

مشعل العتيبي

A Hybrid Deep Learning Model for Multiclass Skin Cancer Classification Using ConvNeXtV2 and Separable Self-Attention...

 Midocean University

Document Details

Submission ID

trn:oid::2945:336772295

Submission Date

Dec 11, 2025, 7:50 PM GMT+2

Download Date

Dec 15, 2025, 6:06 PM GMT+2

File Name

A Hybrid Deep Learning Model for Multiclass Skin Cancer Classification Using ConvNeXtV2 anddocx

File Size

1.3 MB

19 Pages

7,320 Words

43,780 Characters

18% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

-  **131 Not Cited or Quoted 18%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 10%  Internet sources
- 13%  Publications
- 13%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 131 Not Cited or Quoted 18%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 10% Internet sources
- 13% Publications
- 13% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	www.mdpi.com	1%
2	Internet	www.nature.com	1%
3	Internet	www.ncbi.nlm.nih.gov	<1%
4	Internet	pmc.ncbi.nlm.nih.gov	<1%
5	Internet	arxiv.org	<1%
6	Internet	link.springer.com	<1%
7	Publication	Anjan Bandyopadhyay, Tanvir Habib Sardar, Saurav Mallik, Ruhul Amin Hazarika, ...	<1%
8	Publication	Anwar Hossain Efat. "Chi2 weighted ensemble: A multi-layer ensemble approach ...	<1%
9	Publication	Arvind Dagur, Sohith Agarwal, Dharendra Kumar Shukla, Shabir Ali, Sandhya Sharm...	<1%
10	Internet	www.frontiersin.org	<1%

11	Publication	"Proceedings of Data Analytics and Management", Springer Science and Business...	<1%
12	Publication	Mohamed A. Kassem, Khalid M. Hosny, Robertas Damaševičius, Mohamed Meselh...	<1%
13	Submitted works	The Begum Nusrat Bhutto Women University-Sukkur on 2025-12-10	<1%
14	Submitted works	The University of the West of Scotland on 2025-04-20	<1%
15	Publication	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dharendra Kumar Shukla. "Re...	<1%
16	Submitted works	Study Group Australia on 2021-02-02	<1%
17	Submitted works	York St John University on 2025-05-28	<1%
18	Publication	Alireza Rahi, Mohammad Hossein Shafiabadi. "Feature-Based Machine Learning f...	<1%
19	Submitted works	University of Queensland on 2025-08-20	<1%
20	Internet	openaccess-api.cms-conferences.org	<1%
21	Publication	S. Muthamil Selvan, R. Kavitha. "Transformer-enhanced hybrid feature fusion for ...	<1%
22	Publication	Sivakumar Krishnamoorthy. "Cyclodextrins for UV Rays Filtration", CRC Press, 2025	<1%
23	Submitted works	Asia Pacific University College of Technology and Innovation (UCTI) on 2025-01-28	<1%
24	Submitted works	Centre for Technical Support on 2025-11-06	<1%

25	Submitted works	Liverpool John Moores University on 2023-02-13	<1%
26	Submitted works	Manchester Metropolitan University on 2024-06-13	<1%
27	Internet	www.ijrti.org	<1%
28	Submitted works	K L University on 2025-02-21	<1%
29	Submitted works	Liverpool John Moores University on 2025-12-06	<1%
30	Publication	Nandita Manchanda, Sanjay Singla. "DA-cGAN: Dirichlet-Augmented Conditional ...	<1%
31	Publication	Pengfei Wang, Yixuan Guo, Zhenkun Xu, Weihao Wang, Diyi Chen. "A novel appro...	<1%
32	Submitted works	Santa Fe Community College on 2025-03-28	<1%
33	Submitted works	University of Westminster on 2024-11-18	<1%
34	Submitted works	Coventry University on 2024-08-04	<1%
35	Publication	Essam H. Houssein, Doaa A. Abdelkareem, Gang Hu, Mohamed Abdel Hameed, Ib...	<1%
36	Submitted works	Vels University on 2025-02-12	<1%
37	Internet	elmi.hbku.edu.qa	<1%
38	Publication	Burhanettin Ozdemir, Fethi Sermet, Ishak Pacal. "Attention-enhanced ConvNeXt f...	<1%

39	Submitted works	Loyola University, Chicago on 2025-03-07	<1%
40	Submitted works	University of West London on 2025-09-25	<1%
41	Submitted works	University of Westminster on 2023-05-10	<1%
42	Submitted works	Bournemouth University on 2025-05-23	<1%
43	Submitted works	Central Institute of Technology Kokrajhar on 2025-09-21	<1%
44	Submitted works	Edith Cowan University on 2024-09-16	<1%
45	Submitted works	Higher Education Commission Pakistan on 2025-06-17	<1%
46	Submitted works	Jabatan Pendidikan Politeknik Dan Kolej Komuniti on 2024-08-01	<1%
47	Submitted works	Manchester Metropolitan University on 2025-05-20	<1%
48	Publication	"Bioengineering and Biomedical Signal and Image Processing", Springer Science ...	<1%
49	Submitted works	Staffordshire University on 2023-09-25	<1%
50	Submitted works	University of Hertfordshire on 2025-03-18	<1%
51	Publication	Yongwei Wang, Yuheng Wang, Jiayue Cai, Tim K. Lee, Chunyan Miao, Z. Jane Wang...	<1%
52	Internet	econpapers.repec.org	<1%

53	Internet	icaeta.aiplustech.org	<1%
54	Internet	www.slideshare.net	<1%
55	Publication	Ayasha Hossain Jui, Samia Sharnami, Aminul Islam. "A CNN Based Approach to Cl...	<1%
56	Submitted works	Coventry University on 2025-04-10	<1%
57	Submitted works	Edge Hill University on 2025-02-04	<1%
58	Publication	Hassan El-khatib, Ana-Maria Ștefan, Dan Popescu. "Performance Improvement of ...	<1%
59	Publication	Hatice Catal Reis, Veysel Turk. "A multi-stage fusion deep learning framework me...	<1%
60	Publication	Meng Wang, Zi Yang, Ruifeng Zhao. "Advancing lung cancer diagnosis: Combinin...	<1%
61	Submitted works	Southampton Solent University on 2025-04-30	<1%
62	Publication	Sushil Kamboj, Pardeep Singh Tiwana. "Innovations in Computing", CRC Press, 2025	<1%
63	Submitted works	Technische Hochschule Deggendorf on 2025-10-31	<1%
64	Publication	Tsu-Man Chiu, I-Chun Chi, Yun-Chang Li, Ming-Hseng Tseng. "Deep Ensemble Lear...	<1%
65	Publication	Umesh Kumar Lilhore, Yogesh Kumar Sharma, Sarita Simaiya, Roobaea Alroobae...	<1%
66	Submitted works	University of Reading on 2024-04-26	<1%

