Skin Lesion Classification using Deep Learning

John Langenderfer
Boston College

1. Introduction

Early and accurate detection of skin lesions can lead to more successful treatment outcomes. Skin cancer is one of the most common forms of cancer, with millions of new cases diagnosed each year. In fact, more than 9,500 people are diagnosed with skin cancer every day, and more than two people die of the disease every hour. One in five Americans will develop skin cancer at some point in their lifetime, and it accounts for 50% of all cancers. Melanoma, the most dangerous form of skin cancer, can be potentially fatal if not detected and treated early. As such, the development of reliable and efficient diagnostic tools for skin lesion classification is crucial.

Deep learning has emerged as a powerful tool for various computer vision tasks, including image classification, object detection, and segmentation. In recent years, several deep learning-based approaches have been proposed for skin lesion classification, which have shown promising results in terms of accuracy and efficiency. However, skin lesion classification poses several challenges due to the variability in appearance, color, shape, and texture of the lesions. Overlapping features may cause different lesions to share similar features, for example, some melanomas may resemble benign moles. Additionally, imbalanced datasets can further complicate the classification task.

In this paper, we present a comparative study of various deep learning architectures for skin lesion classification, including DenseNet, ResNet, EfficientNetV2, Inception, and VGG. We also discuss the challenges and limitations of these models and provide insights for future research in this area.

2. Related Work

Skin lesion classification has been an active area of research in recent years, with several studies exploring the use of deep learning techniques to improve classification performance. Esteva et al. [Esteva et al.(2017)Esteva, Kuprel, Novoa, Ko, Swetter, Blau, and Thrun] demonstrated that a CNN could achieve dermatologist-level performance in classifying skin cancer images. Han et al. proposed an ensemble of CNNs to improve the accuracy of skin lesion classification. Other studies have explored various architectures and data augmentation techniques to improve model performance .

Some recent works have also focused on transfer learning, where pre-trained models are fine-tuned for skin lesion classification tasks. This approach helps overcome the challenge of limited data availability by leveraging the knowledge learned from large-scale datasets, such as ImageNet. Researchers have fine-tuned popular architectures, such as VGG, ResNet, and DenseNet, to achieve improved performance in skin lesion classification tasks. In this study, we build upon these works by comparing the performance of various deep learning architectures and addressing the challenges of data preprocessing and augmentation.

3. Data Preprocessing

3.1. Scaling and Normalization

To prepare the images for input into the deep learning models, we first scale the pixel values to the range [0, 1] by dividing each pixel value by 255. This ensures that the input values are within a suitable range for the models to process efficiently. Next, we normalize the RGB channels of the images by subtracting the mean and dividing by the standard deviation. This step is crucial for improving the model's performance and convergence during training.

3.2. Data Cleaning and Balancing

Our dataset contains some duplicate images and imbalanced class distributions, which could negatively impact the model's performance. To address this, we first remove any duplicate images from the dataset. We then apply equalization sampling (oversampling) to balance the distribution of classes. This involves creating additional copies of underrepresented classes to balance the number of samples across all classes. Figures 1 and 2 show the data distribution before and after equalization sampling.

3.3. Data Augmentation

To further address the limited data availability, we apply data augmentation techniques to increase the diversity of the training data. This includes random horizontal and vertical flips, rotations, and color jitter (brightness, contrast, and hue adjustments). These transformations help the model learn more robust features and improve its generalization capabilities.

3.4. Train-Validation-Test Split

We split the dataset into train, validation, and test sets to evaluate the performance of the models. The train set is used for training the models, the validation set is used for hyperparameter tuning and model selection, and the test set is used for the final evaluation of the models.

