Deep learning-based skin cancer classification: a comparative study of pre-trained CNN models*

Mukhriddin Arabboev[1,*,†], Shohruh Begmatov[1,†], Mokhirjon Rikhsivoev[1,†], Sukhrob Bobojanov[2,†], Gayrat Tangriberganov[3,†] and Jamoliddin Uraimov[4,†]

[1] *Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Amir Temur 108, 100084 Tashkent, Uzbekistan*

[2] *Urgench Branch of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Al-Khorazmi 110, 220100 Urgench, Uzbekistan*

[3] *Gmission, Ttukseom-ro 1-gil 25, 04778 Seoul, Republic of Korea*

[4] *Chungnam National University, Daehak-ro 99, 34134 Daejeon, Republic of Korea*

Abstract

Skin cancer classification using deep learning techniques plays a vital role in facilitating early diagnosis and improving patient outcomes. This study presents a comprehensive performance evaluation of various pre-trained convolutional neural network (CNN) models for classifying benign and malignant skin cancer types. Specifically, it focuses on VGG19, InceptionV3, and DenseNet169, analyzing their effectiveness in feature extraction and classification accuracy over 30 training epochs. Model performance is assessed using key evaluation metrics, including accuracy, precision, recall, and F1-score. The results indicate that DenseNet169 achieved the highest accuracy (87.92%), precision (86.02%), recall (90.56%), and F1-score (88.23%) in distinguishing benign from malignant cases after 30 epochs, outperforming VGG19 and InceptionV3. These findings highlight the trade-offs between model complexity and performance, offering valuable insights into selecting the most suitable model for real-world clinical applications. The study underscores the potential of deep learning in dermatological diagnostics, emphasizing its role in enhancing early detection and decision-making in benign and malignant classification.

Keywords

benign, malignant, classification, VGG19, InceptionV3, DenseNet169

1. Introduction

Effective disease monitoring and prediction play a crucial role in modern healthcare, enabling early intervention and improved patient outcomes [1,2]. With the increasing availability of medical imaging data, artificial intelligence (AI)-driven systems have become valuable tools for detecting, classifying, and predicting disease progression. By leveraging deep learning techniques, these systems can assist healthcare professionals in identifying patterns, reducing diagnostic errors, and enhancing clinical decision-making. One such application is in the detection and classification of skin

IVUS2025: Information Society and University Studies 2024, May 15, Kaunas, Lithuania

* Corresponding author.
† These authors contributed equally.
m.arabboev@tuit.uz (M. Arabboev); sh.begmatov@tuit.uz (S. Begmatov); suxrob.w@gmail.com (S. Bobojanov); gayrat.tangriberganov@gmail.com (G. Tangriberganov); uraimovuzbkor@gmail.com (J. Uraimov)
0000-0001-5733-5889 (M. Arabboev); 0000-0002-2441-916X (S. Begmatov)

cancer, a disease that has seen a significant rise in incidence over recent years.

Skin cancer is one of the most prevalent types of cancer worldwide, with early detection playing a critical role in ensuring successful treatment outcomes. According to the World Health Organization (WHO) [3], skin cancers are the most frequently diagnosed cancers globally, with an estimated 1.5 million new cases in 2022. An early and accurate diagnosis significantly improves survival rates and reduces the financial burden of treatment. However, traditional manual diagnosis by dermatologists can be time-consuming and subjective, leading to variations in diagnostic accuracy across practitioners.

The rapid advancements in artificial intelligence (AI) and deep learning have paved the way for computer-aided diagnosis (CAD) systems, which serve as powerful tools to assist medical professionals. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in image classification tasks, particularly in medical imaging applications. A key approach in deep learning-based medical diagnostics is transfer learning, which utilizes pre-trained models trained on large-scale datasets like ImageNet. This method is particularly advantageous in medical image analysis, where labeled data is often scarce.

