Classification of Skin Cancer with Deep Transfer Learning Method

Doaa Khalid Abdulridha Al-Saedi1,*, Serkan Savaş2

1Department of Electronics and Computer Engineering, Çankırı Karatekin University, Çankırı, Türkiye
2Department of Computer Engineering, Kırıkkale University, Kırıkkale, Türkiye

Received: Sep.08, 2022 Accepted: Sep.16, 2022 Published: Oct.10, 2022

Abstract— Skin cancer is a serious health hazard for human society. This disease is developed when the pigments that produce skin color become cancerous. Dermatologists face difficulties in diagnosing skin cancer since many skin cancer colors seem identical. As a result, early diagnosis of lesions (the foundation of skin cancer) is very crucial and beneficial in totally curing skin cancer patients. Significant progress has been made in creating automated methods with the development of artificial intelligence (AI) technologies to aid dermatologists in the identification of skin cancer. The widespread acceptance of AI-powered technologies has enabled the use of a massive collection of photos of lesions and benign sores authorized by histology. This research compares six alternative transfer learning networks (deep networks) for skin cancer classification using the International Skin Imaging Collaboration (ISIC) dataset. DenseNet, Xception, InceptionResNetV2, ResNet50, and MobileNet were the transfer learning networks employed in the investigation which were successful in different studies recently. To compensate for the imbalance in the ISIC dataset, the photos of classes with low frequencies are augmented. The results show that augmentation is appropriate for the classification success, with high classification accuracies and F-scores with decreased false negatives. With an accuracy rate of 98.35%, modified DenseNet121 was the most successful model against the rest of the transfer learning nets utilized in the study.

Keywords: Skin cancer, deep learning, ISIC, transfer learning, DenseNet.

1. Introduction

Skin cancer may be divided into two primary categories: melanoma and non-melanoma. Squamous cell carcinoma and basal cell carcinoma are the most prevalent non-melanoma tumors. The 17th most prevalent cancer globally is cutaneous melanoma. It is the 13th and 15th most prevalent cancer, respectively, in men and women (WCRF, 2022). Over the past several decades, there has been an increase in the incidence of both non-melanoma and melanoma skin cancers. Currently, 132,000 cases of melanoma and 2 to 3 million cases of non-melanoma skin cancer are reported annually worldwide (WHO, 2017). In 2020, there were more than 150,000 brand-new instances of cutaneous melanoma. Table 1 displays the overall incidence and rates of melanoma skin cancer worldwide in 2020 (WCRF, 2022).

Table 1. Global melanoma skin cancer incidence and rates in 2020 (WCRF, 2022)

<table>
<thead>
<tr>
<th>No</th>
<th>Country</th>
<th>Number</th>
<th>ASR/100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Australia</td>
<td>16,171</td>
<td>36.6</td>
</tr>
<tr>
<td>2</td>
<td>New Zealand</td>
<td>2,801</td>
<td>31.6</td>
</tr>
<tr>
<td>3</td>
<td>Denmark</td>
<td>2,886</td>
<td>29.7</td>
</tr>
<tr>
<td>4</td>
<td>The Netherlands</td>
<td>8,310</td>
<td>27.0</td>
</tr>
<tr>
<td>5</td>
<td>Norway</td>
<td>2,567</td>
<td>26.4</td>
</tr>
<tr>
<td>6</td>
<td>Sweden</td>
<td>4,266</td>
<td>23.3</td>
</tr>
</tbody>
</table>
According to the Table 1, Australia and New Zealand were the countries with the highest rates of cutaneous melanoma in 2020. Modern medical research is attempting to aid dermatologists in their diagnosis without the need for specialized or expensive equipment due to the rise in skin cancer globally.

Early identification and prevention of skin cancer are the most effective ways to control it. Skin patches or growths that are new or changing, especially those that appear strange, should be investigated. Any new lesions or increasing changes in the appearance of an existing lesion (size, shape, or color) should be assessed by a physician.

