Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Deep Learning for Skin Cancer Classification: A Study of Model Accuracy, Generalization, and Ensemble Learning

Kratika[1], Nandini Tyagi[2], Riya Sharma[3], Monika Sharma[4]

[1-4]Department of Computer Science and Engineering (Data Science), ABESEC, Ghaziabad.

E-mail: [1]kratikak8@gmail.com, [2]nandinityagi4282@gmail.com, [3]riyasharma7800726@gmail.com, [4]monika.sharma@abes.ac.in

Abstract

Skin cancer is one of the most common types of cancer, primarily caused by unmutated DNA changes influenced by both environmental and genetic factors. Early and accurate identification is crucial for reducing mortality and improving treatment outcomes. This study utilizes the HAM10000: MNIST dataset, which consists of 10,015 high-resolution dermoscopic images, to evaluate various CNN models. The images were preprocessed and standardized to a consistent resolution to ensure uniformity. Among the models tested, Inception-ResNet-V2 exhibited the lowest accuracy at 51.22%, while VGG19 achieved the highest accuracy at 94.26%. This was followed by DenseNet121 at 93%, Xception at 93%, and ResNet50 at 92%. To further enhance predictive performance, an ensemble learning technique was employed, combining VGG19, DenseNet121, Inception-ResNet-V2, ResNet50, and Xception, resulting in an impressive accuracy rate of 98%. These findings highlight the potential of deep learning and ensemble methods to significantly improve early skin cancer detection, paving the way for more reliable and effective clinical decision-making.

Keywords: Skin Cancer, Machine Learning, Ensemble Learning, VGG19, DENSENET121, Inception-RESNET-V2, RESNET50, Xception Model.

1. Introduction

"Early detection saves lives." This adage underscores the importance of prompt and accurate cancer diagnosis. Cancer is the second leading cause of death worldwide, accounting

DOI: https://doi.org/10.36548/jscp.2025.2.005
Received: 04.04.2025, received in revised form: 05.05.2025, accepted: 23.05.2025, published: 06.06.2025

for over 10 million fatalities annually. Early detection is vital for effective treatment across various cancer types, including skin, oral, and pancreatic cancers, with survival rates increasing by more than 90% when diagnosed early. Skin cancer is one of the most lethal forms of the disease due to its high mortality rate globally. Clinical screenings typically represent the initial step in the diagnostic process. Following this, a biopsy, histological analysis, and dermoscopy are conducted to confirm the diagnosis. Skin cancer arises when the natural process of skin cell proliferation is disrupted, leading to DNA mutations that result in malignant growths. UV radiation is a significant contributing factor to the development of skin cancer. Various risk factors include fair skin, prolonged exposure to radiation or harmful chemicals, severe burns or skin injuries, aging, smoking, and a compromised immune system. While skilled dermatologists remain essential to traditional diagnostic approaches, advancements in machine learning have introduced new possibilities for automated skin cancer detection and classification. Skin cancer can be classified as either benign (non-cancerous) or malignant (cancerous). Benign tumors are typically harmless, slow-growing, and do not invade surrounding tissues or metastasize. In contrast, malignant tumors such as melanoma, basal cell carcinoma, and squamous cell carcinoma are aggressive and invasive, posing significant health risks.

Figure 1. Skin Lesion Types

Figure 1 presents the classification of skin lesions, which includes the following designations: BCC (basal cell carcinoma), SEK (seborrheic keratosis), ACK (actinic keratosis), DEF (dermatofibroma), MEL (melanoma), NEVI (nevus), and VASCL (vascular lesion).

1.1 Research Objective

This research compares the accuracy of various deep learning models for the categorization of skin cancer, with an emphasis on additional assessment metrics including precision, recall, F1-score, and AUC-ROC. To determine how well these cutting-edge architectures, recognize benign and malignant skin lesions, the study utilizes models such as VGG19, INCEPTION-RESNET-V2, DENSENET121, Xception, and RESNET50. Traditional diagnostic procedures rely significantly on trained dermatologists; however, they can be subjective, time-consuming, and limited in availability, particularly in impoverished areas. Machine learning and deep learning approaches have demonstrated tremendous potential for automating skin cancer categorization, increasing diagnostic speed and accuracy. [4] The results of this study will help build more effective and dependable automated systems for the early detection and diagnosis of skin cancer, which will ultimately improve patient outcomes by shedding light on the advantages and disadvantages of various models. According to WHO data, Figure 2 shows a steady rise in skin cancer cases in both men and women between 2022 and 2025.

