Sustainable Machine Intelligence Journal

Journal Homepage: sciencesforce.com/smij

Sustain. Mach. Intell. J. Vol. 10 (2025) 86–95

Paper Type: Original Article

A Robust Skin Cancer Classification using Deep Learning

Abdelrahman Abbas [1] and Hussam Elbehiery [2,*]

[1] Faculty of Information Systems and Computer Science, October 6th University 12585, Egypt; 212103294@o6u.edu.eg.

[2] Vanridge University, USA; drhussam@vru-edu.net.

Received: 11 Nov 2024 **Revised**: 07 Feb 2025 **Accepted**: 08 Mar 2025 **Published**: 10 Mar 2025

Abstract

In the medical field, detecting skin cancer from pictures can be difficult. Skin cancer identification is a laborious process in modern medicine that could eventually result in patient death. Early detection of skin cancer is essential to the effectiveness of full treatment. It is difficult to diagnose skin cancer effectively. Consequently, there are not enough qualified dermatologists around the globe to meet the demands of modern healthcare. Data imbalance problems are caused by the wide diversity in data from several healthcare industry categories. Deep learning models are frequently trained in one category more than others due to issues with data imbalance. Using an unbalanced dataset, this study suggests a revolutionary deep learning-based skin cancer detector. Data augmentation was used to balance different skin cancer categories and overcome data imbalance. The MNIST: HAM 10000 skin cancer dataset, consisting of seven skin lesion categories, was used. Deep learning models are widely used in image-based disease diagnosis. Deep learning-based models (Custom CNN, DenseNet201, ResNet-101, and Xception) have been used to classify skin cancer. The proposed framework was fine-tuned using different sets of hyperparameters. The results show that Xception outperformed ResNet-101, DenseNet201, and Custom CNN in terms of accuracy, Fl score, and receiver operating characteristic (ROC) curve. It achieved an accuracy of 0.9148. Our proposed framework could help in disease identification, which could save lives, reduce unnecessary biopsies, and reduce costs for patients, dermatologists, and healthcare workers.

Keywords: Skin Cancer; Deep Learning; Custom CNN; DenseNet201; ResNet-101; Xception.

1 | Introduction

In contemporary medicine, identifying and diagnosing skin cancer using dermoscopic pictures continues to be a significant difficulty. Identification of skin cancer is a difficult and time-consuming procedure that can result in treatment delays and a higher risk of patient morbidity if it is not done correctly and quickly. Because it enables prompt action and raises the chance of a successful course of therapy, early identification is crucial for improving patient outcomes. However, because of the wide range of lesion types and the apparent resemblance between benign and malignant lesions, it is difficult to make an accurate diagnosis of skin cancer. Furthermore, the global scarcity of qualified dermatologists makes it more challenging for healthcare systems to meet the rising need for precise diagnosis [1].

Corresponding Author: drhussam@vru-edu.net

https://doi.org/10.61356/SMIJ.2025.10507

In the field of medical image processing, deep learning has made impressive strides and provided a solid answer to the skin cancer diagnosis problem. There is less need for intensive feature engineering because deep convolutional neural networks (DCNNs) have shown remarkable ability in automatically extracting and learning hierarchical features from enormous datasets. However, issues like data imbalance, which result from the intrinsic variety and diversity of medical images, might impair deep learning models' performance. Usually, DCNNs do less well on underrepresented classes because they overfit to more common categories in a dataset.

In this work, we leverage the HAM10000 dataset, a well-known and often-used dataset in dermatological research, to present a novel deep learning-based system for automatically identifying skin cancer. 10,015 dermoscopic images representing seven types of skin lesions—actinic keratosis, basal cell carcinoma, benign keratosis-like lesions, dermatofibromas, melanoma, melanocytic nevi, and vascular lesions—are included in the HAM10000 dataset. The information was gathered from two separate locations over 20 years, guaranteeing a representative and varied sample of pigmented lesions.

