Erciyes Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 40, Sayı 2 , 2024

Erciyes University
Journal of Institue Of Science and Technology
Volume 40, Issue 2, 2024

# Benchmarking Deep Learning Models for Dermatological Image Analysis: EfficientNet Takes the Lead

**Rukiye Nur Kaçmaz**[*1], **Refika Sultan Doğan**[2],

[*1]Erciyes University Faculty of Engineering Software Engineering, KAYSERİ
[2]Abdullah Gül University, Faculty of Life and Natural Sciences, Bioengineering, KAYSERİ,

(Alınış / Received: 31.07.2024, Kabul / Accepted: 26.08.2024, Online Yayınlanma / Published Online: 30.08.2024)

**Anahtar Kelimeler**
Deri Kanseri,
Transfer Öğrenme,
Ablasyon

**Öz**: Hızlı yayılan ve ölümcül olan cilt kanserine melanom denir. Cilt kanserinin erken evrelerinde tedavi edilmezse ölüm oranı çok yüksektir ancak erken evrelerinde doğru bir şekilde tanımlandığında hastaların hayatları kurtarılabilir. Doğru ve hızlı bir teşhis ile hastanın hayatta kalma şansı artabilir. Bilgisayar destekli bir tanı destek sisteminin oluşturulmasını gerekir.. Bu çalışmada Dense201, DarkNet19, EfficientNet melanom sınıflandırması için 3 farklı derin transfer öğrenme modeli sunmaktadır. Buna ek olarak, transfer öğrenmesinde kullanılan filtre boyutu açısından ablasyon çalışması yapılmıştır. Filtre boyutunun etkisine bakmak için her bir modelde farklı sayıda filtre boyutu oluşturulup sonuç alınmıştır. Çalışmada 1792 iyi huylu ve 1464 kötü huylu görüntü içeren ISIC veri seti kullanılmıştır. Bu çalışmaya göre, DenseNet201, boyutlarına bakılmaksızın farklı filtre boyutlarında doğru ve güvenilir sonuçlar sağlamıştır. Bu nedenle, cilt lezyonlarının sınıflandırılmasını içeren çalışmalarda DenseNet201 kullanımının seçilmesi önerilir.

# Dermatolojik Görüntü Analizi için Derin Öğrenme Modellerinin Karşılaştırılması: EfficientNet Zirvede

**Keywords**
Skin Cancer,
Transfer Learning,
Ablation

**Abstract:** Skin cancer that spreads quickly and is deadly is called melanoma. If skin cancer is not treated in its early stages, the mortality rate is very high, but when it is correctly identified in its early stages, patients' lives can be saved. With an accurate and fast diagnosis, the patient's chance of survival can be increased. A computer-aided diagnostic support system needs to be created. In this study, Dense201, DarkNet19, and EfficientNet offer 3 different deep transfer learning models for melanoma classification. In addition, an ablation study was conducted in terms of the filter size used in transfer learning. To look at the effect of the filter size, different filter sizes were created in each model and the results were obtained. The ISIC dataset containing 1792 benign and 1464 malignant images was used in the study. According to this study, DenseNet201 provided accurate and reliable results at different filter sizes regardless of their size. Therefore, it is recommended to use DenseNet201 in studies involving the classification of skin lesions.

*İlgili Yazar, email: rukiyekacmaz@erciyes.edu.tr

## 1. Introduction

Melanoma is the least common but most deadly type of skin cancer, accounting for only 1% of all skin cancer cases. It is estimated that 200,340 cases of melanoma will be diagnosed in the United States in 2024 [1]. Of these cases, 99,700 will be noninvasive (in situ) and 100,640 will be invasive. Of these invasive cases, 59,170 will be in men and 41,470 in women. It is estimated that 8,290 deaths from melanoma will occur, 5,430 in men and 2,860 in women. Melanoma is associated with UV exposure, especially during adolescence, such as sunburn and indoor tanning [2]. People with fair skin and those with lighter eyes and skin tones are at higher risk. The incidence of melanoma has been steadily increasing in recent years. The number of new invasive melanoma diagnoses per year has increased by 46% in the last 15 years, 32% in the last 10 years, and 16% in the last 5 years [3]. The incidence of melanoma varies by age. It is higher in women before the age of 50, twice as high in men as in women after the age of 65, and three times as high in women at the age of 80 [4], [5]. In young adults, especially those between the ages of 15 and 29, melanoma is the second most common cancer. Pediatric melanoma is rare and is usually due to different causes, such as large, hairy moles present at birth [6]. The relationship between UV exposure and melanoma is well established. Even one severe sunburn during childhood or adolescence doubles the risk of developing melanoma later in life [7]. The use of indoor tanning devices is also a significant risk factor, and the risk increases significantly in those who use these devices before the age of 35 [6], [8]. Indoor tanning is more common among women, and this habit significantly increases the risk of melanoma [9].

