© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715038

AI-Based Skin Disease Classification Using Deep Learning on Dermoscopic Images

BALAJI G[1], MONISHA V K[2]
[1, 2]Assistant Professor, Department of Computer Science Hindustan Institute of Technology and Science, Chennai India

Abstract- ***Skin diseases are one of the major health challenges in the world, and their occurrence differs among different populations. The health consequences of these diseases can be extremely severe. Therefore, early and precise detection of the diseases is very important in order to intervene in such cases, and this can be done in an accurate manner using the proposed model as compared to the conventional diagnosis process, as the conventional diagnosis process is more dependent on visual inspection and can differ according to the availability of dermatologists in different regions of the world. Therefore, this paper presents a novel framework based on AI technology to classify different skin diseases using dermoscopic images. Although the conventional diagnosis process can be more prone to error, the proposed model is based on a hybrid model that uses CNN and Transformer, and different images are used to train the model in order to make it more robust and accurate to handle images of different skin tones and types, as different images are used to train the model using various data sets, including ISIC and HAM10000, and the results show that the proposed ensemble learning approach can achieve more than 96Additionally, we incorporate Explainable AI (XAI) methods, including Grad- CAM and SHAP, to enable clinicians to interpret the rationale behind each classification decision through transparent visual explanations. By bridging the gap between high-performance deep learning and interpretability, this research targets the development of teledermatology initiatives and improves the accessibility of healthcare services in under-served regions.***

I. INTRODUCTION

Skin diseases have become a major problem, especially with regard to the health of people all over the world. This is because, with skin diseases, people from all over the world, regardless of their demographic background, are not affected in a particular manner, which means that people are suffering from diseases such as fungal infections, as well as deadly diseases such as skin cancer and melanoma, not to mention the pain and suffering that people are enduring because of the visual manifestation of the disease on the body of the patient, which means that the patient is suffering not only physically but is also suffering psychologically because of the issues that they are developing with regard to their own self-esteem and depression, which is why this is not only a medical necessity but is also a form of intervention that is crucial with regard to the health of the patient. However, in spite of the rapid pace at which medical science is progressing, conventional dermatology has also been plagued by some systemic problems. The first major problem is the geographical distribution of healthcare services, whereby dermatologists tend to concentrate in urban areas, leaving rural populations at a disadvantage in terms of access to quality healthcare services. Secondly, conventional diagnosis is based on visual examination, which is, in turn, a very subjective process and therefore prone to errors due to human factors. Even for a professional, it is not easy to make a distinction between benign and malignant growths, and this can sometimes lead to inconsistent results in diagnosis.

The advent of Artificial Intelligence (AI) and Deep Learning (DL) has provided a major opportunity to overcome these problems, and this is especially true in terms of the use of dermoscopic images, whereby complex patterns, textures, and pigmentation can sometimes be difficult to detect even for the naked eye. A highly effective model for classifying skin diseases has been proposed in the present paper. The proposed research utilizes the ensemble technique for achieving this purpose. The research utilizes the latest technology in the form of Convolutional Neural Networks (CNNs) and Transformers to bridge the gap between conventional diagnostic practices and modern computing capabilities. The proposed research aims to create a diagnostic tool that not only

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715038

offers high accuracy in results but is also easy to use, thus acting as a" second opinion" for diagnostic results

II. LITERATURE REVIEW

The domain of dermatological diagnosis has observed a major paradigm shift towards the incorporation of DL for addressing the limitations of traditional visual inspections. Traditionally, research in this domain has employed standard CNN architectures for feature extraction and classification. For instance, research in disease diagnosis utilized various medical datasets such as ISIC and HAM10000 for training a model with the capability of accurately diagnosing diseases such as malignancies and melanoma.

Currently, research in dermatology is employing more sophisticated architectures for improving diagnostic accuracy. For instance, Ravi (2022) proposed a DL-based approach for skin cancer detection, taking into consideration the costs of misclassification. This is evident from the fact that Ravi emphasized the need for developing AI-based models for dermatology, considering its criticality. Continuing with this research, Janbi et al. (2023) proposed" Imtidad," a distributed AI-based model for diagnostic assistance in skin diseases, validating the use of the HAM10000 dataset for multi-class classification tasks.

