Determination of Skin Cancer Types with Deep Learning Methods and Patient Management

Ulku Veranyurt¹, Betul Akalin¹, Arzu Gercek^{2*} and Ozan Veranyurt³

¹University of Health Sciences, Health Care Management, Istanbul, Turkey ²University of Health Sciences, Dean of Health Science Faculty, Istanbul, Turkey ³Bahcesehir University, Cyber Security, Istanbul, Turkey

ABSTRACT

Background: Melanoma, a malignant type of skin tumor, can suddenly appear on normal skin without warning or develop on a pre-existing mole. Therefore, moles should be carefully monitored.

Objective: In this study, the aim was to use deep learning methods to successfully classify skin cancer cases and apply this approach to patient diagnosis.

Materials and Methods: The study used a publicly available dataset of 10,015 images. For training and testing, 80% of the dataset is divided into training and validation sets, and 20% into test sets. In the methodological sequence, we first trained a customized Convolutional Neural Network and then applied fine-tuning on ResNet50, VGG16, VGG19, and DenseNet101 deep learning models. Afterwards, each model was assessed using test accuracy and test F1-score metrics.

Results: At the end of the trials, we observed that the VGG16 fine-tuned deep learning model achieved 96% accuracy in training and 86% in the test set.

Conclusion: Artificial intelligence in health management and services is gaining popularity. In today's literature, there are different use cases of deep learning and artificial intelligence applied in the diagnosis and treatment of other diseases. This study suggests that deep learning methods can be used to speed skin cancer diagnosis and applied in the patient diagnosis process.

changes [2].

Keywords: Skin cancer, patient management, clinical decision support systems, deep learning.

INTRODUCTION

Melanoma, a malignant type of skin tumor, can suddenly appear on normal skin without warning or develop on a pre-existing mole. Therefore, moles should be carefully monitored.

The source of melanoma, the most life-threatening skin cancer, is the melanocyte cell. Melanocytes produce the pigment melanin, which gives our skin its color and enables us to tan. Melanoma results from abnormal, excessive, and uncontrolled proliferation of melanocytes and can spread to other organs.

While melanoma can appear suddenly on normal skin, it can also occur on a pre-existing mole. As a result, it is essential to identify the location and appearance of moles on the body so they can be appropriately monitored and diagnosed. In some cases, melanoma can be mistaken for a normal mole, but the mole could also have been melanoma from day one [1].

Pigmented lesions on the skin are the first sign of whether a mole is potentially malignant or benign. Generally,

important. One of the key areas that new algorithms and artificial intelligence methods are being focused on is clinical decision support systems. These systems aim to improve the quality of health care services and minimize foreseeable practitioner errors. These systems help

chemicals increases exponentially.

patients do not pay attention to skin pigmentation, and even when they do, practitioners may misinterpret early

Basal and squamous cell type skin cancers are the most

widespread types of skin cancer, and according to research, there are 5.4 million skin cancers diagnosed

every year, only in the United States of America. These

numbers show an increasing trend as people are more exposed to the sun, allergen chemicals are used more

frequently on the skin, and our daily exposure to

In various areas of health services and patient care, using

machine learning and deep learning methods to support

diagnosis and treatment has become increasingly

practitioners be more productive in diagnosing and

treating various illnesses.

The diagnosis of pigmented lesions poses a challenge for

dermatology practitioners, and the use of deep learning methods to support their decision-making can be a vital tool for early detection of skin cancer. This kind of

Email: arzu.gercek@sbu.edu.tr

Received: November 15, 2025; Revised: December 1, 2025; Accepted: December 1, 2025 DOI: https://doi.org/10.37184/lnjcc.2789-0112.7.9

^{*}Corresponding author: Arzu Gercek, University of Health Sciences, Dean of Health Science Faculty, Istanbul, Turkey,

application can serve as a decision-support mechanism for early diagnosis and contribute to the overall diagnostic process for skin cancer. In the following sections, we aimed to introduce the different deep learning approaches we applied to images of pigmented skin lesions to successfully diagnose the type of mole and confirm whether or not it is malignant.

LITERATURE REVIEW

While there are various applications of artificial intelligence methods in health services and patient management, we analyzed studies focused on the diagnosis and treatment of cancer, specifically skin cancer.

In one study, researchers focused on mobile health and machine learning *via* cloud-based services. They presented an application running on a wearable device that processes a skin image and uses cloud services to suggest a diagnosis for the problematic skin lesion. A pretrained Convolutional Neural Network (CNN) was used for analyzing the images [3].

