# Machine Learning Based Classification of Medical Images of Melanoma Using Transfer Based Learning

Ved Barakam[1] and Morteza Sarmadi[#]

[1]Monta Vista High School, USA
[#]Advisor

## ABSTRACT

Over 7650 people died from the United States died from melanoma just within 2022, with this number projected to increase by 4.4% in 2023 according to the Skin Cancer Foundation. By accurately diagnosing this disease and implementing an accurate model into the healthcare space for clinical decision-making, early classification of skin lesions may increase the likelihood of treatment before cancer metastasis. Machine learning diagnosis has gained attention in the previous years and has been proven to contribute to the early diagnosis of various diseases. In the context of skin cancer diagnosis, there is a limited amount of medical images, making it challenging to utilize typical machine learning approaches for classification. Therefore, in this work, we utilize transfer learning for the automated classification of medical images of melanoma into benign and malignant. Accordingly, we develop a transfer based algorithm based on a pre-trained InceptionV3 and a VGG16 model in Python. We compared the performance of these two models in order to evaluate the optimal model. The number of epochs and the learning rate were optimized for both models. In order to assess the model, we utilize a variety of metrics, including confusion matrix, ROC, AUC, accuracy, sensitivity, specificity, precision, and F1 score. The results of this study demonstrated that an optimized VGG16 model outperformed VGG16 and was able to successfully classify at least 90% of testing images. The developed model can potentially contribute to the early and automated diagnosis of melanoma in clinical settings.

## Introduction

Potentially claiming many lives, melanoma skin cancer is an disease where the uncontrolled growth of pigment-producing cells in the skin marks its presence (Wang et al., 2023). To ensure successful treatment and improved survival rates, early detection and diagnosis are indispensable. The American Cancer Society highlights how widespread it is amongst different types of cancers with projections indicating 106,110 new cases alongside 7,180 deaths anticipated this year alone within the United States. Furthermore, the management of melanoma incurs significant expenses, with an estimated annual cost surpassing $3 billion encompassing all diagnoses. This includes various expenditures related to treatments and follow-up care that are also taken into account simultaneously (Guy & Ekwueme, 2011). Identifying melanoma on time not only ameliorates patient outcomes but additionally reduces healthcare costs through regular self-examination supported by routine check-ups (Safran et al., 2018). Recent technologies have attempted to do this but improvements are still needed in order to increase the accuracy of skin cancer detection (Johansen et al., 2020). Specifically, machine learning has proved to be a useful tool of diagnosis in health care and has also been proven to help with the diagnosis of various diseases. (Painuli & Bhardwaj, 2022)

Amongst various machine learning tools, image classification algorithms have been a useful tool for diagnosis based on a patient's pathological images (Goyal et al., 2020). For image recognition and classification applications, convolutional neural networks (CNNs), a form of deep learning neural network, are frequently employed. CNNs are made up of many layers of interconnected, image-processing-specific neurons. They employ convolutional layers, which can identify edges, lines, and other picture features, as well as pooling layers, which decrease the image's dimensionality while preserving the crucial details of the image (Cagnetta et al., 2023).

While there are various image classification algorithms developed for this purpose, they suffer from the need for an extensive number of images (several 100,000's to millions) needed for training, which is a limitation in clinical settings. Transfer learning appears to be a promising and viable option in clinical settings due to it not needing an excessive number of images for training purposes. Transfer learning uses a pre-trained model as a starting point for training a new model on a separate but related task. Because the pre-trained model has already learnt common characteristics and patterns from a big dataset, this method can save time and computational resources. Transfer learning includes optimizing the pre-trained model on a smaller, domain-specific dataset as opposed to creating a new model from scratch (Neyshabur et al., 2020). The concept behind this is that since the pre-trained model has already acquired relevant characteristics and patterns, it can be used to complete the new assignment. In computer vision applications including picture classification, object recognition, and segmentation, transfer learning is frequently utilized (Brodzicki et al., 2020). Skin cancer diagnosis can be aided by CNN's with Keras and transfer learning. A pre-trained CNN model is utilized as a starting point for a new model in the process of transfer learning. This makes it possible to train the model on fresh data more quickly and precisely (Ribani & Marengoni, 2019). Using a large dataset of photos of skin lesions to train the model, then applying it to the classification of new images as benign or malignant, is a typical method for using CNNs for skin cancer diagnosis.