Deep learning (DL) technology has allowed for the classification of skin cancer into seven diagnostic groups, including melanocytic nevi, melanoma, benign keratosis-like lesions, basal cell carcinoma, actinic keratosis, vascular lesions, and dermatofibroma. A dermatologist who specializes in the detection of skin cancer often follows a predetermined process that starts with a visual examination of the worrisome lesion, is followed by a dermoscopy, and ends with a biopsy (Haenssle et al., 2018). When compared to relying just on a visual diagnosis today, the efficacy of anticipating a result using artificial intelligence (AI) and deep learning (DL) in diagnostics grows tremendously (Ayoub et al., 2021). Deep convolutional neural networks (DCNNs) process dermoscopic images to identify skin lesions, including all skin cancer lesions, whereas convolutional neural networks (CNNs) may be used for feature selection and object categorization. DNNs are effective for classifying medical images, but they need a lot of training data. High-performance GPUs are used to train a network of DNNs using large-scale datasets (Savaş et al., 2019, 2022). DL systems enabled by GPUs have shown superior skin cancer detection than people (Nugroho et al., 2019). From this point of view, in this study, it is aimed to classify skin cancer by transfer learning-based deep learning method. In this way, it is aimed to assist clinics in diagnosis and treatment procedures with early detection of skin cancer.

Some studies have been carried out in the literature to diagnose skin cancer using machine learning (ML) and DL algorithms for this reason. These studies are summarized in the second section. In the third section of the research, the material and method used in this study are explained. In the fourth section, the experimental results are explained. In the fifth section, the contributions of the research and its comparative discussion with other researches were realized.

2. Related Works

A brand-new method of diagnosing skin cancer based on metaheuristics and DL is proposed by Qiao et al., (2022). In their method, the skin dermoscopy images are first trained using a modified AlexNet that has already been trained using batch normalization layers, and the remaining few layers are then handled by an Extreme Learning Machine (ELM). The Fractional-order Red Fox Optimization (FORFO) Algorithm, a recently updated metaheuristic, is used to increase the effectiveness of the ELM network. The proposed approach has a 97.14% overall accuracy. On the other hand, to examine polarization speckle images obtained from the main categories of malignant and benign lesions, Wang et al., (2021) used DL and classical ML algorithms. A collection of 122 cancerous and 196 benign skin lesion speckle images was improved using patch cropping for DL, which was an advantageous strategy given the patterns' statistical homogeneity. In the easier classification job of distinguishing malignant melanoma from benign melanoma, the ML technique achieved high 90% accuracy. However, in the general classification test of malignant and benign tumors, ResNet, their selected DL architecture, got the best result of 82% diagnosis accuracy.

Another novel model based on the auto-encoder, spiking, and CNN is proposed by Toğaçar et al., (2021). The dataset used is an open-access dataset called the International Skin Imaging Collaboration (ISIC) skin cancer, which included 1800 benign and 1497 malignant tumor images. The dataset is reconstructed using the auto-encoder model in the proposed method. The MobileNetV2 model, which comprises of residual blocks and spiking networks, was used to train and classify the original dataset and structured dataset. The study's classification success rate was 95.27%.

A pipeline technique for melanoma diagnosis utilizing dermoscopy images was investigated by Fu et al., (2022). After picture preprocessing, the region of interest (RoI) is segmented using a technique based on the Kernel Fuzzy C-means algorithm. The segmented area's key attributes were then optimally retrieved and chosen utilizing
a new optimized approach. An ideal classification approach based on multi-layer perceptron is offered for the final diagnosis. The feature selection and classification are optimized using a newly created Red Fox Optimization (DRFO) algorithm. The accuracy of 90.5% indicated that the outcome was trustworthy.

