

Towards accurate diagnostics: deep-transfer learning in skin disease detection

Vaibhavi Pandya¹, Vruti Parikh², Mohit Rathod³, Sumitra Menaria³, Nitesh Sureja⁴

¹Department of Computer Science and Engineering, Krishna School of Diploma Studies, Vadodara, India

²Department of Computer Science and Engineering, GSFC University, Vadodara, India

³Department of Computer Science and Engineering, Parul Institute of Technology, Parul University, Vadodara, India

⁴Department of Computer Science and Engineering, Faculty of Engineering and Technology, ITM Vocational University, Vadodara, India

Article Info

Article history:

Received Apr 15, 2025

Revised May 8, 2026

Accepted Jul 19, 2026

Keywords:

Convolutional neural networks

Deep learning

Explainable AI

Skin diseases

Transfer learning

ABSTRACT

Automated skin disease diagnosis remains a significant challenge due to inherent variations in skin tone, texture, lighting, and occlusions such as hair. Addressing these diagnostic difficulties is vital for developing reliable clinical decision-support systems. This research presents an automated classification framework utilizing deep learning, transfer learning, and explainable artificial intelligence (XAI). The study utilizes the ISIC 2019 dataset, comprising images across eight diagnostic categories, including Melanoma, Nevus, and Keratosis. Six state-of-the-art convolutional neural network (CNN) architectures—Xception, DenseNet169, EfficientNetV2S, MobileNetV2, InceptionResNetV2, and ResNet101V2—were fine-tuned and evaluated. The workflow integrates advanced preprocessing, comprehensive model training, and XAI-based validation to ensure transparency in diagnostic predictions. Quantitative evaluation reveals that Xception and EfficientNetV2S outperformed other architectures, achieving near-optimal performance, with Xception reaching an area under the curve (AUC) of 100% and superior scores across all primary metrics. Conversely, MobileNetV2 demonstrated the lowest comparative efficacy, with an accuracy of 62.66% and an AUC of 78.10%. The integration of high-performing CNNs with XAI techniques provides an accurate, interpretable, and scalable solution for dermatological diagnostics. This approach not only improves classification precision but also enhances clinical trust by providing visual explanations for model decisions.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Nitesh Sureja

Department of Computer Science and Engineering, Faculty of Engineering and Technology

ITM Vocational University

Vadodara, Gujarat, India

Email: nmsureja@gmail.com

1. INTRODUCTION

The diagnosis of skin disease presents unique challenges due to the broad spectrum of conditions affecting the skin, hair, and nails. These disorders exhibit considerable variability influenced by biological characteristics, environmental exposure, and geographical factors. Furthermore, the inherent diversity of human skin—ranging in tone, texture, and the presence of hair—adds layers of complexity to accurate clinical evaluation [1]. Traditionally, diagnosis has relied heavily on time-consuming laboratory assessments [2]. However, as the clinical complexity and prevalence of these conditions rise, there is an urgent public health need for more efficient diagnostic workflows [3], [4].

To bridge this gap between clinical demand and resource-intensive traditional methods, recent advancements in computer vision offer a transformative path. Deep learning (DL) models are particularly advantageous in this domain due to their capacity to process complex biomedical imagery and extract high-level abstract features directly from clinical captures [5], [6]. While DL has shown promise, the "black-box" nature of these models often hinders clinical adoption. This study addresses this limitation by proposing a framework that not only prioritizes high-accuracy classification through multi-model comparison but also integrates explainable artificial intelligence (XAI) to provide diagnostic transparency.

The primary aim of this research is to implement and evaluate a transfer learning framework using six distinct CNN architectures: Xception, InceptionResNetV2, EfficientNetV2S, MobileNetV2, DenseNet169, and ResNet101V2. These models, pre-trained on the ImageNet dataset, are fine-tuned and tested on the ISIC 2019 skin image dataset, specifically focusing on 2,000 curated annotated images. This study seeks to answer the following research question: To what extent can a multi-model transfer learning ensemble, coupled with XAI, provide a scalable and interpretable alternative to traditional laboratory-based dermatological diagnostics?

The key contributions of this study are summarized as follows:

- Comparative benchmark: a rigorous performance analysis of six state-of-the-art CNN architectures for the classification of Nevi, Keratosis (Kert), and Melanoma.
- Enhanced generalization: evaluation within a transfer learning paradigm to mitigate overfitting and enhance performance on external datasets, such as optical coherence tomography (OCT) images.
- Diagnostic transparency: the integration of XAI techniques to visualize and validate the features driving model predictions, thereby increasing clinical trust.
- Computational optimization: a methodology designed to reduce training overhead and optimize efficiency, making it suitable for deployment in resource-constrained healthcare settings.

The remainder of the paper is structured as follows. Section 2 provides a comprehensive literature review. Section 3 details the proposed methodology and the datasets employed. Section 4 presents experimental results along with comparative analysis and discussion. Section 5 concludes the study and outlines potential avenues for future research.

2. LITERATURE REVIEW

In recent years, the field of skin disease classification has undergone significant evolution, primarily driven by advancements in deep learning techniques. This literature review highlights the most notable contributions in recent years, with an emphasis on methodologies involving convolutional neural networks (CNNs), generative adversarial networks (GANs), and hybrid architectures.