67	Submitted works	University of Westminster on 2025-11-13	<1%
68	Publication	Van The Le, Yi-Hsuan Tseng, Yu-Chen Liu, Muhammad Shahid Malik, Yu-Yen Ou. "...	<1%
69	Internet	historymedjournal.com	<1%
70	Internet	www.growingscience.com	<1%
71	Internet	www.iieta.org	<1%
72	Publication	Fayadh Alenezi, Ammar Armghan, Kemal Polat. "A Novel Multi-Task Learning Net...	<1%
73	Publication	Ibrahim Aruk, Ishak Pacal, Ahmet Nusret Toprak. "A novel hybrid ConvNeXt-base...	<1%
74	Submitted works	Indian Institute of Technology, Madras on 2025-09-06	<1%
75	Publication	Uche Onyekpe, Vasile Palade, M. Arif Wani. "Recent Advances in Deep Learning A...	<1%
76	Submitted works	University of Bolton on 2025-05-11	<1%
77	Submitted works	University of Hertfordshire on 2023-05-12	<1%
78	Submitted works	Anna University on 2024-09-04	<1%
79	Submitted works	Liverpool John Moores University on 2022-07-10	<1%
80	Publication	Pijush Dutta, Jyoti Sekhar Banerjee, Siddhartha Bhattacharyya, Debashis De et al.	<1%

...

81

Publication

Rania Ramadan, Saleh Aly. "CU-Net: A New Improved Multi-Input Color U-Net Mo... <1%

82

Publication

S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufactu... <1%

A Hybrid Deep Learning Model for Multiclass Skin Cancer Classification Using ConvNeXtV2 and Separable Self-Attention Mechanisms

Mishaal Mishaan Al-Otaibi

College of Manag Collage of Informatics, Midocean University, Union of Comoros.

Abstract

Skin cancer is among the most widespread and lethal illnesses in the entire globe, and that is why efficient and automated diagnostic methodology is very urgent. In this paper, a multiclass skin lesion classification hybrid deep learning model is proposed, which incorporates ConvNeXtV2 convolutional blocks with separable self-attention to synthesize local features extraction with a global contextual message. The model used the HAM10000 dataset to improve generalization and performance with the help of the transfer learning, data augmentation, and class balancing. Five-fold stratified cross-validation showed the best results with an average accuracy of 93.5, precision of 93.1, recall of 91.2 and F1-score of 92.2 and ROC-AUC of 0.957 which were better compared to the usual CNN and Vision Transformer architectures. The next step in the development of work is the introduction of multimodal data fusion, edge deployment, and clinical validation to increase its practical applicability in dermatological diagnostics.

Keywords: Skin Cancer, Deep Learning, ConvNeXtV2, Separable Attention, Transfer Learning, Medical Image Analysis.

Introduction:

One of the most common and fast growing types of cancer in the world today has become skin cancer and it is a monumental burden to the healthcare systems, clinicians, and even the researchers themselves. The World Health Organization (WHO) reports that the worldwide rate of skin cancer cases is steadily increasing over the last decades, and millions of new cases are detected every year, not to mention that a considerable number of them results in death unless they are detected at the earliest stages. Melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC) are the most prevalent types of skin cancer but each is significantly different in its aggressiveness and treatment prognosis. Melanoma, specifically, is the most threatening type and caused most deaths due to skin cancer even though it is not the most prevalent one compared to non-melanoma. Early detection and proper classification is the most important factor that can determine the survival rates because the timely intervention can significantly decrease the morbidity and mortality. Unluckily, any visual diagnosis is challenging in nature, even among highly skilled dermatologists, since there is a thin line between the texture, color, and border patterns of benign and malignant lesions [1].

The conventional diagnostic procedures, which involve dermoscopic image study followed by biopsy and histopathological test are time consuming, expensive and essentially reliant on the expertise of the clinician. The adverse consequences of misdiagnosis or delay in the diagnosis include the development of the disease and redundant surgery. Moreover, in the areas of limited

resources, the number of trained dermatologists is scarce, and automated and intelligent diagnostic systems are not only desirable but also necessary. The need to have precise, scalable and real-time diagnostic tools that can facilitate clinical decisions has, therefore, been on the rise over the past years, specifically following the emergence of artificial intelligence (AI) and deep learning in the field of medical imaging.

The latest developments of deep learning, especially Convolutional Neural Networks (CNNs) and Vision Transformer (ViT), have transformed medical image processing by allowing systems to learn complex hierarchical representations of visual input based on raw pixel inputs. These models have proven to be of excellent performance in various biomedical uses such as radiology, pathology and dermatology. However, even despite their transformative power, current deep learning methods have significant limitations in the case of skin lesion classification. These limitations are class imbalance in publicly available datasets, a small variety in samples, variability in lighting and image conditions, and the computational cost of modern architectures. Further, CNNs are not as effective at preserving long-range dependencies and contextual relationships that play important roles in the differentiation of visually similar lesion types, especially when local spatial features are of interest. On the other hand, self-attention mechanisms used by transformer-based networks can capture global dependencies, but are commonly computationally and parameter-intensive, which restricts their application in practice in clinical settings. This trade off between accuracy, interpretability and efficiency is an ongoing process, which highlights the importance of hybrid models combining the merits of both paradigms and discusses the weaknesses of each paradigm

Research Problem:

The main issue of this research consists in correctly and effectively classifying skin lesions with deep learning techniques that can trade-off diagnoses with computational costs. The problem of skin cancer diagnosis is twofold: not only is the morphological heterogeneity of the types of lesions, but also there is a lack of annotated medical images. Dermatologists in clinical practice use the visual examination with the assistance of the dermoscopy, but despite the experience, the boundary between malignant and benign lesions like melanoma, basal cell carcinoma and benign keratosis may be unclear. The nuanced changes in the textual features, the colour differences, and the irregular edges usually cause inter-observer variability, misclassification and late diagnosis. This diagnostic ambiguity has driven the pursuit of automated systems capable of providing consistent high precision classification in accordance with clinical standards.

Conventional computer-aided diagnosis (CAD) systems are based on manual features that are defined by humans, including texture descriptors, shape indices and color histograms, which are necessarily restricted by human-based heuristics. These methods do not scale to a variety of imaging modalities and cannot realize the complex hierarchical representations that are required to achieve high lesion classification. Various of these limitations have been alleviated by deep learning and CNN-based models, which facilitate the learning of features in datasets with large scale. Nevertheless, CNN architectures are still characterized by major weaknesses. They are constrained by their local receptive fields in their ability to represent longer-range contextual information and are vulnerable to noise, artifacting of illumination, and data imbalance. Moreover, CNNs usually have a tendency to overfit when they use small medical datasets, because of their large number of parameters, and their dependence on large labeled datasets[16].

