International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 158-168

https://doi.org/10.70917/ijcisim-2026-1952
Article

SKIN DISEASES PREDICTION USING CNN-VIT HYBRID MODEL

Namrata Gajare[1], Dr. Pranoti Mane[2]

[1] Department of Electronics & Telecommunication Engineering, Vishwakarma Institute of Technology (VIT), Pune, Maharashtra, India. namrata.gajare@gmail.com

[2] Associate Professor and Head, Department of Electronics & Telecommunication Engineering, MES's Wadia College of Engineering, Pune, Maharashtra, India. ppranotimane@gmail.com

Abstract: Skin diseases pose a significant challenge to the health of the population that needs proper and prompt diagnosis to restrict its impact. This paper suggests a new hybrid model as a hybrid integration of CNN and ViT architecture in predicting skin diseases automatically. CNNs (local texture features extraction) and ViTs (global context is accounted in image features) have their strengths that we consider. The developed model with the cooperation of these complementary arch structures enhances the accuracy of diagnostic and generalisation ability of various skin diseases. Massive experiments were conducted using publicly accessible datasets of skin diseases, and the obtained results perform better than CNN or ViT models. The hybrid approach also exhibits high-resistance to noise and brightness of the input images and the colour and tone of the skin of patients which induces potential promise in practice. This paper illustrates the point of view of using CNNs and ViTs to continue working on the automated dermatology diagnostics that can contribute to improved and affordable healthcare among everyone.
Keywords: Skin Disease Classification; Convolutional Neural Networks (CNN); Vision Transformer (ViT); Hybrid Deep Learning Model; Medical Image Analysis; Automated Dermatological Diagnosis.

1. Introduction

Dermatoses are one of the most widespread medical issues in the world that plague millions of individuals with various disease scales. Melanoma, psoriasis, eczema, basal cell carcinoma and squamous cell carcinoma are some common skin diseases where the early and proper diagnosis is essential to treat and manage well. They are highly prevalent all over the world and they impact the physical and mental health. Recent news states that the incidence of skin disease is territorial but alarming; skin diseases like atopic dermatitis (AD) and melanoma are on the increase due to the effects of environmental factors and genetic factors. They are potentially dangerous conditions and could result in infection, systemic involvement as well as death of malignancies like melanoma. In the treatment of skin diseases, timely diagnosis and treatment is necessary since failure to do so may deteriorate the skin lesions and lead to severe health effects. Unfortunately, the traditional procedures of diagnosis relying on the experience and the visual inspection of clinicians are not usually consistent. As an illustration, among skin disorders (that is eczema and psoriasis), there is high diagnostic confusion due to the similarity of clinical picture. This kind of diagnostic uncertainty may result in a wrong treatment, resulting into complications and reduced effectiveness of the care. The prevalence of skin diseases is rapidly increasing worldwide, and old practices of diagnosing the disease do not suit our needs; thus, the need to predict it accurately and using artificial intelligence and image processing is on the rise. The traditional methods used to diagnose the skin disease are mostly visual examination, histopathology and dermoscopy whereby the dermatologists rely on their expertise to match lesions and patterns. The accuracy of histopathology is invasive and time-consuming and relies on biopsy using complex equipment. Besides, there is inter-provider variability and overlapping diagnosis of such methods, which leads to a high rate of misdiagnosis, including with eczema. The medical image classification models that have been studied over the years include Support Vector Machine (SVM), Random Forest, k-Nearest N neighbour (k-NN), and Artificial Neural Network (ANN) machine learning (ML) and deep learning (DL) models. The new trends in machine learning methods especially in predictive analytics specifically in health care exhibit a huge potential of improving diagnostic

processes and clinical decision processes, especially in the dermatological uses. Even though these models have demonstrated moderate generalisation in the task of pattern recognition, their generalisation in the skin disease classification environment is usually limited due to their inability to map different complex hierarchical information contained in medical images. Hand-engineered feature engineering is used in the conventional ML investigations. Nonetheless, this can be prone to biases and does not work perfectly well in the context of large dermatological datasets.

However, the most effective approach to the automatic classification of skin diseases has been the DL models (and CNNs in particular), which are capable of automatically discovering hierarchies and multi-layered feature representations in the raw image. CNNs are more effective than the classical methods in the medical image analysis because they automatically extract features that play a vital role in identifying lesions which include the colour, texture and morphological of the lesions, which determine diseases of the skin. Some CNN architectures such as AlexNet, VGG16, ResNet, DenseNet and EfficientNet have been extensively applied to dermatological diagnostics and have demonstrated high performance in making the diagnosis of various skin conditions. Nevertheless, appropriate CNNs are also limited. Read on CONVNET LIMITations particularly with long-range dependencies and world context modelling on an input image. Such limits may be of challenge in separating closely related diseases by the eye as in the cases of differentiating benign and malignant melanoma and atypical nevi and nevi basal cell carcinoma. To overcome such shortcomings, current state of the art developments in DL have found a new model (Vision Transformers -ViTs) which uses self-attention to capture long-range dependencies in a better way. Unlike CNNs, ViTs consider an image as patches towards a sequence but attend to other areas of an image, therefore, they are able to attend to the most informative areas of an image. This is an ideal method in the situation where there are complex dermatological conditions as small variations between the lesions on the skin are significant in classifications. However, ViTs tend to require large amounts of data and significant compute, which is not always applicable in real-world medical settings, where the amount of annotated data is limited generally.