Melanoma is usually diagnosed by biopsy. Dermoscopy and other imaging techniques are used to evaluate lesions. Pathologic examination provides information about tumor thickness and other prognostic factors[10]. Accurate staging of melanoma is vital to determining appropriate treatment strategies [11]. An exhaustive analysis emphasizes the potential of artificial intelligence (AI) in the field of dermatology for the early diagnosis of skin cancer [12], [13]. The study highlights the importance of validating AI systems in clinical environments to guarantee their efficacy and safety. Furthermore, it explores the significance of taking into account genetic and ethnic variation in patients when using AI applications. The utilization of the MobileNetV2 network for melanoma classification has demonstrated encouraging outcomes. The primary objective of this study is to utilize transfer learning to accurately categorize melanoma photos as either benign or malignant. The study proposes that deep learning models can be efficiently employed in mobile applications to offer accessible diagnostic tools [14]. The findings suggest that Vision Transformers (ViTs) hold significant promise in the detection of skin cancer, particularly in improving the identification of melanoma. The report provides suggestions for future research, including investigating the potential of hybrid models that integrate CNN and ViT technologies, as well as applying ViTs to various medical image categorization tasks [15], [16].

The study [17] examines many publicly accessible skin imaging datasets and their corresponding algorithms for the diagnosis of skin cancer. Another study employs convolutional neural networks (CNNs) to predict skin cancer at an early stage, integrating approaches for data augmentation [18]. This article examines the most advanced deep-learning models used for detecting skin cancer, with a particular focus on the utilization of the ISIC dataset [19]. A new study has integrated many advanced deep-learning models to improve the accuracy of melanoma detection using ISIC datasets [20]. In the literature on ablation research, commonly employed models include Googlenet, Inception3, Densenet201, Inception-ResNetV2, and a deep convolutional neural network (CNN) model called Darknet19. The work utilizes fine-tuning techniques on the ISIC 2019 skin cancer dataset, which comprises eight distinct classifications. The study presents a comparison of the performance of different models on the dataset, with Darknet19 demonstrating competitive outcomes [21][22]. This study conducts a comparative analysis of DenseNet201, DarkNet-19, and other neural networks in the context of skin cancer detection and classification. The analysis is conducted on a sample of 4000 images from the ISIC archive collection. DenseNet201 exhibited robust performance, while DarkNet-19 and other models also displayed noteworthy outcomes [23]. This article examines the utilization of networks such as AlexNet, DarkNet19, GoogleNet, and DenseNet201 that have been trained on the ISIC 2017 and HAM1000 datasets to detect skin cancer. The DenseNet201 model achieved the highest accuracy rate of 82.9% and had a balanced precision and recall, as indicated by its F1 score [24]. The paper compares DenseNet201 and an enhanced version of DarkNet-19 for the classification of skin lesions using the ISIC 2019 and PH2 datasets. DenseNet201 had superior performance in terms of both accuracy and resilience compared to other models [25]. Different research examines different convolutional neural networks, such as DenseNet201 and Darknet-19, to diagnose melanoma using the ISIC 2017 dataset [26]. The paper provides precise metrics that demonstrate the great performance of DenseNet201 and other deep-learning models. This article examines the current developments in melanoma detection by comparing the performance of DarkNet-19, EfficientNet, and other models using ISIC datasets. The utilization of EfficientNets and ensemble approaches resulted in significant enhancements in classification accuracy [27]. The importance of artificial intelligence (AI) in skin cancer diagnosis is becoming increasingly evident, particularly in addressing the limitations of traditional methods. AI offers significant advantages, including the ability to analyze large datasets quickly and efficiently, which accelerates the diagnostic process and increases the capacity to serve more patients. Deep learning models, trained on extensive and diverse datasets, can achieve high accuracy and precision, crucial for the early detection of conditions like melanoma. AI systems capture subtle and complex details within skin lesions that might be overlooked by the human eye, providing a more comprehensive diagnostic tool. Furthermore, AI can standardize the diagnostic process, reducing the likelihood of human error and ensuring a more consistent approach across various cases. In contrast, traditional diagnostic methods such as dermoscopy and biopsy, while effective, have several limitations. These methods can be time-consuming and require specialized expertise, leading to delays and potential accessibility issues, especially in areas with a shortage of dermatologists. Human error is another concern, as factors like fatigue and distraction can impact diagnostic accuracy. Traditional approaches also struggle with the analysis of large volumes of data, as each image must be evaluated manually, which can be inefficient and prone to oversight. Additionally, variability in imaging techniques and equipment can affect the consistency of results. Overall, while traditional methods have been foundational in skin cancer diagnosis, AI presents a promising advancement by addressing these limitations and offering enhanced accuracy, efficiency, and standardization in the diagnostic process. A review of the literature reveals that no studies have compared only these three models and additionally performed ablation analysis—where parameters are altered in various layers of the models to observe changes in results. This study's originality lies in its unique approach of both comparing these specific models and conducting ablation experiments to gain deeper insights into their performance. There is no study in the literature examining the impact of filter size on skin cancer detection in the applied models.