This study investigates the effectiveness of various pre-trained CNN models in classifying benign and malignant skin cancer cases. A comprehensive performance evaluation is conducted on state-of-the-art architectures, including VGG19, InceptionV3, and DenseNet169, analyzing their accuracy, precision, recall, F1-score, and computational efficiency over 30 training epochs. Additionally, the study explores the potential of deep learning techniques for predicting disease progression, which could further aid in early diagnosis and long-term patient monitoring. The objective is to identify the most optimal deep learning model for automated skin cancer detection, balancing classification performance and computational feasibility. The findings of this study aim to contribute to the deployment of AI-driven diagnostic tools in clinical practice, enhancing early detection, disease monitoring, and improving treatment outcomes for skin cancer patients.

The remainder of the study is structured as follows: Section 2 reviews related work, providing an overview of previous research on skin cancer classification using deep learning techniques. Section 3 details the methodology, including data preprocessing, model selection, training procedures over 30 epochs, and evaluation metrics. Section 4 presents the results and discussion, analyzing the performance of VGG19, InceptionV3, and DenseNet169 in classifying benign and malignant skin cancer cases. Finally, Section 5 concludes the study by summarizing the key findings, discussing their implications, identifying limitations, and providing recommendations for future research directions.

2. Related work

Over the years, extensive research in medical image processing has significantly contributed to advancements in disease diagnosis and treatment. Recently, deep learning architectures have been widely applied in cancer detection, utilizing various models and datasets to enhance early diagnosis and improve clinical outcomes. The integration of deep learning in medical image analysis has demonstrated remarkable potential in automating and refining diagnostic accuracy. This section provides a comprehensive review of prior studies in skin cancer diagnosis and classification, that are closely related to the scope of this research.

In [4], the VGG19 pre-trained model was enhanced with max pooling and dense layers for skin cancer prediction and compared with other models like ResNet152v2, InceptionResNetV2, DenseNet201, ResNet50, and InceptionV3. The extracted features were classified using machine learning methods, demonstrating that combining E-VGG19 with traditional classifiers significantly improves classification performance. In [5], an automated system for skin disease classification was developed using a customized AlexNet-based Deep Convolutional Neural Network (DCNN). Trained on 1920 images and validated on 480 images, the model achieved 87.1% accuracy, outperforming traditional classification methods. In [6], the research addresses early skin cancer prediction by developing a CNN model within the PyTorch framework to classify skin moles as benign or malignant. Trained on 3600 ISIC-Archive images using an RTX 3080 GPU, the model achieved an accuracy of 85%, providing a practical solution for early diagnosis. In [7], transfer learning with CNNs is applied for skin cancer classification using InceptionV3 and Xception architectures InceptionV3 outperforms state-of-the-art methods, demonstrating superior performance in early detection and classification of benign and malignant skin lesions. In [8], a deep learning-based system is developed for skin lesion classification using a CNN and hand-coded features processed by a random forest classifier. The final classification result is obtained by averaging both outputs, while a convolutional-deconvolutional model is used for segmentation, demonstrating effective performance in melanoma screening. In [9], a deep learning-based solution using the VGGNET-16 architecture is proposed for classifying dermoscopic images into seven lesion types. Utilizing the HAM10000 dataset and K-Fold Cross Validation, the model achieves improved accuracy and reliability in melanoma detection, with a validation accuracy of 85.62%. In [10], a CNN-based deep learning method is used for classifying malignant and benign from the ISIC2018 dataset, incorporating ESRGAN for image enhancement. Multiple transfer learning models, including ResNet50, InceptionV3, and Inception ResNet, are fine-tuned, with the proposed approach demonstrating competitive classification accuracy. In [11], a CNN-based model using an enhanced VGG-16 architecture is proposed for early skin cancer detection from dermoscopic images. Comparative analysis on the ISIC dataset shows that the proposed model outperforms existing techniques in accuracy. In [12], a comparative study evaluates six classifiers with seven feature extraction methods and four data preprocessing steps on the two largest skin cancer datasets. The best results on the HAM10000 dataset were achieved using Linear Normalization, HSV feature extraction, and Balanced Random Forest, yielding 74.75% accuracy. In [13], deep-learning-based tools for data purification and augmentation are developed to enhance lesion classification. The system utilizes image occlusion removal and generative adversarial networks to balance lesion classes and create virtual patients with predefined lesions. In [14], a computer-aided detection and diagnosis system is proposed for classifying skin lesions using a fusion of the bag-of-features method with speeded-up robust features for extraction and a quadratic support vector machine for classification. The approach is evaluated on the PH^2 dataset, demonstrating effective lesion classification performance. In [15], the ABCD rule-based approach is applied for automatic skin cancer detection using the PH^2 dataset. The system consists of a preprocessing stage for image enhancement and a feature extraction stage using Total Dermoscopic Score (TDS) to classify lesions as benign or malignant, demonstrating satisfactory performance. In [16], two methods are proposed for automatic skin disease classification using the HAM10000 dataset: a standalone Convolutional Neural Network (CNN) and a combination of CNN with a one-versus-all approach. The latter method outperforms the standalone CNN, demonstrating promising results for classifying seven skin disease types from dermoscopy images. In [17], a CNN model is developed for early skin cancer detection using Keras and TensorFlow, incorporating various network topologies and transfer learning for improved convergence. The model is trained and tested on the ISIC challenge archive dataset to enhance classification accuracy. In [18], an automated skin lesion classification method is proposed using a pre-trained AlexNet model with transfer learning, fine-tuning, and data augmentation. Trained on the PH2 dataset, the modified model outperforms existing methods in classifying melanoma, common nevus, and atypical nevus. In [19], a CNN-based model is proposed to classify seven types of skin lesions, with a web-based interface for real-time predictions. Trained on the HAM10000 dataset, the model provides the three most probable lesion types for a given image. In [20], a skin lesion classification approach using a modified MobileNet model is proposed, incorporating data up-sampling and augmentation for improved performance. Trained on the HAM10000 dataset, the modified model outperforms the traditional MobileNet in accuracy, specificity, sensitivity, and F1-score. In [21], a deep learning model using transfer learning with AlexNet is proposed for skin cancer detection from HAM10000 dermoscopy images. The model classifies lesions as benign or malignant without segmentation or manual feature extraction, demonstrating high potential for assisting dermatologists in diagnosis.