### Overview

In this study, a melanoma skin cancer dataset from 2022 is used, which consists of 10600 images and is obtained from Kaggle (Javid, 2022). Of this dataset, 9600 images (90% of dataset) are used for training the model while 1000 (10% of dataset) are used for testing. We utilize transfer learning with CNN's in order to classify pathological images into classes of benign and malignant skin cancer. Transfer learning with Keras and the Inception model is used to train and test the model in Python. We test various learning rates in order to optimize our model and also compare two common deep learning structures, namely, InceptionV3 and VGG16, with InceptionV3 performing better. The result of this study can highlight a successful use of a transfer based learning algorithm for the early diagnosis of skin cancer.

## Methods

Half of the training and testing images were equally split between benign and malignant, meaning 50% of the images were malignant while the other 50% is benign. An example of a benign and malignant image of melanoma is presented in Figure 1. The ImageDataGenerator is a data augmentation technique used to process the images during training. In our case, it rescaled the pixel values of the images to a range of 0 to 1. This normalization step is crucial as it standardizes the input data, ensuring that all images have consistent numerical values, which can help the model's accuracy and convergence during training. The input data was also randomly flipped horizontally, which creates a mirror image of the original. The input data was also randomly rotated by a maximum of 20% of the original image's total angle. These data augmentation layers can help generate more robust data for our model to learn from, possibly resulting in increased accuracy and generalization. The image was also shrinked to 150x150 pixels to reduce the computational cost and memory requirements of the model. This size was also suitable for the image classifier and can also make training faster and efficient.

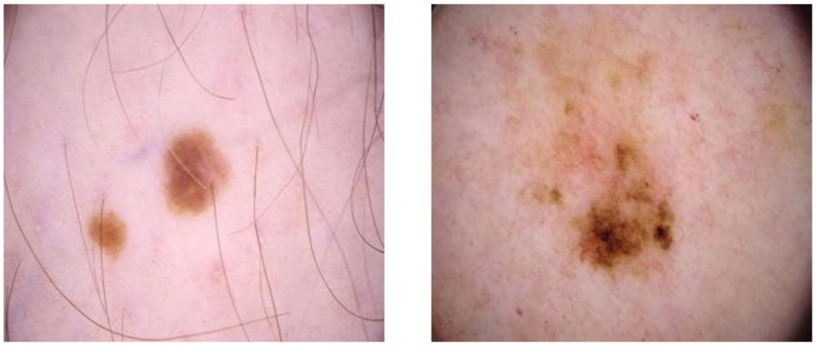

**Figure 1.** Benign (left) and malignant (right) image of melanoma shown respectively.

Custom classification layers are added to the VGG16 and InceptionV3 model to provide binary classification, which separates benign and malignant skin samples as shown in Table 1. A global average pooling 2D layer and two dense layers with ReLU activation make up the custom layers. The first Dense layer comprises of 1024 neurons, while the second has 512 neurons. A sigmoid activation function is used in the last layer to create a probability score that indicates the possibility of melanoma existence. The final layer has a sigmoid activation function to produce a probability score representing the likelihood of the melanoma presence. The base model's weights are frozen during training to retain the knowledge gained from the dataset. The model is then trained on a dataset of skin images, achieving promising results in melanoma classification.

**Table 1.** Machine Learning Architecture of Model

| Input |
|---|
| VGG16/InceptionV3 Base Model |
| GlobalAveragePooling2D |
| Dense (ReLU, 1024) |
| Dense (ReLU, 512) |
| Dense (Sigmoid) |
| Output (Binary Classification Probability) |

In order to assess the model, we utilize a variety of metrics, including accuracy, sensitivity, specificity, precision, F1 score, and ROC AUC score. The accuracy is the proportion of all the samples in the dataset that were properly predicted, which is also shown by $\frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{True\ Positives\ (TP) + False\ Positives\ (FP) + True\ Negatives\ (TN) + False\ Negatives\ (FN)}$. The sensitivity measures the model's ability to correctly predict positive images from the dataset, which is given by $\frac{TP}{TP + FN}$. The specificity measures the model's ability to correctly negative positive images from the dataset, which is given by $\frac{TN}{TN + FP}$. Precision measures the accuracy of positive predictions made by the model while minimizing the number of false positive predictions, which is shown by $\frac{TP}{TP + FP}$. The F1 score measures the model's accuracy by combining its precision and recall scores, which is shown by $\frac{TP}{TP + 0.5(FP+FN)}$. The ROC AUC score tells us how well the model is at distinguishing between positive and negative classes (Carvalho et al., 2019).