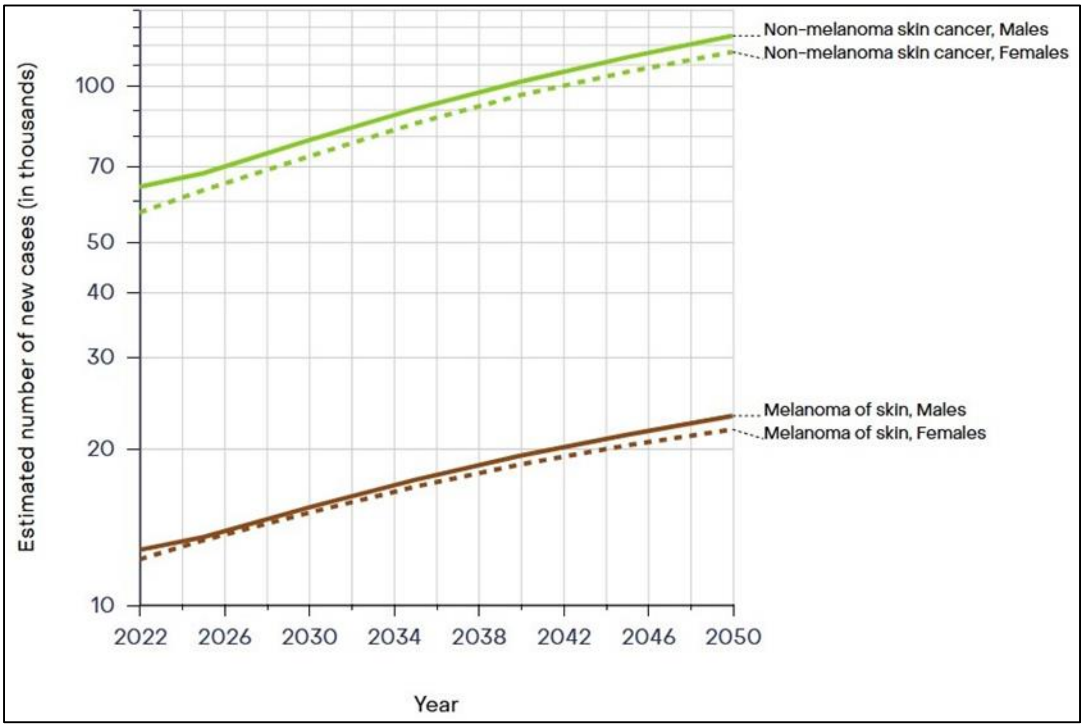

Figure 2. Anticipated Number of New Cases in the Upcoming Years

This study seeks to address this issue by exploring various learning paradigms.:

1. **RESNET50:** Residual Networks (ResNet) are pivotal in addressing computer vision challenges. ResNet architectures utilize residual connections to mitigate the issue of vanishing gradients during backpropagation.
2. **VGG19:** VGG19, originally developed for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), is frequently employed in neural style transfer, transfer learning, and classification due to its accessibility and pretrained weights in frameworks like Keras. The architecture features a consistent arrangement of max-pooling layers and convolutional layers, culminating in two fully connected layers followed by a softmax layer. The name VGG19 reflects its 19 layers, each with varying parameters.
3. **Xception:** Xception is a convolutional neural network that excels in the accurate and efficient classification of skin cancer through its implementation of depthwise separable convolutions. This architecture enhances the Inception model by substituting traditional convolutions with depthwise separable convolutions.
4. **DenseNet121:** DenseNet121 is a powerful deep convolutional neural network noted for its efficacy and feature reuse. Each of its 121 layers is interconnected to subsequent layers within the same dense block, facilitating enhanced information flow and gradient propagation.
5. **Inception-RESNET-V2:** Inception-RESNET-V2 merges the skip connections of RESNET with Inception modules, demonstrating exceptional performance in skin cancer classification. This architecture effectively extracts features, alleviates the vanishing gradient problem, and achieves high accuracy by enhancing depth and feature representation, particularly through transfer learning on datasets such as HAM10000 and MNIST.

2. Problem Statement

Skin cancer, a significant public health issue, poses considerable challenges due to its rapid progression and high mortality rates. Early and accurate detection is essential for effective treatment and improved patient outcomes. The rising incidence of skin cancer, combined with an increasing demand for simple and effective diagnostic technologies, underscores the necessity for innovative solutions.

Deep learning models have emerged as a promising tool for the automatic and precise diagnosis of skin cancer. However, with a diverse array of deep learning architectures available, including VGG19, Inception-RESNET-V2, and other advanced models, it is imperative to analyze and compare their performance to identify the most effective model for this task.

The objective of this study is to evaluate which deep learning model demonstrates the highest accuracy, precision, recall, and robustness for skin cancer classification, utilizing benchmark datasets such as HAM10000 and MNIST.

Key challenges include variation in skin tone, lesion shape, and imaging quality. This study addresses a critical gap by comparing multiple learning models, including VGG19, DENSENET121, Inception-RESNET-V2, RESNET50, and Xception. Our aim is to identify the most efficient and adaptable methodologies for skin cancer classification and detection, particularly in scenarios with limited labeled data. We seek to enhance skin cancer diagnostic methods by assessing the strengths and weaknesses of each approach.

3. Literature Survey

This research focuses on the development of skin cancer detection and classification using digital images. This document reviews years of studies conducted on skin cancer detection, detailing various technologies employed to diagnose and classify skin cancer, with ongoing research in this area. This section serves as a comprehensive review of studies on skin cancer detection and categorization, highlighting the differing methodologies of researchers in terms of feature extraction, training methods, and classification models.

DEMİR A. et al. (2019) and his team emphasized the importance of both malignant and benign images. The dataset utilized is balanced, containing images of both benign and malignant skin moles. Each image in this dataset measures 224x224x3 pixels. The dataset comprises two types of skin cancer images: benign and malignant moles. The training dataset includes 2,437 images—1,330 benign and 1,107 malignant. The testing phase involves 660 images (360 benign and 300 malignant), while 200 images (110 benign and 90 malignant) are randomly selected from the training dataset for validation purposes. The classification performance was evaluated using the RESNET-101 and Inception-v3 neural network architectures. Our goal is to enhance the early detection of skin cancer by classifying images as benign or malignant. The classification task employs the RESNET-101 and Inception-v3 deep learning architectures, achieving an accuracy rate of 84.09% for RESNET-101 and 87.42% for Inception-v3 [7].

Pham T. C. et al. (2019) and his team presented a method for classifying pigmented skin lesions from dermoscopy images based on color, texture, and shape attributes. Color and texture features are segmented to emphasize clinically relevant lesions. An optimization framework is employed to rank the extracted feature data, identifying the optimal subset of characteristics. The approach delineates lesions from background skin by detecting lesion borders for shape analysis. The identified borders are used to extract shape details from the lesions. This method achieved 92.34% specificity and 93.33% sensitivity on a dataset of 564 examined images [8].