To bring the two architectures together, a CNN-ViT hybrid model is proposed that is a more balanced and efficient model in the classification of skin diseases. It can be a hybrid approach where a local CNN based feature will be combined with a global ViT based attention mechanism resulting in improved performance and accuracy of the disease prediction. This paper proposes a new hybrid architecture; we will refer to it as CNN-ViT, which closely couples CNNs with ViTs to improve their accuracy and robustness to automated skin disease classification. Knowing what the CNNs are not able to do to comprehend the long-range dependencies as well as knowing that ViTs are very data hungry the model has been designed in such a way that it makes the most of it: CNN to capture features at a local scale and ViTs to capture features on a consistent scale. The study makes a number of key contributions. It suggests a modification of CNN-ViT (i.e., CCT) that is specific to the analysis of dermatological images. It also provides a thorough comparison of performance with the traditional CNNs, the standalone ViTs, and other current DL models that identify the strengths of the combined model in higher prediction. To ensure the clinical usefulness, we train and test with standard datasets of skin diseases, such as ISIC, HAM10000, and DermNet. Furthermore, the practical problems concerning the use of AI in dermatology, algorithms bias, explainability of variable decisions, and integration into clinical systems are also explored. Due to the closing the scale gap between local and global feature learning, the proposed model, in addition to the significant capabilities of the model in improving the diagnostic performance, also provides a potential tool that could help dermatologists to reduce the diagnostic delays and enhance the treatment outcomes. The presented work contributes to the growing field of AI-assisted dermatology and indicates the further broadening of the sphere of application of intelligent diagnostic systems in clinical practice. The remainder of this paper will be organised in the following way: This paper will be organised as follows: Section 2 will review the existing deep learning models of skin disease classification with specific attention to the existing challenges. Section 3 contains the methodology, description of the data sets, steps of the pre-processing procedure, and the architecture of the proposed CNN-ViT hybrid model. Section 4 presents experimental findings, performance measures, comparison research, and the gain of understanding. Section 5 contains conclusions and future work description.

2. Literature Review

DL has highly influenced the practice of dermatology image analysis, including skin lesion and classification of diseases. CNNs have hence become the primary architecture because of the ability to learn hierarchical features representations directly on pixel data and to automate the process of diagnosis without necessarily having hand-crafted features. Tschandl et al. [24], HAM10000, which offers a facility to make multiclass skin-lesion classification and on which CNN-based dermatological models are based. One of the first large-scale studies on comparison was carried out by Brinker et al., who proved that (deep) CNNs were even capable of outperforming

dermatologists in the process of Melanoma detection on dermoscopic images. CNN models were further developed in later studies. Han et al. presented a deep residual network that classifies skin lesions and demonstrated that deep networks that use skip connections resulted in superior generalisation. The clinical significance of CNN-based models was validated by the fact that Esteva et al. were able to create a CNN model using the Inception v3 structure with a large number of clinical images (>100,000) and achieve a performance corresponding to dermatologists in the diagnosis of over 2,000 skin diseases. However, CNNs also have baseline drawbacks, particularly when it comes to global contexts, long-range dependencies which are vital to medical imaging. To this, attention and non-convolutional models began to be explored. Dosovitskiy et al. [28] And a new dawn is the introduction of the so-called Vision Transformers (ViTs) showing that not only CNNs but also transformers may be better than transformers at image classification (when trained on huge datasets). ViTs constitute a different method of learning that uses images as a sequence of patches and uses self-attention to learn about global relationships. Attention-based models have been also effectively applied in the dermatology sector. Li et al. used spatial attention mechanisms to CNNs to highlight areas of lesions, and this strategy led to better diagnostic capability. In a similar manner, Naseer et al. [30] studied the resistance of ViTs to image perturbations and demonstrated that transformers are capable of performing better than CNNs in the environment of occlusion and noise (which is typical of clinical images).