## Results

We sought to optimize the number of epochs and the learning rate of our model. Both the InceptionV3 and VGG16 model were tested with a number of epochs ranging from 0 to 20, incremented by 2. The number of epochs was also optimized in regards with loss and accuracy in the training set. We found 10 epochs to yield a training accuracy of about 0.9 and loss of 0.3 for the validation dataset for Inception, and an accuracy of 0.8 and a loss of 0.4 for VGG16. Figure 2 demonstrates the performance of the model as a function of the number of epochs, ranging from 0 to 10 for training and validation datasets. As shown in Figure 3, the model with 10 epochs was also found to be the optimal model in terms of the testing dataset, yielding a testing accuracy of approximately 90%.

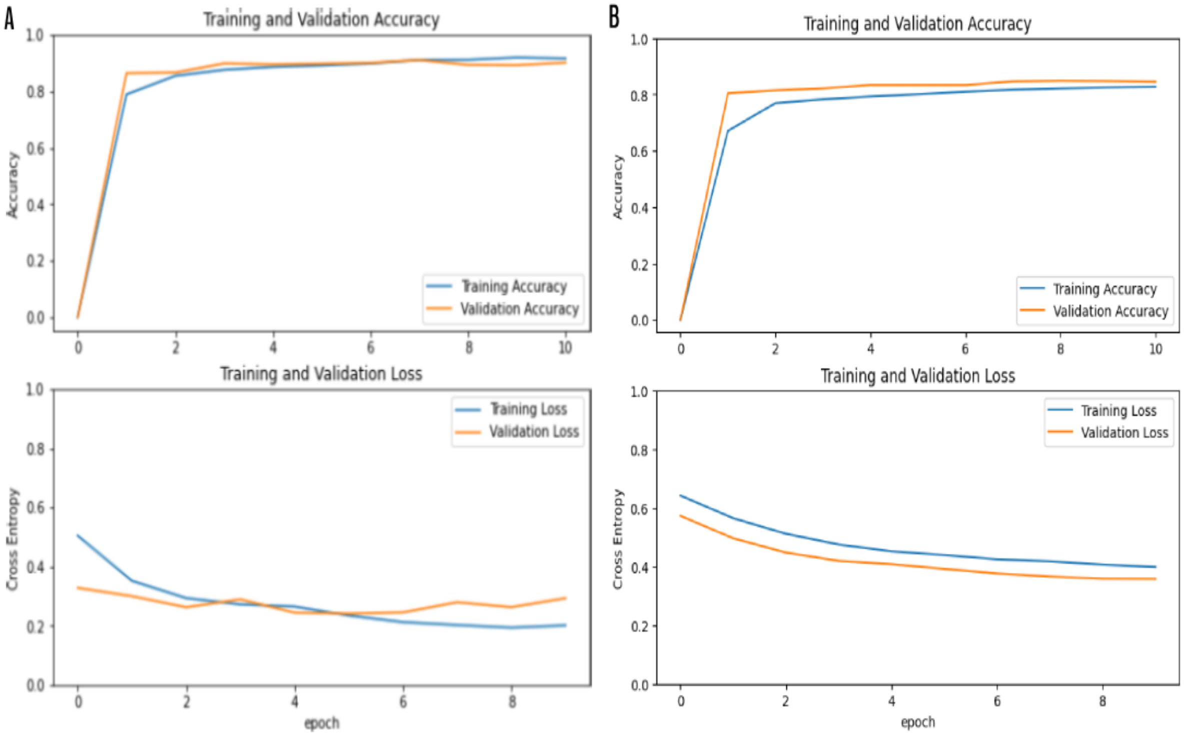

**Figure 2.** The model's performance using an InceptionV3 (A) and VGG16 (B) model using a learning rate of 0.001 and 10 epochs for training and validation datasets