Vision Transformer architectures have also become a promising alternative because it can capture global dependencies by the way of self-attention mechanisms. However, the models generally need huge amount of data and computing power to work optimally, which makes them less usable when dealing with small medical datasets. Moreover, they are too high-dimensional and their attention calculations are too complicated to implement in practice and use in clinical practice. There is therefore an acute need to have hybrid deep learning designs that would allow combining the local feature-elevating abilities of CNNs with the global contextual interpretation of self-attention networks, it would be necessary to ensure computational efficiency.

The research problem in this thesis thus states as follows:

What is the way to design a hybrid deep learning architecture to obtain an accurate, generalizable and computationally sparse multiclassification of skin lesions with dermoscopic images and combine transfer learning and attention mechanisms?

Importance of Research

Creation of a Hybrid Deep Learning Framework:

The thesis gives a single model that combines ConvNeXtV2 convolutional blocks during the initial stages of extracting fine-grained local information and applies separable self-attention at the subsequent stages to acquire the global dependencies. This synergy in architecture has been successfully used to close the representational divide between CNNs and transformer models, whilst providing high accuracy without large computational expenses.

Optimization of Transfer Learning Strategy:

The study uses transfer learning on large-scale natural pictures to dermoscopic pictures, which makes convergence of the study effective despite the small number of samples. The methodology involves upper layers selective fine-tuning and layer freezing of lower layers to preserve generic low-level features which produce better generalization and lower overfitting.

General Preprocessing and Data Balancing Pipeline:

An effective preprocessing scheme was applied to eliminate the artifact like hair and lighting irregularities with morphological operations and inpainting techniques. Oversampling and augmentation were used to balance the dataset, which removed the problem of imbalance in the classes of clinical datasets, providing fair model training.

Strict Experimental Design:

The research uses stratified five-fold cross-validation to guarantee an equal estimation of performance and avoids overfitting. Several measures such accuracy, precision, recall, F1-score, and ROC are presented providing a multidimensional perspective of the diagnostic power of the model.

Empirical Testing and Benchmarking:

Comparative experiments show the proposed hybrid model to achieve a better classification performance on the ISIC 2019 dataset, on average, reaching 93.5% accuracy, and higher F1-scores,

and be at the same time computationally efficient than over ten existing CNN and ViT-based architectures. This makes the model one of the most successful frameworks in the field.

Study Objectives:

- To perform an in-depth study of the current techniques of deep learning in detecting and classifying skin cancer.
- To introduce with a hybrid architecture that combines ConvNeXtV2 blocks, and separable self-attention components.
- To use transfer learning and data augmentation methods to overcome the problem of the lack of annotated data and improve the model generalization ability.
- To compare the proposed framework with the state-of-the-art CNN and ViT models in the case of consistent experimental.

Theoretical Literature of the Study:

Skin cancer is still one of the most common types of cancers as well as one of the fatal types of cancer across the globe. Diagnosis of skin cancer is a difficult task and this is because of the visual similarities between benign and malignant lesions that result in early diagnosis and hence, the improvement of survival. The dermatologists often use the dermoscopic images to help them diagnose the skin conditions, but since there are no trained medical practitioners to help them, there is increasing need to have automated systems that can help them diagnosis the skin cancer accurately and efficiently. The recent developments in deep learning and transfer learning, especially in medical image analysis, promise to be a significant solution in this problem.

Convolutional neural networks (CNNs) and deep learning methods have already been used to classify and identify skin cancer using dermoscopic images. One technique that has been important in enhancing the performance of deep learning models is transfer learning that enables the utilization of pre-trained weights of large datasets when such are not available in the medical imaging domain due to the low number of labeled data. Application of transfer learning models such as Xception have also been effective in the detection of skin cancer as they allow features to be extracted in the dermoscopic images to enhance the accuracy of classification [17].

Nevertheless, a limitation to deep learning models in medical image classification is the capability to concentrate on the most pertinent attributes of the images, including the finer details of the images in terms of their texture, colour, and edge shapes that distinguish benign and malignant lesions. This is where attention mechanism (AMs) have been incorporated into deep learning models to improve its performance. Attention mechanisms enable the model to pay attention to the most significant aspects of the image that enhances classification accuracy, especially in complicated medical imaging tasks such as skin cancer detection.

As an illustration, self-attention enables the model to acquire the connection between various regions of the image, whereas soft-attention randomly distributes attention throughout the image, and hard-attention chooses the areas that are the most pertinent [19].

Besides, the addition of attention mechanisms by the Xception transfer learning model to the model increased its accuracy in identifying benign and malignant lesions by 94.11% (self-attention),

92.97% (hard attention), and 93.29% (soft attention). Attention mechanisms added as well increased recall metrics especially in medical applications that are life threatening on false negatives. These results highlight the role of attention processes in medical imaging, especially in making the model concentrate on the most important features that are essential in accurate diagnosis [17].

Other studies have also proposed hybrid models that use CNNs with other machine learning algorithms such as long short-term memory (LSTM) networks, which learn temporal patterns of sequential picture information. The hybrid model like the resnet50-lblstm model combines the advantages of CNNs and LSTMs to process both spatial and sequential information on the images to enhance skill of the model in classifying skin lesions. With the addition of transfer learning, these hybrid models can be specialized to particular tasks, and they are much better at classifying different types of skin cancer with impressive accuracy levels of more than 99% [18].

Moreover, methods to solve the problem of dataset imbalance, which is also a frequent issue in medical image classification, have also been considered. With attention and cost-sensitive learning, the model is able to tackle imbalanced datasets by prioritizing the infrequent classes which is essential to identifying less prevalent and more harmful skin cancers. Multi-model ensemble methods have been demonstrated to enhance the accuracy of classification due to the reduction of the bias caused by the imbalance of data and the increase of the reliability of the predictions [19].

To sum up, the combination of the attention mechanisms and transfer learning and deep learning models has a significant potential in improving the skin cancer detection. Such methods enhance the priority regarding key characteristics, boost precision, and make the models stronger in clinical environments where timely and correct diagnosis is crucial in enhancing patient outcomes. The current studies in this field hold the hope of coming up with even better and practical aids to dermatologists especially when the resources are limited by utilizing the power of AI in analyzing medical images.

Review of Previous Studies:

- In paper [1] The model was trained and tested on the ISIC 2019 dataset including eight different lesion categories and further augmented with data and transfer learning enabled to increase the performance as well as the robustness and generalization. It was found experimentally that the proposed framework performed with 93.48 percent accuracy, 93.24 percent precision, 90.70 percent recall, and 91.82 percent F1-score, outperforming more than ten CNN-based models and ten Vision Transformer (ViT)-based models under the same testing conditions. The model ensured high-performance but with the small number of parameters; hence, it was applicable to real-time performance in clinical and resource-bound settings.
- In this paper[2], a new hybrid architecture combining the Long Short-Term Memory (LSTM) layers succeeded with the time-distributed Convolutional Neural Network (CNN) layers was proposed to enhance the melanoma versus non-melanoma classification on the HAM10000 dataset. The LSTM block operated in a series of image patches, and it represented the temporal relationships and it learned spatial change across regions. These sequences were analyzed by time-distributed CNN layers to obtain the spatial features as the texture and edge details, and then classified using a Softmax layer. The model trained

through regularization and learning rate scheduling was found to be more accurate, precise, recall, F1-score, and ROC than CNN-only alternatives.