In another attempt at classifying skin cancer, researchers proposed a method for early detection. In his paper, he proposed a 3-step approach comprising feature extraction, dimensionality reduction, and classification. The research used discrete wavelet transforms for feature extraction, feed-forward neural networks for classification, and k-nearest neighbors. The research achieved 95% accuracy [4].

In another study, authors utilized deep convolutional neural networks for automated classification of skin lesions. They trained their model using 129,450 clinical images from 2021 of different diseases. The basic CNN model they used was able to diagnose the most common skin cancer types [5].

In one study, it was used to support a vector machine learning algorithm for the correct classification of melanoma skin cancer cases. In this study, dermoscopy images, pre-processing, and statistical feature-extraction techniques were used for early cancer detection. Afterwards, principal component analysis (PCA) and Support Vector Machine (SVM) algorithms were used for classification. The model achieved 92.1% accuracy [6].

In another research effort, the authors focused on therapeutic techniques for treating skin cancer. Surgery is one of the main treatments presented for skin cancer, but it cannot be applied to all cases. In this study, the researchers review different radiation therapy techniques and discuss their effects on skin tumors [7].

METHODS

In the study, a secondary dataset of 10,015 publicly shared skin pigmentation images was used [8]. For

training and testing, 80% of the dataset is divided into training and validation sets, and 20% into test sets. In the methodological sequence, we first trained a customized Convolutional Neural Network and then applied finetuning on ResNet50, VGG16, VGG19, and DenseNet101 deep learning models. Afterwards, each model was assessed using test accuracy and test F1-score metrics.

Data Collection and Pre-processing

Data from 10015 skin lesions across seven skin cancer types were used in this study. The gender distribution of the data was 54% male, 45% female, and 1% unknown.

The distribution of 7 different cancer types and the number of samples for each type are summarized in Table 1.

Table 1: Case type distribution.

Cancer Type	Abbreviation	Count	
Melanocytic nevi	nv	6705	
Melanoma	mel	1113	
Benign keratosis- like lesions	bkl	1099	
Basal cell carcinoma	bcca	514	
Actinic keratoses	kiec	327	
Vascular lesions	vasc	142	
Dermatofibroma	df	115	

The highest number of cases in the dataset was from melanocytic nevi. As shown in Fig. (1), the distribution of cases per sex showed a similar pattern, with more images from male patients than from female patients.

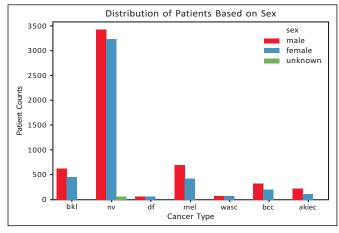


Fig. (1): Distribution of patients based on sex.

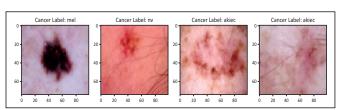


Fig. (2): Sample images from the data set.

For the deep learning models, RGB images of skin lesions were taken, resized, and normalized. In Fig. (2), we can observe sample images of skin lesions and the types of cancer.

Convolutional Neural Networks-Based Deep Learning Model

The first step of the study was to train a model from scratch and evaluate its performance. For each detection method, the same image size, enhancer, number of epochs, and batch size were used. The images were rescaled to a 75x100 size and normalized for the deep learning model. As hyperparameters, 20 epochs and a batch size of 32 were used in all different method trials. SGD (Stochastic Gradient Descent) optimizer was used. The architecture of the CNN model is summarized in Fig. (3). The normalized images were processed by a convolutional block comprising two hidden layers with 256 and 128 filters, respectively. The features extracted by the first block were then processed with a Max pooling and dropout layer, and the same process was repeated for another convolutional block. At the end, the obtained feature matrix was flattened and fed into a 4-layer fully connected neural network.

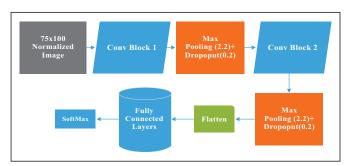


Fig. (3): CNN architecture.

Fine-tuning Application on Pre-trained Models

In the second part of the experiments, we used pretrained CNN models: VGG16, VGG19, ResNet50, and DenseNet101.

VGG16 and VGG19 are both deep convolutional neural networks developed at Oxford University. Versions of the network trained on more than one million images from the ImageNet database are open to methods such as transfer learning in various deep learning studies, and the pre-trained network can categorize images into 1000 object categories, such as keyboard, mouse, pen, and many animals. This pre-trained object classifier can be reused for different study purposes [9, 10].