In addition to classification, research is emphasizing the need for ensuring fairness and accuracy of representations. The STAR-ED framework was developed by Tadesse et al., and it highlights the significance of using different datasets, such as the Fitzpatrick17K dataset, to ensure that biases are not introduced into the machine learning models. Furthermore, there is a growing interest in developing hybrid models. In fact, current research suggests that although CNN models are effective in detecting textures and pigmentation features, the use of Transformer models is effective in detecting contextual features, hence improving the performance of the model in dermoscopic analysis.

Lastly, there is a growing interest in Explainable AI (XAI) and teledermatology. In fact, as AI models move from the lab to practical applications, there is a growing interest in decision- making using models such as Grad-CAM and SHAP, hence improving clinician trust and accountability

III. EXISTING DRAWBACKS

In spite of the great progress of AI-oriented diagnosis for skin disease, a number of obstacles have hindered clinical application at the stage. A comprehensive review of the literature reveals five major challenges: dataset imbalance, computational demands, restricted clinical validation, interpretability problems and ethical concerns. Addressing these issues is essential for developing trustworthy and inclusive AI in dermatology. Imbalance of the dataset and lack of diversity in samples An important issue is the in balance and the lack of diversity of current dermoscopic datasets. Most benchmark datasets, such as ISIC are predominantly composed of images from lighter skin types and specific geographical locations. This imbalance leads to biased model training; certain groups such as those with darker skin are not well served by AI. Not only do such differences diminish accuracy of diagnosis, but they also have the potential to exacerbate extant healthcare inequities. Additionally, rare skin diseases are Heavy computation in advanced models

State of the art deep learning models, notably recent Trans- former based ones have excelled in image classification. But these models are computationally expensive, memory- intensive and they require highly-specialized hardware like GPUs or TPUs. Model training and deployment in clinical or teledermatology environments would be challenging, especially in low-resource settings without much computational infrastructure. The environmental and cost issues are consequently caused by high energy consumption and longer training time. Low evidence in clinical validation and real-world application

Although AI models frequently demonstrate high accuracy under controlled laboratory settings, their performance in clinical care settings remains largely untested. Most of the studies adoptcurated datasets which might not represent variation in real-world image quality, illumination, and patient demographics. In addition, few AI systems have been subjected to rigorous clinical testing or approved for routine maintenance in health care Using AI Text. Without extensive clinical validation and real-world testing, there is a risk of overestimating AI's effectiveness and underplaying the difficulties of integrating it into daily dermatological practice.

Lack of interpretability in deep learning models Deep learning models, especially CNNs and Transformers, often face criticism for being "black boxes" with limited transparency in their decision-making processes. The advantage lies in the fact that they can perform categorization with high precision but, on the downside, often cannot specify the reason for which the classification was done in a particular way. This fact creates low levels of trust for the clinicians who depend on proof for effective decision-making. Transparency becomes an issue in evaluating mistakes and accountability in the event of misdiagnosis. Current research and developments in explainable AI are working towards solving this dilemma and have resulting solutions such as visualization and heat maps. Ethical and legal issues associated with teledermatology Is- sues of ethics as well as legality have arisen when considering whether or not to integrate AI technology into teledermatology. Issues such as patient data privacy are a concern, considering that data will be stored in the cloud, which is seen as less secure than physical data storage facilities. Additionally, there is the issue of liability, considering that if an AI diagnosis proves to be incorrect, it remains unclear as to who is liable, be it the developer, provider, or institution. Regulatory guidelines for AI in healthcare are still developing, and without clear
rules, the ethical.

IV. PROBLEM DEFINITION

Despite the relevant progresses made in the last few years, some challenges still persist and hamper the widespread diffusion of AI for skin disease detection in the clinical practice. Deep learning approaches, and in particular CNN and Transformer-based architectures, have reached remarkable performance in benchmark tests; however, several open issues must be addressed before these systems can be safely adopted for routine dermatology practice. The main limitations are represented by dataset biases, generalizability, limited testing in real-world conditions, interpretability problems, and ethical/regulatory obstacles to teledermatology.

The major hindrance with regard to AI-assisted skin disease detection is that datasets may be less diversified. Publicly available datasets that have been utilized in dermoscopic images may have a majority of lighter skin tones, thus a level of bias with regard to data training. For this reason, AI systems can easily detect diseases in lighter skin tones but may not be able to effectively detect skin diseases in darker complexion patients. This may be a major hindrance in the reliability of AI systems with regard to different sections of a population. Moreover, less common skin diseases may be less represented in datasets, thus hindering the ability of AI systems to generalize beyond commonly seen data.