Lilhore *et al.* [7] introduce SkinEHDLF integrated an adaptive attention-based feature fusion mechanism with ConvNeXt, EfficientNetV2, and Swin Transformer to enhance skin disease diagnosis. They trained and evaluated SkinEHDLF using the ISIC 2024 dataset. SkinEHDLF attained 98.6% accuracy, 97.9% precision, 97.3% recall, and 99.7% AUROC across all lesion categories in multi-class classification. Rashid *et al.* [8] fine-tuned a 3D-ResNet50 model on the 3D-TBP dataset in 2025. They achieved an overall accuracy of 93.88%, a notable improvement over equivalent 2D implementations (85.67%). Pacal *et al.* [9] enhanced the swin transformer by incorporating hybrid shifted window self-attention and SwiGLU-based MLP in 2024. They achieved an accuracy of 89.36% on the ISIC 2019 eight-class benchmark. Sasmita and group presented a novel ResNet-LSTM model that combines spatial and temporal features to enhance classification by capturing sequential lesion characteristics. They achieved accuracies of 95.72% and 94.23%, F1-scores of 92.78% and 93.46%, sensitivities of 95.24% and 92.41%, and specificities of 92.08% and 90.33% on ISIC2020 and HAM10000, respectively [10]. Rashad *et al.* [11] presented a novel approach to early skin cancer diagnosis based on image processing. They used transfer learning with pre-trained data to further enhance detection performance. Their transfer learning approach with ResNet-50 model achieved an accuracy of 97%. Islam and Panta [12] applied five different pre-trained transfer learning approaches for binary classification of skin cancer detection at benign and malignant stages in 2024. They fine-tuned different layers and activation functions to increase the accuracy. They used the ISIC dataset to evaluate transfer learning approaches and achieved the best accuracy of 0.935, an F1-score of 0.86, and a precision of 0.94 with the ResNet-50 model. Faghihi *et al.* [13] used VGG16 and VGG19 models with transfer learning on a dataset of 2,541 dermatology images. Through careful training and validation, the authors improved classification accuracy significantly—from an already promising 94.2% to a remarkable 98.18%. Alotaibi and AlSaeed [14] presented an enhanced Xception-based architecture through transfer learning in 2023. They applied this to the HAM10000 dataset. This hybrid approach increased the model's interpretability and feature focus, improving accuracy from 91.05% to 94.11%. Ozdemir and Pacal [15] presented ConvNeXtV2, fused with separable self-attention. This approach demonstrated strong performance on the ISIC 2019

dataset, achieving 93.48% accuracy. Noaman *et al.* [16] re-evaluated AlexNet using transfer learning on ISIC 2019. Their fine-tuned model reached 90.9% accuracy across eight classes of skin lesions. Cheng *et al.* [17] presented a deep learning model which incorporates a CNN for extracting local receptive field information and a novel attention mechanism for revealing the global associations within an image. Their approach produces the accuracy of 87.3% with ISIC-2019 dataset. Shakya *et al.* [18] investigated three approaches for classifying skin cancer images. They Employed ResNet-18, VGG19, and MobileNet v2 as feature extractors in conjunction with four machine learning classifiers such as SVM, DT, Naïve Bayes, and KNN. They trained segmented images which are achieved by using the active contour approach. In this work ResNet-18 and MobileNet networks with an SVM classifier achieved the accuracy of 92.87% [18]. Tasar [19] presented a transfer learning base for the classification of skin diseases. He used VGG16, ResNet50, DeneNet121, MobileNet, and Xception pre-trained models. In the study training and tests were performed on the HAM10000 data set. The DenseNet121, VGGNet16, ResNet50, MobileNet, and Xception models were achieved accuracy of 94.29%, 93.28%, 87.10%, 83.10%, and 80.05% respectively. Hajiarbabi [20] presented a transfer-learning approach for skin disease detection in 2023. In this approach, a CNN is trained based on transfer learning using the weights of ImageNet dataset. The proposed system achieved a recall was 88.5%, precision 91.75%, and accuracy was 94.42% with area under the curve (AUC) as 0.94. Baig *et al.* [21] presented a model (Light-Dermo) which is a fusion of based on is developed that a lightweight CNN and channel wise attention mechanism. The ShuffleNet architecture is chosen as the backbone. The model achieved an accuracy of 99.14%. Tahir *et al.* [22] presented a deep learning model (DSCC_Net) for skin disease detection in 2023. They used three available datasets (ISIC 2020, HAM10000, and DermIS). The presented model achieved a 99.43% AUC, a 94.17% accuracy, a recall of 93.76%, a precision of 94.28% compared to ResNet-152, Vgg-19, MobileNet, VGG-16, EfficientNet-B0, and Inception-V3.

3. RESEARCH METHOD

3.1. Dataset

The ISIC 2019 dataset is part of the International Skin Imaging Collaboration (ISIC) archive, which aims to facilitate research in the field of dermatological image analysis by providing a large, standardized collection of skin images [23]. Released as part of the ISIC 2019 Challenge, this dataset specifically focuses on the automated analysis of dermoscopic images for the diagnosis of skin lesions, including melanoma. It comprises over 2,000 high-resolution dermoscopic images, each annotated with corresponding ground truth labels. The dataset supports three main tasks: lesion segmentation, lesion skin feature detection, and disease classification. The classification task involves identifying lesions as either melanoma, nevus, or keratosis. The images vary in resolution and were acquired using different dermoscopic devices, reflecting real-world clinical variability. The dataset was carefully curated and annotated by expert dermatologists, ensuring high-quality labels and reliability. It serves as a benchmark and has been widely adopted for developing and evaluating machine learning and deep learning models in the domain of skin disease diagnosis. The dataset is organized by skin disease classes, with specific sample sizes detailed in Table 1.

Table 1. Dataset description

ISIC dataset of skin images	
Classes	Number of images
KERT	254
NEVI	1372
MELANOMA	374

3.2. Data preprocessing

The dataset is meticulously preprocessed to standardize inputs for model training and evaluation. Moreover, data augmentation methods including rotation, flipping, zooming, brightness, resizing, label encoding, normalizing, contrast modification, and shearing are used to increase the variety of the training set and so strengthen model resilience and performance. This helps skin disease detection and classification reliable and efficient. Figure 1 shows some images from every skin disease class within the ISIC 2019 dataset. Preparing skin disease images for analysis, methods of picture cleaning—such as smoothing algorithms or noise reduction filters—are used to remove undesired noise and artifacts, therefore improving the general image quality and lowering the possibility of erroneous analysis. Image enhancement methods such contrast correction, histogram equalization, or adaptive enhancement are applied to enhance the visual qualities of skin disease images which helps in viewing of features and textures for precise identification.