ResNet50, as the name suggests, is a 50-layer CNN trained on the ImageNet dataset and has 25.6 million parameters. The pre-trained CNN model has achieved first place in image segmentation and detection on the COCO dataset. Architecture-wise, the model has 48 convolution layers, combined with one max pooling and

one average pooling layer [11, 12]. Model architecture is summarized in Fig. (4).

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer	
conv1	112 x 112	7 x 7, 64, stride 2					
		3 x 3 max pool, stride 2					
conv2.x	56 x 56	$\left[\begin{array}{c} 3 \times 3, 64 \\ 3 \times 3, 64 \end{array}\right] \times 2$	$\left[\begin{array}{c}3\times3,64\\3\times3,64\end{array}\right]\times3$	\[\begin{array}{c} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{array} \] x 3	$\begin{bmatrix} 1 & x & 1 & 64 \\ 3 & x & 3 & 64 \\ 1 & x & 1 & 256 \end{bmatrix} x & 3$	1 x 1, 64 3 x 3, 64 1 x 1, 256	
conv3.x	28 x 28	\[\begin{array}{c} 3 \ x \ 3, 128 \ 3 \ x \ 3, 128 \end{array} \] x 2	\[\begin{array}{c} 3 \ x & 3 & 128 \ 3 \ x & 3 & 128 \end{array} \] x 4	1 x 1, 128 3 x 3, 128 1 x 1, 512 x 4	1 x 1, 128 3 x 3, 128 1 x 1, 512 x 4	1 x 1, 128 3 x 3, 128 1 x 1, 512	
conv4.x	14 x 14	\[\begin{array}{c} 3 \times 3, 256 \\ 3 \times 3, 256 \end{array} \] \times 2	$\left[\begin{array}{c} 3 \times 3, 256 \\ 3 \times 3, 256 \end{array}\right] \times 6$	1 x 1, 256 3 x 3, 256 1 x 1, 1024 x 6	\[\begin{array}{c} 1 \ x \ 1, 256 \\ 3 \ x \ 3, 256 \\ 1 \ x \ 1, 1024 \end{array} \] \ x \ 23	1 x 1, 256 3 x 3, 256 1 x 1, 1024	
conv5.x	7 x 7	$\left[\begin{array}{c} 3 \times 3, 512 \\ 3 \times 3, 512 \end{array}\right] \times 2$	$\left[\begin{array}{c} 3 \times 3, 512 \\ 3 \times 3, 512 \end{array}\right] \times 3$	1 x 1, 512 3 x 3, 512 1 x 1, 2048 x 3	1 x 1, 512 3 x 3, 512 1 x 1, 2048 x 3	\[\begin{array}{c} 1 \ x \ 1, 512 \\ 3 \ x \ 3, 512 \\ 1 \ x \ 1, 2048 \end{array} \] \[x \ 1 \]	
	1 x 1	average pool, 1000-d fc, softmax					
FLO)Ps	1.8 x 10 ⁹	3.6 x 109	3.8 x 10 ⁹	7.6 x 10°	11.3 x 109	

Fig. (4): ResNet50 architecture (Src: https://iq.opengenus.org/resnet50-architecture/)

The DenseNet101 architecture is another famous pretrained deep learning model that consists of Dense blocks, each containing convolution and batch normalization layers. Each dense block is connected to a convolutional and pooling layer to reduce the feature map size. Unlike ResNet50, which operates on correlated features, DenseNet models are built on diverse features [13, 14].

In our application, these four pre-trained models were fine-tuned, meaning all hidden and pre-trained layers of each model were retrained from scratch. In our customized training method, the fully connected layers of each model were removed, and a 6-layer fully connected neural network was added for output and classification, with a softmax activation function at the end. For this part of the experiments, our hyperparameters were as follows: 20 epochs, a batch size of 32, and the SGD optimizer. Accuracy and F1 scores were used for evaluation. The training set was divided into 80% for training and 20% for validation.

STATISTICAL ANALYSIS

In each experiment, the dataset was split into training, validation, and test sets to ensure reliable model evaluation. To assess the stability and statistical significance of model performance, **a** five-fold cross-validation procedure was conducted across all deep learning architectures.

For comparing the performance of VGG16 with the baseline CNN and other pre-trained models, **a** paired ttest was applied to the cross-validation accuracy scores. The resulting p-values were consistently below 0.05, indicating that the VGG16 model's superior performance was statistically significant and unlikely to be due to random variation. These findings confirm that the observed improvement in classification accuracy reflects a genuine performance advantage rather than noise introduced by dataset splits or training variance.