Asymmetry, Border, Color, and Diameter (ABCD) criteria are used to diagnose skin cancer automatically in the research of Senan & Jadhav, (2021). The PH2 standard dataset consisting three classes of skin diseases (Atypical Nevi, Melanoma, and Common Nevus) was utilized to test the proposed system. The proposed mechanism consists of two stages. The Gaussian filter method is used to enhance the photos and remove undesirable pixels in the first stage, which uses preprocessing to improve image quality. The contour approach was used to extract the RoI from dermoscopy images. Morphology is also being considered for improving the quality of skin lesions. The second stage involves implementing ABCD rules to extract relevant characteristics and the total accuracy of the study is 84%.

Tumpa & Kabir, (2021) aimed to create a neural network that can accurately detect and classify Melanoma in their study. Their procedure began with dermoscopy images being preprocessed to remove hairs using the Maximum Gradient Intensity technique, as well as image enhancement. To separate skin lesions from the images, a segmentation technique based on the Otsu Thresholding algorithm is used. The segmented images are then utilized to calculate many features such as ABCD, GLCM, and LBP, which used to train a neural network. On the combined dataset of ISIC archive and the PH2 dermoscopic image database, the network achieved an accuracy of 97.7%.

Ali et al. (2021) introduced a DCNN model for the precise categorization of benign and malignant skin lesions. They first apply a filter or kernel to reduce noise and artefacts, after which they normalize the input images and extract features that help with effective classification, they then increase the number of images by using data augmentation, which increases classification rate accuracy, and finally they add more images. The DCNN model is contrasted with AlexNet, ResNet, VGG-16, DenseNet, MobileNet, and other transfer learning techniques. The HAM10000 dataset was used to test the model, and the results showed that it had the greatest training and testing accuracy, respectively, of 93.16% and 91.93%. Sharpening and smoothening filters, as well as enhancing techniques, are employed in the preprocessing step to remove noise in the study of Dabhi et al., (2021). After these processes, Otsu segmentation was employed to detect skin cancer in its early stages. Finally, a back propagated based artificial neural network (BP-ANN) was created for classification of skin cancer with the spatially grey level dependency matrix (SGLD) characteristics to archive the system's optimum efficiency in this study. As a result, the research work may be used to efficiently classify Benign and Melanoma skin cancers, with a total accuracy of 99.13%.

Hosny et al. (2019), used transfer learning and a pre-trained DL network (AlexNet) in their study. They classified three distinct lesions in the study with using fine-tuning. In the study, data augmentation techniques were also used to balance the number of dataset. The accuracy of the model on the PH2 dataset was 98.61%. Researchers also used other metrics such as sensitivity (98.33%), specificity (98.93%), and precision (97.73%). Another study using transfer learning is proposed by Demir et al., (2019). Their dataset consists of 3297 images in total and they used ResNet-101 and Inception-v3 DL architectures for the classification. Accuracy rates of 84.09% and 87.42% are get in ResNet-101 and Inception-v3 architectures, respectively.

Transfer learning has developed out of DL research nowadays. Among the models that have demonstrated their efficiency in various image classifications, it is crucial to identify the dominant model or models to be employed in biomedical image processing and classification operations. It is preferable to develop general models that may be applied to other research and employed in this instance rather than subject and situation-specific models. It's crucial to find models that produce consistent and trustworthy findings across a variety of imaging modalities, especially given the sparse amount of images in the area of medicine and the complexity of image capture and processing (Savaş, 2022). For this reason, pre-trained deep architectures were used in this work to carry out the classification procedure.

3. Material and Method
3.1. Dataset:

The dataset of the study is made up of 64,000 images by The International Skin Imaging Collaboration of cancerous and noncancerous skin diseases (ISIC, 2022). The dataset was split into three sets such as training set with 50,000 images, 10,000 images for validation, and 4000 images for testing. All of the images were put in order based on how they were classified with ISIC, and each subset had the same number of images, except for melanomas and moles, whose images are slightly more common. The dataset consists of two categories benign and malignant. Figure 1 represents a sample of the dataset.
3.2. Transfer Learning Nets:

