



Deep Learning-Based Automated Diagnosis of Skin Cancer from Thermoscopic Images

Safa Amer Baker Alsultan^{1*}

1 Computer Department, Faculty of Engineering, Islamic Azad University, Isfahan, Iran.

*Corresponding author email: Safa.am.21996@gmail.com

التشخيص الآلي القائم على التعلم العميق لسرطان الجلد من الصور الحرارية

صفا عامر باقر السلطان¹

1 قسم الحاسوب، كلية الهندسة، جامعة آزاد الإسلامية، أصفهان، إيران

Accepted:

8/11/2024

Published:

31/12/2024

ABSTRACT

Background:

Given the importance of early and accurate detection of skin cancer in preventing its risks and improving its treatment, the study aims to develop automated models based on deep learning to detect skin cancer and accurately classify skin lesions into benign and malignant and identify different types of lesions, which helps doctors in early diagnosis and decision-making.

Materials and Methods:

Relying on pre-trained convolutional neural networks, which makes it possible to train models using only a small amount of available training data. Two datasets of images of skin tumors were used, one to detect the presence of a tumor and the other to classify the tumor among different types of skin tumors. Both datasets underwent a set of initialization and processing steps and were then trained using two pre-trained convolutional neural network models.

Results:

The evaluation results of randomly selected images from the set of test images demonstrate the accuracy and high effectiveness of the models in prediction. DenseNet121 hybrid model achieved an impressive 94% accuracy, with an average precision, recall, and F1 score of 0.94. The EfficientNet B0 model achieved a weighted accuracy of 93%, with a weighted average precision, recall, and F1 score of 0.93 across nine classes.

Conclusion:

Pre-trained deep convolutional neural networks have made significant progress in improving the ability to detect skin cancers using dermoscopic images. This study showcases the capability of artificial intelligence in swiftly and precisely diagnosing skin cancer, as well as detecting (pre)malignant skin cancer.

Key words: Convolutional Neural Networks; Dermoscopy Images; Deep Learning; Skin Cancer; Transfer Learning.



INTRODUCTION

Skin cancer is the 19th most prevalent cancer worldwide and rising [1]. Both sexes of white individuals get skin cancer, especially in the United States, where incidence have skyrocketed [2]. Skin cancer is diverse, complex, and rapidly evolving, making it difficult to recognize. Despite Ultra-violet (UV) exposure regulations and screening efforts, global skin cancer rates are rising [3]. Early detection of skin cancer opens the door to a longer life and better treatment.

Traditional diagnostic methods such as dermoscopy are time-consuming and require self-diagnosis, which has limited their use and stimulated the development of alternative treatments [4]. Dual spectroscopic and imaging techniques, such as real-time Raman spectroscopy, have shown promising results. However, their ability to differentiate between malignant and benign lesions, especially those with overlapping characteristics such as malignant melanoma and seborrheic keratosis, still falls short of expectations [5]. Furthermore, changes in size, color and shape within the same category exacerbate the problem [6]. Machine learning has revolutionized skin cancer diagnosis by combining image classification techniques with deep learning models to produce fast, low-cost, and accurate results [7]. Recent advances in pre-trained deep learning algorithms provide a potentially transformative option for skin cancer detection. Due to their computational efficiency, these models are ideal for real-time applications, making them a promising route for rapid and cost-effective diagnosis.

Some previous research studies have dealt with classification and segmentation of skin tumors through artificial intelligence models. In [8] three different mobile learning models were used using the ISIC2018 dataset. In [9] a deep learning algorithm of convolutional neural network (CNN) was used to detect the two basic categories of tumors: benign and malignant. A data augmentation model for successful diagnosis of melanoma skin cancer was presented in [10]. The trial results for the binary classification scenario indicate a significant improvement in skin cancer identification. In [11] an automated solution for skin cancer prediction, detection and diagnosis was presented that combines deep learning methods with non-invasive and non-ionizing procedures for diagnosis. The study in [12] provided a description of the performance of four unified convolutional neural networks, specifically YOLOv3, YOLOv4, YOLOv5, and YOLOv7, in the categorization of skin lesions. The performance of each model was compared based on lesion localization, classification accuracy, and inference time after being trained on a reference dataset. A Deep convolutional neural networks (DCNN)-based model was created in [13] to precisely and automatically categorize various types of skin cancer as either melanoma or non-melanoma skin cancer. After analyzing the most important studies and works of literature and reviewing the results achieved through the proposed systems, we notice a decrease in the accuracy achieved through the performance experiments of each study. Each study relied on proposing a neural model and training it without delving into the idea of hybridizing the models and obtaining a hybrid model with better characteristics.