RESULTS

In the first part of the experiments, we trained the CNN model and monitored its accuracy and loss throughout training, in Figs. (5) and (6), the accuracy and loss behavior for training and validation are shown.

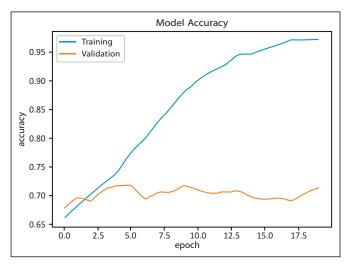


Fig. (5): CNN model training & validation accuracy.

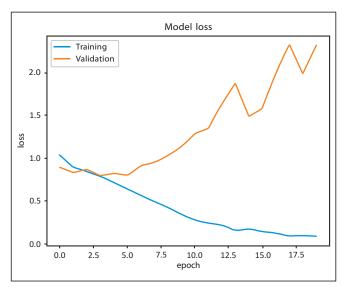


Fig. (6): CNN Model training & validation loss.

As shown in both figures, the training accuracy exceeds 90%, whereas the validation accuracy does not exceed 70%. Similar behaviour is observed for the loss values. Training loss decreased gradually, whereas validation loss increased after a certain number of epochs. On the test set, the CNN model achieved 0.69 accuracy and 0.66 F1 score on a scale of 0 to 1.

In the following experiment, we evaluated the training and validation results for the VGG16 model.

As shown in Fig. (7), the training accuracy for the VGG16 model was 95%, and the validation accuracy climbed

over 75% in the last epochs. In the test set, the model performed 0.86 accuracy and 0.81 F1 score.

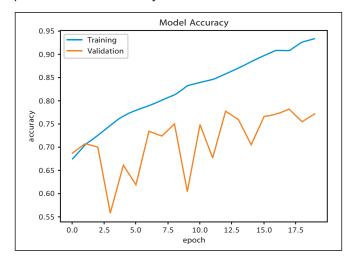


Fig. (7): VGG16 training & validation accuracy.

In the ResNet50 experiment, the training accuracy stabilized after the 15th epoch, and the validation accuracy did not exceed 77% (Fig. 8). The model achieved 0.81 accuracy and 0.77 F1 score on the test set.

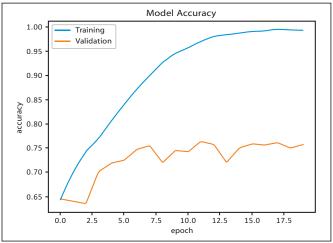


Fig. (8): ResNet50 training and validation accuracy.

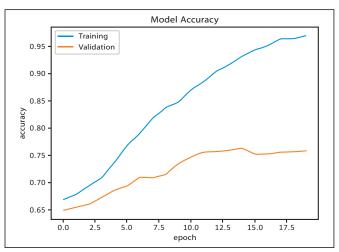


Fig. (9): DenseNet101 training and validation accuracy.

In the experiments with DenseNet101, we observed behavior similar to that of ResNet50 (**Fig. 9**). The training accuracy was around 95%, whereas the validation accuracy remained below 75%. In the test set, the model performed 0.77 accuracy and 0.75 F1 score.

In the final attempt, we analyzed the behavior of the VGG19 model and observed training behavior similar to that of the VGG16 model, as both are derived from the same architecture. The test results for VGG19 were 0.81 accuracy and 0.79 F1 score.

The comparison of all deep learning methods on the test set is displayed in Fig. (10).

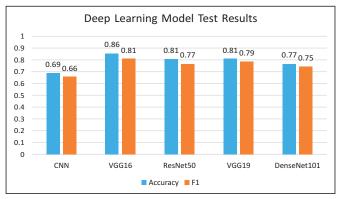


Fig. (10): Deep learning model test results.

As shown in the figure above, the VGG16 model achieved the best accuracy and F1 score. Pre-trained CNN models showed improved performance compared to the CNN model, and there was a small gap in test results between VGG16, VGG19, and ResNet50.

In this study, we propose using an application to aid in diagnosing skin cancer. As practitioners face challenges in identifying lesion type and deciding whether biopsy is required, an application grounded in deep learning can display the probabilities of each potential cancer type and support the practitioner's diagnosis.

In Fig. (11), a potential computer-aided image processing application using a VGG16-based model is demonstrated. The application displays the probability of each cancer type for the input image.