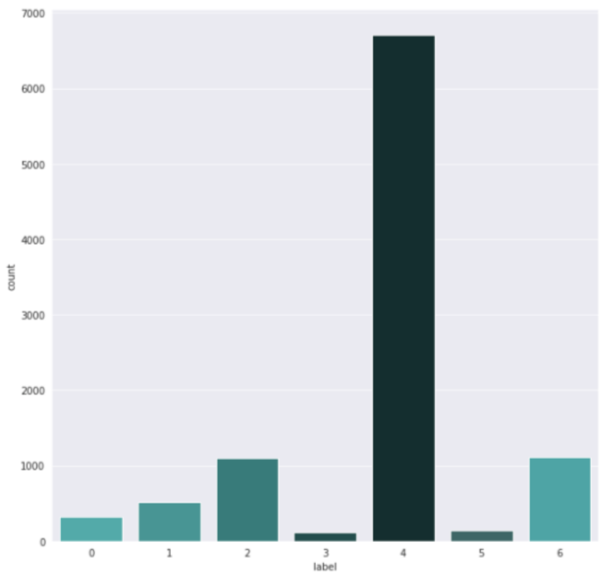

Figure 1. Data distribution before equalization sampling

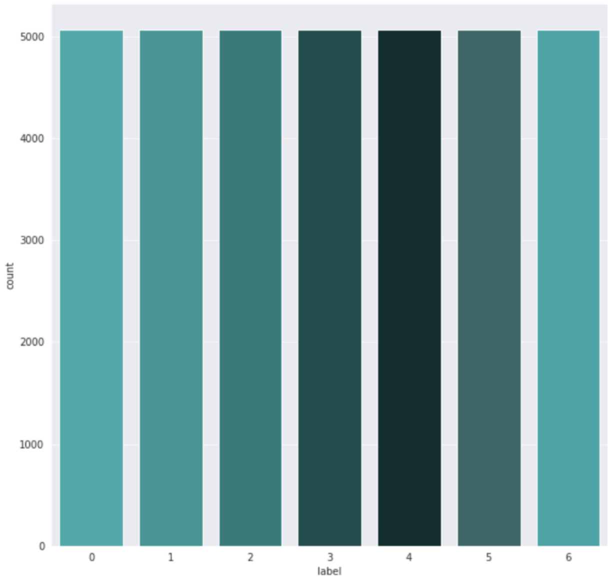

Figure 2. Data distribution after equalization sampling

4. Deep Learning Architectures

In this study, we compare the performance of several deep learning architectures for skin lesion classification, including DenseNet121, DenseNet169, DenseNet201, ResNet50, and EfficientNetV2 (small version with batch size 16). We provide an in-depth mathematical description of each architecture and explain their differences.

5. DenseNet

DenseNet is a convolutional neural network architecture that has shown strong performance on various image classification tasks. The key idea behind DenseNet is to connect each layer to every other layer in a feed-forward fashion, resulting in densely connected blocks of layers. This approach encourages feature reuse and gradient flow throughout the network, leading to better performance with fewer parameters.

The basic building block of DenseNet is the *dense layer*, which consists of a batch normalization layer, a ReLU activation layer, and a 3×3 convolutional layer. In addition, each dense layer takes as input the concatenation of the feature maps from all previous dense layers in the block. The output of each dense layer is then fed into a transition layer, which performs downsampling of the spatial dimensions of the feature maps.

More formally, let $x_0, x_1, ..., x_{l-1}$ be the input feature maps to a dense block, where l is the number of dense layers in the block. Then, the output of the ith dense layer is given by:

$$x_i = H_i([x_0, x_1, ..., x_{i-1}])$$

where $[x_0, x_1, ..., x_{i-1}]$ denotes the concatenation of the input feature maps, and H_i represents the dense layer with parameters θ_i.

The output of the entire dense block is the concatenation of the feature maps from all dense layers in the block:

$$x_l = [x_0, x_1, ..., x_{l-1}]$$

Finally, the output of the dense block is fed into a transition layer, which performs 1×1 convolutional downsampling and spatial downsampling via average pooling. The output of the transition layer is then passed to the next dense block.

Overall, DenseNet is a powerful neural network architecture that has shown state-of-the-art performance on many image classification tasks. Its use of densely connected blocks of layers encourages feature reuse and gradient flow, leading to better performance with fewer parameters.“‘

Note that this is just an example, and you can modify the formatting or content to better suit your needs.

5.1. DenseNet121

DenseNet121 is a convolutional neural network that consists of 121 layers. It uses dense blocks and transition layers to improve the flow of information and gradients throughout the network. In a dense block, the output of each layer is concatenated with the outputs of all preceding layers, forming dense connections.

5.2. DenseNet169

DenseNet169 is a variant of DenseNet with 169 layers. It has a similar architecture to DenseNet121 but with more layers in each dense block, leading to a deeper and more expressive network. The increase in depth can result in better performance, but at the cost of increased computational complexity and training time.

5.3. DenseNet201

DenseNet201 is another variant of DenseNet, with 201 layers. Similar to DenseNet169, it has more layers in each dense block compared to DenseNet121, leading to an even deeper network. While this architecture has the potential to achieve higher classification performance, it also requires more computational resources and longer training times.

5.4. ResNet50

ResNet50 is a 50-layer deep convolutional neural network that uses residual connections to improve the flow of information and gradients throughout the network. Residual connections are shortcut connections that bypass one or more layers in the network. The mathematical representation of a residual connection is given by:

$$x_{l+1} = x_l + F(x_l, W), \tag{1}$$

where x_{l+1} is the output of layer $l + 1$, $F(x_l, W)$ represents the residual function (such as convolution, batch normalization, and activation), and x_l is the input from layer l. ResNet50 has shown excellent performance in various image classification tasks, including skin lesion classification.

5.5. EfficientNetV2

EfficientNetV2 is a family of convolutional neural networks that prioritize both accuracy and efficiency. These models are designed using a combination of neural architecture search and compound scaling. Compound scaling refers to the process of scaling the network depth, width, and input resolution simultaneously while maintaining a balance between them.