Overall, the studies mentioned above ([4]–[21]) emphasize the importance of achieving high accuracy in skin cancer classification using deep learning techniques (see Table 1). However, none of these contributions ([4]–[21]) have specifically explored the comparative analysis of pre-trained convolutional neural networks (CNNs), including VGG19, InceptionV3, and DenseNet169, for distinguishing benign and malignant skin cancer cases.

Table 1
Comparative analysis of related works

Reference	Model/Algorithm	Epochs	Metrics
[4]	DenseNet201	50	Accuracy: 82.44% AUC: 0.93
[5]	DCNN	20	Accuracy: 87.1%
[6]	CNN	20	Accuracy: 85%
[7]	Xception	50	Accuracy: 86.97% Precision: 87.06% Recall: 86.97% F1-score: 86.99% AUC: 0.8703
[8]	CNN	N/A	Accuracy: 80.3%
[9]	CNN	10	Accuracy: 85.62%
[10]	InceptionV3	50	Accuracy: 85.7%
[11]	Improved VGG16	50	Accuracy: 87.05%
[12]	Random Forest	N/A	Accuracy: 74.75%
[13]	Deep convolutional GAN	N/A	Accuracy: 86.1%
[14]	SVM	N/A	Accuracy: 85.7%
[15]	ABCD rule	N/A	Accuracy: 84%
[16]	CNN	N/A	Accuracy: 86.5%
[17]	ResNet50	N/A	Accuracy: 85%
[18]	CNN	32	Accuracy: 80%
[19]	CNN	50	Accuracy: 78%
[20]	MobileNet	30	Accuracy: 83.93%
[21]	CNN	40	Accuracy: 84% AUC: 0.91
Proposed work	VGG19	30	Accuracy: 80% Precision: 76.09% Recall: 87.5% F1-score: 81.40% AUC: 0.8879
	InceptionV3	30	Accuracy: 83.06% Precision: 85.84% Recall: 79.17% F1-score: 82.37% AUC: 0.9245
	DenseNet169	30	Accuracy: 87.92% Precision: 86.02% Recall: 90.56% F1-score: 88.23% AUC: 0.9558

3. Methodology

The methodology of this study focuses on detecting and classifying skin cancer using deep learning models. The performance of these models is evaluated and compared using various metrics, including accuracy, precision, recall, and F1-score.