Hasan R. M. et al. (2021) and his team highlighted a study wherein several researchers have utilized CNN architectures on skin cancer datasets to develop a more effective early detection method. Convolutional Neural Networks (CNNs) were proposed by the authors as a technique for examining skin lesions in dermoscopy images. A deep CNN applied to the International Skin Imaging Collaboration (HAM10000: MNIST) dataset achieved an accuracy of 80.3%. They employed a convolutional-deconvolutional architecture for data segmentation. In a separate study, researchers applied CNNs to extract symptomatic features from the same dataset using a feature extraction method to segment dermoscopic images and extract characteristics of affected skin cells, achieving an accuracy of 89.5%. A report was provided on utilizing vision to classify skin tumors [9].

Ghosh H et al. (2024) and his team conducted a series of studies focused on the categorization of skin cancer, utilizing large and diverse datasets. The researchers implemented class weights during model training and employed various techniques to address issues such as class imbalance. The study combined Convolutional Neural Networks (CNNs) utilizing DenseNet121 and RESNET50 architectures to create a hybrid deep learning model, which demonstrated superior performance compared to either model used independently. These findings underscore the potential of deep learning models to enhance classification performance, thereby providing valuable assistance to dermatologists in clinical settings [10].

Abdullah et al. (2024) and his team highlighted the necessity for more precise analysis in the context of melanoma skin cancer. Researchers favor convolutional neural networks due to their extensive application in image processing. Through the analysis of skin images, deep learning techniques—particularly CNNs—are capable of delivering rapid and accurate predictions of melanoma skin cancer. This approach could offer a more effective and non-invasive diagnostic alternative, significantly reducing the reliance on invasive biopsies. The versatility of deep learning, combined with access to vast amounts of data, positions it as a crucial resource for timely and accurate disease recognition. To enhance early diagnosis, decision-making, and ultimately save lives, researchers have developed computer systems that analyze skin scans for skin cancer detection [11].

4. Proposed System

We present an ensemble learning strategy for skin cancer classification utilizing models including VGG19, DenseNet121, ResNet50, Xception, and Inception-ResNet-V2. To ensure precise and reliable skin cancer diagnoses, the system preprocesses dermoscopic images, refines models using pretrained weights, and integrates predictions through ensemble learning. These architectures were selected for their diverse strengths: some excel at feature extraction (VGG19, DenseNet121), while others address limitations in deep learning (ResNet50, Inception-ResNet-V2), and some enhance efficiency (Xception). By combining their individual strengths in an ensemble learning approach, we aim to develop a robust, accurate, and clinically reliable system for skin cancer classification. The accompanying figure illustrates the workflow for image classification. A soft-voting classifier aggregates the predictions made by each model within the ensemble, selecting the class with the highest mean probability by calculating the average probability for each class across all models. This method enhances robustness and mitigates the bias inherent in individual models.

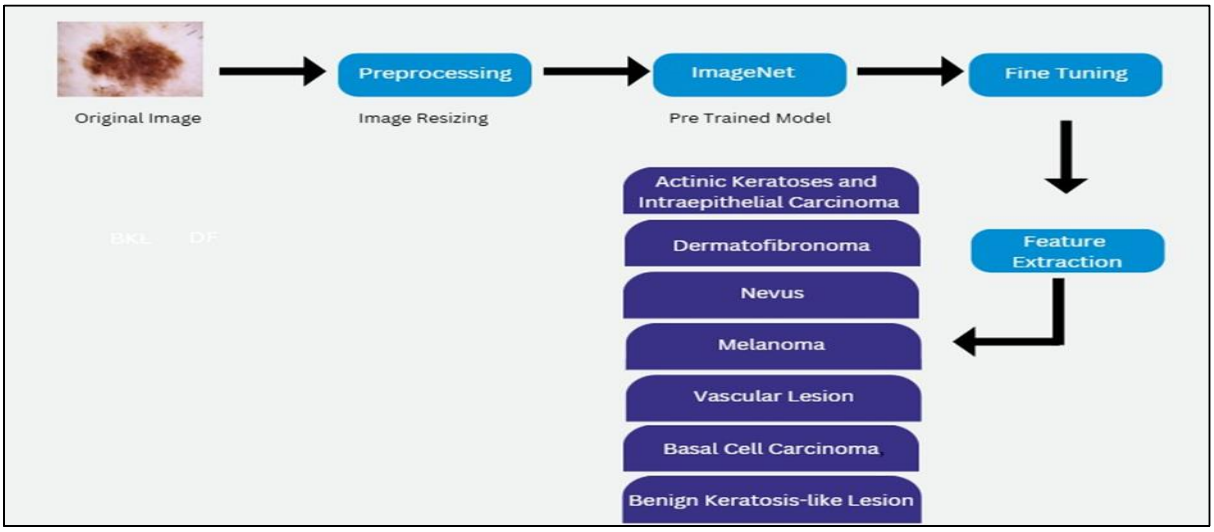

Figure 3. Proposed System

- **Original Image:** The original input image is likely a digital capture of a skin lesion.
- **Preprocessing:** A specific dimension is applied to the original image to ensure uniform input size for subsequent modeling. This step is crucial for maintaining consistency across images.
- **ImageNet:** The term "pre-trained model" refers to a convolutional neural network (CNN) that has undergone extensive training on the ImageNet dataset. This model is capable of recognizing general visual features such as edges, textures, and shapes.
- **Fine Tuning:** In this phase, the pre-trained model is adapted to specifically classify skin lesions. A new dataset consisting of images of skin lesions is utilized to adjust the model's weights accordingly.
- **Feature Extraction:** The refined model extracts features from the input images. These high-level representations encapsulate the essence of the skin lesions.
- **Classification:** A classifier is developed to accurately detect skin diseases by predicting the type of skin lesion (e.g., AKIEC, DF, BKL, VASC, NV, BCC, MEL) based on the collected diagnostic data. This classifier may take the form of a simple linear model or a more complex neural network.