## 2. Material and Method

### 2.1. Data

In this study, 3256 images were obtained from the ISIC archive [skin cancer detection using pytorch | Kaggle](). The distribution of data from the training, validation, and test sets used to diagnose skin cancer is shown in this table. There are 924 malignant and 1144 benign samples in the training set. We analyze the model's performance during training using the validation set, which consists of 288 benign and 240 malignant samples. The test set evaluates the final performance of the model and consists of 360 benign and 300 malignant samples. Understanding and evaluating this distribution is critical to assess the model's performance across various data sets. The model can be thoroughly tested thanks to the total of 2068 samples in the training set, 528 samples in the validation set, and 660 samples in the test set.

TABLE 1 SAMPLE OF SKİN CANCER DATASET

| | BENIGN | MALIGN |
|---|---|---|
| TRAIN | 1144 | 924 |
| VALIDATION | 288 | 240 |
| TEST | 360 | 300 |
| TOTAL | 1792 | 1464 |

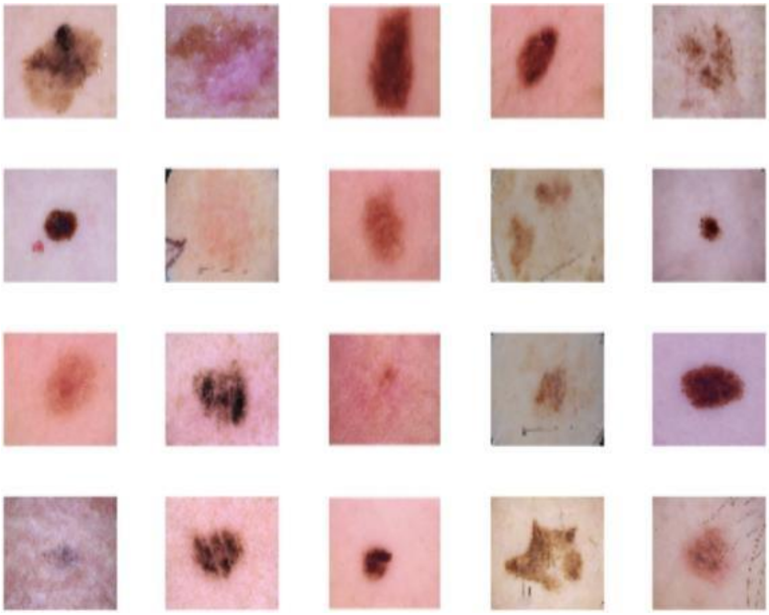

Figure 1. Examples from the dataset.

### 2.2. Model and Parameters

The categorization of diseases based on medical images was thoroughly examined by employing various recognized transfer learning architectures. The designs were pre-trained on the ImageNet dataset, and any modifications to the models and networks were made using the MATLAB environment [28]. Darknet19, EfficientNet, and DenseNet-201 are models used in the field of deep learning, each with different architectural designs. Each of these models has certain advantages and is optimized for various applications. Darknet19 was developed by Joseph Redmon and Ali Farhadi as part of the YOLO (You Only Look Once) object recognition system [29]. Darknet19 is a 19-layer convolutional neural network (CNN) and is optimized for fast and efficient object detection. Darknet19 has 19 convolutional layers followed by max pooling layers. Convolutional layers are used to extract features from the image while pooling layers provide more efficient computation by reducing the size of these features. The last layers contain fully connected layers and a softmax activation function for classification. This structure enables Darknet19 to perform fast and accurate object detection [29]. EfficientNet is an approach that optimizes model scaling and was developed by Google. This architecture aims to achieve high performance using combinations of width, depth and resolution. EfficientNet has a basic network architecture and this basic network is extended using systematic scaling methods to create larger and more complex models. EfficientNet-B0 is a small and efficient model as a starting point and the larger EfficientNet models (B1, B2, ... B7) are versions of this basic model at different scales. This systematic scaling method minimizes computational costs while improving performance [30]. DenseNet-201 is a densely connected convolutional neural network and was developed by Gao Huang et al. DenseNet-201 has 201 layers, and each layer receives information from all the previous layers. This densely connected structure prevents gradient loss and increases the information flow, allowing for more efficient learning. Another important feature of DenseNet-201 is that it reduces the number of parameters, making the model more compact and efficient. Each layer uses the output of the previous layers to create a stronger knowledge base, which increases the model's overall performance [31]. When these three methods were compared all of them had advantages and disadvantages.