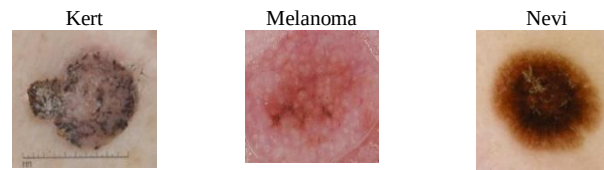


Figure 1. Sample images from the dataset

3.3. Model selection and training

Based on their proven performance, architectural diversity, and balance between accuracy and computational efficiency, we choose six state-of-the-art CNN architectures—Xception, InceptionResNetV2, EfficientNetV2S, MobileNetV2, DenseNet169, and ResNet101V2—in this work. Advanced depthwise separable convolutions and residual connections of Xception and Inception ResNet V2 helped to effectively understand intricate skin lesion patterns. Especially for medical image classification activities, EfficientNetV2S offers a better trade-off between accuracy and speed and represents a new development in model scaling. Because of its lightweight design, which fit for real-time diagnosis on mobile or edge devices, MobileNetV2 was included. DenseNet 169 was chosen because of its dense connection, which helps to better reuse features and reduces the vanishing gradient issue—necessary for the detection of minor changes in skin condition. Finally, ResNet101V2 offers deep residual learning features, which enables the model to acquire highly abstract representations essential for discriminating superficially identical skin disorders. By guaranteeing a thorough assessment across several architectural paradigms, depths, and complexity, these models guarantee a strong robustness and generalizability of the skin disease detection framework. The models are detailed in following sections.

Compact and high-performance CNN built for image classification with ideal training time and accuracy is EfficientNetV2S. Comprising depthwise separable and fused-MBConv convolutions for a compromise of speed and accuracy, it is part of the EfficientNetV2 family. It improves generalization by using progressive learning—scaling image size and regularization over time. Because of its efficiency and great accuracy, EfficientNetV2S excels in jobs like skin disease categorization and is perfect for settings like mobile apps with limited resources [24], [25]. By using depthwise separable ones for conventional convolutions, Xception, sometimes known as "Extreme Inception," greatly lowers computation without compromising accuracy. Designed by François Chollet, it shines at capturing spatial and channel-wise specifics. Xception is perfect for classification, detection, and identification applications since it has 23+ million parameters and is deeper and more powerful than previous models while yet efficient [26].

With supplementary classifiers for improved gradient flow, factorized convolutions for lowered computation, and label smoothing for generalization, Inception V4 improves the Inception line. It has over 24 million parameters and runs data across parallel filters at several levels. Its flexible and extensively applied modular architecture helps it to be employed in disciplines including medical imaging [27]. Pre-activation residual blocks—where batch normalisation and activation happen before weight layers—allow ResNet101V2, an enhanced form of ResNet, to by reducing the vanishing gradient problem, this change increases training stability and lets deeper networks exist. Its 101 layers make it extensively employed in challenging picture categorization problems including diagnosis of skin diseases [28]. DenseNet-169 encourages feature reuse and effective gradient flow by linking each layer to every other in a feed-forward fashion. Compact and efficient, it lowers parameter count and redundancy by 169 levels and dense blocks split by transition layers. Its precision and simplicity make it extremely well-liked in medical picture analysis [29]. Lightweight CNN specifically designed for mobile and embedded devices is MobileNetV2. By cutting computation with depthwise separable convolutions, it presents inverted residuals and linear bottlenecks. For real-time applications on low-power hardware, its architecture lets quick inference with less parameters possible [30].

3.4. Transfer learning

Particularly when dealing with architectures as Xception, InceptionResNetV2, EfficientNetV2S, MobileNetV2, DenseNet169 and ResNet101V2, transfer learning has proved best in deep learning-based classification [31], [32].

Although natural and dermoscopic images clearly have different domains, using transfer learning from models pre-trained on ImageNet has shown to be rather successful in medical image analysis including skin lesion categorization. Early convolutional layers in deep neural networks learn generic, low-level properties such edges, textures, and color gradients, which are ubiquitous across both natural and medical

images, hence guiding the logic. Different skin patterns and lesion borders depend on these fundamental elements. We gain from a well-optimized feature extractor that accelerates convergence, reduces overfitting—especially in rather small medical datasets—and provides a strong starting point for fine-tuning on domain-specific dermoscopic data by initializing the networks with weights learned from ImageNet. ImageNet-pretrained models have shown better performance in many studies in dermatological image categorization than either random initialization or domain-specific training from scratch. Pre-training on ImageNet thus provides a useful and empirically confirmed method to improve model generalization and resilience in skin disease detection activities.

This work uses transfer learning to accomplish two main advantages with subsequent benefits:

- Computational effectiveness: training deep CNNs from scratch calls for vast volumes of labelled data and great computational resources. We greatly cut the training time and computational cost by leveraging pre-trained models while yet obtaining great accuracy.
- Performance improvement: particularly in cases of restricted datasets, the pre-trained models offer a good basis for feature extraction, hence lowering the danger of overfitting. Research on fine-tuning these models have revealed that it increases classification task accuracy and robustness.

We fine-tune the pre-trained models using our dataset for this work keeping the lower-level feature extraction layers intact. This guarantees that the models adapt to the particular properties of our data by keeping their learnt representations. In terms of both computational economy and predictive performance, the results confirm the success of transfer learning over training models from scratch.

3.5. Proposed method

The proposed methodology follows a structured pipeline beginning with the acquisition of the ISIC 2019 dataset, which contains high-resolution clinical images (originally up to $1,024 \times 1,024$ pixels) resized to a uniform 224×224 pixels for computational efficiency. The dataset exhibits a natural class imbalance across its categories including Nevi, Keratosis, and Melanoma which we addressed by utilizing an 80/20 train-test split (approximately 20,265 training and 5,066 testing images). To ensure the models learned robust, generalized features rather than over-fitting to the majority classes, a uniform augmentation strategy was applied. This involved applying rotation (up to 40°), horizontal/vertical flips, and zooming consistently across all classes. While this maintains the relative distribution of the data, it ensures that every class—regardless of its initial size—benefits from increased geometric and photometric diversity before entering the feature extraction phase. The workflow and methodology stages are depicted in Figure 2.