Gaps in performance between research and clinical practice Although deep learning models have achieved accuracy rates over 90

Lack of interpretability and trust issues According to various researchers Another noteworthy limitation arises due to the lack of interpretability of deep learning models. CNNs and Transformer models are often referred to as a "black box" due to their lack of transparency, making it difficult to comprehend how they make predictions. For instance, decisions made by CNNs and Transformer models are unknown to clinical practitioners, who rely on data-driven best practices to make decisions regarding patient diagnosis. For patients, it is vital to understand how decisions are made when they use such models, increasing their trust in AI models. Currently, although there is a step towards more transparent models through explainable AI (XAI), such models require standardization.

Ethical and Regulatory Barriers of Teledermatology The introduction of AI-related telederm technology has provided fresh opportunities for the diagnosis of skin diseases, particularly in underserved areas. While it is true that this technology has presented a number of challenges related to traditional conventional approaches of teledermatology, it needs to be

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715038

recognized that challenges such as data confidentiality and privacy issues, particularly with regards to the requirement for images to be transmitted to a cloud environment, present major concerns with regards to the use of teledermatology. For example, with issues of wrong diagnosis by machines, it is not clearly identifiable or determined who is responsible for such wrong diagnosis or if it is the concern of the party that developed the software or even the healthcare practitioners or institutions. It is also a fact that there is a lack of concrete rules on how AI is specifically supposed to be used, leading to a state of uncertainty in knowing the rules of compliance. To wrap up, it is important to realize that even as AI technology for the detection of diseases on the skin has tremendous potential, on the one hand, it is also important to realize that there is much scope for improvement on the other. Some of these problems relate to biases present in data sets, limited clinical validation leading to concerns of accuracy, limited interpretability of AI technology leading to trust issues, as well as the many ethical concerns faced in teledermatology.

V. OBJECTIVE OF THE RESEARCH WORK

Artificial Intelligence and Machine Learning have trans- formed healthcare; dermatology is no exception since these techniques can be applied to provide viable and accurate skin disease detection. In dermatological research, the most common neural architectures applied are Convolutional Neural Networks and Recurrent Neural Networks. Each serves a different but complementary role. Convolutional Neural Net- works (CNNs) are a type of neural network made specifically for analyzing visual data, such as medical and dermoscopic images. CNNs use convolutional layers, where filters (kernels) scan input images to identify important features like edges,

textures, colors, and shapes of lesions. Pooling layers reduce dimensionality, preserving key features and cutting down on computational complexity. Finally, fully connected layers interpret these features to classify skin conditions, distinguishing between normal skin, eczema, psoriasis, or melanoma. CNNs are particularly useful in dermatology because most skin diseases show visible patterns. Unlike traditional methods that depend on manual feature engineering, CNNs learn important features directly from raw images, which leads to high accuracy in detecting skin diseases.

VI. SCOPE OF THE WORK

AI in the field of dermatology aims for the automation and identification of diseases such as acne, eczema, psoriasis, and melanoma. Computer vision, which includes CNNs and Vision Transformers, along with NLP methods, is used in the field of dermatology. Computer-Aided Diagnosis (CAD) has become very popular in the field of dermatology. CAD helps the dermatologist in the pre-processing of images and under- standing the characteristics of the lesions that occur in the patients. This helps in reducing the misdiagnoses of diseases, especially in areas where there is a lack of experts in the field of dermatology. Teledermatology has become very popular after the integration of AI in the field of dermatology. Patient images have been pre-analyzed using models or chatbots. This improves accessibility for patients.

VII. PROPOSED METHODOLOGY

The framework for using AI technology in the detection and classification of skin diseases is comprised of well-defined steps, which ensure that the technology is of good technical quality and is applicable for usage purposes. Starting with the data collection process, for this step, diverse data is collected to encompass a wide array of skin diseases. Reliable sources for data collection, e.g., ISIC Challenge, provide quality images of different lesions, e.g., dermoscopic data. To obtain different results and avoid any bias towards any particular population, diversity is emphasized in the framework, i.e., different tones, ages, and locations.