To improve the representation of underrepresented classes and create a more balanced dataset, data augmentation techniques were used to solve the problem of data imbalance. We assessed how well many deep learning architectures performed in the categorization of skin cancer, including Custom CNN, DenseNet201, ResNet-101, and Xception. To maximize these models' predicted accuracy, they were fine-tuned using several sets of hyperparameters.

When it came to classification accuracy, F1-score, and the area under the receiver operating characteristic (ROC) curve, the experimental findings showed that the Xception model performed better than ResNet-101, DenseNet201, and Custom CNN. With 91.48% total accuracy, the Xception model demonstrated its excellent performance. Early detection of skin cancer could be facilitated by this suggested framework, which would cut down on needless biopsies and lower total healthcare expenses for both patients and providers. The results of this study highlight how well deep learning models work to improve dermatology diagnostics and support continued attempts to create AI-based medical instrument development.

2 | Literature Review

The difficulty of differentiating between benign and malignant lesions has long presented difficulties in the field of dermatological imaging, especially for the diagnosis of skin cancer. Since delays might result in higher morbidity, an accurate and timely diagnosis is essential. Conventional approaches rely significantly on dermatological knowledge, which is constrained by a worldwide scarcity of experts. A trustworthy automatic diagnosis system is, therefore, crucial.

2.1 | Advances in Deep Learning for Skin Cancer Detection

The detection of skin cancer is one of the many medical imaging tasks in which Convolutional Neural Networks (CNNs) have demonstrated remarkable performance. Research indicates that CNNs can do better than dermatologists on some categorization tasks, indicating the possibility of incorporating deep learning into medical care. Skin cancer detection has made extensive use of CNN architectures like ResNet, VGG, and Inception. ResNet models have achieved excellent accuracy levels because of their deep layers and strong feature extraction capabilities [2].

2.2 | Challenges in Data Imbalance and Model Optimization

Data imbalance is a typical problem in medical imaging, where the underrepresentation of some skin lesion types causes models to overfit more prevalent classes. The distribution of different skin cancer-affected cell types is shown in Figure 1, highlighting the variation in case frequencies across different classes. To overcome this and improve model generalization, Data augmentation, and transfer learning have shown to be crucial tactics to overcome this and improve model generalization. Higher accuracy and robustness have been achieved through research using transfer learning with pre-trained models (e.g., ImageNet) and the HAM10000 dataset [3].

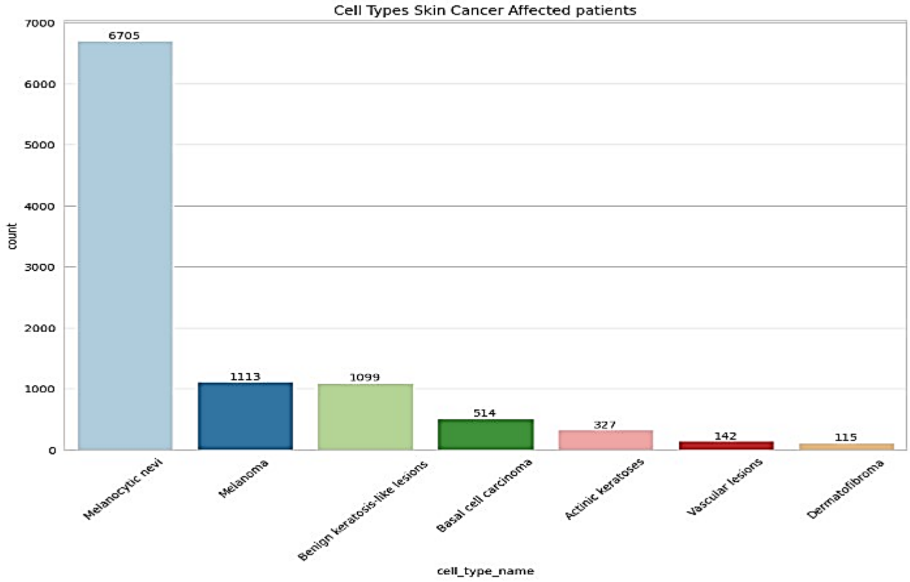

Figure 1. Distribution of seven different classes.