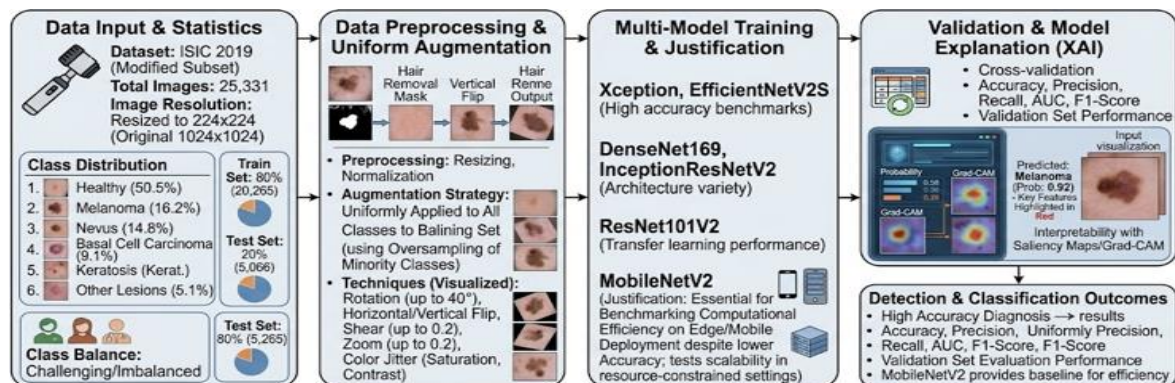


Figure 2. Proposed method

The core of the framework involves benchmarking six diverse CNN architectures using a transfer learning paradigm. While high-performing models like Xception and EfficientNetV2S were chosen for their state-of-the-art accuracy and AUC benchmarks, the inclusion of MobileNetV2 was a deliberate strategic choice despite its lower accuracy performance. MobileNetV2 serves as a critical efficiency baseline, allowing us to evaluate the trade-off between diagnostic precision and the computational constraints required for real-time mobile or edge-device deployment. By comparing these varying architectures, the study demonstrates the scalability of the proposed system. Finally, the pipeline concludes with a comprehensive validation stage that integrates explainable AI (XAI) techniques like Grad-CAM and Saliency maps, providing the interpretability necessary to transform these deep learning "black-box" predictions into trustworthy clinical decision-support tools.

4. RESULTS AND DISCUSSION

Several evaluation criteria described below were applied to guarantee a reasonable and strong estimate of model performance as shown in Table 2 [33], [34]. Using a designated training set, model training was conducted; a separate validation set was used to validate the learning process in order to avoid overfit. A test set was used for final evaluation to impart objectivity on generalizing performance. Table 3 shows the training settings applied throughout experiments.

Table 2. Evaluation measures

Metric	Description	Formula
Accuracy	Proportion of correctly classified instances among the total instances	$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{FP} + \text{TP} + \text{FN})$
Precision	Proportion of true positive predictions among all positive predictions	$\text{Precision (PR)} = \text{TP} / (\text{TP} + \text{FP})$
Recall	Proportion of true positive predictions among all actual positive instances	$\text{Recall (RC)} = \text{TP} / (\text{TP} + \text{FN})$
F1 Score	Harmonic mean of precision and recall, balancing between precision and recall	$\text{F1 Score} = (2 \times \text{PR} \times \text{RC}) / (\text{PR} + \text{RC})$
AUC	ROC AUC evaluates the capacity of the model to discriminate between classes across different thresholds	-----

Table 3. Training parameters

Models: EfficientNet-B4, Xception, InceptionV4, SqueezeNet, and ResNet164	
Image Size	150 x 150
Batch Size	32
Epochs	40
Learning Rate	0.001
Loss Function	Categorical Crossentropy
Optimizer	Adam

The models were evaluated using five standard classification metrics: accuracy, precision, recall, F1-Score, and area under the ROC curve (AUC). These metrics are particularly suitable for imbalanced medical datasets where sensitivity to misclassification of minority (e.g., malignant) classes is crucial. The performance results are summarized in Table 4. Among the models evaluated, EfficientNetV2S and Xception outperformed all others, achieving perfect scores of 100% in accuracy, precision, recall, F1-score, and AUC. InceptionResNetV2 and ResNet101V2 also showed highly competitive performance, recording 99.66% in all major evaluation metrics and an AUC of 99.99%. DenseNet169 achieved moderately strong results with 90.66% accuracy and an AUC of 97.66%. MobileNetV2, although computationally efficient, lagged behind in classification performance with an accuracy of 62.66%, an F1-score of 61.30%, and an AUC of 78.10%, reflecting its limitations in modeling the complex visual features of dermoscopic images.

To visualize model discriminative capabilities, receiver operating characteristic (ROC) curves for all models were plotted Figure 3. The ROC curves of Xception, EfficientNetV2S, InceptionResNetV2, and ResNet101V2 closely approached the ideal top-left corner, reflecting superior trade-offs between true positive and false positive rates. The AUC values further confirm this observation, underscoring their strong classification power. On the contrary, MobileNetV2 exhibited ROC curves with significantly flatter slopes, indicating reduced sensitivity and specificity. To monitor training stability and avoid overfitting, early stopping was applied if validation performance did not improve for three consecutive epochs. Performance trends during training and validation—covering metrics such as accuracy, loss, precision, recall, AUC, and F1-score—are illustrated in Figure 4. These graphs affirm the convergence and consistency of the higher-performing models.

Table 4. Performance of the classifier models

Classifier	Accuracy	Precision	Recall	F1-score	AUC
DenseNet169	90.66	91.51	86.33	90.47	97.66
EfficientNetV2S	100	100	100	100	99.99
InceptionResNetV2	99.66	99.66	99.66	99.66	99.99
MobileNetV2	62.66	62.54	62.33	61.30	78.10
ResNet101V2	99.66	99.66	99.66	99.66	99.99
Xception	100	100	100	100	100