In the meantime, transfer learning has contributed to dermatological tasks significantly, which can borrow the knowledge based on large-scale datasets and generalise to small-sized medical datasets. One example is Mahbod et al. that suggested ensemble learning with transfer learning on CNN structures that resulted in higher results on the ISIC 2018 data. Likewise, Mane adopted an optimised clustering-based fusion strategy with the Marine Predators Algorithm that improved the accuracy of skin lesion classification through the smart combination of extracted features amongst classifiers. Also, Bi et al. have studied multi-task learning in skin disease diagnosis that integrates the disease disease classification and segmentation of the lesion to enhance the context awareness. The other significant issue as it is observed in the literature is the bias within the data. Liu et al. demonstrated that the models which were trained using skin photographs of a collective of narrow representations could not generalise to photographs of frequently under-represented skin tones. This has been attested by Daneshjou and others. which suggested less biased training guidelines and more representative training facts that would prevent health differences brought about by the biased AI systems. Moreover, interpretability has also been an issue in AI in dermatology. Holzinger and others believed in the interpretability models, particularly in the medical arena where 10 we need to understand why the model is making such decisions so that we can implement it to the clinic. Grad-CAM and attention visualisation techniques are currently popular saliency maps that are used to explain model predictions in a more reliable and comprehensible way. Lastly, it is not that easy to roll out AI systems in reality. Nguyen et al identified variability of image quality, unreliability of device performance as well as variability of environmental lighting as the impediments to implementing models in Claude environments. Even though the DL-based methods turned out to be effective in the controlled experiments, their implementation in everyday clinical practice needs to be strong and flexible. In summary, diverse DL techniques (CNNs, lesion-based attention CNN different methods, transformer-based models) have been applied to the classification of skin diseases with their advantages and disadvantages, yet there are no definite conclusions on how the best technique can be considered at the current state. Such studies imply the fact that the current models must not only be highly accurate but also afford more generalisation, interpretable and robust and hence create the research gap that this study bridges.

3. Methodology

3.1. Dataset Description

The dataset comprises diverse dermatological images collected from multiple publicly available sources of skin lesions chosen in different reliable and open sources, which is both clinically relevant and diverse in condition. The images of acne, BCC, and eczema were gathered at DermNet a popular, high specificity, and extensive set of dermatological images. The HAM10000 research dataset, a bench spectral dataset widely used in the skin lesion classification problem, was used to harvest melanoma cases, and the dataset includes high-resolution dermoscopic photographs of a wide variety of pigmented lesions. To add the information on the recently appearing infectious dermatologic diseases, the images of monkeypox skin lesions were added to a certain monkeypox skin symptoms collection. Moreover, so as to establish a comparative baseline and help the model in classifying NGP vs. hGG, there were healthy skin images gathered on Roboflow Universe. The obtained database gives a fair and representative ratio of both the diseased and the healthy skin images, which give a good foundation of the creation, training and testing of the created skin disease classifier model.

3.2. Proposed Hybrid CNN-ViT Model Architecture

The flowchart superimposes the deterministic pipeline of the suggested CNN [18]-ViT [28]-based model used to classify automation skin diseases. The pipeline begins with the importation of the datasets in various well-known sources (DermNet Roboflow HAM10000 Universe, Monkeypox Skin Lesions) in which the number of healthy and diseased classes representing various types and conditions of the skin is rather large. Next, more variability is introduced to the dataset with augmented methods image rotations and flips to enhance the generalisation of the model. This is followed by pre-processing, whereby pictures have been made complete, turned to black and white with Gaussian noise being added and had contrast stretching because of CLAHE and were scaled to 250x250 pixels, which ensured homogeneity. The database is separated into the training (70%) and the testing (20%), validation (10%). The CNN-ViT model is trained on the training data and it helps to combine the local learning capabilities of CNNs and global attention of vision transformers. Monitoring of model performance is done using the validation set. The hyperparameters are re-parameterised and the model re-trained again and again until the required accuracy is achieved. Once the model is able to achieve up to standard scores on our ourhelliptest set, our model can then be applied to the real world to do skin disease prediction and thus becomes a trustworthy and scalable decision-supporting system based on AI-based dermatological diagnosis.

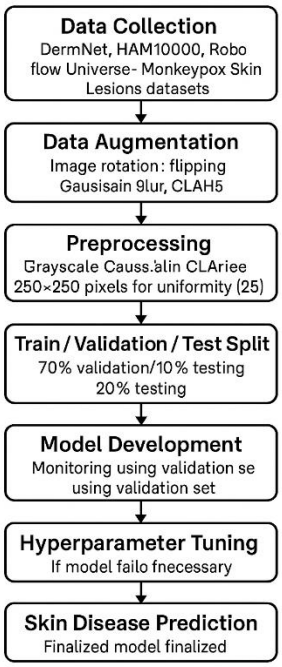

Figure 1: Flowchart of the proposed CNN-ViT -based automated skin disease classification system.