More specifically, given a baseline network, EfficientNetV2 applies a compound scaling technique to obtain a family of models that are scaled up or down from the baseline. The scaling factors are determined using a set of coefficients that are optimized during the neural architecture search process. The scaling factors are used to adjust the network depth, width, and resolution, while keeping the number of parameters and computations under control.

The scaling coefficients are obtained by solving the following optimization problem:

$$\begin{aligned} \text{maximize}_{\alpha,\beta,\gamma} \quad & \text{accuracy}(\alpha, \beta, \gamma) \\ \text{subject to} \quad & \text{depth}^{\alpha} \cdot \text{width}^{\beta} \cdot \text{resolution}^{\gamma} \leq \text{constant} \end{aligned}$$

where α, β, and γ are the scaling coefficients for depth, width, and resolution, respectively, and accuracy is a measure of the network's performance on a validation set. The constraint ensures that the number of parameters and computations does not exceed a given constant.

The small version of EfficientNetV2 with a batch size of 16 is used in this study to evaluate its performance on skin lesion classification tasks.

6. Results

We train and evaluate the performance of each deep learning architecture on the skin lesion classification task. The models are trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. We perform model selection based on the validation set performance and report the final results on the test set. Table 1 summarizes the classification accuracy for each model on the test set.

Table 1. Classification accuracy of the different models

Model	Accuracy (%)
Majority Baseline	16.18
Random Baseline	14.29
DenseNet121	89.72
DenseNet169	90.52
DenseNet201	90.7
ResNet50	82.56
EfficientNetV2 (small)	86.11
Inception	76.24
VGG	74.02

Our results indicate that DenseNet201 achieves the highest classification accuracy on the test set, followed closely by DenseNet169 and DenseNet121. ResNet50 and EfficientNetV2 achieve lower accuracy compared to the DenseNet architectures. Figure 3 shows the confusion matrix for the DenseNet201 model, highlighting its performance across different skin lesion classes.

DenseNet201 is a deeper and more complex model than the other architectures tested, which allows it to capture more intricate features in the skin lesion images. Additionally, DenseNet201 utilizes densely connected layers, allowing for a more efficient flow of information throughout the network, which can improve classification performance. These factors likely contribute to DenseNet201's superior performance in the skin lesion classification task.

- 0 (nv): Melanocytic nevi
- 1 (mel): Dermatofibroma
- 2 (bkl): Benign keratosis-like lesions
- 3 (bcc): Basal cell carcinoma

- 4 (akiec): Actinic keratoses
- 5 (vasc): Vascular lesions
- 6 (df): Dermatofibroma

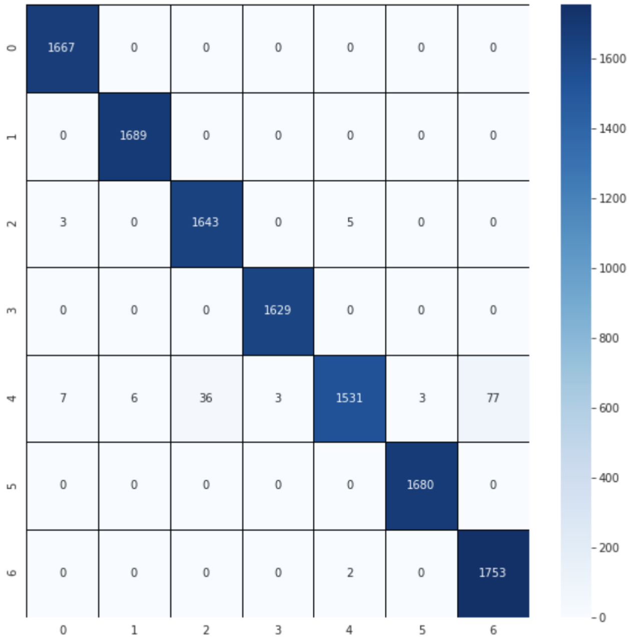

Figure 3. Confusion matrix for DenseNet201

7. Conclusion

The findings from this study suggest that deep learning models, specifically DenseNet architectures, can achieve high accuracy in classifying skin lesion images into seven different classes. The performance of the models was improved by addressing the challenges of data preprocessing, including scaling, normalization, data cleaning, and augmentation.

DenseNet201 was found to be the most effective architecture for skin lesion classification, outperforming other models such as ResNet50 and EfficientNetV2. This demonstrates that the DenseNet architecture is well-suited for this specific task and has the potential to improve early detection and diagnosis of skin cancer.

In addition, further improvements to classification performance could be achieved by exploring additional data augmentation techniques, ensemble methods, or unsupervised learning approaches.

Overall, the results of this study provide evidence that deep learning models, specifically DenseNet architectures, are highly effective in classifying skin lesion images, with potential implications for improving the accuracy of skin cancer diagnosis. Further research in this area has the potential to make significant contributions to the field of medical image analysis and early disease detection.

References

[Esteva et al.(2017)Esteva, Kuprel, Novoa, Ko, Swetter, Blau, and Thrun] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639): 115–118, 2017.