Deep Learning for Skin Disease Detection: Hybrid CNN-Transformer Architectures with Explainable AI

Mr. Y.B. Nawale
ME-II Student,
Nashik, India

Prof. S.A. Gade,
Asst Prof SNDCOE,
Yeola, India

Abstract—**Skin diseases are among the most common global health conditions, and early detection is crucial for reducing complications and enabling timely medical intervention. In this paper, we present a novel, methodology-driven study for predicting skin diseases using deep learning with Convolutional Neural Networks (CNNs).**
Unlike traditional machine learning approaches that rely on manual feature extraction, our CNN-based methodology automatically learns complex spatial and textural features directly from images, making it exceptionally well-suited for dermatoscopic analysis. The abstract outlines our complete framework, including the robust dataset used for training, a detailed preprocessing pipeline to enhance image quality and consistency, a custom-designed CNN architecture optimized for this specific task, and the end-to-end workflow of the system.
We provide an in-depth discussion on how our CNN model processes raw image data, systematically extracts meaningful patterns, and effectively adapts to the nuanced classification of various skin lesions. We also introduce a data augmentation strategy to increase the diversity of our dataset and improve the model's generalization capabilities. Furthermore, we briefly discuss the selection of activation functions and optimizers that contribute to the model's performance.
This research focuses on detailing a comprehensive methodological framework, emphasizing the "how" and "why" behind each architectural choice. While this paper establishes the foundational system, we defer performance evaluations and comparative analyses to a future study.

Keywords — **Skin disease prediction, deep learning, convolutional neural networks, HAM10000, medical image analysis, data augmentation, CNN architecture, image preprocessing.**

I. INTRODUCTION

Skin disorders affect individuals of all age groups and are among the most prevalent medical conditions worldwide. According to the World Health Organization (WHO), nearly 900 million people globally are affected by some form of skin condition at any given time. Dermatologists typically rely on visual inspection and biopsy tests to identify abnormalities; however, this process is subjective, time-consuming, and prone to misdiagnosis. Early detection is critical, especially in malignant cases such as melanoma, which can spread aggressively if not treated in time. The burden of skin diseases is particularly high in developing countries where there is a shortage of trained dermatologists. The advent of computer-aided diagnostic systems, particularly those powered by Artificial Intelligence (AI), presents an opportunity to democratize healthcare by enabling accurate, accessible, and scalable diagnosis.

Over the past few years, deep learning has become a groundbreaking approach in the field of medical imaging. Convolutional Neural Networks (CNNs), a type of deep neural network, have demonstrated outstanding results in detecting complex patterns in images. Unlike traditional machine learning models that require manual feature extraction, CNNs automatically extract features such as color, shape, and texture. This ability makes them particularly well-suited for analyzing dermatoscopic images of skin lesions.
This research focuses on methodology—dataset preparation, CNN architecture, and workflow—while deferring quantitative results for future work. By concentrating on methodology, we aim to provide a clear roadmap for researchers intending to apply CNNs in dermatological diagnostics.

II. RELATED WORK

Numerous studies have explored the application of CNNs in medical image classification, particularly in dermatology. Esteva et al. (2017) demonstrated that CNNs could achieve dermatologist-level accuracy in skin cancer classification, marking a significant milestone in medical AI. Similarly, Tschandl et al. introduced the HAM10000 dataset, which quickly became a benchmark dataset for

training and testing skin disease classifiers. Other works have applied transfer learning with pre-trained architectures such as VGG16, ResNet50, and InceptionV3 to achieve higher accuracy while reducing training time. Transfer learning enables models to leverage patterns learned from large-scale image datasets like ImageNet and adapt them to medical image classification. For example, Harangi (2018) employed ensembles of pre-trained CNNs for skin lesion classification, reporting improved robustness.

Mobile-based diagnostic applications have also been proposed, where CNN-based models are integrated into smartphone applications to allow real-time skin disease screening. This is particularly important for low-resource settings. Despite these advances, a methodological gap exists in literature that thoroughly explains the design, training, and reasoning behind CNN architecture for dermatology. Our work addresses this by focusing on methodological aspects.

III. ETHODOLOGY

A. ataset

The HAM10000 dataset is utilized, containing 10,015 dermatoscopic images categorized into seven different types of skin lesions. This group encompasses melanocytic nevi, melanoma, benign keratosis-like growths, basal cell carcinoma, actinic keratoses, vascular abnormalities, and dermatofibroma. The dataset is diverse and balanced to an extent but still exhibits class imbalance, which we address during preprocessing.

B. ata Preprocessing

Preprocessing plays a vital role in ensuring that the data is consistent and suitable for CNN training. The preprocessing steps include:

- Resizing all images to 128×128 pixels.
- Normalizing pixel values between 0 and 1 for faster convergence.
- Data augmentation through rotations, flips, zooming, and brightness adjustments to handle class imbalance and improve generalization. -Splitting the dataset into training, validation, and test sets.
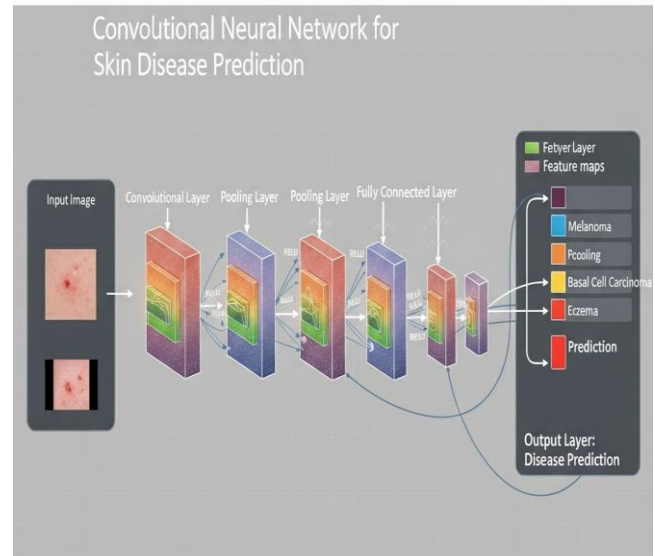
C. NN Architecture