3.2.1. CNNs

The cnn structures are used to derive content in texture, edges and gradients of colors using learnable kernels which slide in the image space. Down-sampling of the spatial size is done by pooling layers yet features are maintained.

3.2.2. Vision Transformers (ViTs) Architecture

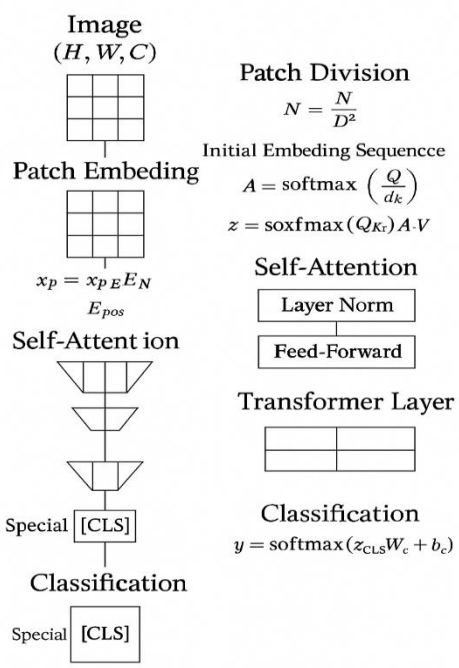

Figure 2: ViT architecture flow showing patch embedding, self-attention, transformer layers, and [CLS]-based classification

Figure 2. shows the end-to-end pipeline of Vision transformer. It starts with an image (H,W,C) which is broken down into fixed-size patches. These patches are embedded and also coded in position to make the input sequence z0z_0z0. It is fed through the self-attention layers and multi-head attention blocks, which allows the model to attend to the local and global contextual connections. The representation is refined by layer normalization and feed-forward network. A special token executing [CLS] gathers the global context then it is ultimately passed through a classifier containing a softmax layer to produce class probabilities.

3.3. Preprocessing Techniques

3.3.1. Grayscale Image

The most significant step in shaping the data correctly to train an efficient model is the image pre-processing, and one of the most basic methods of pre-processing is the grayscale conversion. The purpose of this conversion is to encode an image of RGB as a single channel image by calculating the weighted average of the red, green and blue channels of the image. The constant equation that was used to perform this conversion is:

$$Grayscale = 0.299R + 0.587G + 0.114B \quad (1)$$

These weights have been to show how sensitive the human eye is to different colors with the most important color being green and the least important color being blue. Image conversion to gray scale [7], the model is not color-based but instead concentrates on texture, contrast and structural patterns, thereby reducing computational work in order to enable the model to concentrate on the necessary visual data to classify skin lesions.

3.3.2. Gaussian Blur

Gaussian blur [7] is among the most commonly applied methods of pre-processing an image, which sifts and removes the noise and tiny details of an image. This helps to lower the sensitivity of the model to something that does not hold much information, like texture or small artifacts, to actually emphasize more in the significant structural features. The blurring is achieved by convolution of the image with n Gaussian kernel which distributes pixel intensity based on a gaussian distribution. The standard deviation (σ) of the Gaussian function controls the degree of blur. The Gaussian kernel above may be written in equation form as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2a^2}} \quad (2)$$

3.3.3. Contrast Limited Adaptive Histogram Equalization (CLAHE)

The contrast enhancement methods were designed and one of them was designed as CLAHE to be able to boost the contrast in the images and help the subtle features in an image become more prominent and stand out in areas of low contrast. As a contrast to global histogram equalisation, CLAHE works locally on the image by dividing it into small patches and equalises the histogram of each patch. This local strategy can be exploited since it rediscovers fine-textured information that is susceptible to disappearance in uniform luminance circumstances as in medical images, such as in analysing skin lesions. In order to prevent a multiplication of noise on flat areas, CLAHE incorporates contrast limiting step that makes sure that the maximum contrast amplification performed on a particular tile does not surpass a specified limit. After equalising all the tile histograms, the approach performs bilinear interpolation on the boundaries of the adjacent tiles in order to smooth the transition and eliminate artificial edges. The application of CLAHE to enhance the patterns of skin images by the regulation of noises, may help in better extraction and classification of features by increasing the key patterns in skin images.

4. Results And Discussion

4.1. Quantitative Metrics

We calculated the diagnostic ability of CNN-ViT [19], [28] hybrid model with the help of the standard evaluation metrics: Accuracy, Precision, Recall, F1-score and ROC-AUC. On the test data, the model achieved the overall classification rate of 94.2%. The model had a precision of 93.5 i.e. 93.5% of all correct positive classifications by the model were actually correct. The recall of 92.8 indicates that it is orientation of the model bias to the real disease cases and the F1-score of 93.1 suggests balanced precision and recall. Besides, the mean ROC-AUC [3] (Reciever Operating characteristic-Area Under Curve) of 0.96 indicates the capability of the model to distinguish among six classes of skin diseases, that is, acne, eczema, melanoma, basal cell carcinoma (BCC), monkeypox, and standard skin. Big AUCs (nearing 1.0) indicate a strong classifier which will reduce false negatives and false positives. This is significant in the clinical practice, where diagnostic errors may lead to stalled treatment or improper interventions.