3.1. Dataset

The dataset used in this study, titled "Skin Cancer: Malignant vs. Benign", is available on Kaggle [22] and serves as a valuable resource for research and development in skin cancer detection using machine learning and deep learning techniques. It consists of skin lesion images categorized into malignant (cancerous) and benign (non-cancerous) classes, representing common clinical conditions. The images, provided in JPEG format with varying resolutions, enable training, validation, and testing of classification models. This dataset facilitates research in image processing, computer vision, and AI-based dermatological diagnostics. Table 2 presents an overview of the dataset.

Table 2
Details of used dataset

Classes	Train	Test	Total
Benign	1440	360	1800
Malignant	1197	300	1497
Total	2637	660	3297

Figure 1 presents a set of randomly selected training images representing both benign and malignant cases.

Figure 1: Samples from trained benign and malignant images.

3.2. Preprocessing and data Augmentation

All skin lesion images were normalized and resized to 224×224 pixels to ensure consistency and compatibility with deep learning models. Data augmentation techniques were applied to improve model generalization and prevent overfitting. The augmentation process included random rotations (±40°), width and height shifts (20%), shear transformations (20%), and zoom adjustments (20%) to introduce variations in image orientation, positioning, and scale. Additionally, horizontal flipping accounted for asymmetry in skin lesions, while the fill mode ('nearest') preserved image quality by filling missing pixels with nearby values. These preprocessing and augmentation steps enhanced data diversity, making the model more robust in classifying benign and malignant skin lesions.

Table 3
Details of used dataset after data augmentation

Classes	Train	Test	Total
Benign	1440	360	1800
Malignant	1440	360	1800
Total	2880	720	3600

3.3. Architectures of used models

In our approach to malignant melanoma classification and detection, we employ three well-established pre-trained deep learning models: VGG19, InceptionV3, and DenseNet169. These models, known for their advanced feature extraction capabilities, have been pre-trained on extensive image datasets, enabling them to capture intricate visual patterns essential for melanoma identification. To enhance their performance for this specific task, we apply fine-tuning techniques, carefully adjusting model parameters using a domain-specific dataset. This process allows the models to adapt to the subtle variations in melanoma characteristics while preserving the generalizable knowledge acquired during their initial pre-training.

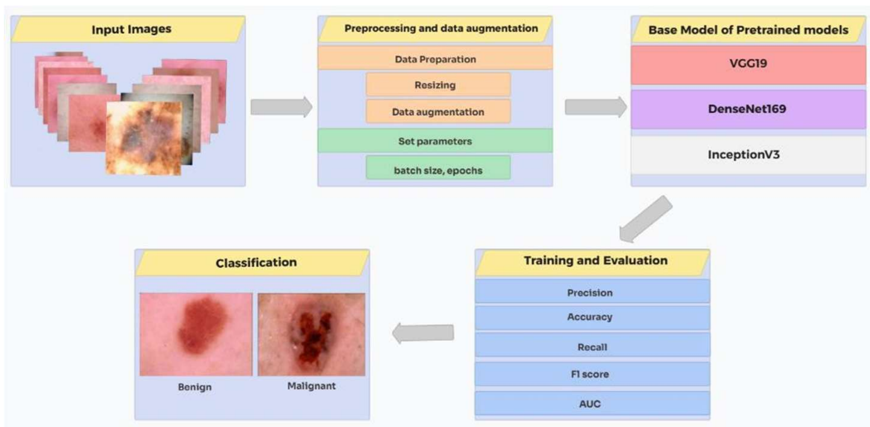


Figure 2: The design of the study.

3.4. Evaluation metrics

The evaluation metrics employed in this study include accuracy, recall, precision, and F1 score, with the confusion matrix used for further analysis. The performance evaluation is conducted using the following equations:

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN}, \tag{1}$$

$$Recall=\frac{TP}{TP+FN}, \tag{2}$$

$$Precision=\frac{TP}{TP+FP}, \tag{3}$$

$$F1=2\times\frac{Precision\times Recall}{Precision+Recall}, \tag{4}$$

where TP denotes true positive values, FP denotes false positive values, TN denotes true negative values, and FN denotes false negative values.