Consequently, we conducted a performance comparison of two methods. One of these methods operates on the same principle as the studies that were presented, which is to develop an architecture that is based on a deep learning model (EfficientNetB0) that has not been previously



proposed for skin disease classification tasks. The second approach involves the development of a deep learning model that utilizes both the algorithmic architecture and the (DenseNet121) and additional layers from recurrent neural network (RNN) algorithm. This study evaluates the DenseNet121 and EfficientNetB0 algorithms in terms of performance and efficiency to investigate the efficacy of potent pre-trained deep learning methods for the detection of skin cancer.

MATERIALS AND METHODS

Transfer learning is a type of deep learning where a model trained for one task is used for another similar task where the model is pre-trained and its weights are stored, so there is no need to retrain the entire model. There are different ways to implement transfer learning, including fine-tuning the model or using the model as a fixed feature extractor. This results in faster training times and higher accuracy using a smaller data set. There are a number of well-known and popular pre-trained models that are widely used in various applications due to their robust architecture and high performance in feature detection and extraction. DenseNets, which stands for “densely connected convolutional networks”, are characterized by a large number of fully connected layers. The numbers assigned usually reflect the number of layers in the network, including convolutional and translational layers. DenseNet-121 was pre-trained on the ImageNet dataset, which includes 1.2 million color images belonging to 1,000 categories. The initialization of the DenseNet-121 weights comes from the pre-trained model, and uses the SoftMax function to predict the probability of each class as output [14]. A comprehensive balance between model size and performance was sought when developing the EfficientNet B0 core model in the EfficientNet series so it is well-suited for resource-constrained applications, such as those involving mobile devices or edge devices [15].

The proposed system aims to use transfer learning to support clinical decision making in skin cancer diagnosis through binary classification (benign (or malignant) and multiple classification of different types of skin lesions. Figure 1 describes the basic stages of building the model for skin cancer detection.

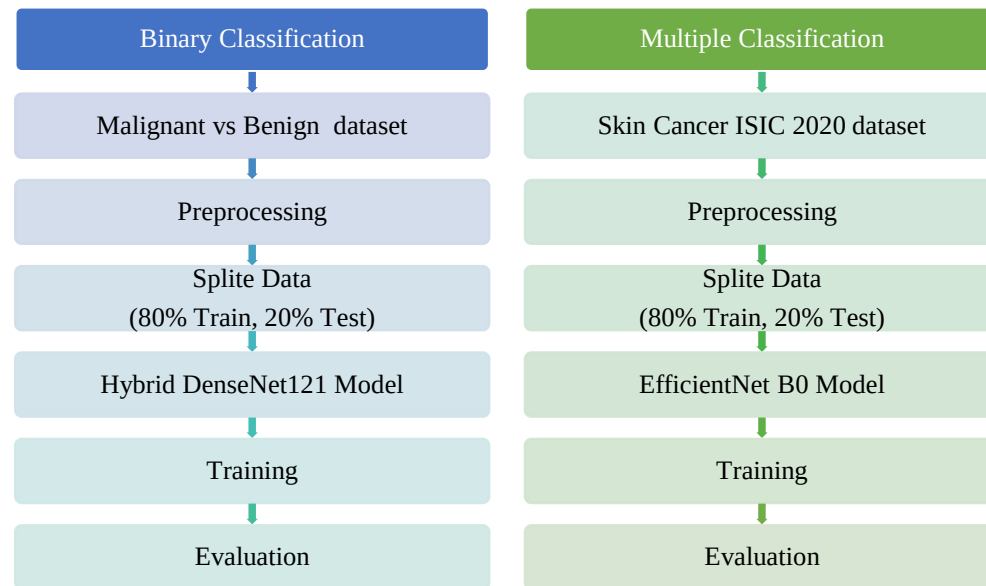


Figure 1. Basic stages of building the proposed model for skin cancer detection in the current study