4.1 Dataset

The HAM10000: MNIST dataset, provided in the form of hmnist_8_8_RGB.csv for this study, is an essential part of the comparison of machine learning methods for classifying skin cancer. A unique dermoscopic image of a skin lesion is represented by each row in this CSV file, which is essentially image data. Pixel values, which range from pixel0000 to pixel0191, are organized in 192 columns in the dataset. This particular number of pixel columns is compatible with 8x8 pixel images, in which red, green, and blue color channels are incorporated into each pixel, for a total of 64 pixels (8 × 8) times three channels (64 × 3 = 192 features). These pixel data columns are followed by a label column that, for each image, provides the ground truth categorization of the particular kind of skin lesion. For training and assessment, this standardized format makes it easier to directly feed image data into different Convolutional Neural Network (CNN) models. When evaluating the accuracy with which various deep learning architectures, including VGG19, DenseNet121, Inception-ResNet-V2, ResNet50, and

Xception, categorize benign and malignant skin lesions based on their visual characteristics, the dataset's determined resolution and RGB format are crucial.

4.2 Tools and Software

The design and implementation of the deep learning-based skin cancer classification system, along with its evaluation, were conducted using Python due to its simplicity, versatility, and the abundance of libraries available for machine learning and image processing. SciPy and Pandas were utilized for key scientific and data manipulation operations. SciPy was specifically employed for image resizing and interpolation, ensuring that all dermoscopic images conformed to a standardized input shape required by the deep learning models. Pandas facilitated the efficient manipulation of the HAM10000 dataset, enabling structured reading, manipulation, and transformation of the image and label data stored in CSV format.

The model training process was established using TensorFlow and its high-level API, Keras, which provided streamlined access to robust pre-trained convolutional neural network models such as VGG19, DenseNet121, ResNet50, Xception, and Inception-ResNet-V2. Transfer learning techniques were applied to fine-tune these models for skin lesion classification. Matplotlib was used to visualize model performance through confusion matrix plots, ROC curves, training-validation accuracy/loss plots, and bar charts comparing performance, thereby enhancing the interpretation of results.

On the hardware front, the system leveraged an NVIDIA GPU with CUDA support, significantly accelerating the computationally intensive training process. The use of a GPU enabled parallel processing and effective management of large volumes of image data, which is essential for deep model training and the successful application of ensemble techniques. In summary, this integration of hardware and software provided a scalable and robust platform for conducting effective experiments and achieving high-speed skin cancer classification.

4.3 System Model

Evaluations of the HAM10000: MNIST dataset, which comprises 10,015 RGB images with a resolution of 600x450 pixels, were conducted for this study. This dataset, developed by multiple organizations following a comprehensive investigation, is specifically designed to facilitate the analysis of skin lesion cases categorized as either benign or malignant [15]. The HAM10000: MNIST dataset also supports research on seven distinct types of skin lesions:

vascular skin lesions, basal cell carcinoma, dermatofibroma, melanoma, actinic keratoses, and benign keratosis [16].

Five deep learning algorithms—VGG16, ResNet50, DenseNet121, Inception-RESNET-V2, and Xception—were evaluated on smaller test datasets after being trained on various combinations of training data. To enable a comparative analysis and demonstrate the efficacy of different architectures in the classification of skin lesions, these models were retained. Performance metrics such as F1-score, recall, accuracy, and precision were employed to guide model selection. The system offers a reliable, efficient, and clinically relevant automated approach for classifying skin cancer [17].

4.3.1 VGG16

VGG19 was developed as an enhanced version of VGG16. It is a 19-layer deep convolutional neural network (CNN) that is widely employed in the classification of skin cancer due to its capability to extract detailed spatial and hierarchical features from images. Datasets such as HAM10000 and MNIST are commonly utilized, comprising preprocessed images of malignant and benign skin lesions that conform to the 224x224 pixel input size of VGG19. To classify lesions into categories such as benign or malignant, the pretrained VGG19 model (originally trained on ImageNet) is refined using transfer learning on the skin cancer dataset. This process involves replacing its fully connected layers. Following training with optimization algorithms like Adam or SGD, the model is evaluated using metrics such as accuracy, precision, recall, and AUC-ROC. Despite its high accuracy and performance, VGG19 is computationally intensive and may be prone to overfitting on small datasets if data augmentation is not implemented. To mitigate overfitting, data augmentation techniques such as rotation, flipping, scaling, and contrast adjustments are typically employed, enhancing model generalization. Overall, VGG19 demonstrates considerable success in classifying skin cancer due to its deep architecture and transfer learning capabilities.

4.3.2 DenseNet121

A densely connected convolutional neural network, known as DENSENET121, has been developed to enhance gradient flow and feature reuse, making it particularly suitable for medical image classification applications, such as skin cancer diagnosis. DENSENET121 serves as a highly effective model for skin cancer classification due to its efficiency and ability to reuse features. Each of its 121 layers is interconnected within the same dense block, which is structured into multiple dense blocks. This innovative architecture reduces the number of parameters compared to conventional models, addressing issues of overfitting and vanishing gradients while facilitating improved feature propagation and reuse. DENSENET121 excels at identifying subtle characteristics in dermoscopic images, including irregular patterns, textures, and color gradients, all of which are crucial for distinguishing between benign and malignant lesions in skin cancer classification. The model's capability to retain features from previous layers allows it to incorporate fine-grained information in its decision-making process. Consequently, DENSENET121 demonstrates exceptional proficiency in differentiating between various skin lesions by effectively extracting both global and local information from dermoscopic images.