TABLE 2 PARAMETERS OF MODELS

| COMPONENT/HYPERPARAMETER | PARAMETERS |
|---|---|
| Test train split ratio | 80-20 |
| Input size (image) | 227x227x3 |
| 'Initiallearnrate', | 3e-4 |

| | |
|---|---|
| Pooling layers | Max |
| Filter size | 2x2<br>3x3<br>5x5 |
| Epoch number | 6 |
| Activation function | ReLu |
| Batch size | 10 |
| Optimizer | SGDM |

In the ablation study, we have tested several kernel sizes, including 2 × 2, 3 × 3, and 5 × 5, to evaluate their impact on performance. The skin lesion classification task was performed using the models DarkNet19, EfficientNet, and DenseNet201. The following provides specifics on the hyperparameters that were used to train and assess the models: Test-Train Split Ratio: 20% was utilized for testing and 80% of the dataset was used for training. 3 channel (RGB) images with an input size of 227 x 227 pixels were utilized. Initial Learning Rate: 3e–4 was the calculated value. This controls the speed at which the model updates its weights. As layers of pooling Max Pooling was applied. By doing this, significant features are preserved while the feature map size is decreased. For the filter size, three different filter sizes—2x2, 3x3, and 5x5—were used to compare the performances of the models. In the epoch number six epochs were used to train the model. This indicates the number of times the entire dataset was processed by the model. Rectified Linear Unit, or ReLU[32], activation function was applied. To improve the model's capacity for learning, this function zeroes out negative values and passes positive values through. For the batch size, ten was found to be the value. This is a reference to the quantity of data that is fed into the model at every training phase. Stochastic Gradient Descent with Momentum (SGDM) [33] was the optimization algorithm that was applied. To avoid local minima and accelerate learning, this method adds momentum. MATLAB software in version 2024 was utilized for this research. MATLAB [34] is an effective tool for processing data, building and evaluating models, and supporting a variety of deep-learning techniques. Carefully selected hyperparameters and training methods were employed to maximize model performance and achieve beneficial classification accuracy for skin lesions.

## 2.2. Performance Metrics

These metrics are used to evaluate the performance of the model in classifying skin cancer images. They help measure the correct diagnosis rates, false diagnosis rates, and the overall accuracy of the model.

TABLE 3 CONFUSİON MATRİX REPRESENTATİON

| | Actual | |
|---|---|---|
| Predicted | TP | FP |
| | FN | TN |

In machine learning and statistical models, it is a fundamental tool used to evaluate the performance of a classification algorithm. Confusion matrix is the most commonly used evaluation method for binary classification problems. It can also be used for multiple classification problems, but its interpretation is more complex. The table 3 explain TP: True positive (Refers to cases where a disease is present and accurately identified), TN: A true negative refers to a situation where a person is healthy and is appropriately diagnosed as such. FP: False positive (A situation when a person is healthy but is wrongly classified as having a medical condition), FN: False negative (Refers to cases where a person has a disease but is wrongly identified as not having it).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Accuracy indicates how accurately the model classified. It is the ratio of true positive (TP) and true negative (TN) results to the total number of tests.

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \quad (2)$$

Sensitivity is a statistical measure that quantifies the proportion of genuine positive outcomes in relation to the combined number of true positives and false negatives. It provides insight into the accuracy of accurately identifying individuals with the condition.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

Specificity is a statistical measure that quantifies the proportion of genuine negative findings in relation to the total of true negatives and false positives. It provides information on the accuracy of identifying healthy individuals.

$$\text{Positive Predictive Value} = \frac{TP}{TP + FP} \quad (4)$$

The positive predictive value measures the proportion of those categorized as positive who are indeed affected by the disease. The term "precision" refers to the ratio of genuine positive findings to the combined number of true positives and false positives.

## 3. Results

The results obtained at the end of this study can be evaluated in two different steps. One step is to compare the classification success of three different types of transfer learning methods, and the second step is to apply ablation to these three different transfer learning methods. The results obtained by changing the filter size as the ablation application and the effect of the filter size are as follows. Accuracy performances of Darknet19, EfficientNet, and DenseNet models in classifying benign and malignant tumors for skin cancer using different filter sizes. The Darknet19 model generally has a lower accuracy rate compared to the other two models. Particularly, as the filter size increases, the performance of Darknet19 noticeably decreases; the accuracy rate, which is 71.97% with a 2x2 filter size, drops to 67.88% with a 3x3 filter size and to 61.82% with a 5x5 filter size. This shows that the model performs better with smaller filters but is overall less effective compared to other models. The EfficientNet model stands out with its consistent and high accuracy rates. The accuracy rate, which is 83.48% with a 2x2 filter size, slightly decreases to 83.03% with a 3x3 filter size, but exhibits its highest performance with 85.91% with a 5x5 filter size. These results demonstrate that EfficientNet performs better with larger filter sizes and has a significantly superior classification ability compared to Darknet19. The DenseNet model stands out as the model that achieves the highest accuracy rates. It starts with an accuracy rate of 84.39% at a filter size of 2x2 and achieves its highest performance with an accuracy rate of 87.73% at a filter size of 3x3. If the filter size is 5x5, it still provides a significantly high accuracy rate of 86.82%. It is often observed that when the filter size of DenseNet increases, its performance tends to improve, but it reaches its peak at a filter size of 3x3. B This study shows that DenseNet gives the most effective results with medium-sized filters and outperforms Darknet19 and EfficientNet in skin cancer classification.