The system begins by collecting skin lesion images as inputs and applying pre-processing steps, including quality enhancement, color adjustment, and dimension normalization, which help reduce noise and standardize data for easier pattern recognition. The system utilizes two CNN models: one for binary classification and the other for multi-class classification. In binary classification, the model determines whether a lesion is benign or malignant, while the other model identifies multiple types of lesions. To achieve higher accuracy in binary classification, the system integrates CNN with a recurrent neural network (RNN) to analyze repetitive patterns, improving the classification of benign and malignant lesions. The model is then trained on a 80% of data and evaluated using metrics such as accuracy, precision, recall, and F1-score. Finally, the system provides a report containing the final classification for each lesion, assisting doctors in making accurate diagnostic decisions.

Datasets

Two skin cancer datasets were selected. The first dataset the "Skin Cancer: Malignant vs. Benign" dataset from Kaggle [16], sourced from the ISIC (International Skin Imaging Collaboration) Archive, contains balanced and preprocessed images for binary classification of skin lesions into benign and malignant categories. It comprises of 1,800 images for each class. Figure 2 shows the graphical analysis of the Malignant vs. Benign dataset. The hybrid DenseNet121 model will be trained on this dataset.



For the purpose of detecting multiple skin cancer lesions, Skin Cancer ISIC 2020 dataset was used. This collection includes 2357 photos of malignant and benign oncological illnesses gathered from ISIC. All photographs were sorted according on the ISIC categorization, and each subset was divided into the equal number of images, with the exception of melanomas and moles, of which images are slightly dominant. The dataset includes the following diseases: possible skin conditions including actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, and vascular lesions. The dataset is available on the Kaggle [17]. Figure 3 shows some samples from the dataset on which the EfficientNetB0 model will be trained.



Page | 204



Hybrid DenseNet121 Model

A pre-trained DenseNet121 model was prepared and tuned for the binary classification task, and the document parameters were adjusted to improve the training process. During the model training phase, validation accuracy was monitored to determine when the learning rate decreased, which decreased by 30% if performance did not improve over two periods. The improvement in performance must be greater than 0.001 to be taken into account. The actual model training phase begins over 20 epochs with 32 batch sizes. Model performance was validated using 10% of the training data. Figure 4 shows the 20 epochs for training the DenseNet121 hybrid model.

Epoch	Accuracy	Loss	Val Accuracy	Val Loss	Learning Rate	Time	Steps/Second
1/20	0.7857	0.4575	0.6615	0.7718	0.0010	724s	3s/step
2/20	0.8201	0.3916	0.8451	0.3631	0.0010	660s	3s/step
3/20	0.8280	0.3779	0.8060	0.4602	0.0010	601s	3s/step
4/20	0.8436	0.3462	0.8216	0.4055	0.0010	4750s	22s/step
5/20	0.8627	0.3120	0.8542	0.3091	0.0003000000142492354	622s	3s/step
6/20	0.8748	0.2782	0.8763	0.2931	0.0003000000142492354	643s	3s/step
7/20	0.8852	0.2585	0.8529	0.2942	0.0003000000142492354	650s	3s/step
8/20	0.8995	0.2398	0.8815	0.2645	0.0003000000142492354	650s	3s/step
9/20	0.8913	0.2385	0.8620	0.2810	0.0003000000142492354	621s	3s/step
10/20	0.9138	0.2064	0.8945	0.2410	0.0003000000142492354	619s	3s/step
11/20	0.9247	0.1953	0.9206	0.2009	0.0003000000142492354	617s	3s/step
12/20	0.9291	0.1734	0.8971	0.2599	0.0003000000142492354	626s	3s/step
13/20	0.9347	0.1697	0.8581	0.3598	0.0003000000142492354	648s	3s/step
14/20	0.9607	0.1032	0.9193	0.2081	0.0000900000427477062	669s	3s/step
15/20	0.9802	0.0613	0.9466	0.1639	0.0000900000427477062	674s	3s/step
16/20	0.9823	0.0576	0.9401	0.2074	0.0000900000427477062	627s	3s/step
17/20	0.9896	0.0387	0.9440	0.1943	0.0000900000427477062	636s	3s/step
18/20	0.9920	0.0313	0.9492	0.1572	0.00002700000040931627	2801s	13s/step
19/20	0.9951	0.0196	0.9505	0.1786	0.00002700000040931627	863s	4s/step
20/20	0.9950	0.0213	0.9557	0.1667	0.00002700000040931627	863s	4s/step