4.3.3 Inception-ResNet-V2

Inception-RESNET-V2 is a hybrid deep learning model that effectively integrates the capabilities of Inception modules with the skip connections of RESNET, demonstrating exceptional performance in skin cancer classification. This architecture leverages the rapid feature extraction of Inception networks alongside RESNET's residual connections, facilitating deeper network training while mitigating vanishing gradient issues. Skin cancer classification employs datasets such as HAM10000 and MNIST, with images preprocessed to meet the model’s input specifications (e.g., 299x299 pixels). Transfer learning is employed to fine-tune the pretrained Inception-RESNET-V2 model, which has been trained on ImageNet, for the classification of skin lesions as benign or malignant. The model's depth, efficient convolutional blocks, and residual learning significantly enhance feature representation, yielding impressive accuracy and robustness [23].

4.3.4 ResNet50

RESNET50 is a deep convolutional neural network that employs residual learning to address the vanishing gradient problem, enabling the effective training of exceptionally deep networks. With its 50 layers and shortcut connections, RESNET50 facilitates smoother gradient flow during backpropagation, thereby enhancing the comprehension of complex features. This architecture efficiently extracts hierarchical characteristics from dermoscopic images, allowing for the differentiation between benign and malignant skin lesions and the classification of skin cancer. By implementing transfer learning to utilize pre-trained weights from extensive datasets, RESNET50 demonstrates improved performance on smaller, domain-specific datasets. Its ability to capture fine-grained information in images makes it a reliable model for accurate classification in skin cancer detection. Furthermore, RESNET50's performance on smaller datasets is further enhanced through the application of transfer learning, leveraging pre-trained weights from large datasets. This model is particularly well-suited for real-time applications, striking an effective balance between accuracy and computational efficiency.

4.3.5 Xception

The Xception (Extreme Inception) convolutional neural network employs depthwise separable convolutions to achieve exceptional accuracy and efficiency, making it particularly well-suited for tasks such as skin cancer classification. This architecture, an extension of the Inception model, replaces traditional convolutions with depthwise separable convolutions, significantly reducing computational complexity while preserving performance. Depthwise convolutions are utilized for spatial filtering, while pointwise convolutions are employed to combine features. Xception's capability to detect intricate and fine-grained patterns in dermoscopic images enhances its effectiveness in skin cancer categorization, as it can differentiate between benign and malignant tumors based on subtle features such as texture, asymmetry, and color variations.

5. Results and Discussion

5.1 Evaluation Procedure

A training, validation, and testing split of 70%, 15%, and 15% was applied to the dataset, respectively. Class balance was ensured through a stratified sampling approach. To mitigate bias and assess generalizability, five-fold cross-validation was employed. The average performance metrics accuracy, precision, recall, and F1-score—were calculated across all folds to guarantee robustness.

This study evaluates the performance of RESNET50, VGG19, DenseNet121, Xception, and Inception RESNET-V2 for skin cancer classification. The accompanying graph displays the accuracy scores of these five distinct deep learning models. The results indicate that VGG19 achieved the highest accuracy at 94%, followed closely by DenseNet121 and Xception, both with 93% accuracy. In contrast, InceptionResNetV2 performed significantly worse with an accuracy of 51.22%, while ResNet50 recorded a slightly lower accuracy of 92%. VGG19 is particularly effective in distinguishing between benign and malignant conditions due to its sequential architecture featuring small 3×3 convolutions, which adeptly capture fine textures, color variations, and irregular patterns in skin lesions. Its usability facilitates effective feature transfer from pre-trained ImageNet weights, thereby reducing the risk of overfitting on smaller medical datasets.

While suboptimal precision can lead to unnecessary medical procedures, high precision minimizes false positives, thus preventing unwarranted biopsies and alleviating patient anxiety. A high recall rate is crucial for timely intervention, as it ensures that malignant tumors are not overlooked. The F1-score provides a balanced measure between precision and recall, making it more suitable for practical clinical applications by ensuring both accuracy and reliability. The integration of multiple models through ensemble learning enhances generalization, accuracy, and robustness. This research demonstrated improved predictive performance with an ensemble comprising VGG19, DenseNet121, Inception-RESNET V2, RESNET50, and Xception, achieving an accuracy of 98%. It underscores the precision in diagnosing benign and malignant cases, with fewer errors, attributable to the collective strengths of these models. Figure 4 illustrates the comparison of all classification techniques based on accuracy, precision, recall, and F1-score.