4. Results and Discussion

The training process employed the Stochastic Gradient Descent (SGD) optimizer with an initial learning rate set at 0.001 and a momentum of 0.9, maintaining a batch size of 16 over 50 epochs. This approach aimed to ensure a fair and consistent comparison between models, with identical sets of hyperparameters and adaptation strategies applied to all.

The training pipeline involved transfer learning with three pre-trained convolutional neural networks (VGG19, DenseNet169, and InceptionV3) as feature extractors. The base models were frozen except for the batch normalization layers, which remained trainable to maintain adaptive learning. A custom classification head, consisting of a Global Average Pooling layer, a fully connected layer with 256 neurons and ReLU activation, a dropout layer (0.5 probability), and a sigmoid output layer for binary classification, was added.

The experimental results illustrating the training and testing behaviors of the utilized models are presented in Figures 3.

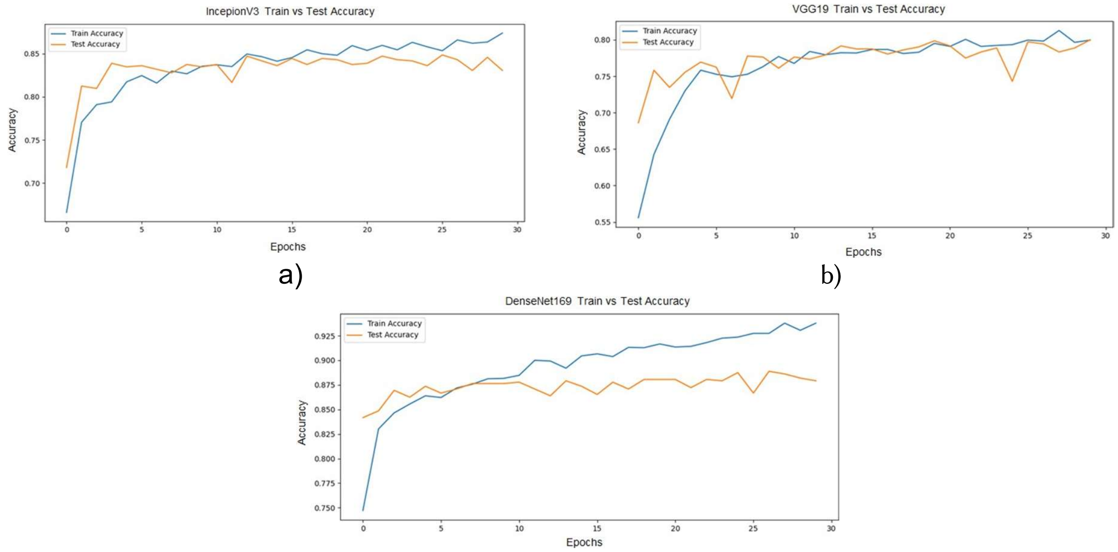


Figure 3: Training and testing accuracy of the proposed models: a) InceptionV3, b) VGG19, c) DenseNet169.

Accuracy measures the overall correctness of a model's predictions. However, it is essential to recognize that accuracy alone may not be a sufficient indicator of performance, particularly in the context of imbalanced datasets. A more comprehensive evaluation is achieved through multiple performance metrics, including precision, recall, F1 score, accuracy, and the ROC/AUC. These metrics range in value from 0 to 1, providing a nuanced assessment of the model's predictive capabilities. The experimental results corresponding to these metrics are detailed in Tables 4.