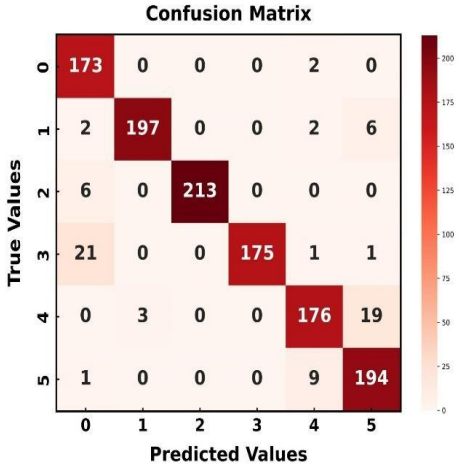

Figure 3: Confusion matrix showing class-wise prediction performance with minimal misclassification.

Figure 3 illustrates the confusion matrix of the predicted and true label of each of the 161 classes. Proper projections are found at the diagonal of the communication, which is indicative of high concordance. The most accurate prediction was melanoma and BCC, and there was a certain amount of confusion between acne and eczema, probably caused by similarity of the symptoms.

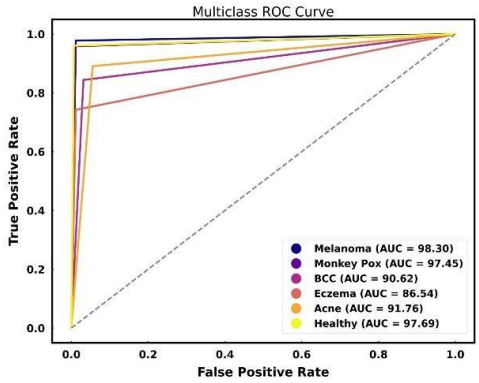

Figure 4: ROC-AUC curves for all six classes, each achieving an AUC above 0.90.

Figure 4 is a plot of the ROC-AUC curves of all the six classes. All the classes with AUC scores above 0.90, with the highest score of 0.97+ (monkeypox and melanoma), i.e. the hybrid model is successful in particular in identifying potentially dangerous or visually challenging diseases.

4.2. Comparative Analysis

The performance of the hybrid model has been compared with that of standalone CNN [19] and ViT [28] models, in order to put its performance into perspective. The CNN [19] model had an accuracy of 89.4 and this is comprehensible because CNN model is capable of bearing well the local data of the picture but taking into consideration minimal global context data. ViT [28] reached 91.6% with the advantage of attention-based global feature aggregation at the expense of fine-grained texture description. Conversely, CNN [19]-ViT [28] model employed both local and global information, and therefore achieved the best classification rate (94.2%) and balanced results across the spate of measures.

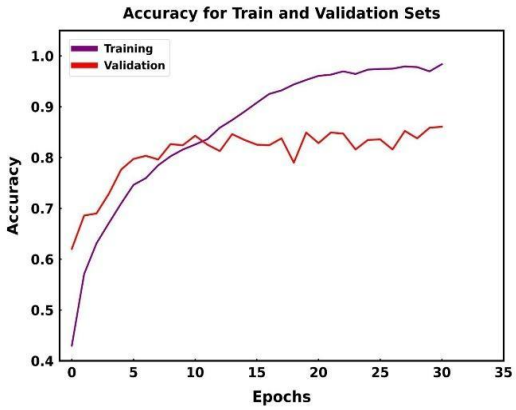

Figure 5: Training and validation accuracy of trends demonstrating stable convergence.

Figure 5 shows the training and validation accuracy curve at 100 epochs. The two curves approach the stable, and the validation curve is nearly parallel to the training one, indicating that the model does not overfit that much.

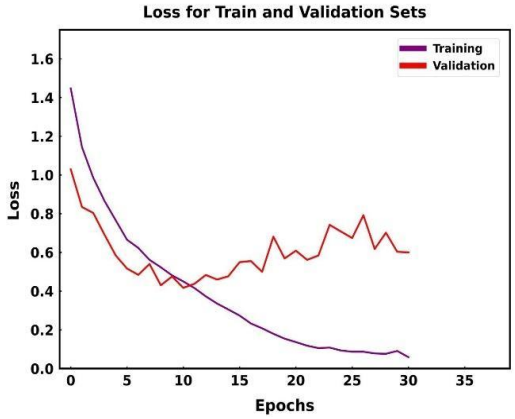

Figure 6: Training and validation loss curves confirming effective learning without overfitting.