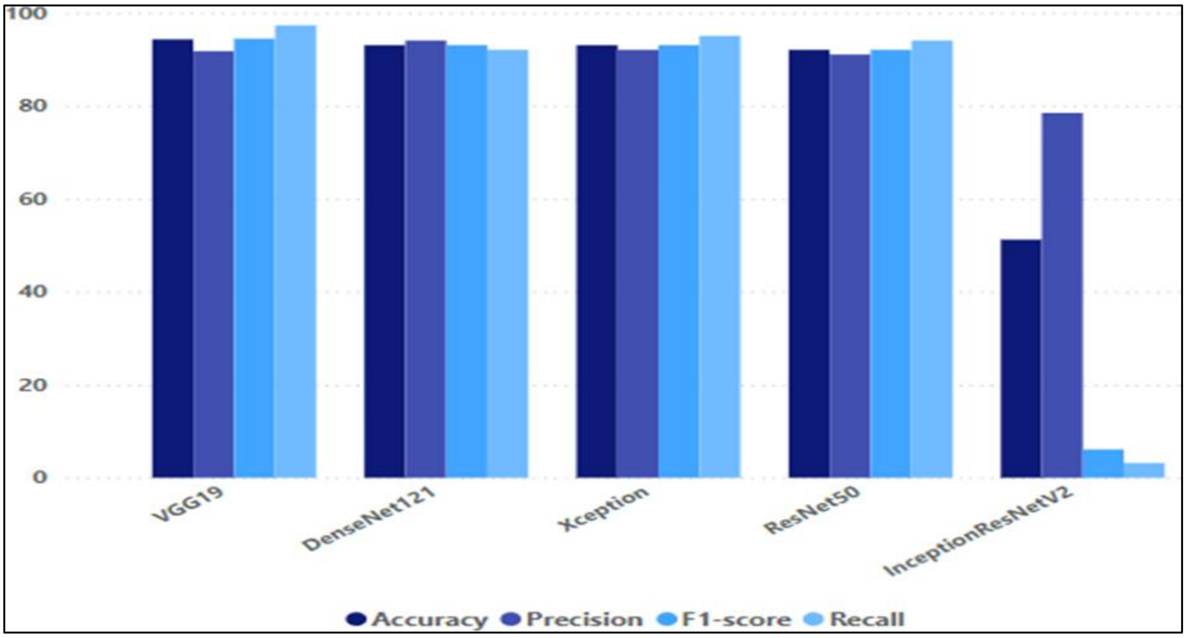

Figure 4. Comparison of Classification Techniques based on Accuracy, Precision, Recall, and F1-Score

The table 1 below presents the performance metrics of various deep learning models assessed in a classification task. The evaluation metrics employed are Accuracy, Precision, Recall, and F1-score, which offer a comprehensive understanding of each model's classification performance.

Table 1. Comparative Analysis of All Implemented Models on the Basis of Accuracy, Precision and Recall

Model	Accuracy	Precision	Recall	F1-score
VGG19	94.26	91.74	97.27	94.42
DenseNet121	93	94	92	93
Inception-RESNET-V2	51.22	78.46	3.17	6.09
RESNET50	92	91	94	92
Xception	93	92	95	93

- **Accuracy:** This metric reflects the overall correctness of the model's predictions. A high accuracy suggests that the model is predominantly classifying data correctly.

- **Precision:** This measures the proportion of true positive predictions among all positive predictions made by the model. A high precision indicates that the model is less likely to misclassify negative instances as positive.

- **Recall:** This metric assesses the percentage of true positive occurrences that were accurately predicted by the model. A high recall signifies that the model is proficient in identifying all positive cases.

- **F1-score:** The F1-score serves as a balanced measure of precision and recall, representing the harmonic mean of the two. A high F1-score indicates that the model achieves an appropriate balance between precision and recall.

The confusion matrix for the ensemble model presented in Figure 5, which includes VGG19, Inception-ResNet-V2, DenseNet121, ResNet50, and Xception, illustrates the model's effectiveness in identifying skin cancer.

The ensemble model demonstrates a low incidence of false negatives, a critical factor in medical diagnostics, as indicated by the confusion matrix (Figure 5). Benign nevi and melanoma, due to their visual similarity, were the conditions most frequently misclassified. The true positive rate for melanoma exceeded 95%, underscoring the model's reliability in potentially life-threatening cases.

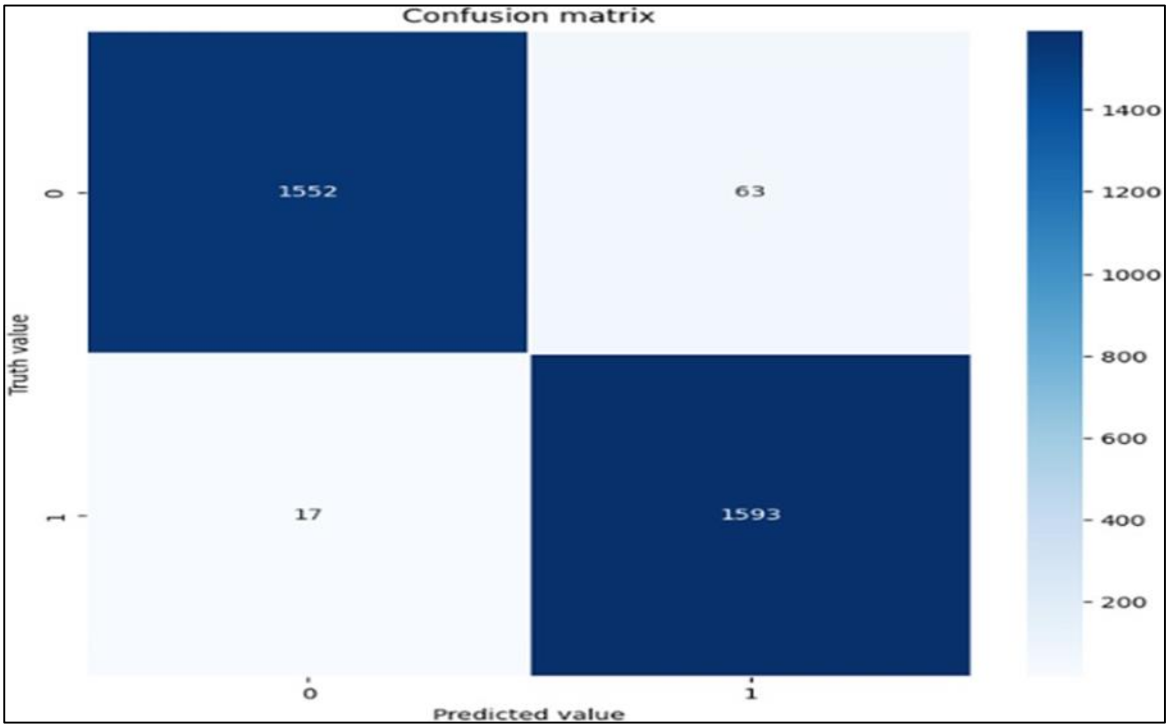

Figure 5. Confusion Matrix of the Ensemble Model for Skin Cancer Classification

6. Conclusion

This study, while yielding encouraging results, has several limitations. The geographical constraints of the HAM10000 dataset may affect its generalizability. Additionally, the high computational demands of the ensemble method restrict its real-time clinical deployment. Furthermore, the predictions may be biased towards more common lesion types, such as nevi, due to an imbalanced class distribution.