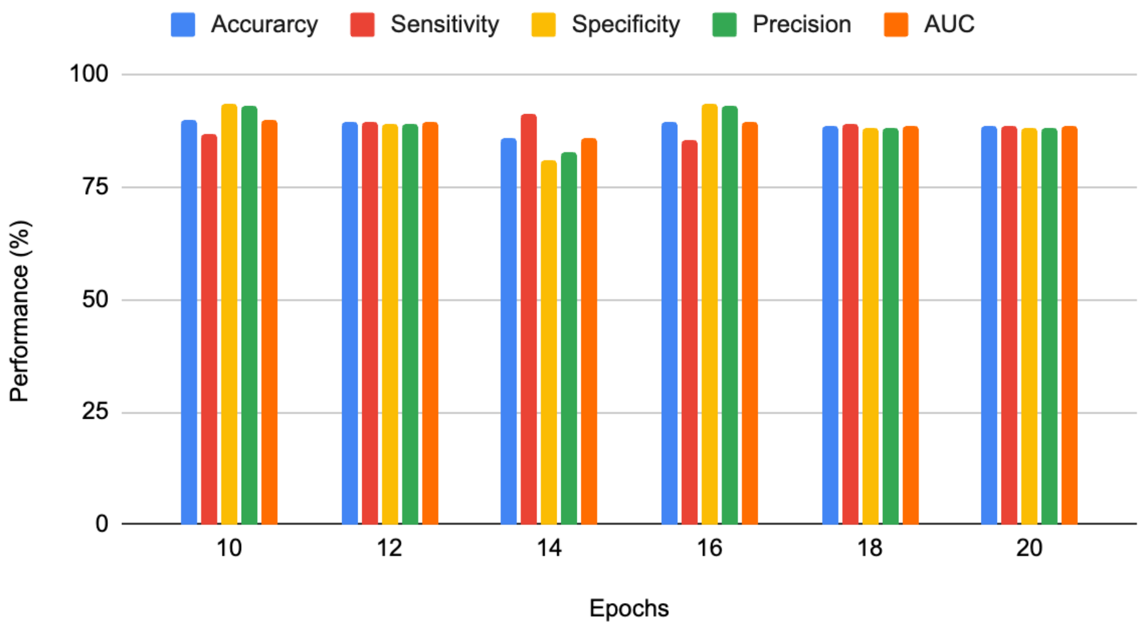

**Figure 3.** The model's performance using a InceptionV3 model (optimized) with a learning rate of 0.001 and 10 epochs for the testing set

The VGG16 and InceptionV3 models were compared in order to evaluate which model would provide a better performance for this structure. The performance was evaluated based on metrics such as the accuracy, precision, sensitivity, AUC, and specificity for both models. As the next step, various learning rates were recorded to optimize the model's performance, which is represented in Figures 4 (InceptionV3) and 5 (VGG16).

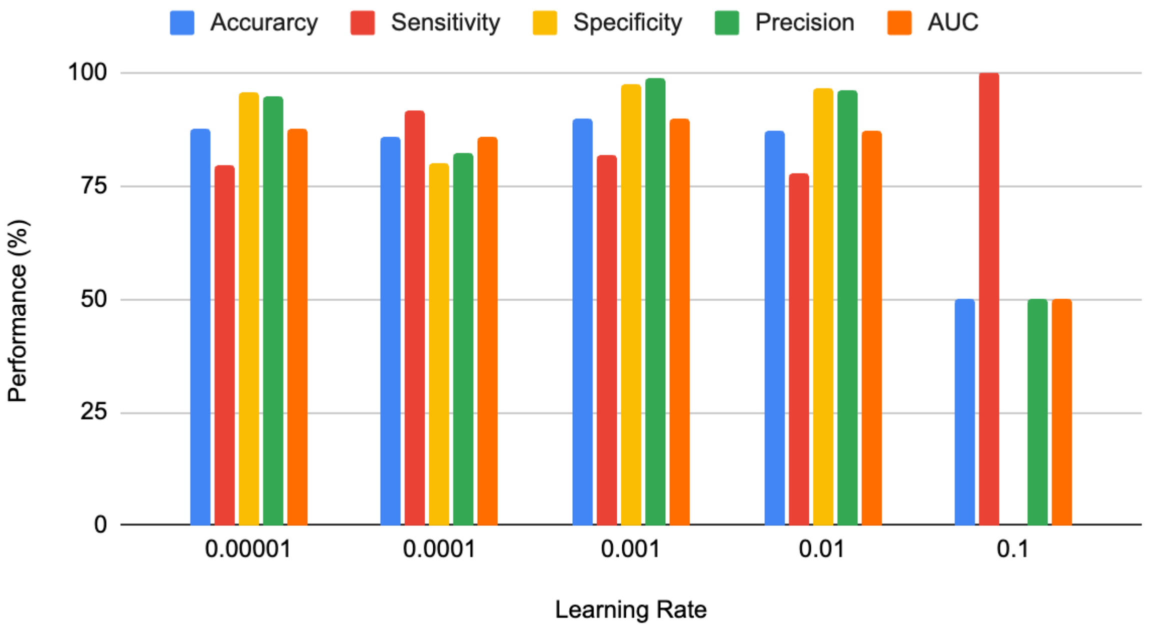

**Figure 4**. The model's performance using a InceptionV3 model with various learning rates and 10 epochs for the testing set