Overall, this comparison demonstrates that the EfficientNet and DenseNet models are superior to Darknet19 in classifying skin cancer, particularly with DenseNet yielding the best results with a 3x3 filter size and EfficientNet achieving the highest accuracy rate with a 5x5 filter size. This also demonstrates that the filter size is a significant factor in model performance, and each model responds differently to different filter sizes. EfficientNet and DenseNet get higher accuracy with larger filter sizes, but Darknet19 has lower accuracy rates even when working with smaller filters and generally lags behind the other two models. The results of our analysis are summarized in the Table 4 and 5. We compared the performance of three different deep learning models (DarkNet19, EfficientNet and DenseNet201) using three different filter sizes (2x2, 3x3 and 5x5). The metrics used for evaluation are sensitivity, specificity, positive predictive value (PPV) and accuracy.

TABLE 4 PERFORMANCE METRİCS ESULTS OF MODELS

| Model | FilterSize | Sensitivity | Specificity | PPV | Accuracy |
|---|---|---|---|---|---|
| **DarkNet19** | 2x2 | 0.67 | **0.92** | **0.97** | 0.72 |
| | 3x3 | 0.63 | **0.91** | **0.97** | 0.68 |
| | 5x5 | 0.56 | 0.63 | **0.96** | 0.56 |
| **EfficientNet** | 2x2 | 0.81 | 0.87 | 0.91 | 0.83 |
| | 3x3 | 0.81 | 0.87 | 0.91 | 0.83 |
| | 5x5 | **0.87** | 0.85 | 0.88 | 0.86 |
| **DenseNet201** | 2x2 | 0.82 | **0.89** | 0.92 | 0.84 |
| | 3x3 | 0.9 | 0.85 | 0.88 | 0.88 |
| | 5x5 | **0.89** | 0.84 | 0.87 | 0.87 |

TABLE 5 THE ACCURACY RESULTS FROM DİFFERENT FİLTER SİZES OF MODELS

| MODEL | 2x2 | 3x3 | 5x5 |
|---|---|---|---|
| **Darknet19** | 0.7197 | 0.6788 | 0.6182 |
| **Efficient-Net** | 0.8348 | 0.8393 | 0.8591 |
| **Dense201** | 0.8439 | **0.8773** | 0.8682 |

TABLE 6 THE CONFUSİON MATRİX RESULTS FROM DİFFERENT FİLTER SİZES OF MODELS

| | 2x2 | | 3x3 | | 5x5 | |
|---|---|---|---|---|---|---|
| DarkNet19 | 349 | 11 | 350 | 10 | 344 | 16 |
| | 174 | 126 | 202 | 98 | 273 | 27 |
| EfficientNet | 326 | 34 | 327 | 33 | 315 | 45 |
| | 75 | 225 | 79 | 221 | 48 | 252 |
| DenseNet201 | 332 | 28 | 315 | 45 | 312 | 48 |
| | 75 | 225 | 36 | 264 | 39 | 261 |

TABLE 7 THE F-MEASURE RESULTS FROM DİFFERENT FİLTER SİZES OF MODELS

| | 2x2 | 3x3 | 5x5 |
|---|---|---|---|
| DarkNet19 | 0.79 | 0.76 | 0.70 |
| EfficientNet | 0.86 | 0.85 | 0.87 |
| DenseNet201 | 0.72 | 0.67 | 0.66 |