This research evaluates the performance of various deep learning models in classifying skin cancer, specifically focusing on the differentiation between benign and malignant lesions. Among the models tested, VGG19 demonstrated the highest accuracy at 94%. Its sequential architecture, characterized by small 3x3 convolutional filters, effectively captures fine-grained and subtle features in dermoscopic images. This characteristic is particularly advantageous when working with small medical datasets, as it reduces the risk of overfitting while facilitating the extraction of significant patterns from the input data.

In contrast, Inception-ResNet-V2 exhibited the lowest performance, achieving an accuracy of only 51.22%. This underperformance is primarily attributed to its excessive parameter count (approximately 55 million), which introduces unnecessary complexity and limits its generalization ability when trained on smaller datasets. Moreover, the deep architecture of the model heightens the risk of challenges such as vanishing gradients, complicating the training process and jeopardizing the model's reliability in clinical settings.

To address these limitations and enhance prediction accuracy, an ensemble learning approach was employed, combining the outputs of VGG19, DenseNet121, Inception-ResNet-V2, ResNet50, and Xception. This ensemble method significantly increased accuracy to 98%, underscoring the potential of hybrid models to deliver reliable and clinically valid diagnoses of skin cancer. The findings emphasize the importance of selecting architectures that align with the scale and complexity of datasets, particularly in critical healthcare applications.

7. Future Scope

To mitigate bias and enhance model generalization across diverse patient demographics, the next phase of this project aims to improve skin cancer classification by leveraging larger and more varied datasets. We will explore hybrid and ensemble learning approaches that integrate the strengths of multiple deep learning architectures to further enhance diagnostic accuracy. This strategy not only improves performance but also minimizes the risk of results being influenced by the limitations of any single model.

Given that explainable AI (XAI) can provide transparent insights into decision-making processes, it will play a crucial role in increasing medical professionals' confidence in AI-driven systems. Furthermore, the development of lightweight models tailored for mobile and edge devices will facilitate real-time diagnostics in remote or resource-constrained environments, thereby expanding access to care.

We will also investigate multi-modal data integration, which involves combining dermoscopic images with genetic and patient data, to enhance the comprehensiveness of AI-based diagnostics. This holistic approach may yield deeper insights into risk factors and disease progression. Additionally, we will implement continuous learning processes to ensure that the model can adapt to new data while maintaining its accuracy and relevance in evolving clinical settings.

Finally, to address data scarcity and improve model robustness, we will employ generative AI to augment datasets by generating synthetic yet realistic skin lesion images. This technology also has the potential to automatically annotate and produce interpretable outputs, further empowering clinicians to utilize AI for informed decision-making.

Acknowledgement

First and foremost, we would like to express our heartfelt gratitude to everyone who contributed to this project, making it a success. We truly appreciate the support, guidance, and collaboration that helped us achieve remarkable results. Nandini Tyagi obtained the dataset and completed the final formatting and modifications of the model. Kratika was responsible for the experimental analysis and model design. Riya Sharma worked on the graphical representation and contributed to the model’s design. Working on this project was an enriching experience, and we are grateful for the opportunity to learn and grow throughout the process

References

[1] Islam, Md Khairul, Md Shahin Ali, Md Mosahak Ali, Mst Farija Haque, Abhilash Arjan Das, Md Maruf Hossain, D. S. Duranta, and Md Afifur Rahman. "Melanoma skin lesions classification using deep convolutional neural network with transfer learning." In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), IEEE, (2021): 48-53.

[2] Gouda, Niharika, and J. Amudha. "Skin cancer classification using ResNet." In 2020 IEEE 5th International conference on computing communication and automation (ICCCA), IEEE, (2020): 536-541.

[3] Salian, Abhishek C., Shalaka Vaze, Pragya Singh, Gulam Nasir Shaikh, Santosh Chapaneri, and Deepak Jayaswal. "Skin lesion classification using deep learning architectures." In 2020 3rd International conference on communication system, computing and IT applications (CSCITA), IEEE, (2020) 168-173.

[4] Milton, Md Ashraful Alam. "Automated skin lesion classification using ensemble of deep neural networks in isic 2018: Skin lesion analysis towards melanoma detection challenge." arXiv preprint arXiv:1901.10802 (2019).

[5] Abuared, Nour, Alavikunhu Panthakkan, Mina Al-Saad, Saad Ali Amin, and Wathiq Mansoor. "Skin cancer classification model based on VGG 19 and transfer learning." In 2020 3rd International conference on signal processing and information security

(ICSPIS), IEEE, (2020): 1-4.

[6] Sreedhar, B., Manjunath Swamy BE, and M. Sunil Kumar. "A comparative study of melanoma skin cancer detection in traditional and current image processing techniques." In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), IEEE, (2020): 654-658.

[7] Demir, Ahmet, Feyza Yilmaz, and Onur Kose. "Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3." In 2019 medical technologies congress (TIPTEKNO), IEEE, (2019): 1-4.

Pham, Tri Cong, Giang Son Tran, Thi Phuong Nghiem, Antoine Doucet, Chi Mai Luong, and Van-Dung Hoang. "A comparative study for classification of skin cancer." In 2019 International Conference on System Science and Engineering (ICSSE), IEEE, (2019): 267-272.

[8] Hasan, M. R., Fatemi, M. I., Monirujjaman Khan, M., Zaguia, A., Kaur, M. (2021). Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. Journal of Healthcare Engineering, 2021(1), 5895156.