Table 4

Training and testing results of the proposed models

Model	Accuracy (Train/Test)	Precision (Train/Test)	Recall (Train/Test)	F1-score (Train/Test)	AUC (Train/Test)
InceptionV3	0.8993/ 0.8306	0.9396/ 0.8584	0.8535/ 0.7917	0.8945/ 0.8237	0.9711/ 0.9245
VGG19	0.8142/ 0.8000	0.7774/ 0.7609	0.8806/ 0.8750	0.8258/ 0.8140	0.8842/ 0.8879
DenseNet169	0.9587/ 0.8792	0.9448/ 0.8602	0.9743/ 0.9056	0.9593/ 0.8823	0.9927/ 0.9558

Table 4 presents the performance metrics of three models (InceptionV3, VGG19, and DenseNet169) based on various evaluation criteria. As indicated by the results in Table 3, DenseNet169 achieves the highest performance across most metrics, with accuracy (0.8792), precision (0.8602), recall (0.9056), F1-score (0.8823), and ROC/AUC (0.9558). These results suggest that DenseNet169 excels in identifying positive instances while maintaining a strong balance between precision and recall.

In contrast, InceptionV3 demonstrates competitive performance, particularly in precision (0.8584) and AUC (0.9245), but with a lower recall (0.7917) and F1-score (0.8237), which indicates a tendency to miss some positive instances. Similarly, VGG19 shows strong recall

(0.8750), meaning it effectively identifies positive cases, but its lower precision (0.7609) suggests a higher number of false positives, leading to a moderate F1-score (0.8140).

Among the three models, DenseNet169 achieves the best overall performance, particularly in recall and AUC, making it the most suitable choice for applications requiring high sensitivity and robust predictive capability. However, depending on the specific application, InceptionV3 may be preferable for scenarios prioritizing precision, while VGG19 could be advantageous when recall is more critical.

Based on the comparison of ROC/AUC values, the DenseNet169 model demonstrated superior performance, indicating its enhanced ability to distinguish between classes and implying a higher discrimination capability.

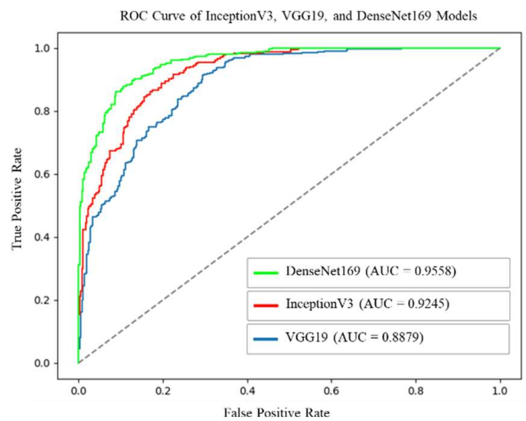
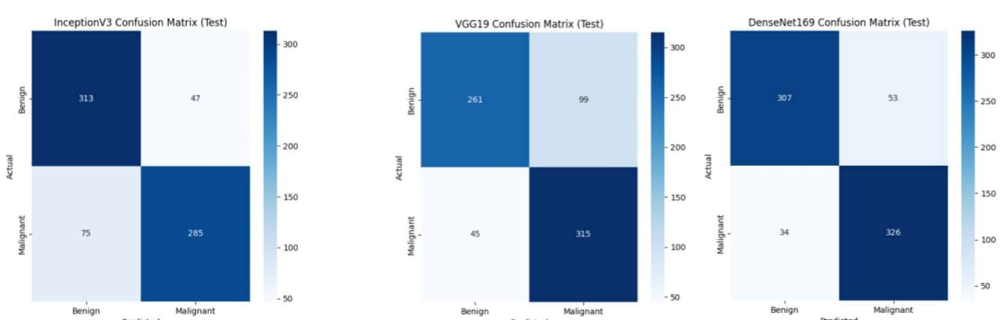


Figure 4: ROC curve of the proposed InceptionV3, VGG19, and DenseNet169 models.

As presented in Table 4 and Figure 4, the ROC/AUC values serve as key indicators of each model's ability to effectively differentiate between classes. Among the evaluated models, DenseNet169 exhibits the highest AUC value (0.9558), signifying its superior discriminatory power. This is particularly critical in binary classification tasks, where accurately distinguishing between positive and negative instances is essential. A higher AUC value denotes enhanced classification performance, with DenseNet169 demonstrating a more robust capability compared to InceptionV3 (0.9245) and VGG19 (0.8879) in distinguishing between classes.