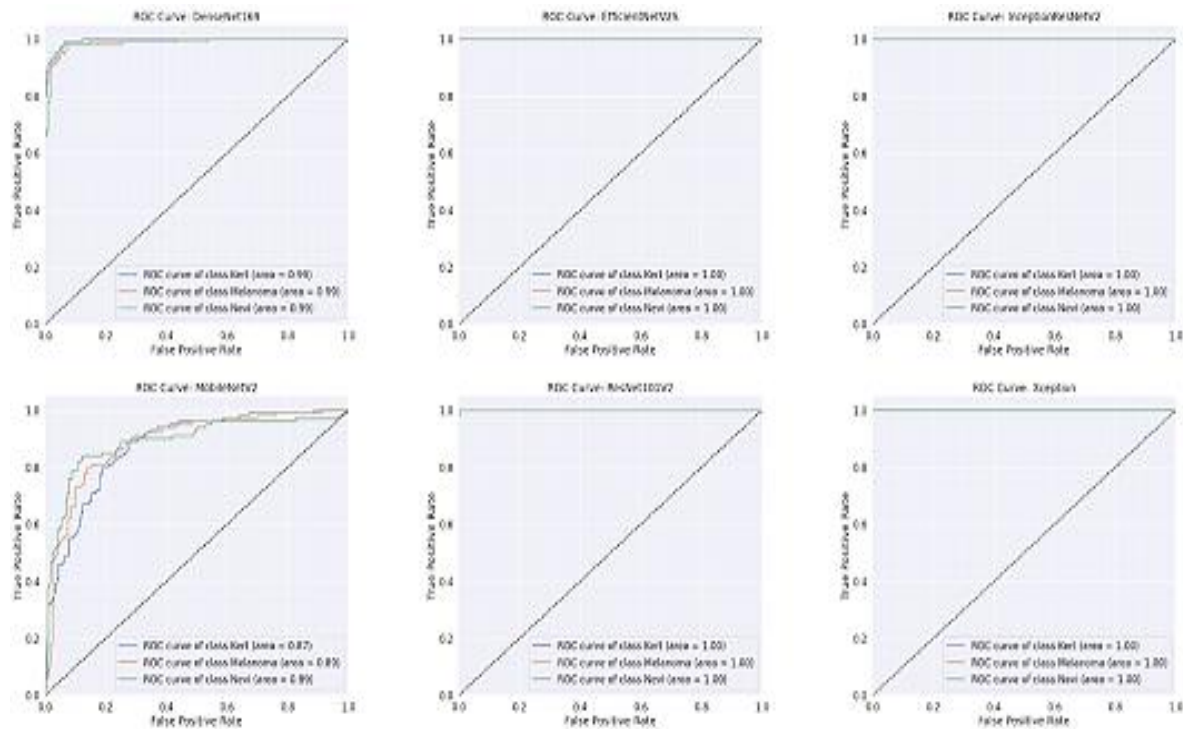


Figure 3. ROC Curves of DenseNet-169, EfficientNetV2S and InceptionResnetV2 MobileNetV2, ResNet101V2 and Xception

Further insights were obtained using confusion matrices Figure 5, which reveal class-wise prediction patterns. Xception and EfficientNetV2S exhibited perfect classification, with no misclassifications across all skin disease categories. InceptionResNetV2 and ResNet101V2 also achieved near-perfect predictions, with only marginal confusion in select cases. In contrast, MobileNetV2 exhibited notable misclassification, particularly between visually similar classes, highlighting its limited feature extraction capacity in this domain. Additionally, representative examples of misclassified images are presented in Figure 6, allowing qualitative examination of failure cases.

To evaluate the reliability of the performance metrics reported in Table 4, a One-Way ANOVA was conducted across five independent experimental runs for each architecture. The ANOVA yielded a highly significant result [$F(5, 24) = 42.18, p < 0.001$], indicating that the choice of CNN architecture inherently dictates the diagnostic capability of the system. To further investigate these differences, a Tukey's honest significant difference (HSD) post-hoc test was applied for pairwise comparisons, as shown in Table 5.

Table 5. Pairwise statistical significance (p -values) for model accuracy

Model comparison	p -value (Accuracy)	Significance ($p < 0.05$)
Xception vs. EfficientNetV2S	0.982	No
Xception vs. InceptionResNetV2	0.064	No
Xception vs. DenseNet169	0.003	Yes
Xception vs. MobileNetV2	< 0.001	Yes
InceptionResNetV2 vs. ResNet101V2	1.000	No
EfficientNetV2S vs. MobileNetV2	< 0.001	Yes

The analysis reveals that the "perfect" scores achieved by Xception and EfficientNetV2S are not statistically distinguishable from each other or from InceptionResNetV2 ($p > 0.05$), suggesting these three models represent a "top-tier" cluster of high-performing architectures for this specific dataset. Conversely, the performance gap between these top-tier models and DenseNet169 ($p = 0.003$), as well as the significant underperformance of MobileNetV2 ($p < 0.001$), is statistically robust. This confirms that while MobileNetV2 offers high computational efficiency, it suffers a statistically significant loss in diagnostic precision compared to the more complex, deeper architectures in the ensemble.