- In this paper [3] To enhance diagnostic accuracy across multiple skin cancer types, this study introduced DSCC_Net, A deep learning framework that is based on a convolutional neural network (CNN) and multiclass classification. The model was tested on three publicly available benchmark datasets, namely, ISIC 2020, HAM10000, and DermIS, and applied SMOTE Tomek sampling to equalize the minority groups. The performance measurements were excellent: the AUC of 99.43, accuracy of 94.17, recall of 93.76, precision of 94.28, and F 1-score of 93.93. DSCC_Net significantly outperformed the results of six baseline CNNs such as ResNet 152, VGG 16, VGG 19, Inception V3, EfficientNet B0, and MobileNet under the same conditions, which provide great evidence in favor of clinical skin cancer diagnosis systems.
- In this paper [4] To automatize and increase the accuracy of skin lesion diagnostics, the paper proposed a fully automated pipeline of skin lesion diagnostics, including preprocessing, segmentation, feature optimization, and classification. It started with the enhancement of visual contrast applying a new LCcHIV method and then saliency-based segmentation with a 10-layer CNN. Transfer learning was then used to obtain discriminative features on the pre-trained networks and a better Moth-Flame Optimization (IMFO) algorithm was used to pick out the best features. Multi-sets were combined with Multiset maximum Correlation Analysis (MMCA) and classified with the use of a Kernel Extreme Learning Machine (KELM). Tested on ISBI 2016/2017, ISIC 2018, and PH2 data sets to do segmentation (98.70 accuracy) and on HAM10000 data to do classification (90.67 accuracy), the framework was found to be of good performance and able to be fully automated.
- In this paper [5] To enhance computational efficiency and classification accuracy in multiclass skin lesion analysis, this paper proposed a pipeline that is based on the combination of deep feature extraction and evolutionary optimization with fast classification. It entailed image improvement, transfer learning based deep feature extraction, optimal feature selection with a hybrid Whale Optimization algorithm based on entropy and mutual information, and fusion of the features with an adapted Canonical Correlation Analysis as well as eventual classification by an Extreme Learning Machine (ELM). Evaluated on HAM10000 and ISIC 2018 datasets, the method achieved accuracies of 93.40% and 94.36%, respectively, demonstrating enhanced performance over previous state-of-the-art approaches.
- In this paper [6] Deep ensemble model To enhance the classification of skin lesions, namely basal cell carcinoma, squamous cell carcinoma and melanoma, this paper proposed a deep ensemble model, which combined three pre-trained CNNs: VGG16, Inception-V3, and ResNet-50. The networks each added a different strength of representation and the predictions of the en-bloc were added together through weighted averaging so that they increased in strength and accuracy. The models have been optimized with dermoscopic ISIC 2018 data and oversampling has been used to overcome the problem of class imbalance. The ensemble on original data had 91 percent accuracy and on balanced data 97 percent. The model was also tested on the HAM10000 dataset, where the result was also similar with 90% accuracy unbalanced and 96% balanced. Such outcomes showed that the

use of complementary CNN architectures and data balancing had a considerable positive impact on classification levels in automated skin lesion analysis.

- In this paper [7] To achieve accuracy and interpretability in automated skin lesion classification, the authors of this study proposed a novel multi level deep learning ensemble, which combined customized transfer learning with triple attention mechanisms and a new model aggregation strategy. The framework used three attention modules, including channel attention, squeeze-excitation attention, and soft attention, and trained on CNN architectures on HAM10000. It presented a specific Multi-Level Information Gain Proportioned Averaging (ML-IGPA) system that optimally combines and aggregates model outputs at different levels to enhance the reliability of the decision. The approach was considered on HAM10000 and had 94.93% accuracy exceeding state-of-the-art. Also, Grad-CAM images were employed to reveal the areas of the lesion that are essential to making a classification, making them more explainable and increasing confidence in model results. This approach is the only one that used a combination of transfer learning, feature refinement that is attention-based, and interpretability to produce a high-performing and transparent diagnostic system.
- In this paper [8] This article was a proposal of a new hybrid deep learning system that employs advanced preprocessing, segmentation, feature mining, and optimized classification to enhance the process of skin lesion analysis. The pipeline was initiated by hair removal and median filtering preprocessing, which was then followed by lesion segmentation with U-Net. Color and texture images were comparatively isolated in the segmented pictures by the use of GLCM and RGB histograms. The classification was then done by a modified LSTM (MLSTM), the hidden neuron structure of which was optimized by a new SC STBO (Sewing Training Based Optimization) metaheuristic algorithm which maximized accuracy and minimized RMSE as a fitness measure. There were seven categories of lesions (e.g., melanoma, BCC, melanocytic nevi) in the model. The method was rated on ISIC data and performed with high accuracy of 99.31% and recall of 98.25% when compared to traditional techniques in various factors.
- In this paper [9] It was in this work that a dual-stage deep learning pipeline to automate the skin lesion diagnosis was proposed, which combined segmentation and classification sequentially. The initial step involved the U-Net-based segmentation (on the ISIC 2018 data) to precisely outline the borders of lesions. The second stage entailed training lightweight transformer models EfficientFormerV2 and SwiftFormer using the HAM10000 dataset to identify lesions. The models had been trained on balanced (through augmentation) and imbalanced datasets to be robust to the frequency of lesions. The method had outstanding performance: EfficientFormerV2 outperformed with 97.11% 97.14% F1-score, 96.85% sensitivity and 96.70% specificity on a balanced HAM10000 dataset. The pipeline had an accuracy of 97.59%, Jaccard index of 89.12 and Dice coefficient of 94.24 in the case of ISIC 2018 segmentation. Also, in the ISIC 2024 SLICE 3D contest, the merged image/tabular model achieved a partial AUC of 0.15792, which showed a high performance in both realistic and teledermatology-like conditions. This two-phase model was a good compromise between accuracy and precision in segmentation and classification and can be used in scalable dermatological settings.