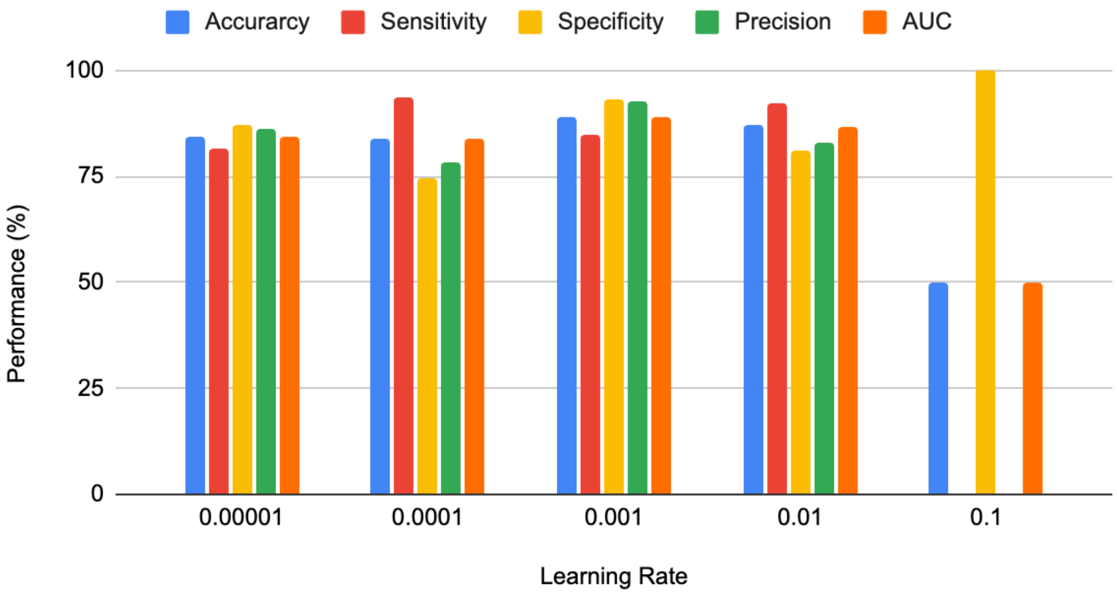

**Figure 5.** The model's performance using a VGG16 model with various learning rates and 10 epochs for the testing set

The use of the InceptionV3 model resulted in higher metrics over the VGG16 model when various learning rates were tested (Table 2). Additionally, the performance of both models went up as the learning rate decreased, reaching optimal values at a learning rate of 0.001. Table 2 illustrates a comparison between the VGG16 and InceptionV3 model with different learning rates and a constant number of epochs of 10. Based on our results, we used a learning rate of 0.001 and an epoch number of 10 for both models. Figure 6 also depicts the ROC graph and AUC score for when an InceptionV3 (left) and VGG16 (right) model is used with a learning rate of 0.001 and 10 epochs, showing how the InceptionV3 model is optimal with its AUC score of 0.9

**Table 2**. Metrics of VGG16 and InceptionV3 model using 10 epochs for the testing set with various learning rates

| **Learning Rate** | **Model** | **Accuracy** | **Sensitivity** | **Specificity** | **Precision** | **AUC** |
|---|---|---|---|---|---|---|
| 0.00001 | VGG16 | 84.2% | 81.4% | 87% | 86.2% | 0.842 |
| | InceptionV3 | 87.6% | 79.6% | 95.6% | 94.8% | 0.876 |
| 0.0001 | VGG16 | 84% | 93.6% | 74.4% | 78.5% | 0.84 |
| | InceptionV3 | 86% | 91.8% | 80.2% | 82.3% | 0.86 |
| 0.001 | VGG16 | 89.1% | 85% | 93.2% | 92.6% | 0.891 |
| | InceptionV3 | **89.7%** | **83%** | **97.4%** | **98.8%** | **0.9** |
| 0.01 | VGG16 | 86.9% | 92.4% | 81.2% | 83.1% | 0.868 |
| | InceptionV3 | 87.4% | 78% | 96.8% | 96% | 0.874 |
| 0.1 | VGG16 | 50% | 0% | 100% | 0% | 0.5 |
| | InceptionV3 | 50% | 100% | 0% | 50% | 0.5 |

