL)

The MySQL database functions as an essential element for reliable data management and consistent storage within the system. It is responsible for storing the user's credentials and their profile for secure access and personalized services. The MySQL database will also store the image of skin lesion uploaded by the user, including metadata such as the date and time it was taken, format of the image, etc. It stores diagnostic outcomes returned from the AI module displaying the skin condition it diagnosed and the confidence it has in the diagnosis. Additionally, the database stores personalized recommendations from every diagnosis, to include treatment options offered, a list of dermatologists, hospitals with dermatology departments nearby and any suggested home treatments. This organized database storage will allow for seamless data search capabilities - enabling the ability to perform extended analytical work, like tracking a user's past diagnoses over time. In this way, the database allows for access to a user's historical data, facilitating the ongoing user insight and increasing the ongoing user experience for returning users. In summary, MySQL would ensure scalability, integrity and responsiveness of the system in handling any number of medical-related data or user-related data.

5. CONCLUSION

To conclude, the project represents a state-of-the-art artificial intelligence powered platform that integrates ViT and CNNs to develop both an effective and precise method for skin disease detection. Built using Python, Flask, MySQL, Bootstrap, and several other modern libraries, provides a high degree of user experience with real-time image analysis, accurate classification, and patient-specific treatment advice. Specilaist referrals, medication advice, and home remedies are also part of a comprehensive approach for patients. The application was designed to be scalable, vernacular, and efficient. As a tool to enable reliable early detection and diagnosis, especially for patients living remotely - our approach means that we can offer an early diagnosis rather than merely waiting for patients to visit full time specialists. Our recommendations and future work may also include obtaining a larger dataset for more accurate learning results, improve our AI models, and look into the use of Blockchain technology for the surgical and secure use of medical data - establishing a more robust and reliable system access to diagnostic tools.

6. FUTURE ENHANCEMENT

Several distinct advancements to the proposed system could be made. First, disease classification can be improved greatly by using more advanced deep learning architectures like Swin Transformer and EfficientNet that strictly outperform others on visual recognition tasks. Second, by employing a much larger dataset with more diverse images covering skin tones, age groups, and uncommon skin diseases, the underlying model's generalization and bias will be improved. Third, blockchain should be used to provide secure, transparent, and tamper-proof storage of sensitive medical information to guarantee patient privacy and trust. Additionally, the system can provide a guided user experience through telemedicine components, such as AI chatbots and live video consultations, with dermatologists to better assist users in real-time. The development of a mobile application is also important to increase accessibility for the proposed system and allow users to complete skin assessments with their smartphone cameras while on-the-go. Finally, a more explicit and dynamic real-time notification feature will enhance users by allowing detection of any urgency with immediate notification with their diagnosis, ultimately enabling better health outcomes through timely intervention.

7. REFRENCES

1. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161.

2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*.

3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

4. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

5. Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.

6. Mahbod, A., Schaefer, G., Wang, C., & Ecker, R. (2021). DermGAN: Synthetic generation of clinical skin lesion images with pathology. *Scientific Reports*, 11(1), 15354.

7.Yadav, S. S., & Jadhav, S. M. (2022). Skin disease detection using deep learning and image processing techniques: A review. *Journal of King Saud University – Computer and Information Sciences*, 34(8), 5099–5110.

8. Zhang, Z., Liu, Q., & Wang, Y. (2020). Road to the classification of skin lesions: Transfer learning from pretrained models. *IEEE Access*, 8, 128025–128038.

9. Haenssle, H. A., Fink, C., Schneiderbauer, R., et al. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, 29(8), 1836–1842.