2.3 | Evaluation of Model Performance

F1-score, accuracy, and ROC curve measures are frequently used to assess these models' performance. For example, models like ResNet and Xception have demonstrated remarkable outcomes; current research has revealed that the Xception model achieves high accuracy. Given these developments, deep learning may be able to benefit dermatology by lowering the need for biopsies and associated medical expenses [4].

3 | Methodology

Training and test sets of the skin cancer dataset were first acquired to create an efficient skin cancer detection system. This study highlights how crucial it is to balance the dataset to increase model accuracy, especially considering the medical industry's difficulties in identifying skin cancer from photographs. Skin cancer detection is time-consuming and prone to mistakes, which could result in treatment delays or, in the worst-case scenario, death. To increase the performance of deep learning models, the problem of data imbalance must be addressed, as early diagnosis is essential for effective treatment outcomes [5].

3.1 | Dataset Description

Meaningful data is a key component of deep learning. In this study, we used the open-source skin cancer dataset MNIST: HAM10000 [33] which consists of 7 types of skin lesions, namely: actinic keratosis and intraepithelial carcinoma/Bowen's disease (apiece), basal cell carcinoma (bcc), benign keratosis-like lesions (seborrheic keratosis/seborrheic keratosis and lichen planus-like keratosis (bkl)), dermatofibroma (pdf), melanoma (mel), melanocytic nevi (NV), and vascular lesions (hemangiomas, angiokeratosis, pyogenic granuloma, and hemorrhage (vasc)). More than 50% of the lesions were confirmed by histopathology (history). The ground truth for the remaining cases was either follow-up examination (follow_up), expert consensus (consensus), or confirmation by confocal microscopy (confocal microscopy) [6].

3.2 | Preprocessing

This study's HAM10000 dataset is extremely vulnerable to unbalanced data problems, as some lesion types—like "melanocytic nevi"—makeup over half of the sample. A major obstacle to training deep learning models for complicated tasks is imbalanced data. Since some lesion kinds are uncommon, real-time medical datasets frequently lack the balanced data that most deep-learning models require to function well. This disparity may result in skewed or biased predictions, impairing the model's functionality.

This was addressed by using data augmentation techniques for the underrepresented classes, which increased the amount and diversity of the dataset by translating, rotating, and zooming photographs. Methods such as Histogram Equalization were also used to improve skin lesions' contrast. These techniques strengthened the model by increasing dataset diversity, which also improved the model's capacity for high accuracy and generalization.

3.3 | Custom Convolutional Neural Network (CNN)

In Here in this study, a Custom Convolutional Neural Network (CNN) model was designed and applied to address the problem of classifying skin lesions from the HAM10000 dataset. As compared to pre-trained models such as AlexNet, ResNet, or Inception with fixed architecture and deep prior training on big datasets like ImageNet, Custom CNN is flexible to develop a model customized to the peculiar nature and requirement of the dataset. This flexibility is particularly valuable when dealing with unique patterns or data distributions that are not well-represented in typical pre-trained models.

The proposed Custom CNN model is made up of a series of convolutional layers, each of which is followed by max-pooling layers for feature extraction and spatial down-sampling. Figure 2 shows the architecture, beginning with an input layer that receives images of dimensions 75×100 with three color channels (RGB). The first convolutional layer applies 32 filters of kernel size 3×3, and then applies a ReLU activation function in order to introduce nonlinearity. This is succeeded by a max-pooling layer whose pool size is 2×2 to reduce the spatial sizes. There then are the succeeding convolutional layers with increasing filter counts (64, 128, and 256 filters) each followed by a ReLU activation and a max pooling so that deeper features are obtained from the reduction in spatial resolution.