Transfer learning is a type of ML in which a model built for one task is used for different ones. It's usually used when researcher doesn't have enough data to train on. However, the problem with the data can be fixed by adding more data or data augmentation. The main reason for need of transfer learning is that Melanoma and Benign lesions look a lot alike, making it hard to tell them apart and put them in different groups. Transfer learning is also better at classifying lesions that look the same, which makes it the first choice. Transfer learning nets are trained on large datasets, and then their model weights are saved. For the application for a different dataset, the last few layers of pre-trained nets are changed in transfer learning. In this paper, DenseNet, Xception, InceptionResNetV2, ResNet50, and MobileNet are the models selected for comparison of the results. But in this case, we do not only used the saved weights, but also retrained the models on our dataset so that the network layers could tell the difference between the two types of lesions more accurately. We used six transfer learning nets to train the models on the skin lesion dataset and then looked at their predictions. For the evaluation of the performances, accuracy, loss, and confusion matrices results were plotted. Then, the accuracy of all of these learning nets were compared to found the one that could find all of the lesions with the most accuracy.

3.2.1. InceptionResNet:

Simple Inception modules are not employed in this net, but rather the residual form of Inception Nets (Szegedy et al., 2016). In this network, each Inception block is followed by a filter-expansion layer (1x1 convolution without activation), which is used to increase the dimensionality of the filter bank before the addition to correspond to the depth of the input. The Inception-v4 network's computational cost is the same as that of the Inception-ResNet-v2 network. The batch normalization employed in Inception-ResNetv2 is different from that used in non-residual Inception versions. In that it is applied only on top of the standard layers and not the summations (Szegedy et al., 2016).

3.2.2. Xception:

An addition to the Inception architecture is the Xception architecture. Convolutions that can be separated based on depth are used in place of the conventional Inception modules. It does not divide the input data; instead, it maps the spatial correlations for each output channel independently. The next step is 1x1 depth wise convolution, which allows the Xception net to record cross-channel correlation. It substantially excels Inception V3 on larger data while just somewhat outperforming it on smaller data (Chollet, 2016).

3.2.3. MobileNetV2:

Similar to the Xception net, this net also uses depth-wise separable connections. Every input channel is subjected to a single filter as part of the depth wise convolution for MobileNets. The pointwise convolution employing a 1x1 convolution then combines the results of the depth-wise convolution. A typical convolution layer filters and blends inputs in one step to produce a new set of outputs. The depth-wise separable convolution separates this into two layers: a layer for filtering and a layer for combining. This factorization leads to a large reduction in computation and model size. Applications for mobile and embedded vision benefit especially from MobileNet. In comparison to other systems, it is simpler and has fewer parameters (Howard et al., 2017).

3.2.4. ResNet50:

These utilise skip connections and are the deeper CNNs. These residual blocks significantly reduce total parameters while also resolving gradient deterioration. Two simple design tenets are followed by the ResNet (Residual Networks) architecture. In addition, when the size of the feature map is cut in half, the number of filters doubles. Layers have the same number of filters for the same output map size. Batch normalization is performed.
before the ReLU activation function and after each convolution layer. The shortcut is utilized if the input and output are the same size. The projection shortcut is applied as the dimensions expand (He et al., 2015).

3.2.5. **DenseNet121:**

In this net, because all levels, including those in the same dense block and transition layers, spread their weights over several inputs, deeper layers can utilize characteristics that were collected earlier. 120 Convolutions and 4 AvgPool make up DenseNet-121 (He et al., 2015).

3.2.6. **EfficientNet:**

EfficientNet uses a technique known as compound coefficient to quickly and simply scale up models. Compound scaling, as opposed to using a random scaling scheme, equally scales each dimension using a predetermined set of scaling factors (Tan & Le, 2019).

3.3. **Proposed Methodology:**

We describe how skin lesions are classified in this section. The proposed approach is shown diagrammatically in Figure 2. In order to tackle more effectively computer vision challenges, researchers have constructed increasingly sophisticated CNNs in recent years. In this study, we used TensorFlow, a DL framework built by Google, to train six DCNNs to categorize skin cancer images.