Figure 4. Training the hybrid DenseNet121 model in the current study

After training the model, the model behavior was analyzed as shown in Figure 5, where the DenseNet121 hybrid model recorded a high training accuracy of 99.5% and a validation accuracy of 95.57%.

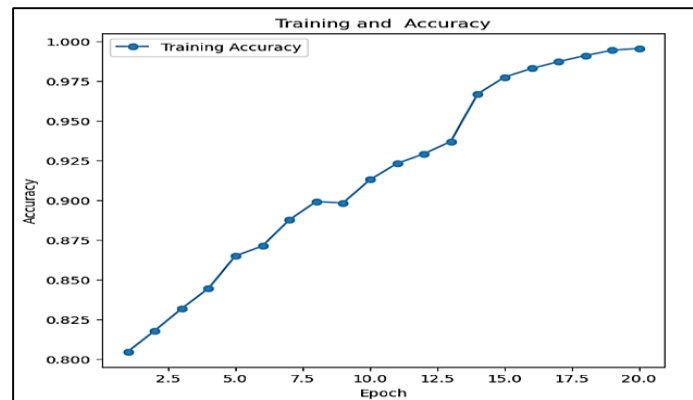


Figure 5. Training curve for the hybrid DenseNet121 model in the current study
EfficientNet B0 Model

EfficientNet B0 model used the ISIC 2020 dataset for multiple classification. The network parameters were set during the training phase to the same values and weights used for the DenseNet121 model. Figure 6 shows the training phase of the EfficientNet B0 model over 20 epochs. The model reached a high training accuracy of 99.84% and validation of 92.42%. The difference between training accuracy and validation accuracy is due to the small size of the validation data, which constitutes 10% of the training data.

epoch	time	Accuracy	Loss	Validation Accuracy	Validation Loss	Learning Rate
1	2s	0.5245	1.3609	0.5565	1.5901	0.0010
2	2s	0.7760	0.6266	0.6832	0.9755	0.0010
3	2s	0.8348	0.4657	0.7702	0.7109	0.0010
4	2s	0.8750	0.3371	0.7540	0.8644	0.0010
5	2s	0.9060	0.2699	0.8037	0.7191	0.0010
6	2s	0.9052	0.2721	0.8149	0.5539	0.0010
7	2s	0.9406	0.1780	0.8099	0.7166	0.0010
8	2s	0.9562	0.1358	0.7714	0.9300	0.0003
9	2s	0.9705	0.0918	0.9081	0.3913	0.0003
10	2s	0.9881	0.0390	0.9168	0.3338	0.0003
11	2s	0.9937	0.0235	0.9106	0.3721	0.0003
12	2s	0.9914	0.0277	0.8932	0.4762	0.0001
13	2s	0.9945	0.0188	0.9193	0.3792	0.0001
14	2s	0.9968	0.0123	0.9280	0.3684	0.0001
15	2s	0.9970	0.0113	0.9205	0.3661	0.0001
16	2s	0.9953	0.0149	0.9155	0.3891	0.00003
17	2s	0.9964	0.0118	0.9255	0.3824	0.00003
18	2s	0.9984	0.0085	0.9205	0.3891	0.00001
19	2s	0.9980	0.0078	0.9217	0.3860	0.00001
20	2s	0.9984	0.0070	0.9242	0.3829	0.000002

Figure 6. Training the EfficientNet B0 model in the current study

Figure 7 shows an increase in accuracy during the training phase, which indicates the success of the model in training and learning with high accuracy from the training data.