The Table 5 compares the performance of DarkNet19, EfficientNet and DenseNet201 models on 2x2, 3x3 and 5x5 grid sizes using confusion matrixes. The correct and incorrect classification numbers of the models are used to evaluate their classification performance using confusion matrix. These results can be interpreted; while DarkNet19 has higher accuracy in small grid sizes such as 2x2 and 3x3, it also exhibits many incorrect classifications. The accuracy rate decreases in a 5x5 grid size as the number of misclassifications rises. These findings indicate that DarkNet19 has a decline in performance as the grid sizes increase. EfficientNet had a well-balanced performance across all grid sizes and achieved decreased misclassification rates compared to other models. This demonstrates that EfficientNet is a reliable model for classifying skin lesions. DenseNet201 achieved high performance when applied to grid sizes of 2x2 and 3x3. Nevertheless, while using a 5x5 grid size, the accuracy rate showed a modest reduction while the number of misclassifications increased. Table 6 shows the confusion metrics of the dataset and table 7 shows the F-measure (F1 score) performance of different convolutional neural network (CNN) architectures using various kernel sizes. DarkNet19 achieves its highest F1 score with a 2x2 kernel size (0.79), and performance decreases as the kernel size increases, with scores of 0.76 and 0.70 for 3 x3 and 5x5 kernels, respectively. This suggests that DarkNet19 performs better with smaller kernels, which might capture local features more effectively. EfficientNet, on the other hand, consistently shows high F1 scores across all kernel sizes, peaking at 0.87 with a 5x5 kernel. This indicates that EfficientNet is more robust to changes in kernel size and benefits from larger kernels, potentially due to its efficient feature extraction capabilities. DenseNet201 exhibits the lowest F1 scores among the architectures, with a decreasing trend as the kernel size increases, scoring 0.72, 0.67, and 0.66 for 2x2, 3x3, and 5x5 kernels, respectively. This suggests that DenseNet201 might struggle with larger kernels and does not perform as well as the other models. Overall, EfficientNet demonstrates superior performance and adaptability across different kernel sizes, while DarkNet19 is effective with smaller kernels and DenseNet201 shows limited performance improvements with increasing kernel size. The results indicate that the EfficientNet model outperforms other models in the task of classifying skin lesions. The EfficientNet model stands out in its high sensitivity, specificity, positive predictive value, and accuracy across all filter sizes. This demonstrates the capability of EfficientNet in accurately classifying skin lesions. The DenseNet201 model likewise achieved excellent performance. The sensitivity and positive predictive value of the system are strong for both 2x2 and 3x3 filter sizes. Furthermore, the system maintains its high performance even with a 5x5 filter size. Nevertheless, the level of specificity is relatively low for certain filter sizes. The DarkNet19 model yielded inconclusive outcomes. Although the 2x2 filter size yields excellent specificity and positive predictive value, the sensitivity and accuracy scores are only average. Although the sensitivity and accuracy numbers are relatively low with a 3x3 filter size, they are significantly lower with a 5x5 filter size.

## 4. Discussion and Conclusion

This study aimed to answer these 2 questions:

1) Firstly, the objective is to compare the outcomes of transfer learning techniques that have not been employed for the simultaneous classification of skin cancer.
2) Secondly, the objective is to conduct ablation experiments in transfer learning approaches by modifying the filter size, and to analyze the impact of the filter size on the classification performance.

This study investigated the classification abilities of DarkNet19, EfficientNet, and DenseNet201 models in the classification of skin lesions. The models were tested using three different filter sizes: 2x2, 3x3, and 5x5. The EfficientNet model demonstrated better results compared across all filter sizes, exhibiting both balance and high efficiency. EfficientNet attracted attention, especially with its low misclassification rates and high sensitivity, PPV, and accuracy values. The DarkNet19 model demonstrated outperformed specifically with a 3x3 filter size but gave inconsistent results with other filter sizes. The EfficientNet model is the preferred option for skin lesion classification investigations. The study demonstrates the impact of various deep-learning models and filter sizes on the effectiveness of skin lesion categorization. It offers significant insights that can inform future research. In general, EfficientNet performs appropriately, although it does not achieve the same level of accuracy in classifying data as DenseNet201. According to these findings, we can conclude that the DenseNet201 model is the most efficient model for classifying skin lesions, as it exhibits low rates of misclassification and a well-balanced performance. DenseNet201 provided accurate and reliable results on different filter sizes, without regard for their size. Thus, it is recommended to use DenseNet201 for studies including the classification of skin lesions.

Future research on data augmentation and diversity should prioritize these recommendations. Techniques for augmenting data are essential to improving the model's capacity for generalization on small data sets. Generative adversarial networks (GANs) [35], [36] can be used to generate synthetic data, which can be assessed in addition to data augmentation techniques like color substitution, random cropping, and vertical and horizontal flipping. Furthermore, a variety of ablation study designs can be used to investigate how particular model elements affect output. The accuracy of the model on a variety of skin types and lesions can be improved by supplementing the data set with information from other sources. The creation of more thorough and broadly applicable models will be aided by such methods.