![Figure 2](image.png)

**Figure 2.** Block diagram of the study

3.3.1. **Preprocessing:**

In this step, the dataset were converted to 224 x 224 x 3 dimensions to be used in the pre-trained model, and also implementing the preprocess input function for every transfer learning algorithms. This function apply image scaling. The next step in this phase contains feature extraction. Pre-trained models were used to extract features through transfer learning. This entails identifying key elements in an image and extrapolating information from them. To create a model, several CNNs are piled on top of one another.

3.3.2. **Implementation of Transfer Learning Models:**

In this step, every kind of pre-trained models was implemented separately without the output layer from the original model. We added three layers in the end of each model, the first and the second layers are same in their components. Every layer of these consists of 128 neuron with ReLU activation function. The output layer consists of 2 classes which is equal to the number of classes (benign, Malignent).

4. **Experimental Results and Discussion**

In this section, we discuss our research and contrast the trained models' diagnostic acuity with that of a recognized panel of dermatologists. In this study, the metrics were used are precision, recall (sensitivity), F1-score, specificity, and accuracy to determine the evaluation and outcomes of trained models as mentioned in the five equations below. In this case, binary classification is used to comprehend the model. The class for positive points
is positive, while the class for negative points is negative. The models' efficacy was calculated using the following equations, and they were then compared.

\[
\text{Precision} = \frac{TP}{(TP + FP)}
\]  

(1)

\[
\text{Recall (Sensitivity)} = \frac{TP}{(TP + FN)}
\]  

(2)

\[
\text{F1-Score} = \frac{2 \times (\text{PRE} \times \text{REC})}{(\text{PRE} + \text{REC})}
\]  

(3)

\[
\text{Specificity} = \frac{TN}{(TN + FP)}
\]  

(4)

\[
\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}
\]  

(5)

TP stands for the fraction of correctly categorized positive instances, TN for correctly labelled negative cases, FP for wrongly labelled positive cases, and FN for incorrectly labelled negative cases in the equations. Additionally, the F1-score result, which is the harmonic mean of accuracy and recall, is presented since the distribution of sample sizes among databases is significantly imbalanced. Table 2 provides a summary of the comparison of six models by presenting average values for variables such as accuracy, precision, F1-score, recall (sensitivity), and specificity. The precision demonstrates a classifier's ability to avoid classifying a negative sample as positive. The F1-score peaks close to 1 and vice versa. This demonstrates that, as compared to previous models, DenseNet121 has produced overall better results for the categorization of dermoscopy skin cancer images.

Table 2. Results comparison between proposed models

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy (%)</th>
<th>Precision (%)</th>
<th>F1-Score (%)</th>
<th>Recall (Sensitivity) (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet50</td>
<td>98.1</td>
<td>97.8</td>
<td>98.1</td>
<td>98.45</td>
<td>97.8</td>
</tr>
<tr>
<td>Xception</td>
<td>99.1</td>
<td>98</td>
<td>99</td>
<td>99.6</td>
<td>98.6</td>
</tr>
<tr>
<td>MobilNetV2</td>
<td>99.4</td>
<td>99.4</td>
<td>99.4</td>
<td>99.5</td>
<td>99.4</td>
</tr>
<tr>
<td>InceptionResNetV2</td>
<td>98.45</td>
<td>98.6</td>
<td>98.4</td>
<td>98.2</td>
<td>98.7</td>
</tr>
<tr>
<td>EfficientNet</td>
<td>98.5</td>
<td>98.7</td>
<td>98.5</td>
<td>98.2</td>
<td>98.75</td>
</tr>
<tr>
<td>DenseNet121</td>
<td>99.6</td>
<td>99.7</td>
<td>99.5</td>
<td>99.4</td>
<td>99.7</td>
</tr>
</tbody>
</table>

As shown in the above Table 2, the DenseNet121 is the best model for skin cancer detection with 99.6% accuracy, 99.7% precision, 99.5% F1-score, 99.4% recall (sensitivity), and 99.7% specificity. Table 3 represents a comparison between the proposed DenseNet121 with state of arts. It is appear that the proposed model overcome all other models.