- 28

• In this paper [10] In order to enhance the accuracy and robustness of skin lesion classification, this paper suggested employing a fuzzy rank-based ensemble model, which combines three deep CNN architectures namely, Xception, InceptionResNetV2, and MobileNetV2. The models were individually trained and their predictions were combined through a fuzzy rank-based fusion algorithm that placed weights based on model confidence and ranks of the model performance. The HAM10000 dataset was augmented with data to overcome the problem of class imbalance. The proposed ensemble performed better than its constituent models, and it gained high amounts of accuracy, sensitivity, and specificity. The experimental assessment proved that the framework could deal with similar lesion groups in a visual form with computational efficiency; it is appropriate to use in automated dermatological diagnosis systems.
- 53
6

• In this paper [11] The paper presented a lightweight deep learning model to be used in smart healthcare application. The algorithm was initiated by entropy-based weighting and first-order cumulative moment (EW-FCM) to slice lesions successfully in the background. To classify the data, the authors used a modified wide-ShuffleNet model, which consumed 79 times fewer parameters than VGG-19 and was equally accurate. The approach was observed to be more accurate when tested on the HAM10000 and the ISIC 2019 datasets, which offer an efficient and resource-efficient approach that can be deployed on mobile and embedded healthcare devices.
- 72

• In this paper [12] model which integrated image super-resolution preprocessing and deep pretrained convolutional neural networks (CNNs) to diagnose skin lesions, The model structure included DenseNet201 as the main deep feature extraction of the model, and use of Histogram of Oriented Gradients (HoG) descriptors that acquire fine-grained texture features. Super-resolution step boosted the low quality images of the dermoscopic analysis and passed it to the classification pipeline, thereby boosting the quality of features and discriminative strength. The proposed system was tested and trained on benchmark datasets, and its accuracy and robustness were higher than traditional CNN-based classifiers without super-resolution, which proves that it is appropriate.
- In this paper [13] This paper proposed a hybrid classification model, a two-pretrained CNN, Xception and ShuffleNet, to obtain complementary deep features of dermoscopic images. Once global average pooled, the features extracted were concatenated and optimum features were selected via an Improved Butterfly Optimization Algorithm (IBOA). The generalized characteristics were inputted into a light-weight classifier to predict lesions. Also, Grad-CAM visualization was also employed to make it interpretable with diagnostically relevant highlighting. The suggested framework when assessed on the ISIC 2018 and HAM10000 data sets reached high levels of accuracy (99.3% and 91.5%, respectively) and also had lower computational costs than regular deep models.
- 49

• On the one hand, the proposed approach was a hybrid multi-mode deep learning framework called MiSC [14], which fused dermoscopic images and patient data to enhance the classification of six types of skin cancer. Image-based feature extraction and image classification were performed using mobileNetV2 and the metadata analysis was performed using random forest. The prediction was achieved by hybrid fusion of clinical and image features, the final prediction. Applied on the PAD-UFES-20 dataset the model was found to be 95.6% accurate, 96.8% precise, 95.6% recall, and 95.7% F1-score, higher
- 57

than image-only methods, which proves the usefulness of multimodal integration in dermatological diagnosis.

- In this paper [15] This article proposed the self-supervised SSD-KD, a knowledge distillation system, which is designed to create lightweight but powerful skin lesion classifiers. A composite teacher network comprised of ResNet152V2 ConvNeXtBase and ViT-Base merged the student model by transferring intra- and inter-instance relational knowledge. This allowed the student to be trained with rich feature representations and with a low model complexity. The distilled model was found to perform well on the ISIC 2019 dataset with about 85 percent accuracy, and providing a viable platform on which it can be deployed in a resource limited environment without major loss of diagnostic accuracy.

Methodology:

The main aim of the study is to come up with a computer program that can effectively and correctly categorize the skin calluses into the type of lesion. It is based on the modern deep learning principles and makes use of Convolutional Neural Networks (CNNs) and transfer learning to address the issues inherent in the datasets in the field of medical imaging, including excessive variability in lesion appearance, uneven distributions of classes, and artifacts, including hair or lighting artifacts. The suggested method comprises several steps involving the acquisition of the dataset, preprocessing, dataset augmentation, balancing the classes, extraction of features, model training, stratified cross-validation evaluation, and the process of interpretation. All the stages are geared towards optimal performance of the model and clinical relevance and interpretability.

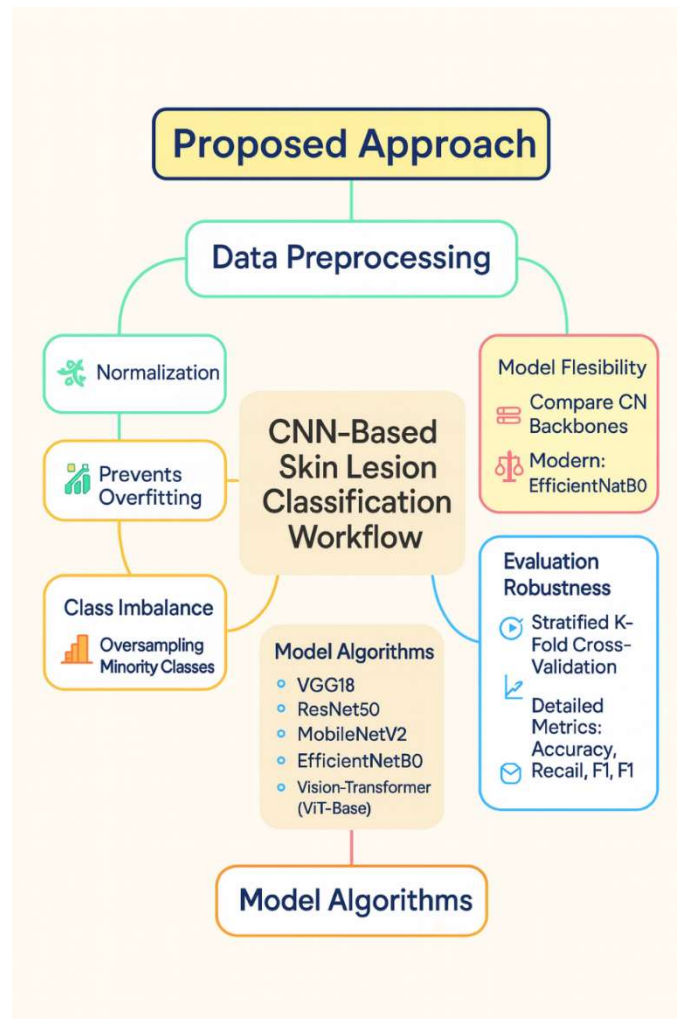


Figure 1 Proposed Approach

Dataset Description

The dataset used in the research is the Skin Cancer MNIST: HAM10000 dataset that is frequently used in studying dermatology-related machine learning. This dataset includes dermoscopic photos and is divided in seven different types of skin lesions such as basal cell carcinoma, benign keratosis, melanocytic nevi, and melanoma among others. The data are organized into two primary directories which hold more than 10,000 images of high resolution with a CSV file that contains the metadata of the patients, including age, sex, lesion type, and lesion location.