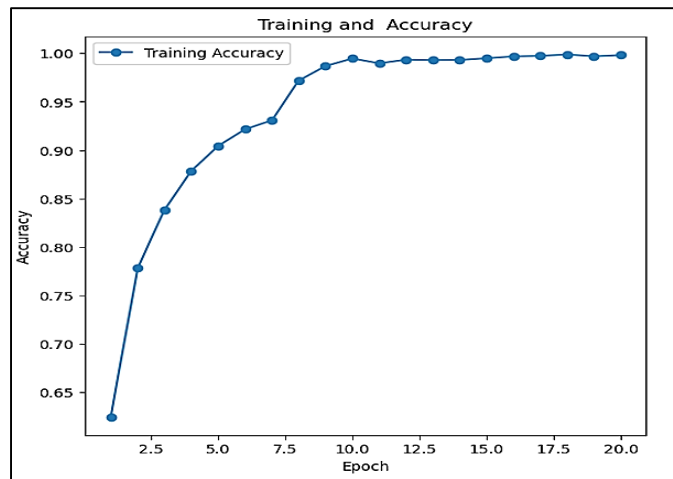


Figure 7. Training curve for the EfficientNet B0 model in the current study

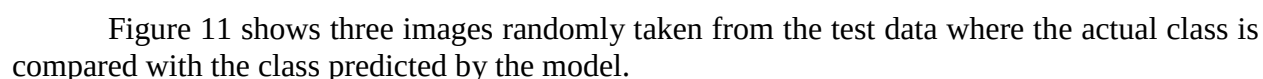
RESULTS AND DISCUSSION

DenseNet121 model's performance was evaluated based on the metrics for classification tasks. Figure 8 displays the results of the rating metrics in the rating report.

	precision	recall	f1-score	support
0	0.95	0.93	0.94	983
1	0.93	0.95	0.94	937
accuracy			0.94	1920
macro avg	0.94	0.94	0.94	1920
weighted avg	0.94	0.94	0.94	1920

Figure 8. Classification report for hybrid DenseNet121 model in the current study

Figure 9 compares the evaluation results of two random images with their real classification. Testing and evaluating the performance of the proposed model shows high detection accuracy, which qualifies the system to work with real-world data.



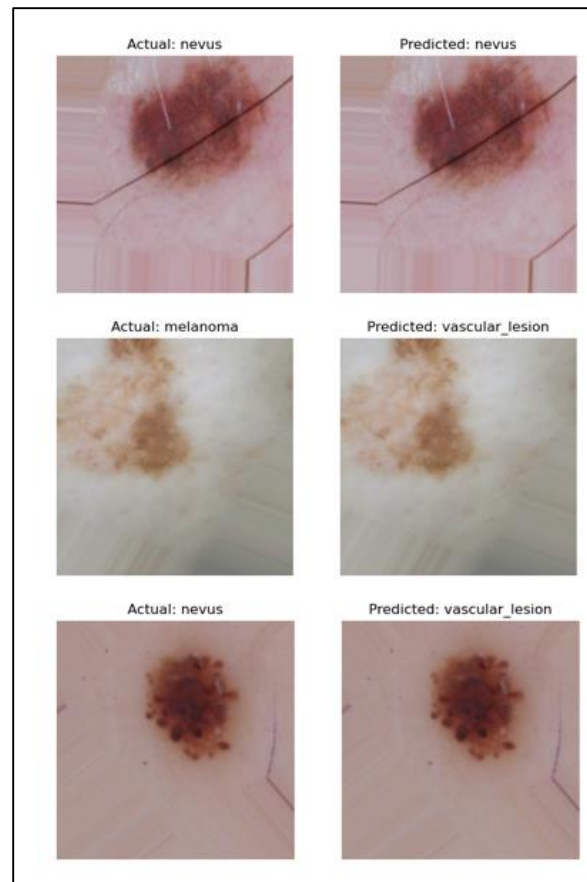


Figure 11. Comparison of actual versus predicted samples for the EfficientNet B0 model in the current study

The models show strong performance in skin cancer detection across binary and multi-class classification tasks making them valuable tools in supporting clinical decision making for skin cancer diagnosis.