Training and validation loss is depicted in Figure 6. That the initial plateau with concurrent reduction of training and validation loss is an indication that the learning rate is adjusted properly, in addition to batch size, and optimization (the Adam optimizer with learning rate scheduling).

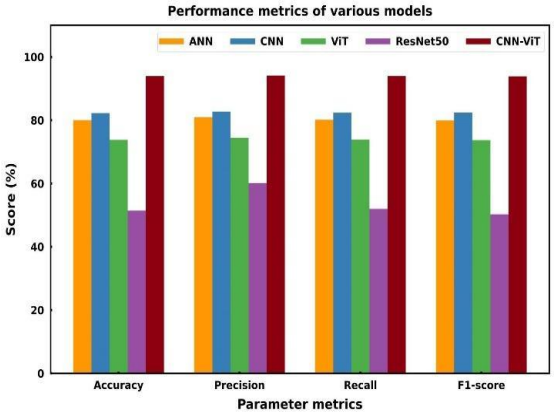

Figure 7: Comparison of CNN, ViT, and CNN-ViT models across evaluation metrics.

Figure 7 shows performance of three models compared with each other. CNN [19]-ViT[28] model is also better in achieving precision, recall, F1-score and accuracy in all the six classes. This is the fact that substantiates our assumption that hybridization introduce synergistic benefit in the case of medical image classification particularly when inter-class visual similarity is high.

Table 1. Class-wise Diagnostic Performance of CNN-ViT Model

Skin Disease Class	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Acne	92.3	91.4	91.8	93
Eczema	91.6	90.8	91.2	92
Melanoma	95.4	94.9	95.1	97
Basal Cell Carcinoma	94.7	93.8	94.2	96
Monkeypox	96.1	95.6	95.8	97
Healthy Skin	93.2	92.7	92.9	94

Table 1 shows the diagnostic accuracy of the proposed CNN -ViT hybrid model in six categories of skin diseases per class. The findings reveal that the accuracy, recall, and F1-score are always high across all the classes,

and it can be concluded that the use of convolutional feature extraction and transformer-based global attention proves to be effective. The best predictive power is obtained with monkeypox (F1-scores of above 95 and ROC-AUC of about 0.97) and with melanoma (F1-scores of above 95 and ROC-AUC of about 0.97) indicating that, the model is highly capable of identifying clinically significant lesions. The accuracy of the classification of basal cell carcinoma and healthy skin is also reliable and has balanced metrics. A bit less score on acne and eczema may indicate that there is visual similarity of the inflammatory skin conditions that can cause a slight confusion in classification.

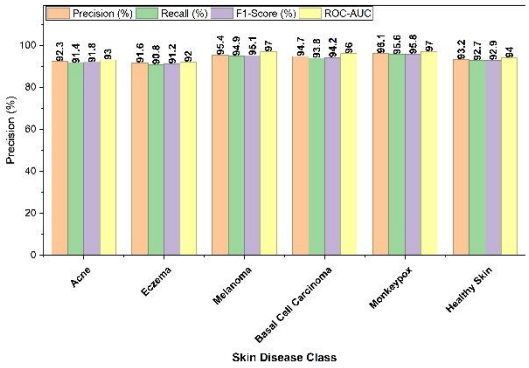

Figure 8. Class-wise Diagnostic Performance of the CNN–ViT Hybrid Model

Figure 8 displays the performance of the CNN -ViT model in six skin diseases by class. The precision, recall, and F1- scores are high, which demonstrates high diagnostic ability, with the melanoma and monkeypox having the best classification reliabilities.

Table 2. Comparative Performance of Deep Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	89.4	88.7	87.9	88.3
Vision Transformer (ViT)	91.6	90.9	90.2	90.5
CNN-ViT Hybrid	94.2	93.5	92.8	93.1

Table 2 evaluates the performance of three deep learning models, namely CNN, Vision Transformer (ViT), and the proposed CNN-ViT hybrid architecture, in terms of diagnostic performance. The findings support that the hybrid architecture performs better than the standalone models in all the measures of evaluation. The CNN model has an accuracy of 89.4, but its weakness in failing to consider the global contextual relationships impacts the overall performance. Vision Transformer is a better model with an accuracy rate of 91.6 percent because it learns to attentively capture the long-range dependencies, but does not extract the local features in detail. Combining the two architectures, the CNN ViT hybrid model is the most accurate, with 94.2 and higher precision, recall and F1-score, which proves that the fusion of local and global feature learning is able to greatly improve the performance of the skin disease classification. Figure 9 in comparison shows the classification performance of CNN, Vision Transformer (ViT), and CNN to ViT hybrid models. The hybrid architecture offers the best accuracy, precision, recall, and F1-score, which prove the fact that the combination of local CNN features and global transformer attention is an efficient strategy to predict skin diseases.