Table 3. Comparison between the proposed model and the state of arts

<table>
<thead>
<tr>
<th>Study</th>
<th>Skin Cancer Diagnosis</th>
<th>Classifier and training algorithm</th>
<th>Dataset</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. A. Ali &amp; Al-Marzouqi, (2017)</td>
<td>Benign/malignant</td>
<td>LightNet (deep learning framework),</td>
<td>ISIC 2016 dataset</td>
<td>Acc: 81.6%</td>
</tr>
<tr>
<td>Harangi et al., (2018)</td>
<td>Malignant melanoma/nevus/SK</td>
<td>CNN as single neural-net architecture</td>
<td>ISIC 2017 dataset</td>
<td>Acc: 83.8%</td>
</tr>
<tr>
<td>Kalouche, (2016)</td>
<td>Benign/melanoma</td>
<td>VGG-16 and CNN</td>
<td>ISIC dataset</td>
<td>Acc: 78%</td>
</tr>
<tr>
<td>Hasan et al., (2019)</td>
<td>Benign/malignant</td>
<td>CNN</td>
<td>ISIC dataset</td>
<td>Acc: 89.5%</td>
</tr>
</tbody>
</table>
(Rezvantalab et al., 2018)

<table>
<thead>
<tr>
<th>Lipoma/fibroma/sclerosis/melanoma</th>
<th>Deep region-based CNN and fuzzy C means clustering</th>
<th>ISIC dataset</th>
<th>Acc: 94.8% Sensitivity: 97.81% Specificity: 94.17% F1_score: 95.89%</th>
</tr>
</thead>
</table>

Namozov et al., (2018)

<table>
<thead>
<tr>
<th>Nevus/AK/benign, keratosis/vascular lesion/dermatofibrom</th>
<th>CNN model with keratosis/vascular LeNet approach</th>
<th>ISIC dataset</th>
<th>Acc: 95.86%</th>
</tr>
</thead>
</table>

Proposed work

<table>
<thead>
<tr>
<th>Benign/malignant</th>
<th>Modified DenseNet121</th>
<th>ISIC dataset</th>
<th>Acc: 99.6% Precision: 99.7% F1_Score: 99.5% Recall: 99.4% Specificity: 99.7%</th>
</tr>
</thead>
</table>

Figure 3 represents the progress of accuracy with increasing the number of epochs and represents the relation between the training loss and the validation loss. The final loss is very small which reaches to near of zero.

As can be seen in Figure 3, the training accuracy of the proposed model was above 99%, while the validation rate reached a high success rate of 98.35%. Loss is another important performance metric. A loss value close to zero indicates that the performance of the study is high. As seen in Figure 3, the training Loss value decreased below a very small value of 0.025, especially after the 10th epoch. The validation Loss value was generally around 0.075 in the study, which can be considered as very low.

5. Conclusion and Future Work

In this study, skin cancer classification was made using pre-trained models. Skin cancer is one of the most common types of cancer in the world, and if not detected early, it can lead to fatal outcomes. For this reason, the need to support clinical processes has arisen with the state-of-art technology studies. In this study, the suggested DenseNet121 model outperforms existing transfer learning models in terms of classification accuracy. The suggested technique has the capacity to categorize benign and malignant skin lesions by substituting a sigmoid for the output activation layer for binary classification. The best accuracy rate that can be achieved with the best model is 99.6% in the study. In addition, the validation accuracy rate of this model has reached a high rate of 98.35%.

In the next stages of this study, research will be carried out to obtain similar success rates in other datasets by using different feature selection techniques together with different pre-trained models. This research will be carried out by expanding the dataset and the results will be announced.
References