The network, as in Figure 2, flattens the output feature maps into one-dimensional vector format and then forwards this into multiple fully connected (Dense) layers. The first dense layer consists of 256 neurons, followed by ReLU activation and a dropout layer with a dropout rate of 0.5 to prevent overfitting. The same process is repeated with dense layers of 128 and 64 neurons followed by ReLU activation and dropout after each layer. The model concludes with an output layer that gives a probability distribution for each of the dataset's seven skin lesion classes with a SoftMax activation function.

The model is created using the categorical cross-entropy loss function, which is the best for multi-class classification problems, and the Adam optimizer to facilitate efficient gradient descent. The structure, as described in Figure 2, is best suited to leverage the heterogeneity and skewness of the HAM10000 dataset through a balance between model complexity and generalization. Several convolutional layers combined with dropout regularization help the network effectively learn hierarchical features with little danger of overfitting.

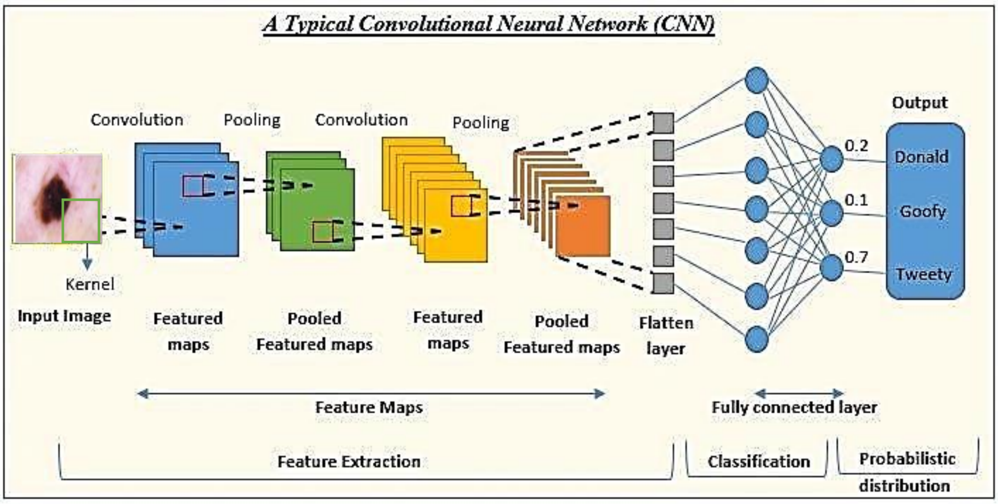

Figure 2. Custom convolutional neural network architecture.

3.4 | DenseNet201

By using dense connectivity patterns, DenseNet201 is one of the top-of-the-line convolutional neural networks employing dense connectivity. Contrary to typical CNNs where a layer feeds input directly only from the previous layer, DenseNet deploys an alternate strategy where all layers have relations to all previously formed layers within a dense block. The mentioned structure, which can be exemplified in Figure 3, enhances feature propagation, encourages the reuse of features, and avoids vanishing gradient phenomena, thereby rendering itself highly effective for medical image use.

DenseNet201 consists of multiple dense blocks, transition layers, convolutional layers, and pooling layers and has a total of 201 layers. From Figure 3, it can be seen that each dense block learns more sophisticated features, whereas ReLU activation and batch normalization in transition layers realize smooth information transfer among the network. Finally, a SoftMax activation function in the output layer supports multi-class classification, which is compatible with datasets like HAM10000, whose skin lesions have had very different appearances and whose sample sizes are unbalanced.

We employed a pre-trained DenseNet201 model in this study, which was pre-trained on ImageNet to leverage its strong feature extraction capability. We applied transfer learning to fine-tune the model for the particular application in skin lesion classification, as presented in Figure 3. We also fine-tuned the final layers by introducing additional dense layers and trained them using the Adam optimizer to allow for rapid gradient descent. Categorical cross-entropy loss, which best fits multi-class classification, was employed for the model training.

Task-adaptive tuning and pre-trained feature extraction in combination, DenseNet201 provides an equitable solution to dermoscopic image analysis. Its ability to detect intricate patterns with minimal computational requirements makes it an ideal model for skin cancer diagnosis, where accurate classification can largely impact patient treatment.