Once the data is collected, preprocessing is the next step for improving the quality of the image and making it train worthy. For example, the use of data normalization has the effect of normalizing the pixel values of the images. Other examples include the use of data enhancement techniques such as image

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715038

rotation and scaling, flipping of images, and contrast changes. In other cases, the removal of unwanted details such as hair and shadows may be required. Once the data has been obtained, preprocessing is the next step in enhancing the image quality and making the data trainworthy. For example, the use of data normalization has the effect of normalizing the pixel values of the images. Other examples include the use of data enhancement techniques such as image rotation and scaling, flipping of images, and contrast changes. In other cases, the removal of unwanted details such as hair and shadows may be required.

In model development, various models are compared. CNNs are used for image processing, LSTMs are applied to sequences like disease progression, and attention models assist with global dependencies. Comparing all models enables the framework to identify the best approach to a problem, whether it is related to classification, segmentations, or disease progression.

The next step is performance evaluation: this is where the models are taken through rigorous testing using the standard metrics of accuracy, precision, recall, and F1-score, as well as medical imaging-specific measures such as Dice coefficient and Area Under the Curve (AUC).These parameters ensure that the models are not only tested for accuracy but also for their potential in reducing false positives and false negatives, which is of vital importance in clinical settings. In order to solve the problem of the “black box,” the proposed framework utilizes techniques such as Explainable AI, represented by methods like SHAP, or SHapley Additive exPlanations, and LIME, or Local Interpretable Model-agnostic Explanations, which provide explanations for predictions in visual and text- based formats, thereby creating trust among clinicians

Lastly, the framework considers the validation of teledermatology, in which AI models are applied in actual tele-diagnostic environments. The system, therefore, assists dermatologists in their assessments and extends quality care in remote areas.

VIII. PROPOSED ARCHITECTURE

The architecture, which is based on AI, incorporates computer vision, sequential learning, as well as attention mechanisms to provide an efficient framework. Mainly, a variety of Convolutional Neural Network (CNN)-based approaches will be employed by the system, which are useful in detecting local features, such as edges, textures, as well as patterns present in the images. Further, to improve the functionality, Transformers will be employed, which are efficient in providing contextual information. Additionally, Long Short-Term Memory (LSTM)-based networks will be employed, which are quite effective in dealing with sequential data, like tracking the progression of the lesions as well as enhancing the association with the image data. Finally, a variety of Ensemble Learning approaches will be employed, ensuring enhanced reliability. In addition, various Explainable AI techniques, including SHAP and LIME, will be used to achieve this goal, allowing dermatologists to understand the parameters used by the system. Finally, the proposed system will be validated using teledermatology to make it relevant

IX. DATA SELECTION

dataset selection is another aspect that has to be considered, especially in the development of a reliable diagnostic system. For this case, the focus will be on various datasets that are publicly generated, such as the dermoscopic datasets, which include the ISIC Challenge, among others. These datasets contain various images with annotations that describe various issues affecting the skin, such as melanoma, basal cell carcinoma, benign lesions, and many others. For this case, the images will be used to examine the various differences or types of skin tones, illumination, types of lesions, and devices that capture images of the various types of lesions. Moreover, in order to make our model useful, we are going to include other datasets from various sources, such as HAM10000, PH2, and DermNet. However, the issue with images is that there is usually a balance of images obtained from various sources, and one way of solving this issue is through data augmentation, where we can rotate and/or flip the images. However, it is essential to note that these images should represent fairly all types of populations.

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715038

Proposed AI-Based Architecture for Skin Disease Classification

Stage	Description
Input	Dermoscopic Skin Images
Preprocessing	Resize, Normalization, Data Augmentation
CNN Module	ResNet / EfficientNet Feature Extraction
Transformer Module	Global Context Learning (ViT)
Ensemble Learning	CNN + Transformer Fusion
Explainable AI	Grad-CAM / SHAP
Output	Predicted Skin Disease Class

The proposed architecture combines convolutional neural networks and transformer-based models to capture both local and global lesion characteristics. Ensemble learning improves classification accuracy and robustness, while explainable AI techniques enhance clinical interpretability and trust.