CONCLUSION

The study indicates how deep learning may be used to automatically diagnose skin cancer from thermoscopic images. Important preprocessing steps like hyperparameter tuning, data augmentation, and normalization enhanced model accuracy and decreased overfitting. Future studies could improve usability by extending datasets, investigating sophisticated model architectures, and adding real-time diagnostic features for real-world use.

Acknowledgments:

Thanks to everyone who helped complete this achievement.



Conflict of interests:

There are non-conflicts of interest.

References

- [1] K. Urban, S. Mehrmal, P. Uppal, R. L. Giesey, and G. R. Delost, "The global burden of skin cancer: A longitudinal analysis from the Global Burden of Disease Study, 1990–2017," *JAAD International*, vol. 2, pp. 98-108, 2021. DOI: 10.1016/j.jdin.2020.12.001.
- [2] J. Chen, C. Ning, Z. Zhou, P. Yu, Y. Zhu, G. Tan, and C. Mao, "Nanomaterials as photothermal therapeutic agents," *Progress in Materials Science*, vol. 99, pp. 1-26, 2019. DOI: 10.1016/j.pmatsci.2018.10.001.
- [3] M. Wolf, A. de Boer, K. Sharma, P. Boor, T. Leiner, G. Sunder-Plassmann, E. Moser, A. Caroli, and N. P. Jerome, "Magnetic resonance imaging T1- and T2-mapping to assess renal structure and function: A systematic review and statement paper," *Nephrology Dialysis Transplantation*, vol. 33, pp. ii41–ii50, 2018. DOI: 10.1093/ndt/gfy216.
- [4] Y. Wang, C. Zu, G. Hu, Y. Luo, Z. Ma, K. He, X. Wu, and J. Zhou, "Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications," *Neural Process. Lett.*, vol. 48, pp. 1323–1334, 2018. DOI: 10.1007/s11063-018-9885-0.
- [5] A. B. Tufail and A. I. Ahmad, "Diagnosis of diabetic retinopathy through retinal fundus images and 3D convolutional neural networks with limited number of samples," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 6013448, 2021. DOI: 10.1155/2021/6013448.
- [6] L. N. Fuzzell, R. B. Perkins, S. M. Christy, P. W. Lake, and S. T. Vadaparampil, "Cervical cancer screening in the United States: Challenges and potential solutions for underscreened groups," *Preventive Medicine*, vol. 144, pp. 106400, 2022. DOI: 10.1016/j.ypmed.2020.106400.
- [7] R. Mehrotra, M. A. Ansari, R. Agrawal, and R. S. Anand, "A transfer learning approach for AI-based classification of brain tumors," *Machine Learning with Applications*, 2020. DOI: 10.1016/j.mlwa.2020.100034.
- [8] A. Yilmaz, M. Kalebasi, Y. Samoylenko, M. E. Guvenilir, and H. Uvet, "Benchmarking of lightweight deep learning architectures for skin cancer classification using ISIC 2017 dataset," arXiv preprint arXiv:2110.12270, 2021.
- [9] 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," *Healthcare*, vol. 10, no. 7, p. 1183, 2022. DOI: 10.3390/healthcare10071183.
- [10] O. O. Abayomi-Alli, R. Damaševičius, S. Misra, R. Maskeliūnas, and A. Abayomi-Alli, "Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 29, pp. 2600-2614, 2021. DOI: 10.3906/elk-2103-81.
- [11] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the ISIC image datasets: Usage, benchmarks and recommendations," *Medical Image Analysis*, vol. 75, p.