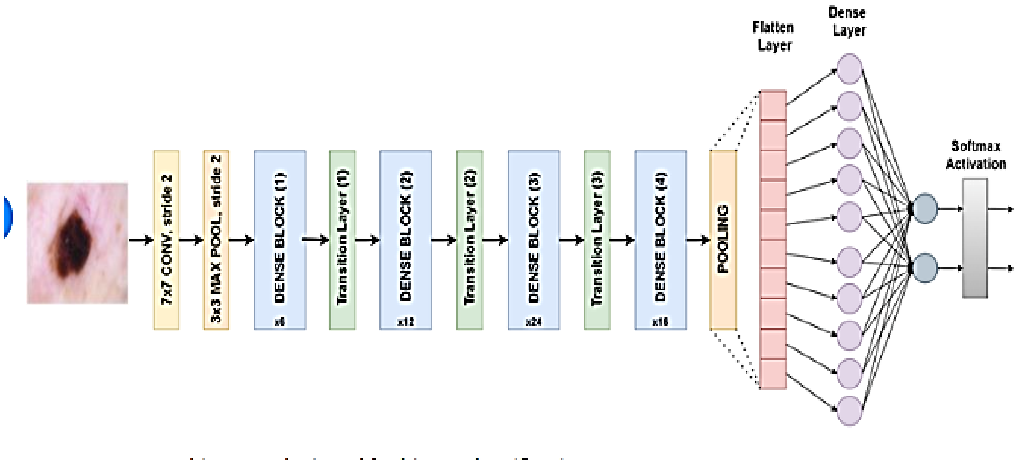

Figure 3. DenseNet201 architecture.

3.5 | ResNet-101

ResNet-101 is a deep convolutional neural network that contains 101 layers and is part of the ResNet (Residual Network) family designed to overcome the degradation problem commonly encountered in very deep neural networks. As networks deepen, the performance may saturate or even degrade due to vanishing gradients. To address this issue, ResNet introduces residual blocks with skip connections or identity mappings, which allow the model to bypass certain layers. This structure helps gradients flow directly through the network, mitigating the vanishing gradient problem and facilitating the training of deeper architectures.

The architecture of ResNet-101 consists of multiple residual blocks, each composed of convolutional layers, batch normalization, ReLU activations, and identity shortcuts. These identity mappings enable the network to learn residual functions (i.e., differences from the input features) instead of direct feature mappings. This approach makes it easier to optimize deeper networks by simplifying the learning process and accelerating convergence. The residual blocks are particularly beneficial in medical imaging tasks, where subtle features and patterns must be captured effectively.

ResNet-101 is designed to extract both low-level and high-level features through its layered structure:

- Early Layers: Capture basic features like edges and textures through convolutional operations.
- Intermediate Layers: Extract more complex features using deeper convolutional layers.
- Deep Layers: Learn high-level, abstract representations of the input images, enabling effective classification even for complex datasets.

The final architecture typically concludes with fully connected layers and a SoftMax activation layer for multiclass classification tasks. ResNet-101's ability to generalize well across different image classification problems makes it an ideal choice for applications in medical imaging, such as skin lesion classification, where distinguishing between subtle patterns is crucial.

3.6 |Xception

Xception, short for "Extreme Inception," is a state-of-the-art deep convolutional neural network architecture that extends the ideas introduced by the Inception model. Instead of using standard convolutions, Xception employs depth-wise separable convolutions, which significantly reduce the computational complexity while preserving high model accuracy. This design choice makes Xception highly efficient and particularly suited for large-scale image classification tasks, including medical image analysis.

3.6.1 |Structural Overview

Xception is organized into three main stages: Entry Flow, Middle Flow, and Exit Flow. Each stage comprises several layers that use depth-wise separable convolutions and point-wise convolutions. As illustrated in Figure 4, Xception processes input tensors using a sequence of specialized layers to extract and refine features before classification:

- Depth-wise Separable Convolutions: These replace the traditional convolution operation by splitting it into two steps:
    - Depth-wise Convolution: Applies a single filter per input channel, capturing spatial features independently for each channel.
    - Pointwise Convolution: Uses a 1x1 convolution to combine the features across different channels.