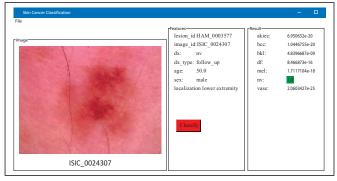


Fig. (11): Skin cancer classification application.

In our proposed process, the practitioner decides to use the help of the deep learning based skin cancer classifier and requests further diagnostic tests, such as a biopsy, if the application shows a potential cancer probability. This approach can reduce the number of additional diagnostic tests requested per patient and enhance the effectiveness of cost and material management within the health institution. Fig. (12) illustrates the proposed process at a high level.

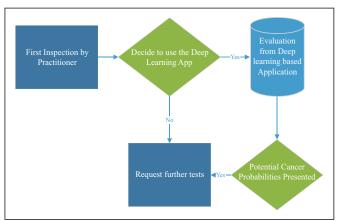


Fig. (12): Skin cancer diagnosis process.

DISCUSSION

Our study indicates that optimized deep learning architectures, particularly VGG16, provide a solid basis for automatic classification of pigmented skin lesions. All pre-trained networks achieved considerably higher test accuracy and F1-score than our custom CNN, in line with the beneficial effect of transfer learning when examining complex visual patterns such as keratosis or melanoma. The relatively small differences in performance among the VGG16, VGG19, and ResNet50 models suggest that the applicability of feature representations from large natural-image datasets is essential for achieving reliable diagnostic performance, rather than the depth of the architecture.

These results agree with previous studies. For instance, ensembles of pre-trained networks such as VGG-19, ResNet, DenseNet, and InceptionResNet achieved accuracies of ~98% on ISIC archive images, well above most single-model baselines, according to a comprehensive review of deep learning methods in skin cancer detection [15]. Along these lines, another recent study, using VGG16 and VGG19 transfer learning on a subject-specific dataset, reported significant improvements in lesion classification accuracy, noting that a well-designed transfer-learning framework can achieve high accuracy even without data augmentation [16].

Besides, a 2025 study combined ResNet-18 and MobileNet with a hybrid classifier and achieved ~92.9% accuracy using segmentation + transfer learning,

outperforming simpler CNN-only pipelines [17]. Transfer learning is superior compared to custom CNNs. Secondly, in the study using the HAM10000 dataset, the best test accuracy, ~77%, was observed with VGG-16, followed by VGG-19 and InceptionV3. These results support our conclusion that VGG-based models remain competitive among transfer learning approaches [18].

Yet, a few works identify limitations similar to those discussed in our study. For example, one work using EfficientNet architectures for dermoscopic images demonstrated outstanding performance, on the order of AUC ~ 0.968 for the task of melanoma vs. non-melanoma classification, but emphasized the importance of class imbalance, heterogeneity of the dataset, and the need for augmentation and metadata use in achieving generalization performance [19].

The imbalance between high training and much lower validation accuracy has been well-documented in the literature for skin-lesion classification studies like ours. Many authors attribute this to limited dataset size or diversity, class imbalance, and other factors that make lesions highly variable both intra- and inter-class [20]. These challenges illustrate how difficult it is to create a generalizable skin-lesion classification model, even with transfer learning, where overfitting can be a problem.

Considering this background, our results support the notion that deep learning, especially transfer learning of pre-trained CNNs, significantly enhances skin lesion classification and may be employed as a screening tool in computer-aided diagnosis systems. However, the success of such systems requires class balance and lesion diversity in the datasets, appropriate regularization/augmentation techniques, and external validation across different patient populations.

CONCLUSION

In this study, to accelerate the diagnosis of skin cancer, the aim was to achieve an accurate classification of skin cancer patients using skin lesion images from different body parts. We applied a CNN model and fine-tuning methods across different CNN architectures to determine the best model for supporting decision-making and diagnostic processes in skin cancer. Our approach is to run an image-processing application to analyze the skin lesions and display potential cancer types to the practitioner. This way, a deep learning-based model can be used not as a stand-alone decision-maker but to indicate the potential cancer type and help the practitioner with further diagnosis.

Compared with other studies in the literature, we aimed to evaluate different deep learning models for skin cancer diagnosis and to propose an application for the diagnostic process.

In the future steps of this study, we plan to apply parameter optimization and image pre-processing techniques to improve our results and feed an improved deep learning model into our proposed solution; moreover, we plan to use these techniques for early diagnosis of other cancer types.