- 102305, 2022. DOI: 10.1016/j.media.2021.102305.
- [12] N. A. AlSadhan, S. A. Alamri, M. M. Ben Ismail, and O. Bchir, "Skin Cancer Recognition Using Unified Deep Convolutional Neural Networks," *Cancers*, vol. 16, p. 1246, 2024. DOI: 10.3390/cancers16071246.
- [13] M. A. Rahman, E. Bazgir, S. M. S. Hossain, and M. Maniruzzaman, "Skin cancer classification using NASNet," *International Journal of Science and Research Archive*, vol. 11, no. 01, pp. 775–785, 2024. DOI: 10.30574/ijrsra.2024.11.1.0106.
- [14] N. Joshi, "A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning," *Towards Data Science*, 2021. Available: <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-4a74d9f62f09>.
- [15] K. Kansal, T. B. Chandra, and A. Singh, "ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning," *Procedia Computer Science*, vol. 235, pp. 70-80, 2024. DOI: 10.1016/j.procs.2024.03.009.
- [16] C. Fanconi, "Skin cancer: malignant vs. Benign," *processed skin cancer pictures of the ISIC archive Kaggle dataset*. Available: <https://www.kaggle.com/fanconic/skin-cancer-malignant-vsbenign>.
- [17] T. Mazhar, I. Haq, A. Ditta, S. A. H. Mohsan, F. Rehman, I. Zafar, et al., "The role of machine learning and deep learning approaches for the detection of skin cancer," *Healthcare*, vol. 11, no. 3, p. 415, 2023. DOI: 10.3390/healthcare11030415.



الخلاصة

المقدمة:

ونظراً لأهمية الكشف المبكر والدقيق لسرطان الجلد في الوقاية من مخاطره وتحسين علاجه، تهدف الدراسة إلى تطوير نماذج آلية تعتمد على التعلم العميق للكشف عن سرطان الجلد وتصنيف الآفات الجلدية بدقة إلى مجموعات حميدة وخبيثة وتحديد الأنواع المختلفة من الآفات، مما يساعد الأطباء في التشخيص المبكر واتخاذ القرار.

طرق العمل:

تم الاعتماد على الشبكات العصبية التلافيفية المدربة مسبقاً، مما يجعل من الممكن تدريب النماذج باستخدام كمية صغيرة فقط من بيانات التدريب المتاحة. تم استخدام مجموعتين من صور أورام الجلد، واحدة للكشف عن وجود ورم والأخرى لتصنيف الورم بين أنواع مختلفة من أورام الجلد. خضعت كلتا مجموعتي البيانات لمجموعة من خطوات التهيئة والمعالجة، ثم تم تدريبهما باستخدام نموذجين من الشبكات العصبية التلافيفية المدربين مسبقاً.

الاستنتاجات:

تبين نتائج تقييم الصور المختارة عشوائياً من مجموعة صور الاختبار مدى دقة النماذج وفعاليتها العالية في التنبؤ. حقق نموذج DenseNet121 الهجين دقة مذهلة تبلغ 94%، مع متوسط دقة واستدعاء ودرجة F1 تبلغ 0.94. حقق نموذج EfficientNet B0 دقة مرجحة بنسبة 93%، مع متوسط مرجح للدقة والاستدعاء ودرجة F1 قدرها 0.93 عبر تسع فئات. حققت الشبكات العصبية التلافيفية العميقة المدربة مسبقاً تقدماً كبيراً في تحسين القدرة على اكتشاف سرطانات الجلد باستخدام صور تنظير الجلد. تسلط هذه الدراسة الضوء على قدرة الذكاء الاصطناعي على تشخيص سرطان الجلد بسرعة ودقة، وكذلك الكشف عن سرطان الجلد (ما قبل) الخبيث.

الكلمات المفتاحية: الشبكات العصبية التلافيفية، صور تنظير الجلد، التعلم العميق، سرطان الجلد، نقل التعلم.