This approach reduces the number of parameters and computational load, making the network more efficient without compromising the ability to learn complex patterns.

- Entry Flow: The initial layers of Xception focus on extracting low-level features from the input images, such as edges and textures. It begins with several convolution and max-pooling layers, which down-sample the input and capture essential details.
- Middle Flow: This section contains most of the depth-wise separable convolution layers. It consists of multiple repeated blocks that capture high-level and abstract features. The use of ReLU activations and batch normalization ensures stable gradient flow and efficient learning.

- Exit Flow: The final layers focus on capturing the most abstract and high-level features necessary for classification. It includes additional depth-wise separable convolution layers, followed by a Global Average Pooling layer that reduces the spatial dimensions of the feature maps before passing them to fully connected layers for final classification.
- Classification Head: The architecture typically concludes with fully connected layers and a SoftMax layer, making it suitable for multi-class classification tasks. In this study, the final classification layer predicts probabilities across seven distinct skin lesion categories.

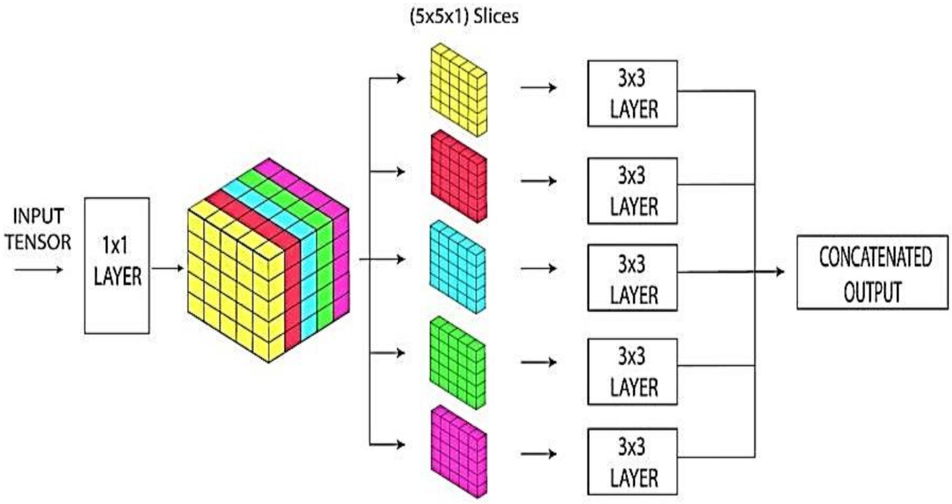

Figure 4. Xception architecture.

3.6.2 | Implementation for Skin Lesion Classification

In this project, the pre-trained Xception model (initialized with ImageNet weights) was adapted using transfer learning to classify skin lesions in the HAM10000 dataset. The following adjustments were made:

- Base Model: We utilized the pre-trained Xception architecture without its top (fully connected) layers, allowing for custom classification layers tailored to the specific task of skin lesion identification.
- Custom Top Layers:
  - A GlobalAveragePooling2D layer was added to reduce the feature map's spatial dimensions.
  - A dense layer with 1024 neurons and ReLU activation was included to enhance feature extraction.
  - The final output layer comprises seven neurons with SoftMax activation, corresponding to the seven lesion categories in the dataset.
- Transfer Learning: Initially, all layers of the base Xception model were frozen to preserve the pre-learned features from ImageNet, focusing the training on the newly added top layers.
- Fine-Tuning: After initial training, the top ten layers of the Xception base model were unfrozen to enable further fine-tuning. This step allows the model to adapt the learned features to the specifics of the skin lesion dataset, improving classification performance.
- Compilation and Optimization: The model was compiled using the Adam optimizer, known for its efficient adaptive learning capabilities. Categorical cross-entropy loss was used due to the multi-class nature of the classification task.