The confusion matrix, also referred to as an error matrix, provides a rigorous quantitative framework for evaluating the accuracy of image classification models. It functions as a structured summary of classification outcomes, detailing the frequency of correct and incorrect predictions across all classes. Figure 5 presents the confusion matrices for the three analyzed models.



a)

b)

c)

Figure 5: Confusion Matrix obtained by the proposed model: a) InceptionV3, b) VGG19, c) DenseNet169.

In the evaluation of the proposed models, confusion matrices provide insights into their classification performance on the test dataset. As depicted in Figure 5, the DenseNet169 model demonstrated the highest classification accuracy, effectively distinguishing between benign and malignant cases. Specifically, DenseNet169 correctly classified 633 out of 660 test images, with 307 benign cases accurately predicted and 34 malignant cases misclassified as benign. Similarly, 326 malignant cases were correctly identified, while 53 benign cases were erroneously classified as malignant.

Comparatively, the InceptionV3 model exhibited slightly lower performance, accurately classifying 598 instances, with 313 benign cases correctly identified, 47 benign cases misclassified as malignant, and 285 malignant cases correctly classified, while 75 malignant cases were misclassified as benign. VGG19, on the other hand, showed a moderate classification ability, correctly identifying 576 images, with 261 benign and 315 malignant cases classified correctly, while 99 benign and 45 malignant cases were misclassified.

These findings highlight the superior discriminatory power of DenseNet169 in medical image classification, demonstrating its robustness in distinguishing between benign and malignant lesions with higher accuracy and fewer misclassifications compared to InceptionV3 and VGG19. This underscores its potential applicability in medical diagnostic tasks where precise classification is crucial.

Table 1 presents a comparative analysis of various deep learning models in terms of accuracy and other performance metrics. Among the existing models, MobileNet, ResNet50, and InceptionV3 demonstrate competitive accuracy scores of 83.93%, 85.00%, and 85.70%, respectively. Furthermore, the improved VGG16 model achieves an accuracy of 87.05%, while the highest-performing model among the listed works is a DCNN with transfer learning (AlexNet), which attains an accuracy of 87.10%.

In comparison, our proposed DenseNet169 model surpasses these benchmarks with an accuracy of 87.92%, while the proposed InceptionV3 model achieves 83.06%.

5. Conclusion

This study evaluated VGG19, InceptionV3, and DenseNet169 models for skin cancer classification using transfer learning. DenseNet169 achieved the best performance with 87.92% accuracy, outperforming the others in precision and recall. The findings highlight deep learning's potential in dermatology, aiding early detection. Future work includes fine-tuning models, expanding datasets, integrating explainable AI (XAI), and developing real-world applications such as mobile diagnostic tools. Enhancing lesion segmentation and exploring hybrid approaches (e.g., transformers, ensemble learning) could further improve performance.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] K. Nosirov, S. Begmatov, and M. Arabboev, "Design of a model for multi-parametric health monitoring system," *2020 Int. Conf. Inf. Sci. Commun. Technol. ICISCT 2020*, 2020. doi:10.1109/ICISCT50599.2020.9351522.

[2] K. Nosirov *et al.*, "Real-time multi parametric human health monitoring and prediction system," in *World Scientific Proceedings Series on Computer Engineering and Information Science, Developments of Artificial Intelligence Technologies in Computation and Robotics*, 2020, pp. 639–646. doi:10.1142/9789811223334_0077.

[3] "Skin cancer," *World Health Organization.* [Online]. Available: https://www.iarc.who.int/cancer-type/skin-cancer/.

[4] I. A. Kandhro *et al.*, "Performance evaluation of E-VGG19 model: Enhancing real-time skin

cancer detection and classification," *Heliyon*, vol. 10, no. 10, p. e31488, 2024.doi:10.1016/j.heliyon.2024.e31488.

[5] T. M. Ghazal, S. Hussain, M. F. Khan, M. A. Khan, R. A. T. Said, and M. Ahmad, "Detection of Benign and Malignant Tumors in Skin Empowered with Transfer Learning," *Comput. Intell. Neurosci.*, vol. 2022, 2022. doi:10.1155/2022/4826892.