## References

[1] R. L. Siegel, A. N. Giaquinto, and A. Jemal, 'Cancer statistics, 2024', CA Cancer J Clin, vol. 74, no. 1, pp. 12–49, Jan. 2024, doi: 10.3322/CAAC.21820.
[2] M. P. Purdue et al., 'Etiologic and other factors predicting nevus-associated cutaneous malignant melanoma', Cancer Epidemiology Biomarkers and Prevention, vol. 14, no. 8, pp. 2015–2022, Aug. 2005, doi: 10.1158/1055-9965.EPI-05-0097.
[3] N. J. Petrelli et al., 'Clinical cancer advances 2009: Major research advances in cancer treatment, prevention, and screening -A report from the American Society of Clinical Oncology', Journal of Clinical Oncology, vol. 27, no. 35, pp. 6052–6069, Dec. 2009, doi: 10.1200/JCO.2009.26.6171.
[4] R. L. Siegel Mph, A. N. Giaquinto, | Ahmedin, J. Dvm, and R. L. Siegel, 'Cancer statistics, 2024', CA Cancer J Clin, vol. 74, no. 1, pp. 12–49, Jan. 2024, doi: 10.3322/CAAC.21820.
[5] F. Morgese et al., 'Gender Differences and Outcomes in Melanoma Patients', Oncol Ther, vol. 8, no. 1, p. 103, Jun. 2020, doi: 10.1007/S40487-020-00109-1.
[6] L. De Ann, R. I. Vogel, M. A. Weinstock, H. H. Nelson, R. L. Ahmed, and M. Berwick, 'Association between indoor tanning and melanoma in younger men and women', JAMA Dermatol, vol. 152, no. 3, p. 268, Mar. 2016, doi: 10.1001/JAMADERMATOL.2015.2938.
[7] J. J. Nordlund and A. B. Lerner, 'On the causes of melanomas.', Am J Pathol, vol. 89, no. 2, p. 443, 1977, Accessed: Jul. 29, 2024. [Online]. Available: /pmc/articles/PMC2032245/?report=abstract
[8] G. P. Guy, Z. Berkowitz, D. M. Holman, and A. M. Hartman, 'Recent Changes in the Prevalence of and Factors Associated With Frequency of Indoor Tanning Among US Adults', JAMA Dermatol, vol. 151, no. 11, p. 1256, Nov. 2015, doi: 10.1001/JAMADERMATOL.2015.1568.
[9] G. P. Guy, Y. Zhang, D. U. Ekwueme, S. H. Rim, and M. Watson, 'The potential impact of reducing indoor tanning on melanoma prevention and treatment costs in the United States: An economic analysis', J Am Acad Dermatol, vol. 76, no. 2, pp. 226–233, Feb. 2017, doi: 10.1016/J.JAAD.2016.09.029.
[10] B. Ahmed, M. I. Qadir, and S. Ghafoor, 'Malignant melanoma: Skin cancer—diagnosis, prevention, and treatment', Crit Rev Eukaryot Gene Expr, vol. 30, no. 4, pp. 291–297, 2020, doi: 10.1615/CRITREVEUKARYOTGENEEXPR.2020028454.
[11] PDQ Cancer Genetics Editorial Board, 'Genetics of Skin Cancer (PDQ®): Health Professional Version', PDQ Cancer Information Summaries, 2002, Accessed: Jul. 30, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26389333
[12] S. R. Jartarkar et al., 'Artificial intelligence in Dermatopathology', J Cosmet Dermatol, vol. 22, no. 4, pp. 1163–1167, Apr. 2023, doi: 10.1111/JOCD.15565.
[13] G. H. Dagnaw, M. El Mouhtadi, and M. Mustapha, 'Skin cancer classification using vision transformers and explainable artificial intelligence', J Med Artif Intell, vol. 7, no. 0, pp. 14–14, Jun. 2024, doi: 10.21037/JMAI-24-6.
[14] B. C. R. S. Furriel et al., 'Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review', Front Med (Lausanne), vol. 10, p. 1305954, Jan. 2023, doi: 10.3389/FMED.2023.1305954/BIBTEX.
[15] C. Xin et al., 'An improved transformer network for skin cancer classification', Comput Biol Med, vol. 149, Oct. 2022, doi: 10.1016/J.COMPBIOMED.2022.105939.
[16] C. Flosdorf, J. Engelker, I. Keller, and N. Mohr, 'Skin Cancer Detection utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer', Jul. 2024, Accessed: Jul. 30, 2024. [Online]. Available: https://arxiv.org/abs/2407.18554v1
[17] D. Wen et al., 'Characteristics of publicly available skin cancer image datasets: a systematic review',
[18] Lancet Digit Health, vol. 4, no. 1, pp. e64–e74, Jan. 2022, doi: 10.1016/S2589-7500(21)00252-1.