ETHICS APPROVAL

Since a public dataset was used in the experiments, further ethical approval was not required for this study.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA

Authors confirm that data supporting the results of this study are available in the article.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

AUTHOR'S CONTRIBUTION

All authors equally contributed.

REFERENCES

- Linares MA, Zakaria A, Nizran P. Skin cancer. Prim Care 2015; 42(4): 645-59. DOI: https://doi.org/10.1016/j.pop.2015.07.006
- Polder KD, Landau JM, Vergilis- Kalner IJ, Goldberg LH, Friedman PM, Bruce S. Laser eradication of pigmented lesions: A review. Dermatol Surg 2011; 37(5): 572-95. DOI: https://doi.org/10.1111/j.1524-4725.2011.01971.x
- Dai X, Spasic I, Meyer B, Chapman S, Andres, F. Machine learning on mobile: An on-device inference app for skin cancer detection. Fourth International Conference on Fog and Mobile Edge Computing (FMEC). Rome, Italy, 2019. DOI: https://doi.org/10.1109/FMEC.2019.8795362
- Elgamal M. Automatic skin cancer images classification. Int J Adv Comput Sci Appl 2013; 4(3): 287-94.
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542(7639): 115-18. DOI: https://doi.org/10.1038/nature21056
- Alquran H, Qasmieh IA, Alqudah AM, Alhammouri S, Alawneh E, Abughazaleh A. Melanoma skin cancer detection and classification using a support vector machine. IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). Aqaba, Jordan, 2017.
- Pashazadeh A, Boese A, Friebe M. Radiation therapy techniques in the treatment of skin cancer: an overview of the current status and outlook. J Dermatolog Treat 2019; 30(8): 831-39. DOI: https://doi.org/10.1080/09546634.2019.1573310

- Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset is a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 2018; 5(1): 1-9. DOI: https://doi.org/10.1038/sdata.2018.161
- Rajinikanth V, Joseph Raj AN, Thanaraj KP, Naik GR. A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection. Appl Sci 2020; 10(10): 3429. DOI: https://doi.org/10.3390/app10103429
- Abuared N, Panthakkan A, Al-Saad M, Amin SA, Mansoor W. Skin cancer classification model based on VGG 19 and transfer learning. 3rd International Conference on Signal Processing and Information Security (ICSPIS). Dubai, United Arab Emirates, 2020.
- Theckedath D, Sedamkar RR. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. SN Comput Sci 2020; 1:79.
 - DOI: https://doi.org/10.1007/s42979-020-0114-9
- Abd Elghany S, Ibraheem MR, Alruwaili M, Elmogy M. Diagnosis of various skin cancer lesions based on fine-tuned Resnet50 Deep Network. Comput Mater Sci 2021; 68(1): 117-35.
 DOI: https://doi.org/10.32604/cmc.2021.016102
- Choudhary M, Tiwari V, Venkanna U. An approach for iris contact lens detection and classification using ensemble of customized DenseNet and SVM. Future Gener Comput Syst 2019; 101: 1259-70.
 - DOI: https://doi.org/10.1016/j.future.2019.07.003
- Pacheco AG, Ali AR, Trappenberg T. Skin cancer detection based on deep learning and entropy to detect outlier samples. arXiv preprint arXiv 2019; 1909.04525.
 - DOI: https://doi.org/10.48550/arXiv.1909.04525

- Naqvi M, Gilani SQ, Syed T, Marques O, Kim HC. Skin cancer detection using deep learning—a review. Diagnostics 2023; 13(11):1911.
 - DOI: https://doi.org/10.3390/diagnostics13111911
- Faghihi A, Fathollahi M, Rajabi R. Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models. Multimedia Tool Appl 2024; 83(19): 57495-510.
 DOI: https://doi.org/10.48550/arXiv.2404.01160
- Shakya M, Patel R, Joshi S. A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification. Sci Rep 2025; 15(1): 4633.
 DOI: https://doi.org/10.1038/s41598-024-82241-w
- Hasan SN. Accurate deep learning algorithms for skin lesion classification. IIETA 2024;29(4):1529-39.
 DOI: https://doi.org/10.18280/isi.290426
- Jaisakthi SM, Mirunalini P, Aravindan C, Appavu R. Classification of skin cancer from dermoscopic images using deep neural network architectures. Multimed Tools Appl 2023; 82(10): 15763-78.
 - DOI: https://doi.org/10.1007/s11042-022-13847-3
- Magalhaes C, Mendes J, Vardasca R. Systematic review of deep learning techniques in skin cancer detection. BioMedInformatics 2024; 4(4): 2251-70.