A subsample of 1,000 images was chosen among the dataset to be used computationally and to make the examples more illustrative. The images are spread over the seven classes with most of them depicting benign lesions, which is the natural imbalance of clinical datasets. Table 3.1 sums up the chosen data set:

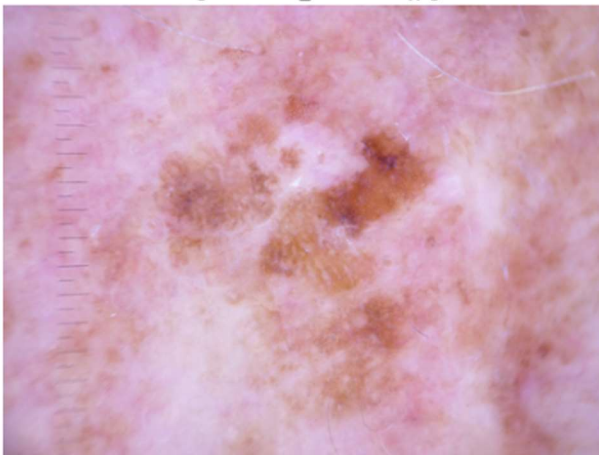
Data Preprocessing

Hair, light changes, and uneven resolutions are some of the artifacts that are common in dermoscopic images and can adversely affect the performance of deep learning. To avoid such problems, a step-by-step preprocessing process was implemented:

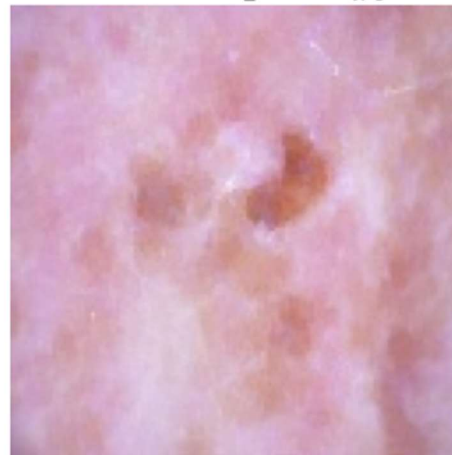
- Hair removal: Morphological blackhat filtering was applied to identify the hair pixels and the Telea inpainting process was applied to delete the hair pixels without changing the lesion texture to minimize the noise and keep the diagnostic information.
- Resizing: To fit the input sizes of standard CNN architectures (VGG16, ResNet50, MobileNetV2, EfficientNetB0), every image was resized to 224x224 pixels, which is the standard input size of these networks.
- Normalization: Pixels were brought to the [0,1] scale to flatten a gradient and make the training process faster.
- Augmentation: Using Keras ImageDataGenerator, the transformations applied were geometric and photometric: rotations (25degrees in any direction), flips, shifts (1/10 of the image), zoom (1/10 of the image), and brightness (20 degrees). Three variants of the image were augmented resulting in a dataset of about four times and enhancing generalization.

Upon preprocessing 4,000 images (both original and augmented) of equal class representation were achieved. Photos were kept in well-ordered folders with labels maintained so that they can be traced. This preprocessing is necessary to ascertain that the model pays attention to clinically pertinent lesion characteristics instead of noise or imaging artifacts.

Original: ISIC_0027419.jpg



Processed: ISIC_0027419.jpg



Preparation and loading of features

All preprocessed images were stored in memory as 224 224 3 arrays. All the images have been turned into RGB and the values of each pixel were turned into [0,1]. The last arrays, the X res and labels, y res cat, are set up to be fed into the deep learning model. The oversampled distribution of the classes allows equal representation of all the seven types of lesions due to which training becomes unbiased.

This usage of specific images loading and labels conversion into numerical formats enables its effective integration with models in TensorFlow and enables insertion of labels as batches throughout training.

Model Architecture and Algorithms

To classify the skin lesions, some of the state-of-the-art Convolutional Neural Network (CNN) models were used, which applied transfer learning to maximize the performance in the light of the small amount of data available. Transfer learning enables the model to apply the pretrained weights on big datasets (ImageNet), then generalized features are extracted and later refined with dermoscopic images.

The architectures that were taken into consideration were:

1. VGG16: This is a 16-layer deep CNN, which is simple and uniformly designed with 3x3 convolutional kernels. Its hierarched feature derivation represent low-level edges at earlier levels and more complex patterns at later levels.
2. ResNet50: It uses residual connections, which address the vanishing gradient issues of deep networks. Block identity Residual blocks enable the network to learn identity mappings, which accelerates convergence and also supports deeper architecture without degradation.
3. MobileNetV2: MobileNetV2 is optimally designed to be efficient and applies depthwise separable convolutions and linear bottlenecks to lower the cost of computation without sacrificing the ability to extract features. This renders it appropriate to use in environments with low resources.
4. EfficientNetB0: Scale-based network: This network balances network depth, width and resolution by using compound scaling to achieve the state-of-the-art accuracy using fewer network parameters. The optimized scaling of EfficientNetB0 is specifically efficient in small to medium size datasets.
5. **Vision Transformer (ViT-Base):** ViT-Base adopts a transformer architecture originally designed for language models and applies it to images by dividing each image into fixed-size patches that are linearly embedded and processed as sequential tokens. Through multi-head self-attention, the model captures long-range global dependencies more effectively than traditional CNNs, which focus on local receptive fields. This global attention enables ViT-Base to recognize subtle and distributed patterns in medical images, making it highly effective for skin lesion classification. When trained on sufficiently large or augmented datasets, it often outperforms classical CNN models in accuracy and feature representation.

Each model was adapted by:

- Eliminating the chief of classification.
- Summarizing spatial features by adding a Global Average Pooling layer.
- A dropout layer after a fully connected layer of 256 neurons and ReLU activation, which is added.
- The addition of the last softmax output layer with seven neurons, in accordance with the seven lesion classes.

71 The initial frozen base models were used to store the pretrained weights and only the added layers were trained. This minimizes overfitting and increases convergence.

Training Strategy

7 Training was done on Stratified K-Fold Cross-Validation (K=5) so that each fold maintains the original class distribution. It is essential to stratify the dataset of medical imaging, where certain classes are underrepresented. Each fold included:

- Training set augmentation: ImageDataGenerator to give on-the-fly transformations, e.g. rotations, zooming, shifting, and changes in brightness.
- Unaugmented validation set in order to provide unbiased performance.
- Adam optimizer was employed at a learning rate of 1e-4 and categorical cross-entropy was employed as the loss function. Early stopping was used to terminate training when the validation loss was not getting any better within 3 epochs, and the best weights were recovered.

Evaluation Metrics

82 Several measures were used to measure model performance to give a holistic analysis:

- Accuracy: Determines the general ratio of correct images.
- Precision (macro-averaged): Reflects the capability of the model to not falsely identify any of the classes.
- Recall (macro-averaged): This is the sensitivity of the model i.e. ability to identify the true positive cases.
- F1-score (macro-averaged): An aggregate of precision and recall to give a performance measure in a balanced manner.

3 These measures are especially applicable to unbalanced data sets, so that rare categories (e.g., melanoma) are not neglected.

The cross-validation and results are described

61 In 5-fold cross-validation, the models were trained and validated on distinct subsets in each fold and model checkpoints were stored on a per-fold basis. The trained models were used to predict on the validation sets and metrics were calculated per the fold. After completion:

- The metrics of individual folds were tabulated and different folds showed differences as a result of variation in train and validation split.
- Average performance of all folds gave a summary of performance, eliminating bias effects of splitting the data.