[19] R. A. Mehr and A. Ameri, ‘Skin Cancer Detection Based on Deep Learning’, J Biomed Phys Eng, vol. 12, no. 6, p. 559, Dec. 2022, doi: 10.31661/JBPE.V0I0.2207-1517.10364-Y/TABLES/9.
[20] A. G. C. Pacheco and R. A. Krohling, ‘Recent advances in deep learning applied to skin cancer detection’, Dec. 2019, Accessed: Jul. 30, 2024. [Online]. Available: https://arxiv.org/abs/1912.03280v1
[21] R. Ali, R. C. Hardie, B. N. Narayanan, and S. De Silva, ‘Deep Learning Ensemble Methods for Skin Lesion Analysis towards Melanoma Detection’, Proceedings of the IEEE National Aerospace Electronics Conference, NAECON, vol. 2019-July, pp. 311–316, Jul. 2019, doi: 10.1109/NAECON46414.2019.9058245.
[22] M. A. Khan, T. Akram, M. Sharif, S. Kadry, and Y. Nam, ‘Computer Decision Support System for Skin Cancer Localization and Classification’, Computers, Materials & Continua, vol. 68, no. 1, pp. 1041– 1064, Mar. 2021, doi: 10.32604/CMC.2021.016307.
[23] K. Thurnhofer-Hemsi and E. Domínguez, ‘A Convolutional Neural Network Framework for Accurate Skin Cancer Detection’, Neural Process Lett, vol. 53, no. 5, pp. 3073–3093, Oct. 2021, doi: 10.1007/S11063-020-
[24] H. Hussein, A. Magdy, R. F. Abdel-Kader, and K. A. El Salam, ‘Binary Classification of Skin Cancer Images Using Pre-trained Networks with I-GWO’, Inteligencia Artificial, vol. 27, no. 74, pp. 102–116, Dec. 2024, doi: 10.4114/INTARTIF.VOL27ISS74PP102-116.
[25] M. Naqvi, S. Q. Gilani, T. Syed, O. Marques, and H. C. Kim, ‘Skin Cancer Detection Using Deep Learning—A Review’, Diagnostics 2023, Vol. 13, Page 1911, vol. 13, no. 11, p. 1911, May 2023, doi: 10.3390/DIAGNOSTICS13111911.
[26] S. Benyahia, B. Meftah, and O. Lézoray, ‘Multi-features extraction based on deep learning for skin lesion classification’, Tissue Cell, vol. 74, p. 101701, Feb. 2022, doi: 10.1016/J.TICE.2021.101701.
[27] S. Alzahrani, B. Al-Bander, and W. Al-Nuaimy, ‘A Comprehensive Evaluation and Benchmarking of Convolutional Neural Networks for Melanoma Diagnosis’, Cancers 2021, Vol. 13, Page 4494, vol. 13, no. 17, p. 4494, Sep. 2021, doi: 10.3390/CANCERS13174494.
[28] D. ; Popescu et al., ‘New Trends in Melanoma Detection Using Neural Networks: A Systematic Review’,
[29] Sensors 2022, Vol. 22, Page 496, vol. 22, no. 2, p. 496, Jan. 2022, doi: 10.3390/S22020496.
[30] ‘Transfer Learning - MATLAB & Simulink’. Accessed: Jul. 31, 2024. [Online]. Available:
[31] https://www.mathworks.com/discovery/transfer-learning.html
[32] J. Redmon and A. Farhadi, ‘YOLO9000: Better, faster, stronger’, Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 6517–6525, Nov. 2017, doi: 10.1109/CVPR.2017.690.
[33] M. Tan and Q. V Le, ‘EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks’, May 24, 2019, PMLR. Accessed: Jul. 30, 2024. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html
[34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ‘Densely Connected Convolutional Networks’, Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 2261–2269, Aug. 2016, doi: 10.1109/CVPR.2017.243.
[35] A. F. Agarap, ‘Deep Learning using Rectified Linear Units (ReLU)’, Mar. 2018, Accessed: Jul. 31, 2024.
[36] [Online]. Available: http://arxiv.org/abs/1803.08375
[37] K. Tang, W. Liu, Y. Zhang, and X. Chen, ‘Acceleration of stochastic gradient descent with momentum by averaging: finite-sample rates and asymptotic normality’, May 2023, Accessed: Jul. 31, 2024. [Online]. Available: https://arxiv.org/abs/2305.17665v2
[38] A. Biswas et al., ‘Generative Adversarial Networks for Data Augmentation’, Data Driven Approaches on Medical Imaging, pp. 159–177, Jun. 2023, doi: 10.1007/978-3-031-47772-0_8.
[39] A. Antoniou, A. Storkey, and H. Edwards, ‘Data Augmentation Generative Adversarial Networks’, J Phys A Math Theor, vol. 44, no. 44, pp. 1–13, Nov. 2017, Accessed: Jul. 31, 2024. [Online]. Available: https://arxiv.org/abs/1711.04340v3