[6] M. S. Al Huda *et al.*, "Enhancing Early Detection of Melanoma: A Deep Learning Approach for Skin Cancer Prediction," *Int. J. Informatics Vis.*, vol. 8, no. 3–2, pp. 1772–1778, 2024.

[7] A. Alrabai, A. Echtioui, and F. Kallel, "Skin Cancer Detection Based on Transfer Learning Techniques," *7th IEEE Int. Conf. Adv. Technol. Signal Image Process. ATSIP 2024*, no. July, pp. 8–13, 2024. doi:10.1109/ATSIP62566.2024.10638982.

[8] V. Singh and I. Nwogu, "Analyzing Skin Lesions in Dermoscopy Images Using Convolutional Neural Networks," *Proc. - 2018 IEEE Int. Conf. Syst. Man, Cybern. SMC 2018*, pp. 4035–4040, 2018. doi:10.1109/SMC.2018.00684.

[9] E. Çevik and K. Zengin, "Classification of Skin Lesions in Dermatoscopic Images with Deep Convolution Network," *Eur. J. Sci. Technol.*, no. October, pp. 309–318, 2019.

[10] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning," *Healthc.*, vol. 10, no. 7, pp. 1–18, 2022. doi:10.3390/healthcare10071183.

[11] H. Tabrizchi, S. Parvizpour, and J. Razmara, "An Improved VGG Model for Skin Cancer Detection," *Neural Process. Lett.*, vol. 55, no. 4, pp. 3715–3732, 2023.

[12] T. C. Pham, G. S. Tran, T. P. Nghiem, A. Doucet, C. M. Luong, and V. D. Hoang, "A Comparative Study for Classification of Skin Cancer," *Proc. 2019 Int. Conf. Syst. Sci. Eng. ICSSE 2019*, pp. 267–272, 2019. doi:10.1109/ICSSE.2019.8823124.

[13] D. Bisla, A. Choromanska, R. S. Berman, J. A. Stein, and D. Polsky, "Towards automated melanoma detection with deep learning: Data purification and augmentation," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 2720–2728, 2019. doi:10.1109/CVPRW.2019.00330.

[14] G. Arora, A. K. Dubey, Z. A. Jaffery, and A. Rocha, "Bag of feature and support vector machine based early diagnosis of skin cancer," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8385–8392, 2022. doi:10.1007/s00521-020-05212-y.

[15] E. M. Senan and M. E. Jadhav, "Analysis of dermoscopy images by using ABCD rule for early detection of skin cancer," *Glob. Transitions Proc.*, vol. 2, no. 1, pp. 1–7, 2021.

[16] K. Polat and K. Onur Koc, "Detection of Skin Diseases from Dermoscopy Image Using the combination of Convolutional Neural Network and One-versus-All," *J. Artif. Intell. Syst.*, vol. 2, no. 1, pp. 80–97, 2020. doi:10.33969/AIS.2020.21006.

[17] A. Abdelaziz and A. N. Mahmoud, "Skin Cancer Detection Using Deep Learning and Artificial Intelligence: Incorporated model of deep features fusion," *Fusion Pract. Appl.*, vol. 8, no. 2, pp. 08–15, 2022. doi:10.54216/FPA.080201.

[18] M. K. Hosny, A. M. Kassem, and M. M. Foaud, "Skin Cancer Classification Using Deep Learning," in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, 2019, pp. 90–93. doi:10.1109/CIBEC.2018.8641762.

[19] K. Pai and A. Giridharan, "Convolutional Neural Networks for classifying skin lesions," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2019-Octob, pp. 1794–1796, 2019.

[20] W. Sae-Lim, W. Wettayaprasit, and P. Aiyarak, "Convolutional Neural Networks Using MobileNet for Skin Lesion Classification," *JCSSE 2019 - 16th Int. Jt. Conf. Comput. Sci. Softw. Eng. Knowl. Evol. Towar. Singul. Man-Machine Intell.*, pp. 242–247, 2019.

[21] A. Ameri, "A deep learning approach to skin cancer detection in dermoscopy images," *J. Biomed. Phys. Eng.*, vol. 10, no. 6, pp. 801–806, 2020.

[22] C. Fanconi, "Skin Cancer: Malignant vs. Benign," *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign/data.