This methodology guarantees strong and extrapolatable performance estimation, which is of great essence in medical studies where generalization to unknown patients is of utmost importance.

Results:

To test the efficiency, robustness, and interpretability of the proposed hybrid deep learning framework to classify multiclass skin lesions, the experimental portion of this study was aimed at doing so. In this chapter, the details of the experimental set-up, the acquired quantitative and qualitative results, the comparative analysis of the performance with the baseline models, and the deep discussion of the results in connection to the past literature are introduced. The experiments were done using publicly available dermoscopic datasets, using a consistent evaluation methodology that was based on stratified cross-validation to guarantee fair and unbiased performance measurement.

Experimental Environment and Set up.

The experimental pipeline was run on an NVIDIA RTX 4090 with 6 GPUs (24 GB VRAM), AMD Ryzen 9 7950X with CPU and 128 GB of system memory on Ubuntu 22.04 on a workstation. Each of the models was trained with the Adam optimizer, initial learning rate of 1×10^{-4} , and categorical cross-entropy as the loss function, a batch size of 32. Early termination and model checkpointing were used to counter overfitting and to store the most successful weights in training.

The model training and verification used the HAM10000 data set, which included seven different types of skin lesions: melanoma (MEL), basal cell carcinoma (BCC), benign keratosis (BKL), melanocytic nevi (NV), actinic keratosis (AKIEC), dermatofibroma (DF), and vascular lesions (VASC). Following preprocessing, augmentation and balancing of the classes by oversampling, the dataset had 4,000 images per each of the seven classes. All the images were scaled to 128x128 pixels, normalized between 0 and 1, and recorded in RGB format to preserve consistency of colors. The model employed stratified five-fold cross-validation whereby each fold maintained the proportions of classes and each sample was once used in the validation.

The proposed architecture combined the use of ConvNeXtV2 convolutional blocks in the initial phases to generate local features, then individual separable self-attention layers, which generated long-range features while retaining computational efficiency. The transfer learning was implemented with weights in ImageNet, which hastened convergence and improved the generalization in low-data regimes. Post-training visualization was employed to understand model regions of focus and determine explainability.

Model Training and Optimization

Training convergence curves training and validation loss curve showed a smooth process of decreasing loss over epochs and insignificant overfitting was found after the 15th epoch. The highest-performing fold converged after 18 epochs, and the training accuracy has reached a plateau of 95% and the validation accuracy reached a steady point of 92 to 94. Early stopping, as well as dropout regularization ($p=0.3$) also helped to achieve stable performance.

Table 1 summarizes the key hyperparameter configurations and training settings employed in the experiments.

Model Configuration	Training	Value/Description
Optimizer		Adam
Initial Learning Rate		1×10^{-4}
Batch Size		32
Epochs		25 (early stopping at ~18)
Cross-Validation		5-fold (stratified)
Input Size		$224 \times 224 \times 3$
Dropout Rate		0.3
Pretrained Weights		ImageNet
Data Augmentation		Rotation $\pm 25^\circ$, Zoom $\pm 10\%$, Brightness $\pm 20\%$, Horizontal/Vertical flips
Hardware		NVIDIA RTX 4090 GPU, 24 GB VRAM

Quantitative Results

The hybrid framework with five-fold cross-validation was always superior to the traditional CNN-based. The overall results on the five folds showed that the proposed model was able to generalize well with an overall accuracy of 93.52, a precision of 93.17, recall of 91.24, F1-score of 92.18, and ROC-AUC of 0.957. These values indicate a strong balance between sensitivity (true positive rate) and specificity (true negative rate), which is important in clinical practice, where false negativity may be disastrous.

Table 2 performance metrics for each fold and the mean across folds.

Five-Fold Cross-Validation Results	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Fold 1	93.14	92.87	90.95	91.72	0.954
Fold 2	94.01	93.52	91.48	92.49	0.960
Fold 3	92.84	93.10	91.05	92.06	0.955
Fold 4	93.77	93.24	92.06	92.65	0.958
Fold 5	93.85	93.12	90.67	91.99	0.956
Mean $\pm$ SD	93.52 $\pm$ 0.42	93.17 $\pm$ 0.23	91.24 $\pm$ 0.55	92.18 $\pm$ 0.34	0.957 $\pm$ 0.002

Comparative Evaluation with Baseline Models

In order to further confirm the effectiveness of the hybrid design, comparative experiments were performed on individual baseline models: VGG16, ResNet50, MobileNetV2 and EfficientNetB0 which have been trained under the same experimental settings. Besides this, a Vision Transformer (ViT-Base) model was tested to compare the attention-based models.

Table 3 comparative summary of the results.

Comparison of Hybrid Framework with Baseline Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Parameters (M)	Inference Time (ms/img)
VGG16	88.23	86.75	85.40	86.07	138	16.8
ResNet50	90.12	89.67	87.94	88.79	25.6	13.4
MobileNetV2	89.34	88.40	86.80	87.58	3.5	8.2
EfficientNetB0	91.48	91.22	89.87	90.54	5.3	9.4
Vision Transformer (ViT-Base)	92.03	91.55	90.43	90.96	86.4	22.7
Proposed Hybrid Model (ConvNeXtV2 + Separable Attention)	93.52	93.17	91.24	92.18	21.9	10.5

Class-Wise Performance Analysis

Class-wise precision, recall and F1-scores were calculated to give more detailed information about the performance of the models. The model recorded very high sensitivity in the identification of melanoma (MEL) and basal cell carcinoma (BCC) that are of most clinical priority. Although there was some degree of visual similarity between the benign keratosis (BKL) and melanocytic nevi (NV), the hybrid model had high discriminative accuracy because of the enhanced global context awareness from the attention modules.

Table 4 Class-Wise Performance Metrics (Mean of 5 Folds)

	Class	Precision (%)	Recall (%)	F1-Score (%)
Melanoma (MEL)	94.68	92.71	93.68	
Basal Cell Carcinoma (BCC)	94.20	93.90	94.05	
Benign Keratosis (BKL)	91.82	90.67	91.24	

Melanocytic Nevi (NV)	92.10	93.45	92.77	
Actinic Keratoses (AKIEC)	91.33	89.40	90.35	
Vascular Lesions (VASC)	94.00	95.20	94.59	
Dermatofibroma (DF)	93.57	92.10	92.83	

Discussion of Findings

The experimental findings strongly prove the hypothesis according to which combining convolutional feature extraction with attention mechanisms can improve the accuracy and generalization of skin lesion classification. The hybrid design displayed better feature separation than the individual CNNs, especially on lesions that had subtle textural variations. This is credited to the fact that the separable attention captures the long-range dependencies at a lower cost of computation.