[9] Ghosh, Hritwik, Irfan Sadiq Rahat, Sachi Nandan Mohanty, J. V. R. Ravindra, and Abdus Sobur. "A study on the application of machine learning and deep learning techniques for Skin Cancer Detection." International Journal of Computer and Systems Engineering 18, no. 1 (2024): 51-59.

[10] Siddique, Ansar, Kamran Shaukat, and Tony Jan. "An intelligent mechanism to detect multi-factor skin cancer." Diagnostics 14, no. 13 (2024): 1359.

[11] Kaya, Volkan, and İsmail Akgül. "Classification of skin cancer using VGGNet model structures." Gümüşhane Üniversitesi Fen Bilimleri Dergisi 13, no. 1 (2022): 190-198.

[12] Diab, Amal G., Nehal Fayez, and Mervat Mohamed El-Seddek. "Accurate skin cancer diagnosis based on convolutional neural networks." Indonesian Journal of Electrical Engineering and Computer Science 25, no. 3 (2022): 1429-1441.

[13] Jain, Satin, Udit Singhania, Balakrushna Tripathy, Emad Abouel Nasr, Mohamed K. Aboudaif, and Ali K. Kamrani. "Deep learning-based transfer learning for classification of skin cancer." Sensors 21, no. 23 (2021): 8142.

[14] Alshehri, Abdullah. "Skin-NeT: Skin Cancer Diagnosis Using VGG and ResNet-Based Ensemble Learning Approaches." Traitement du Signal 41, no. 4 (2024).

[15] Mehra, Anubhav, Akash Bhati, Amit Kumar, and Ruchika Malhotra. "Skin cancer

classification through transfer learning using ResNet-50." In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2, Singapore: Springer Nature Singapore, (2021): 55-62.

[16] Panthakkan, Alavikunhu, S. M. Anzar, Sangeetha Jamal, and Wathiq Mansoor. "Concatenated Xception-ResNet50—A novel hybrid approach for accurate skin cancer prediction." Computers in Biology and Medicine 150 (2022): 106170.

[17] Gururaj, Harinahalli Lokesh, N. Manju, A. Nagarjun, VN Manjunath Aradhya, and Francesco Flammini. "DeepSkin: a deep learning approach for skin cancer classification." IEEE access 11 (2023): 50205-50214.

[18] Budhiman, Arief, Suyanto Suyanto, and Anditya Arifianto. "Melanoma cancer classification using resnet with data augmentation." In 2019 international seminar on research of information technology and intelligent systems (ISRITI), IEEE, (2019): 17-20.

[19] Pradhan, Adarsh, Subhojit Saha, Abhinay Das, and Santanu Barman. "Classification of Skin Lesion Using Image Processing and ResNet50." In International Conference on Big Data, Machine Learning, and Applications, Singapore: Springer Nature Singapore, (2021): 341-353.

[20] Ogundokun, Roseline Oluwaseun, Aiman Li, Ronke Seyi Babatunde, Chinecherem Umezuruike, Peter O. Sadiku, AbdulRahman Tosho Abdulahi, and Akinbowale Nathaniel Babatunde. "Enhancing skin cancer detection and classification in dermoscopic images through concatenated MobileNetV2 and xception models." Bioengineering 10, no. 8 (2023): 979.

[21] Mehmood, Abid, Yonis Gulzar, Qazi Mudassar Ilyas, Abdoh Jabbari, Muneer Ahmad, and Sajid Iqbal. "SBXception: a shallower and broader xception architecture for efficient classification of skin lesions." Cancers 15, no. 14 (2023): 3604.

[22] Chaturvedi, Saket S., Jitendra V. Tembhurne, and Tausif Diwan. "A multi-class skin Cancer classification using deep convolutional neural networks." Multimedia Tools and Applications 79, no. 39 (2020): 28477-28498.

[23] Moataz, Laila, Gouda I. Salama, and Mohamed H. Abd Elazeem. "Skin cancer diseases classification using deep convolutional neural network with transfer learning model." In Journal of physics: conference series, vol. 2128, no. 1, p. 012013. IOP Publishing, 2021.

[24] Singh, Priya, Mayand Kumar, and Aman Bhatia. "A comparative analysis of deep learning algorithms for skin cancer detection." In 2022 6th international conference on

intelligent computing and control systems (iciccs), IEEE, (2022): 1160-1166.

[25] Lu, X., Firoozeh Abolhasani Zadeh, Y. A. (2022). Deep Learning-Based Classification for Melanoma Detection Using XceptionNet. Journal of Healthcare Engineering, 2022(1), 2196096.

[26] Bello, A., Ng, S. C., Leung, M. F. (2024). Skin Cancer Classification Using Fine-Tuned Transfer Learning of DENSENET-121. Applied Sci ences, 14(17), 7707.

[27] Jain, S., Agrawal, K. (2023, November). An Efficient Diagnosis of Melanoma Skin Disease Using DenseNet-121. In 2023 3rd Interna tional Conference on Technological Advancements in Computational Sciences (ICTACS). IEEE - 908-912.

[28] Niranjana, R., Hemadarshana, T., Ilakkya, S., Krishnan, R. S., Epziba, J. J., Preetha, T. (2024, June). Enhanced Skin Diseases Prediction using DenseNet-121: Leveraging Dataset Diversity for High Accuracy Clas sification. In 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC). IEEE - 1270-1278.

[29] Kujani, T., Kumar, V. D., Kavitha, A., Sathya, T. (2024, August). Dermoscopic Skin Lesion Classification using Color Features Based on Fine-Tuning of Resnet50 and Densenet121 Transfer Learning Models. In 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) (pp. 1-7). IEEE.