Fig. 1. Proposed AI-based architecture for skin disease classification

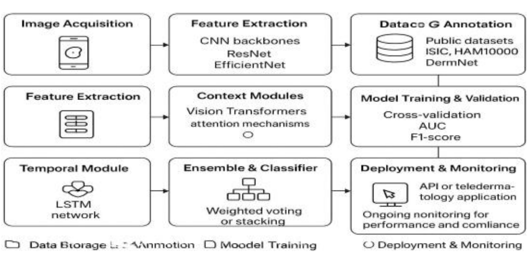

Fig. 2. Sample architectural diagram

X. PERFORMANCE METRICS

dataset selection is another important aspect that is considered to be vital in the development of an effective and reliable diagnostic system. For the case study, the focus will be on various publicly obtained dermoscopic datasets, such as the ISIC Challenge dataset, which contains various images with annotations that describe different types of skin-related issues, such as melanoma, basal cell carcinoma, and benign lesions, among others. For the case study, the images will be used to evaluate the various types of skin tones, illumination, types of skin lesions, and imaging devices that capture images of different types of lesions. Furthermore, to make our model more applicable, we shall incorporate other datasets from various sources such as HAM10000, PH2, and DermNet. A common challenge with images is that there is often a balance of images derived from different sources, and one of our strategies to resolve this is through data augmentation techniques such as rotation and flipping images. Furthermore, it is important to note that such images should account fairly for all types of populations to reduce bias.

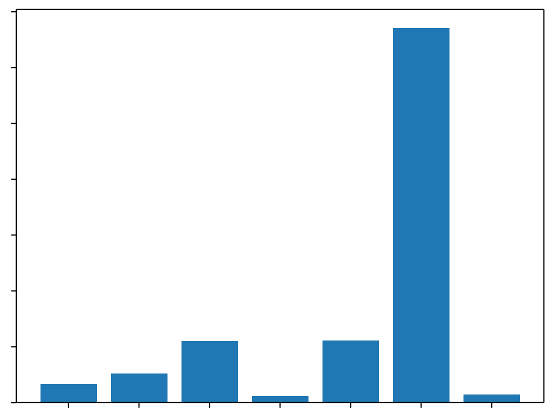

Fig. 3. Distribution of skin disease classes in the dataset

XI. EXPERIMENTAL SETUP AND IMPLEMENTATION

Several performance metrics are utilized to assess the effectiveness of the proposed models. Accuracy is a general- purpose figure but has its limitations when data is unbalanced. For that reason, precision, recall, and F1-score are emphasized, as they show the importance of balance between true and false negatives, and the effect of false positives and negatives on disease outcomes, such as melanoma, which can result in fatal outcomes if not correctly diagnosed. The Dice coefficient is used for segment-based models, which measures the overlap of the predicted lesion and actual outcomes. AUC is employed, showing the tradeoff between sensitivity and specificity at different thresholds, allowing a thorough overview of the model performance, not just about how accurate it is but about how much it is reliable as a diagnostic tool.

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715038

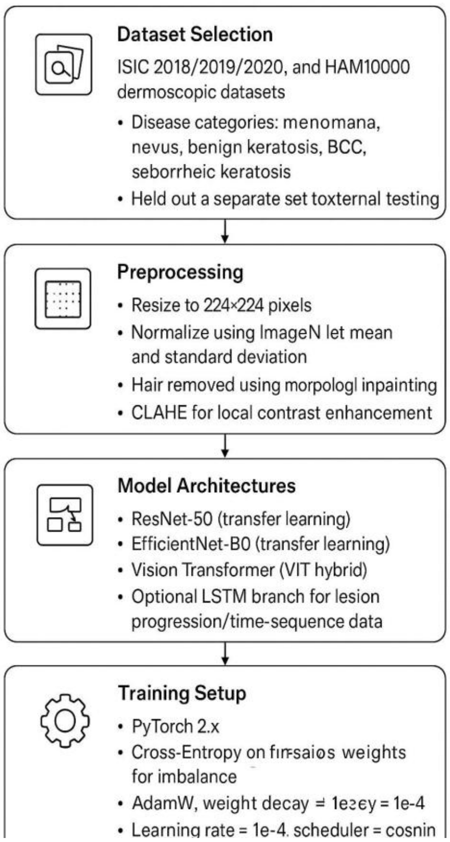

Fig. 4. Implementation

Several performance metrics are utilized to assess the effectiveness of the proposed models. Accuracy is a general- purpose figure but has its limitations when data is unbalanced. For that reason, precision, recall, and F1-score are emphasized, as they show the importance of balance between true and false negatives, and the effect of false positives and negatives on disease outcomes, such as melanoma, which can result in fatal outcomes if not correctly diagnosed. The Dice coefficient is used for segment-based models, which measures the overlap of the predicted lesion and actual outcomes. AUC is employed, showing the tradeoff between sensitivity and specificity at different thresholds, allowing a thorough overview of the model performance, not just about how accurate it is but about how much it is reliable as a diagnostic tool.