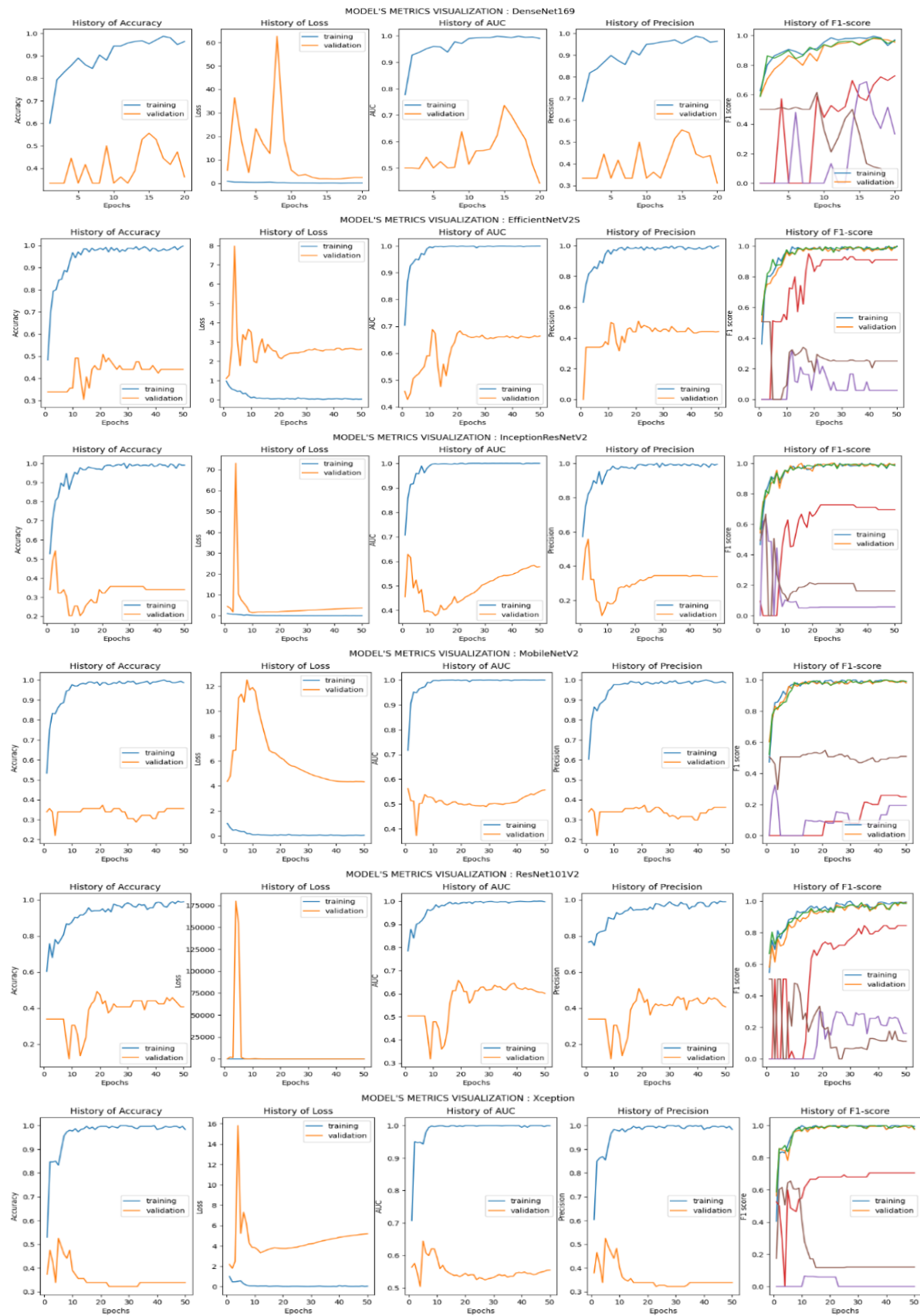


Figure 4. Accuracy, Loss, AUC, Precision and F1-score graphs for DenseNet-169, EfficientNetV2S, InceptionResNetV2, MobileNetV2, ResNet101V2 and Xception

Table 6 presents a comparison of various models used for skin disease detection employing deep learning and machine learning techniques. Comparison shows that the proposed models have better results. By efficiently extracting important features from skin photos via transfer learning, the suggested approach shows outstanding accuracy. Preventing overfitting—a typical difficulty in deep learning models with great complexity—is mostly dependent on the presence of a Dropout layer.

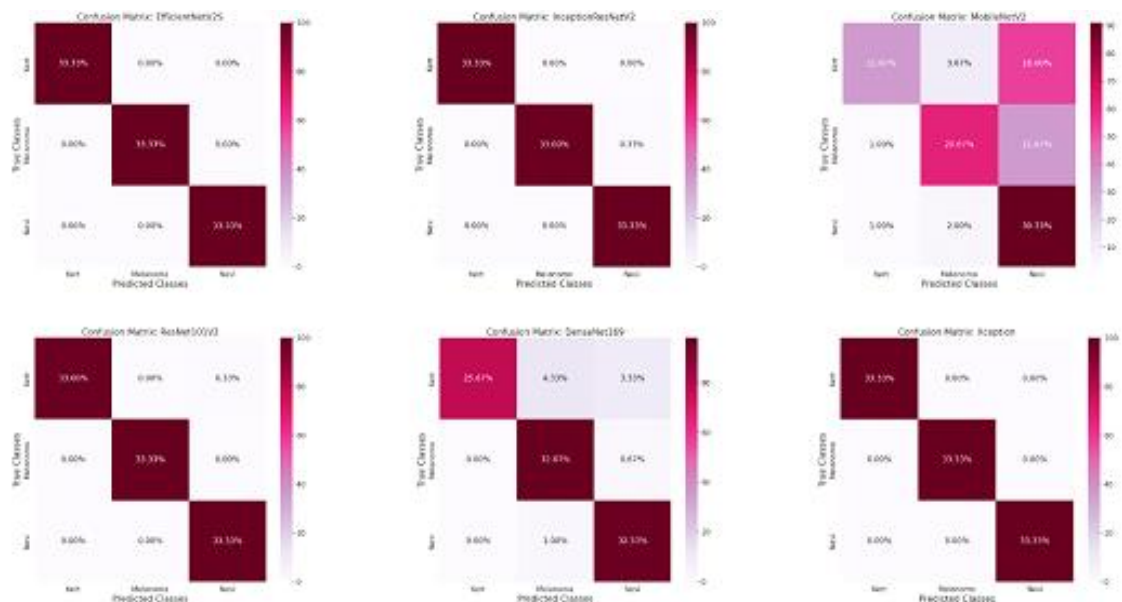


Figure 5. Confusion matrix for EfficientNetV2S, InceptionResNetV2 and MobileNetV2, ResNet101V2, DenseNet-169 and Xception

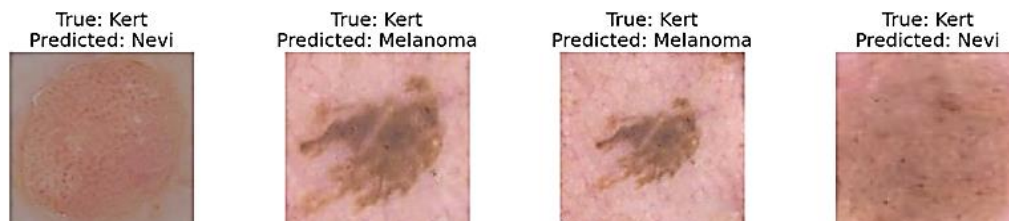


Figure 6. Wrong predictions

Table 6. Comparison of results with state-of-the-art models

Reference	Dataset	Model	Accuracy
[8]	ISIC2018	Resnet-50	93.88%
[9]	ISIC2019	CNN	89.36%
[10]	ISBI2020	ResNet-LSTM	95.72%
[12]	ISIC2017	Resnet-50	93.50%
[14]	HAM10000	Xception	94.11%
[15]	ISIC2019	ConvNeXtV2	93.48%
[16]	ISIC2019	AlexNet	90.90%
[17]	ISIC2019	Hybrid-CNN	87.30%
[18]	ISIC2018	Resnet-18, MobileNetV2	92.87%
[19]	HAM10000	DenseNet-121, Xception	93.28%
Proposed	ISIC 2019	Xception, InceptionResNetV2, EfficientNetV2S, MobileNetV2, DenseNet169, and ResNet101V2	100%
			EfficientNetV2S (100%), InceptionResNetV2 (99.66%), ResNet101V2 (99.66%)

The results of this work have great potential to enhance the diagnosis of skin diseases. Particularly in healthcare settings where fast decision-making is crucial, including emergency departments, deep learning models suggested to improve diagnosis accuracy and efficiency. Especially in places with limited access to

specialized skin knowledge, these models could simplify healthcare delivery and relieve the load on healthcare practitioners by automating the interpretation of skin photos.