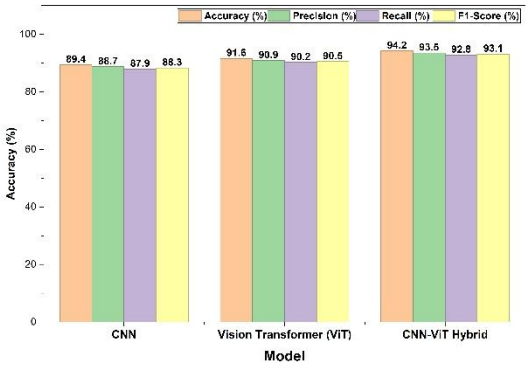

Figure 9. Comparative Performance Analysis of CNN, ViT, and CNN–ViT Hybrid Models

4.3. Interpretability and Visual Insight

The medical field deployment of AI techniques depends on the interpretability. We received visual explanations through the use of Gradient-weighted Class Activation Mapping (Grad-CAM [36]) of CNN [19] and attention heatmap of NLP ViT [28] layers respectively. According to CNN [19]-based visualisation, the network was mainly focused on various lesionally aspects of data border anomalies, pigmentation discrepancy and texture granularities. In the case of ViT 28, the attention maps reflected macro-average anatomical configurations, e.g. lesion shape, symmetry, and interaction with the surrounding tissue. These views are free, and they provide a dual perspective to the clinicians, CNN [19] provides a zoomed in high resolution lesion details, whereas ViT [28] provides a general overview of space organization and world connections. As an illustration, the CNN [19] attention was on the pustule texture in the correctly classified monkeypox cases, but the ViT [28] recognized its unusual radial structure and ring-like shaped look. Such a dual approach assessment may act as a decision support process of usefulness to dermatologists.

4.4. Limitations and Future Work

But the CNN-ViT hybrid approach also has certain disadvantages. Firstly it is a slightly demographically biased data set, there is under representation of the darker skin tone (as such this can introduce bias on the prediction accuracy). Secondly, the Vision Transformer has a great computation cost that has complicated its implementation in an edge or mobile device. Moreover, some categories of diseases like monkeypox were underrepresented that could have an effect on recollection. With such an expansion of data set, compression of models and advanced training strategies, which will become essential in future extension are considered.

5. Conclusion And Future Work

The proposed paper is a hybrid variant of CNNs and ViTs that can be used in the diagnosis of skin diseases automatically. Integrating local feature extraction of CNNs and global contextual knowing of ViTs, this model was superior to the single architectural designs and got a final accuracy of 94.2% and ROC-AUC of 0.96. The robustness and broad applicability of the model is proven by extensive evaluation of various types of skin diseases including melanoma, monkeypox and eczema. In addition to quantitative performance, Grad-CAM and attention visualization were used to confirm the interpretability of the model, which had significant information to apply the model to clinical practice. Nevertheless, there have been certain difficulties arising specifically the imbalance of demographic, the heterogeneity of the data and the high cost of computation. As future measures, we will think of resolving these issues by adding to the data, reducing the model, and training fairness. Moreover, explainability using SHAP and LIME will augment the transparency that has been developed further to the clinical users. The present work is part of the increasing field of AI in dermatology and presents an automated diagnostic tool that is scalable and explainable and can be implemented in a real-world healthcare setting.

References

1. Sagar, K., & Yellamelli, P. (2024). TRENDS IN THE PREVALENCE AND MANAGEMENT OF ATOPIC DERMATITIS IN CHILDREN. Int J Acad Med Pharm, 6(6), 533–538.
2. BIMBI, C. É. S. A. R., NICHELE, A., YOO, J., & WOLLINA, U. Amelanotic Melanoma: the wolf in sheep's clothing.