4 | Results

Using the HAM10000 dataset, we assessed four different Convolutional Neural Network (CNN) architectures for the skin lesion classification task. Custom CNN, DenseNet201, ResNet-101, and Xception are among the models that have been tested. Three main metrics—F1 Score, Accuracy, and ROC AUC—that are frequently employed in multi-class classification issues to offer a thorough assessment of the model's efficacy were utilized to evaluate each model's performance as summarized in Table 1.

- Custom CNN: For this dataset, a specially tailored CNN architecture was created. It obtained an ROC AUC of 0.914, an accuracy of 0.751, and an F1 Score of 0.472. Although the Custom CNN fared very well, as demonstrated in Figure 5, the deeper pre-trained models performed better because of their stronger feature extraction skills from transfer learning.

- DenseNet201: The DenseNet201 model, utilizing dense connections between layers, performed notably well with an F1 Score of 0.534, an Accuracy of 0.766, and an ROC AUC of 0.923. The dense connectivity pattern of this architecture enhanced gradient flow and feature reuse, contributing to improved model performance compared to the Custom CNN.

- ResNet-101: Known for its skip connections and deep residual blocks, ResNet-101 obtained an ROC AUC of 0.888, an accuracy of 0.728, and an F1 Score of 0.356. ResNet-101 performed worse than DenseNet201 or Xception in this job, even though it successfully addressed the vanishing gradient issue. Overfitting or a failure to make fine-tuning modifications specific to the skin lesion data could be the cause of this...

- Xception: Out of all the assessed architectures, the Xception model performed the best overall, obtaining an F1 Score of 0.760, an Accuracy of 0.912, and an ROC AUC of 0.903. While capturing intricate, high-level patterns in the photos, Xception's depth-wise separable convolutions enabled effective parameter utilization. This model's capacity to handle complicated picture data in medical imaging tasks is demonstrated by its outstanding performance.

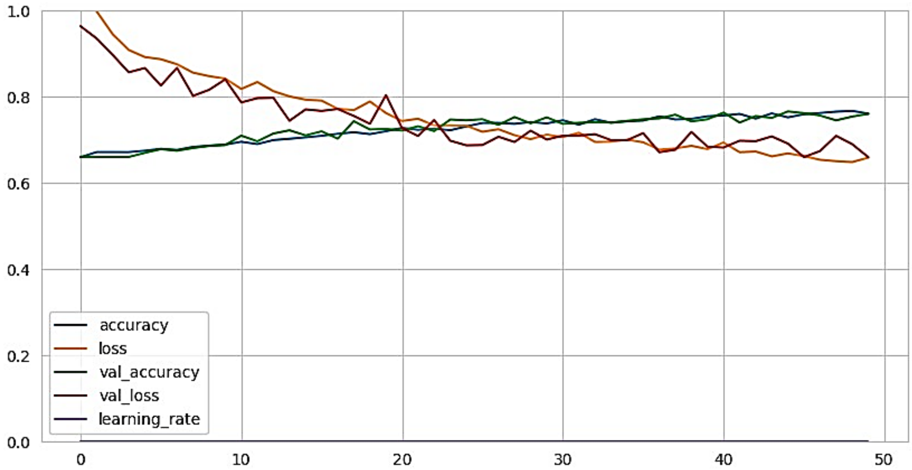

Figure 5. Model performance.

Table 1. Summary of model performance.

Model	F1 Score	Accuracy	ROC AUC
Custom CNN	0.472	0.751	0.914
DenseNet201	0.534	0.766	0.923
ResNet-101	0.356	0.728	0.888
Xception	0.760	0.912	0.903

4.1 | Analysis

The results indicate that the Xception model achieved the highest performance across all metrics, suggesting its effectiveness in capturing detailed and complex features from the dermoscopic images. DenseNet201 also demonstrated robust performance, leveraging its dense connectivity to reuse features across layers effectively. Custom CNN, while competitive, did not achieve the same level of performance as the pre-trained models, highlighting the advantages of using transfer learning with large-scale, pre-trained architectures. ResNet-101, despite its depth and residual connections, showed comparatively lower scores, suggesting the need for further fine-tuning or adaptations for this specific dataset.