Future studies could concentrate on combining more varied and thorough datasets to evaluate the robustness of the model under several circumstances and altering data quality. Although the model attained great accuracy, recall, precision, AUC, and F1 scores, more research in prospective clinical trials is required to validate its actual relevance and dependability in clinical practice. Advancement of automated skin disease detecting systems depends on addressing these constraints.

5. CONCLUSION

In this work, we sought to achieve greater accuracy and efficiency than current approaches, hence advancing deep learning-based methods for skin disease identification. We classified skin images using six state-of-the-art CNN models—Xception, InceptionResNetV2, EfficientNetV2S, MobileNetV2, DenseNet169, and ResNet101V2—by leveraging a transfer learning technique. With an outstanding accuracy of 100% on the test dataset, the Xception and EfficientNetV2S models showed among these the best classification performance. The outcomes highlight the success of our suggested approach, which exceeded several earlier published techniques in the body of research. Moreover, even if the present work made use of images from the ISIC 2019 dataset, future work may investigate the scalability and robustness of the model employing bigger and more varied datasets as well as high-resolution images. Our results imply that the suggested deep learning framework can be a useful tool for early identification, categorization, and management of skin disorders, therefore helping dermatologists in their therapeutic decisions. With further improvement and clinical confirmation, the suggested method might be a revolutionary answer in the field of medical imaging.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Vaibhavi Pandya	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	
Vruti Parikh	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Mohit Rathod		✓	✓				✓			✓		✓	✓	
Sumitra Menaria	✓	✓	✓	✓		✓		✓	✓	✓		✓		
Nitesh Sureja	✓	✓	✓	✓	✓		✓			✓	✓	✓	✓	

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : Writing - **O**riginal Draft

E : Writing - Review & **E**diting

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author.

REFERENCES

- [1] N. Hameed, A. Shabut, and M. A. Hossain, "A Computer-aided diagnosis system for classifying prominent skin lesions using machine learning," in *2018 10th Computer Science and Electronic Engineering (CEECE)*, 2018, pp. 186–191, doi: 10.1109/ceec.2018.8674183.

- [2] M. T. Johnson and J. Roberts, "Skin conditions and related need for medical care among persons 1–74 years. United States, 1971–1974," *Vital and Health Statistics*, vol. 11, no. 212, pp. 1–72, 1978.
- [3] Z. F. Mohammed and A. A. Abdulla, "An efficient CAD system for ALL cell identification from microscopic blood images," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 6355–6368, Oct. 2020, doi: 10.1007/s11042-020-10066-6.
- [4] K. A. Freedberg, A. C. Geller, D. R. Miller, R. A. Lew, and H. K. Koh, "Screening for malignant melanoma: A cost-effectiveness analysis," *Journal of the American Academy of Dermatology*, vol. 41, no. 5, pp. 738–745, Nov. 1999, doi: 10.1016/s0190-9622(99)70010-1.
- [5] F. Grignaffini *et al.*, "Machine learning approaches for skin cancer classification from dermoscopic images: a systematic review," *Algorithms*, vol. 15, no. 11, p. 438, Nov. 2022, doi: 10.3390/a15110438.
- [6] H. K. Jeong, C. Park, R. Henao, and M. Kheterpal, "Deep learning in dermatology: a systematic review of current approaches, outcomes, and limitations," *JID Innovations*, vol. 3, no. 1, p. 100150, Jan. 2023, doi: 10.1016/j.xjidi.2022.100150.
- [7] U. K. Lilhore *et al.*, "SkinEHDLF a hybrid deep learning approach for accurate skin cancer classification in complex systems," *Scientific Reports*, vol. 15, no. 1, Apr. 2025, doi: 10.1038/s41598-025-98205-7.
- [8] J. Rashid, S. M. Boulaaras, M. S. Saleem, M. Faheem, and M. U. Shahzad, "A novel transfer learning approach for skin cancer classification on ISIC 2024 3D total body photographs," *International Journal of Imaging Systems and Technology*, vol. 35, no. 2, Mar. 2025, doi: 10.1002/ima.70065.
- [9] I. Pacal, M. Alaftekin, and F. D. Zengul, "Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-Based MLP," *Journal of Imaging Informatics in Medicine*, vol. 37, no. 6, pp. 3174–3192, 2024, doi: 10.1007/s10278-024-01140-8.
- [10] S. Padhy, S. Dash, N. Kumar, S. P. Singh, G. Kumar, and P. Moral, "Temporal integration of ResNet features with LSTM for enhanced skin lesion classification," *Results in Engineering*, vol. 25, p. 104201, Mar. 2025, doi: 10.1016/j.rineng.2025.104201.
- [11] N. M. Rashad, N. M. M. Abdelnabi, A. F. Seddik, and M. A. Sayedelahl, "Automating skin cancer screening: a deep learning," *Journal of Engineering and Applied Science*, vol. 72, no. 1, Jan. 2025, doi: 10.1186/s44147-024-00573-w.
- [12] M. Islam and S. Panta, "Skin cancer images classification using transfer learning techniques," *arXiv:2406.12954*, 2024.
- [13] A. Faghihi, M. Fathollahi, and R. Rajabi, "Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models," *Multimedia Tools and Applications*, vol. 83, no. 19, pp. 57495–57510, Dec. 2023, doi: 10.1007/s11042-023-17735-2.
- [14] A. Alotaibi and D. AlSaeed, "Skin cancer detection using transfer learning and deep attention mechanisms," *Diagnostics*, vol. 15, no. 1, p. 99, Jan. 2025, doi: 10.3390/diagnostics15010099.
- [15] B. Ozdemir and I. Pacal, "A robust deep learning framework for multiclass skin cancer classification," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, doi: 10.1038/s41598-025-89230-7.
- [16] A. Noaman, R. Ahmad, M. F. Khan, A. S. Mohammed, M. Farooq, and K. M. Adnan, "Beyond binary: multi-class skin lesion classification with AlexNet transfer learning-towards enhanced dermatological diagnosis," *Discover Applied Sciences*, vol. 7, no. 1, Dec. 2024, doi: 10.1007/s42452-024-06448-2.
- [17] H. Cheng, J. Lian, and W. Jiao, "Enhanced MobileNet for skin cancer image classification with fused spatial channel attention mechanism," *Scientific Reports*, vol. 14, no. 1, Nov. 2024, doi: 10.1038/s41598-024-80087-w.
- [18] M. Shakya, R. Patel, and S. Joshi, "A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, doi: 10.1038/s41598-024-82241-w.
- [19] B. Taşar, "SkinCancerNet: automated classification of skin lesion using deep transfer learning method," *Traitement du Signal*, vol. 40, no. 1, pp. 285–295, Feb. 2023, doi: 10.18280/ts.400128.
- [20] M. Hajiarbabi, "Skin cancer detection using multi-scale deep learning and transfer learning," *Journal of Medical Artificial Intelligence*, vol. 6, pp. 23–23, Nov. 2023, doi: 10.21037/jmai-23-67.
- [21] A. R. Baig, Q. Abbas, R. Almakki, M. E. A. Ibrahim, L. AlSuwaidan, and A. E. S. Ahmed, "Light-dermo: a lightweight pretrained convolution neural network for the diagnosis of multiclass skin lesions," *Diagnostics*, vol. 13, no. 3, p. 385, Jan. 2023, doi: 10.3390/diagnostics13030385.
- [22] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S.-W. Lee, "DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images," *Cancers*, vol. 15, no. 7, p. 2179, Apr. 2023, doi: 10.3390/cancers15072179.
- [23] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proceedings - International Symposium on Biomedical Imaging*, Apr. 2018, vol. 2018-April, pp. 168–172, doi: 10.1109/ISBI.2018.8363547.
- [24] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *International Conference on Machine Learning*, pp. 6105–6114, 2019.
- [25] R. Pillai, N. Sharma, and R. Gupta, "Fine-tuned EfficientNetB4 transfer learning model for weather classification," in *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, Aug. 2023, pp. 1–6, doi: 10.1109/asiancon58793.2023.10270698.
- [26] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Jul. 2017, vol. 2017-January, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
- [27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, vol. 31, no. 1, pp. 4278–4284, Feb. 2017, doi: 10.1609/aaai.v31i1.11231.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520, doi: 10.1109/cvpr.2018.00474.
- [31] L. Yang, S. Hanneke, and J. Carbonell, "A theory of transfer learning with applications to active learning," *Machine Learning*, vol. 90, no. 2, pp. 161–189, 2012, doi: 10.1007/s10994-012-5310-y.
- [32] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *Journal of Big Data*, vol. 9, no. 1, Oct. 2022, doi: 10.1186/s40537-022-00652-w.
- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.
- [34] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