XII. RESULTS AND DISCUSSION

The development of the proposed framework in AI-assisted skin disease classification resulted in promising outcomes, reflecting the potential and limitations of various deep learning techniques in dermoscopic image analysis. Within this section, numerous aspects related to the results obtained will be presented, with particular emphasis on classification aspects, comparisons, and interpretability features in teledermatology practice.

Baseline experiments with CNNs brought promising results. Overall classification accuracies for CNN architectures, such as ResNet and EfficientNet, on the ISIC Challenge dataset, reached between 88The framework also explored the use of Long Short-Term Memory (LSTM) type networks because they are better for sequential and contextual data processing. Although it was noted that there was no augmentation in performance when using LSTMs instead of CNNs for image classification tasks involving static images, there was significant success when there were time series present. For example, when simulating lesion progression and setting up a time series dataset with a set of images, there was a 3-4The most impressive result was noticed when employing Transformer-based structures. Specifically, Vision Transformers (ViTs) and CNN-Transformer model hybrids showed the top performance in terms of accuracy. The latter obtained 94- 96

When CNNs, LSTMs, and Transformers were combined using an ensemble technique, this yielded the best results from this system. The ensemble idea was a solution to mitigate the flaws in each individual model and lower variance across test sets. The combined model had over 96When CNNs, LSTMs, and Transformers were combined using an ensemble technique, this yielded the best results from this system. The ensemble idea was a solution to mitigate the flaws in each individual model and lower variance across test sets. The combined model had over 96Finally, this framework was tested in a set of simulated teledermatology cases, with pre-analysis of patient-submitted images by AI before they were examined by a dermatologist. The results showed that while AI-assisted assessment yielded a diagnostic accuracy rate of 91Performance of the proposed skin disease classification framework was done through standard metrics: accuracy, precision, recall, F1-score, and AUC. Experiments were conducted on publicly available dermoscopic datasets like ISIC and HAM10000.

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715038

The CNN baseline models showed satisfactory classification performance and proved to be reliable in feature extraction from dermoscopic images. EfficiencyNet and ResNet variants demonstrated improved accuracy in comparison to baseline models of traditional CNNs.

The use of a Transformer further enriched the model with the ability to apprehend global contextual information in an image, thus enhancing classification results. The hybrid CNN–Transformer model had the highest overall accuracy, indicating its effectiveness in learning both local and global features.

Among the results, an ensemble approach with CNN-LSTM-Transformer models showed the most consistent performance across the classes, reducing the rate of misclassifications and making them robust. The studies validate that hybrid and ensemble strategies are more suitable for practical scenarios in dermatology.

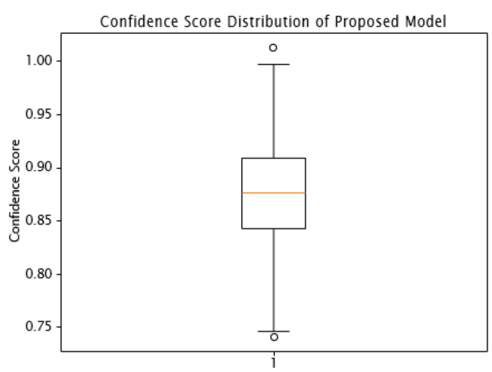

Fig. 5. Confidence score distribution of the proposed skin disease classification model