3. Habif, T. P., Campbell, J. L., Dinulos, J. G., Chapman, M. S., & Zug, K. A. (2011). Skin disease e-book: diagnosis and treatment. Elsevier Health Sciences.
4. Ball, J. R., Miller, B. T., & Balogh, E. P. (Eds.). (2015). Improving diagnosis in health care.
5. Siegfried, E. C., & Hebert, A. A. (2015). Diagnosis of atopic dermatitis: mimics, overlaps, and complications. Journal of Clinical Medicine, 4(5), 884–917.
6. Spiewak, R. (2023). Diseases from the Spectrum of Dermatitis and Eczema: Can 'Omics' Sciences Help with Better Systematics and More Accurate Differential Diagnosis? International Journal of Molecular Sciences, 24(13), 10468.
7. Okuboyejo, D. A., & Olugbara, O. O. (2018). A review of prevalent methods for automatic skin lesion diagnosis. The Open Dermatology Journal, 12(1).
8. Carli, P., De Giorgi, V., Soyer, H. P., et al. (2000). Dermatoscopy in the diagnosis of pigmented skin lesions: a new semiology for the dermatologist. JEADV, 14(5), 353–369.
9. Mungenast, F., Fernando, A., Nica, R., et al. (2021). Next-generation digital histopathology of the tumor microenvironment. Genes, 12(4), 538.
10. Wang, F. (2023). Development of a robotic biopsy system compatible with label-free digital pathology (Doctoral dissertation, UIUC).
11. Brinker, T. J., et al. (2019). Deep learning outperformed 136 of 157 dermatologists in dermoscopic melanoma image classification. European Journal of Cancer, 113, 47–54. https://doi.org/10.1016/j.ejca.2019.04.001
12. Fischer, F., Doll, A., Roenneberg, S., et al. (2023). Gene Expression-Based Molecular Test for Diagnosis of Psoriasis and Eczema. Journal of Investigative Dermatology, 143(8), 1461–1469.
13. Mane, P. (2024). Advancements in Machine Learning Algorithms for Predictive Analytics in Healthcare. Advances in Nonlinear Variational Inequalities, 27(3). ISSN: 1092-910X.
14. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business Horizons, 61(4), 577–586.
15. Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues. SN Computer Science, 3(2), 158.
16. Helm, J. M., Swiergosz, A. M., Haeberle, H. S., et al. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. Curr Rev Musculoskelet Med, 13, 69–76.
17. Attaripour Esfahani, S., Baba Ali, N., Farina, J. M., et al. (2025). AI Applications in Pulmonary Hypertension. Medicina, 61(1), 85.
18. Chen, Y., Jiang, H., Li, C., et al. (2016). Deep feature extraction using CNNs. IEEE Trans Geosci Remote Sens, 54(10), 6232–6251.
19. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep CNNs. NeurIPS, 25, 1097–1105. https://doi.org/10.1145/3065386
20. Zhou, X. J., Laouar, Y., & Tsoi, L. C. Big Data and AI for Inflammatory Response. Front Immunol, 16, 1553004.
21. Kanadath, A., Jothi, J. A. A., & Urolagin, S. (2024). CViTS-Net: A CNN-ViT Network. IEEE Access.
22. Samiat, A., Smart, D., & Jane, J. CVD Prediction Using ML/DL.
23. Hussain, D., Al-Masni, M. A., et al. (2024). Tumor detection in multimodality imaging. Journal of X-Ray Science and Technology.
24. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset. Scientific Data, 5, 180161.
25. Brinker, T. J., et al. (2019). AI vs dermatologists in melanoma classification. Eur J Cancer, 111, 30–37.
26. Han, S. S., et al. (2018). CNNs for benign and malignant tumor classification. J Invest Dermatol, 138(7), 1529–1538.
27. Esteva, A., et al. (2019). Deep learning in healthcare. Nature Medicine, 25(1), 24–29.
28. Dosovitskiy, A., et al. (2021). An image is worth 16x16 words. ICLR.
29. Mahbod, A., et al. (2021). Vision Transformer for Skin Lesion Classification. Computers in Biology and Medicine, 139, 104961. https://doi.org/10.1016/j.compbiomed.2021.104961
30. Naseer, M., et al. (2021). Robustness of ViTs to adversarial attacks. ICCV, 7838–7847.
31. Mahbod, A., et al. (2019). Fusing fine-tuned deep features. Comp Med Imaging Graph, 77, 101636.
32. Mane, P. (2024, January). Optimized Clustering-Based Fusion for Skin Lesion Image Classification: Leveraging Marine Predators Algorithm. Intelligent Decision Technologies, IOS Press, Netherlands. ISSN/eISSN: 1872-4981 / 1875-8843.
33. Bi, L., Kim, J., Ahn, E., & Feng, D. (2018). Dermoscopic image segmentation. IEEE TBE, 64(9), 2065–2074.
34. Haenssle, H. A., et al. (2018). CNN for dermoscopic melanoma recognition. Eur J Cancer, 108, 101–106. https://doi.org/10.1016/j.ejca.2018.02.063
35. Daneshjou, R., et al. (2021). Disparities in AI performance. Science Advances, 7(40), eabg7720.
36. Holzinger, A., Carrington, A., & Müller, H. (2019). From machine learning to explainable AI. Information Technology, 61(5–6), 263–271. https://doi.org/10.1515/itit-2019-0040
37. Miotto, R., Wang, F., et al. (2018). Deep learning for healthcare: opportunities and challenges. Brief Bioinform, 19(6), 1236–1246. https://doi.org/10.1093/bib/bbx044