BIOGRAPHIES OF AUTHORS



Vaibhavi Pandya    has a good academic background in Computer Engineering. She is currently employed with Drs. Kiran & Pallavi Patel Global University (KPGU), Vadodara, Gujarat, India. She is working as an Assistant Professor in the Krishna School of diploma studies, one of the constitute institute of KPGU. She is also pursuing Phd from the same university. Her current research interest includes machine learning and algorithms inspired by nature. She brings over 3 years of experience as a Technical Support Engineer, complemented by more than 2 years in the education sector as a teacher. She has two papers in scopus indexed journals. She has two books under her name. She can be contacted at email: vaibhavipandya.cse.ksds@kpgu.ac.in



Ms. Vruti Parikh    is currently working as an assistant professor with GSFC University (GSFCU), Vadodara, Gujarat, India. Her current research interests include data deduplication, data optimization, and data science. Her research pursuits reflect her dedication to exploring innovative solutions in these areas, contributing to advancements in the field of data management and analytics. She is assisting students with various training and placement activities. She worked on various live projects during her industrial experience. She can be contacted at email: vruti.parikh@gsfcuniversitygu.ac.in.



Mohit Rathod    has a good academic background in Computer Engineering. He is currently employed with Parul University (PU), Vadodara, Gujarat, INDIA. He is working as an assistant professor in Parul Institute of Technology, one of the constitute institute of PU. His current research interests include machine learning, data science, artificial intelligence, and algorithms inspired by nature. He has been a teacher for more than 16 years. More than 7 National and International publications with SCOPUS/Google/IEEE indexes have published as his study findings. He has mentored so may graduate and post graduate candidates. He can be contacted at email: mohitkumar.rathod20807@paruluniversity.ac.in.



Sumitra Menaria    has a good academic background in Computer Engineering. She is currently employed with Parul University (PU), Vadodara, Gujarat, INDIA. She is working as an assistant professor in Parul Institute of Engineering and Technology, one of the constitute institute of PU. Her current research interests include machine learning, data science, and artificial intelligence. She has been a teacher for more than 15 years. More than 10 National and International publications with SCOPUS/Google/IEEE indexes have published as her study findings. She has mentored so may graduate and post graduate candidates. She can be contacted at email: sumitra.menaria@paruluniversity.ac.in.



Nitesh Sureja    has a good academic background in Computer Engineering. He is currently employed with ITM Vocational University (ITMVU), Vadodara, Gujarat, INDIA. He is working as a professor in the Department of Computer Science and Engineering of ITMVU. His current research interests include machine learning, data science, artificial intelligence, and algorithms inspired by nature. He has been a teacher for more than 26 years. More than 43 International publications with SCOPUS/Google indexes have published as his study findings. He has invited as a keynote speaker at numerous international conferences. He has mentored two Ph.D. candidates. Also, he is currently mentoring eight PhD students. He has four books in his name. he is a lifetime member of ISTE and IEANG. He can be contacted at email: nmsureja@gmail.com.