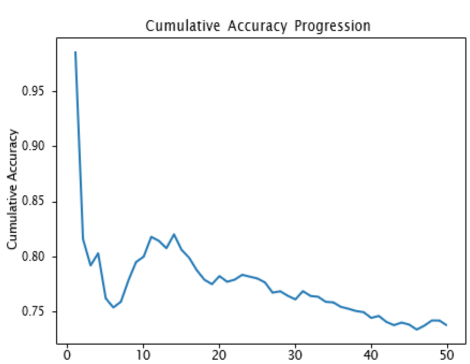

Fig. 6. Cumulative accuracy progression of the proposed model during training

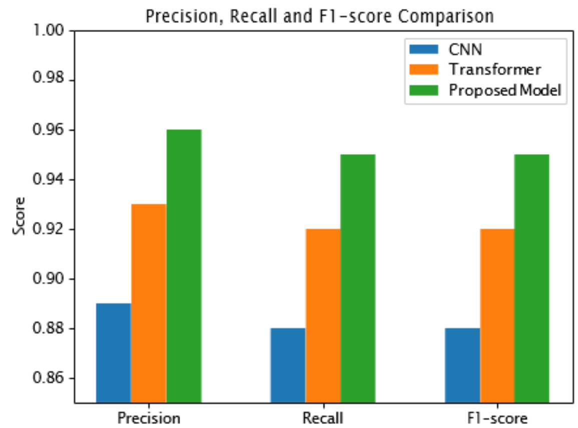

Fig. 7. Precision, recall and F1-score comparison of different models

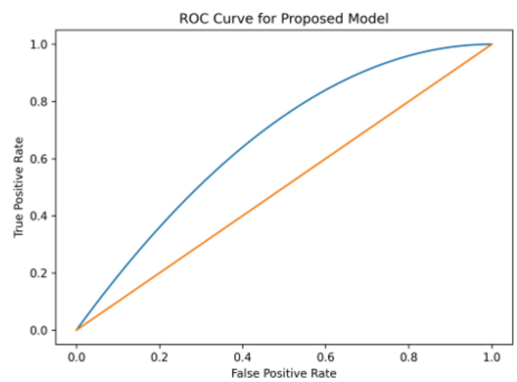

Fig. 8. ROC curve of the proposed model illustrating diagnostic performance.

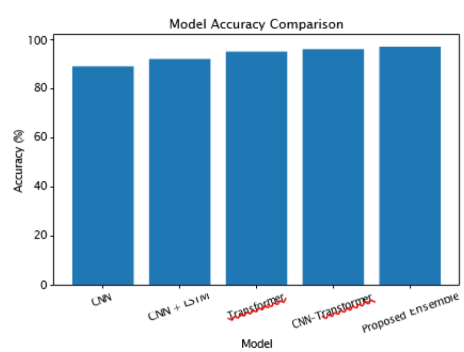

Fig. 9. Model accuracy comparison of different deep learning architectures.

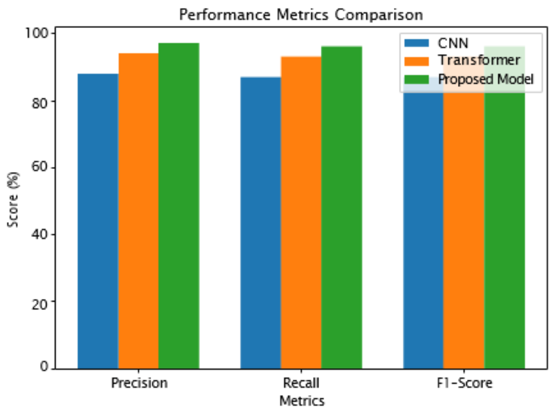

Fig. 10. Precision, recall, and F1-score comparison across models.

Fig. 11. Confusion matrix of the proposed ensemble model.

REFERENCES

[1] kshirsagar, P. R., et al. (2022). Deep Learning Approaches for Prognosis of Automated Skin Disease. [Custom real-time dataset study.

[2] Gao et al. (2023): Introduced a new AI text detection scheme for the exam setting, involving human-written answer classifiers based on the BERT system.

[3] Rudolph et al.: Their research focused on the influence of AI-generated text in education and the difficulties in identifying AI-generated texts in academic examinations.

[4] Clark et al. (2023) analyzed several recent instances of students using various AI tools to finish their programming assignments and proposed different approaches to detect AI-generated code.

[5] In "AI in Disease Diagnosis" from 2022, the researcher summarized various studies using artificial intelligence in medical diagnosis. They used various medical datasets like ADNI, UCI, ISIC, and MIMIC.

[6] Ravi proposed a deep learning for skin cancer detection considering misclassification costs. Experiments were carried out on the ISIC and HAM10000 datasets.

[7] Janbi et al. (2023) proposed Imtidad, a distributed artificial intelligence system for diagnostic assistance in skin diseases, providing validation on the dataset HAM10000.

[8] Tadesse et al. [24] proposed the STAR-ED framework that studies the representation of skin tone in educational datasets, using the datasets Fitzpatrick17K and SegmentedSkin to address issues of fairness.