Overall, these findings underscore the importance of model selection and adaptation when working with medical imaging data, where subtle differences in image features can significantly impact classification performance.

4.2 | Comparative Discussion

The comparison of results indicates that Xception and DenseNet201 are superior in handling the intricacies of dermoscopic image classification compared to the previously reported InceptionV3 in the Paper. The enhanced architecture of Xception, with its efficient feature extraction capabilities, demonstrated significant improvements in accuracy and robustness. Similarly, DenseNet201's dense connectivity structure contributed to a high ROC AUC score, indicating strong discriminatory power between lesion categories.

Overall, our findings suggest that utilizing more advanced architectures like Xception and DenseNet201, coupled with extensive fine-tuning, can lead to substantial improvements in performance for automated skin lesion classification tasks. This comparison highlights the importance of selecting suitable architecture and hyperparameter optimization when applying deep learning models to medical imaging tasks. Notebooks on GitHub: https://github.com/3booood/paper-data-mining [7].

5 | Conclusions

This research presents various deep learning architectures like Custom CNN, ResNet-101, Xception, and DenseNet201. They tried to classify skin lesions. Using the HAM10000 dataset. The target was to compare the models' capabilities in separating seven types of dermatological diseases, which is a critical process in the early detection and diagnosis of skin cancer.

The findings of our study reveal that the Xception model has the most authentic performance of overall accuracy of 91.2%, an F1 Score of 0.760, and a ROC AUC of 0.903. It was able to outperform the other tested architectures, DenseNet201, Resnet-101, and our Custom CNN. The standout reason Xception is better than the others is that it employs depth-wise separable convolutions that are efficient and effective in getting important parts of an image. DenseNet201 also displayed a robust performance with its ingenious technique to transfer features through dense connections, with an accuracy of 76.6% and an ROC AUC of 0.923.

Compared with the existing studies, e.g., the Data that employed InceptionV3, our models revealed a significant improvement in classification performance. The results imply that the selection of more advanced architecture, together with specific fine-tuning techniques, improves the model's capacity to generalize on several types of lesions. The Custom CNN model, which is less complex compared to the pre-trained models, still got competitive results, thus showing that we can design tailored architectures when computational resources are limited.

The research which is the subject of the paper brings to light the significance of model selection and hyperparameter optimization in the medical image classification tasks. The differences in performance between architectures require us to be careful in terms of model complexity, data augmentation techniques, and fine-tuning strategies to get the best performance.

To sum up, our research shows that the use of the advanced deep learning architecture, Xception, can lead to considerable improvements in the automatic classification of skin lesions. These findings can be used by dermatologists for early diagnosis, and they can thereby contribute to improved patient outcomes, showing AI's role in medical imaging.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contributions

All authors contributed equally to this work.

Funding

This research was conducted without external funding support.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Codella, N. C., Nguyen, Q. B., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A. C., & Smith, J. R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development, 61(4/5), 5-1.

[2] Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial intelligence review, 53, 5455-5516.

[3] Batool, D.; Shahbaz, M.; Asif, H.S.; Shaukat, K.; Alam, T.M.; Hameed, I.A.; Ramzan, Z.; Waheed, A.; Aljuaid, H.; Luo, S. A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning. Plants 2022, 11, 1925.

[4] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf.Process. Syst. 2012, 25, 1097–1105

[5] Garg, Rishu, Saumil Maheshwari, and Anupam Shukla. "Decision support system for detection and classification of skin cancer using CNN." Innovations in computational intelligence and computer vision: proceedings of ICICV 2020. Springer Singapore, 2021.

[6] Dataset: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

[7] NOTEBOOKS on Github: https://github.com/3boooood/paper-data-mining