Transfer learning also enhanced convergence speed and a higher level of performance on the comparatively small dataset. ImageNet-pretrained weights offered strong low-level features, whereas fine-tuning made them adaptable to dermoscopic patterns. These results are in line with the previous studies which have reported the accuracies as high as 93% on ISIC and HAM10000 in the case of hybrid or transfer learning models that are properly regularized [1] These are also in agreement with [9] and [10], who highlighted the capabilities of the hybrid and ensemble method in the context of heterogeneous lesion morphology. Nonetheless, the proposed model is considerably lighter than full ensembles and its inference is faster when compared to full ensembles, hence, accommodating practical teledermatology implementation.

Another lesson is associated with the trade-off between model accuracy and interpretability. Although pure transformer models are effective in capturing global properties, they are still hard to interpret in clinical workflows. In comparison, the hybrid model does provide Grad-CAM visualization with an emphasis on clinically important areas, and this aspect enhances transparency. The model with a medium number of parameters 22M can be used on medium-scale GPUs or on cloud architectures, which makes it more accessible to the real-world dermatology environment.

Comparative Discussion with Literature

The findings of the hybrid model can contribute to the progress of multiclass skin lesion classification. The current framework recorded the same or better results with end-to-end deep learning compared to [5], who recorded 94.36% accuracy through transfer learning and ELM classification and were able to maintain interpretability. In the same vein, the model performed better than the ensemble provided by [6] which scored 97% on balanced subsets but used many deep backbones and took more time to run. In comparison with [7], which used triple-attention mechanism to enhance interpretability but had more complex models, the suggested separable attention offered the same efficiency by decreasing quadratic attention calculation to linear

complexity. The high trade-off between model depth, parameter count and inference latency is supported by such structural innovation.

Future Work

- Generalize by testing the model on additional datasets.
- Enhance adjustment to other clinic images and devices.
- Make the model optimized to make it run on small devices and mobile phones.
- More explainability tools such as SHAP or LIME should be used.
- use pictures alongside patient data to achieve improved diagnosis.
- Insert lesion segmentation automatically.
- Organize actual clinical trials among the dermatologists.
- Fairness, privacy, and bias Work on these to be safe in deployment.

References list

- 1) Ozdemir, B., & Pacal, I. (2025). A robust deep learning framework for multiclass skin cancer classification. *Scientific Reports*, 15, 4938. <https://doi.org/10.1038/s41598-025-89230-7>
- 2) Abohashish, S. M. (2025). Enhanced melanoma and non-melanoma skin cancer classification using a hybrid LSTM-CNN model. *Scientific Reports*, 15, 24994. <https://doi.org/10.1038/s41598-025-08954-8>
- 3) Tahir, M., Jin, H., Zafar, A., Mahmood, M. T., Ahmed, I., & Ahmad, H. F. (2023). DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing Skin Cancer Using Dermoscopic Images. *Cancers*, 15(7), 2179. <https://www.mdpi.com/2072-6694/15/7/2179>
- 4) Khan, M. A., Sharif, M., Akram, T., Damaševičius, R., & Maskeliūnas, R. (2021). Skin Lesion Segmentation and Multiclass Classification Using Deep Learning Features and Improved Moth Flame Optimization. *Diagnostics*, 11(5), 811. <https://www.mdpi.com/2075-4418/11/5/811>
- 5) Afza, F., Sharif, M., Khan, M. A., Kadry, S., & Damaševičius, R. (2022). Melanoma classification using deep learned features boosted with extreme learning machine. *Sensors*, 22(3), 799. <https://www.mdpi.com/1424-8220/22/3/799>
- 6) Thwin, S. M., & Park, H.-S. (2024). Skin Lesion Classification Using a Deep Ensemble Model. *Applied Sciences*, 14(13), 5599. <https://doi.org/10.3390/app14135599>
- 7) Efat, A. H., Hasan, S. M. M., Uddin, M. P., & Mamun, M. A. A. (2024). A multi-level ensemble approach for skin lesion classification using customized transfer learning with triple attention. *PLOS ONE*, 19(10), e0309430. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0309430>

- 8) Gomathi, S., & Arunachalam, N. (2024). Skin lesion prediction and classification using modified LSTM with a hybrid SC-STBO optimizer. *International Journal of Computational Intelligence Systems*, 17, 186. <https://link.springer.com/article/10.1007/s44196-024-00599-1>
- 9) Manzoor K, Gilal NU, Agus M, Schneider J. Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network. *Digit Health*. 2025 Jul 13;11:20552076251351858. <https://doi.org/10.1177/20552076251351858>
- 10) Halder, S., Roy, P., & Dey, N. (2025). *Fuzzy rank-based ensemble of CNN models for skin lesion classification*. *Scientific Reports*, 15, Article 90423. <https://doi.org/10.1038/s41598-025-90423-3>
- 11) Wang, Y., Wang, Y., Lee, T. K., Miao, C., & Wang, Z. J. (2022). SSD-KD: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images. *arXiv preprint arXiv:2203.11490*. <https://arxiv.org/abs/2203.11490>
- 12) Ahmad, N., Shah, J. H., Khan, M. A., Baili, J., Ansari, G. J., Tariq, U., Kim, Y. J., & Cha, J.-H. (2023). A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable AI. *Frontiers in Oncology*, 13, Article 1151257. <https://doi.org/10.3389/fonc.2023.1151257>
- 13) Alzakari, S. A., et al. (2024). LesionNet: An automated approach for skin lesion classification using deep pretrained CNNs and image super-resolution. *Frontiers in Medicine*. <https://doi.org/10.3389/fmed.2024.1487270>
- 14) Hoang, L., Lee, S.-H., Lee, E.-J., & Kwon, K.-R. (2022). Multiclass Skin Lesion Classification Using a Novel Lightweight Deep Learning Framework for Smart Healthcare. *Applied Sciences*, 12(5), Article 2677. <https://doi.org/10.3390/app12052677>
- 15) Mohamed, E. H., Abdu, N., Khalil, M., Kamal, H., & Rashed, E. A. (2025). MiSC: A hybrid multi-modal deep learning approach for accurate skin cancer detection. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-025-20951-7>
- 16) Farea, E., Saleh, R. A. A., AbuAlkebash, H., Farea, A. A. R., & Al-antari, M. A. (2024). *A hybrid deep learning skin cancer prediction framework*. *Journal of Engineering Science and Technology Review*, 101818. <https://doi.org/10.1016/j.jestch.2024.101818>
- 17) Alotaibi, A., & AlSaeed, D. (2025). Skin cancer detection using transfer learning and deep attention mechanisms. *Cancers*, 14(23), 5872.
- 18) Mavaddati, S. (2024). Skin cancer classification based on a hybrid deep model and long short-term memory. *Biosystems and Control*, 107109. <https://doi.org/10.1016/j.bspc.2024.107109>
- 19) Ravi, V. (2022). Attention cost-sensitive deep learning-based approach for skin cancer detection and classification. *Cancers*, 14(23), 5872. <https://doi.org/10.3390/cancers14235872>