Our CNN architecture is designed with multiple convolutional and pooling layers, followed by dense layers for classification:

1. nput Layer: Accepts $128 \times 128 \times 3$ dermatoscopic images.
2. Convolutional Layers: Extract features using filters that detect edges, textures, and lesion patterns.
3. Pooling Layers: Reduce dimensionality and retain dominant features.
4. Dropout Layers: Introduced to prevent overfitting.
5. Fully Connected Layers: Combine features for high-level reasoning.

6. Softmax Output Layer: Provides class probabilities for the seven categories.
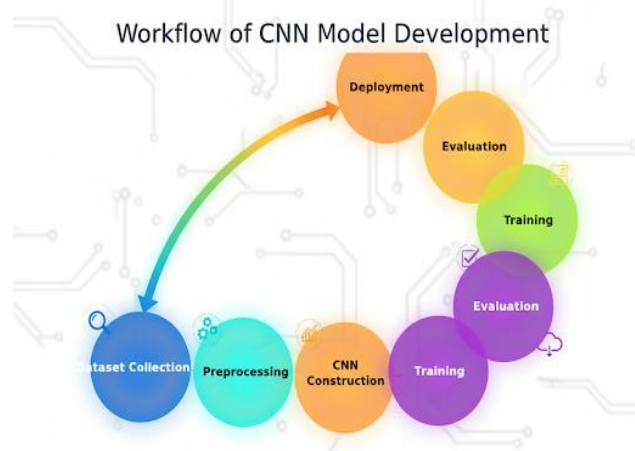
This architecture is inspired by established CNN models such as VGGNet but simplified for efficient training on medical data.

D. Model Workflow

The workflow of the model involves several systematic steps:

1. Dataset Collection → Acquire HAM10000 dataset.
2. Data Preprocessing → Resize, normalize, and augment images.
3. CNN Construction → Define layers and architecture.
4. Training → Optimize weights using backpropagation and gradient descent.
5. Evaluation → Reserved for Version 2 paper.
6. Deployment → Integrate into diagnostic applications.

IV. RESEARCH AND DISCUSSIONS

Convolutional Neural Networks (CNNs) have a significant impact on advancing medical image classification. Their hierarchical learning ability allows the extraction of features at different abstraction levels. Initial layers detect low-level features such as edges and color gradients, while deeper layers capture complex lesion characteristics like asymmetry, irregular borders, and color distribution. One of the primary reasons for choosing CNNs over traditional machine learning methods is their capability to automatically learn features without manual intervention. Classical models such as Support Vector Machines (SVM) or Random Forests require handcrafted features, which are often insufficient to capture the complex variability of skin lesions. CNNs, by contrast, directly process image pixels and adaptively learn distinguishing features. Challenges exist in applying CNNs for medical imaging. Class imbalance remains a persistent issue since malignant cases are rarer compared to benign lesions. Overfitting is another challenge, especially when the dataset size is limited. Methods such as data augmentation, dropout, and transfer learning are frequently applied to overcome these challenges.

Ethical considerations are also important in this domain. CNN-based diagnostic systems should be designed to assist dermatologists, not replace them. Transparency, explainability, and fairness are crucial to ensure clinical acceptance of AI-based tools.

V. CONCLUSION AND FUTURE WORK

This paper presented a detailed methodology for skin disease prediction using CNNs. We described the dataset, preprocessing steps, CNN architecture, and workflow in depth. Unlike prior work that primarily highlights results, this study focuses on methodology to provide a roadmap for future researchers.

In the next phase of this work, we will conduct extensive experiments using CNNs, evaluate the model against state-of-the-art methods, and report quantitative results. We also plan to explore transfer learning, federated learning, and explainable AI to enhance interpretability and applicability in clinical settings.

Our ultimate goal is to integrate CNN-based models into healthcare systems for real-time skin disease diagnosis, bridging the gap between AI research and medical practice.

REFERENCES

A. Esteva et al., 'Dermatologist-level classification of skin cancer with deep neural networks,' Nature, vol. 542, no. 7639, pp. 115–118, 2017.

P. Tschandl, C. Rosendahl, and H. Kittler, 'The HAM10000 dataset, a large collection of multi-sources dermatoscopic images of common pigmented skin lesions,' Scientific Data, vol. 5, p. 180161, 2018.

Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning,' Nature, vol. 521, pp. 436–444, 2015.

K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition,' arXiv:1409.1556, 2015.

G. Hinton, N. Srivastava, and K. Swersky, 'Neural networks for machine learning,' Coursera Lecture Notes, 2014.

M. Harangi, 'Skin lesion classification with ensembles of deep convolutional neural networks,' Journal of Biomedical Informatics, vol. 86, pp. 25–32, 2018.

O. Russakovsky et al., 'ImageNet Large Scale Visual Recognition Challenge,' International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.

S. Rajkomar et al., 'Machine learning in medicine,' New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

J. Ker, L. Wang, J. Rao, and T. Lim, 'Deep learning applications in medical image analysis,' IEEE Access, vol. 6, pp. 9375–9389, 2018.

A. Krizhevsky, I. Sutskever, and G. Hinton, 'ImageNet classification with deep convolutional neural networks,' Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.