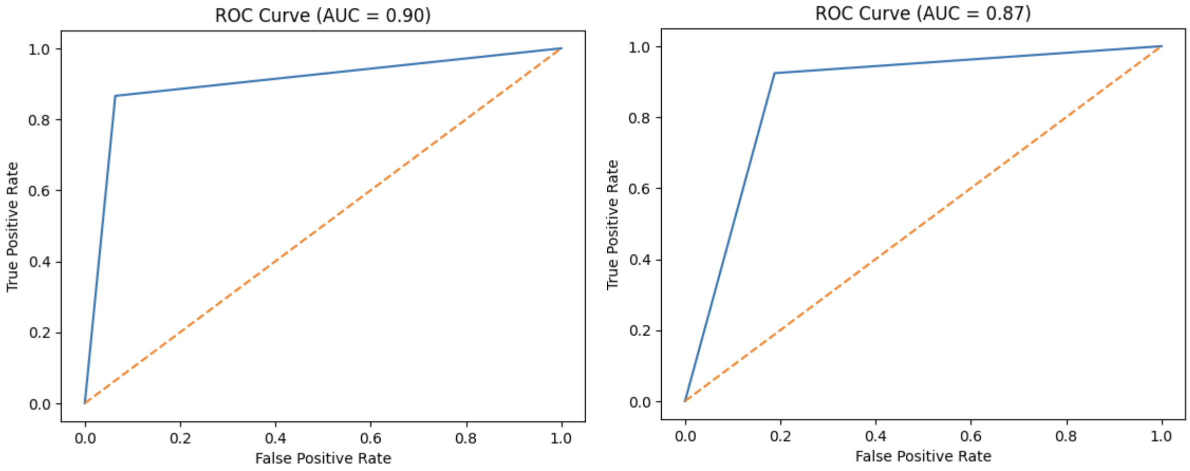

**Figure 6.** The model's ROC AUC using a InceptionV3 (left) and VGG16 (right) model with a learning rate of 0.001 and 10 epochs

## Discussion

The findings of this study demonstrate that transfer learning can be useful for identifying benign from malignant images and for early melanoma diagnosis. With regards to the optimized InceptionV3 model, 90% of the testing set's images were correctly classified by the trained model, which approximately had a 90% accuracy rate on the training set. These results suggest that transfer based learning has to the potential to help in the the diagnosis of melanoma.

The model represented a high sensitivity and specificity in addition to a good accuracy. The proportion of true positives that the model properly identifies is measured by sensitivity, and the proportion of true negatives is measured by specificity. The model in this study was able to properly detect the majority of malignant melanoma images (true positives) while avoiding incorrectly classifying benign images as malignant, with a sensitivity of 82% and a specificity of 97% (a lack of false positives), which means that only 3% was false positives. The AUC value of 0.90 also indicates that the model performed well in differentiating between the two classes. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) for different threshold values, and the AOC measures the area under this ROC curve.

Transfer learning is a powerful approach that can considerably boost model performance when it comes to applying deep learning to identify photos of melanoma cancer as benign or malignant, especially when there is a lack of available data. This feature makes transfer learning an ideal candidate for medical diagnosis due the lack of medical data. The choice of base model architecture in this situation can significantly affect the outcome. The VGG16 and InceptionV3 models are two well-liked options for basic models. A traditional deep learning architecture called VGG16 has 16 layers total, including several convolutional layers and fully connected layers (22). InceptionV3, on the other hand, employs a more intricate architecture with inception modules that support the acquisition of multi-scale information (23). In this study, InceptionV3 was discovered to perform better than VGG16 in terms of accuracy, sensitivity, and specificity. This shows that the more detailed design of InceptionV3 can more effectively capture the nuanced aspects of melanoma images and differentiate between benign and malignant cases. The results of this study demonstrate that InceptionV3 outperforms VGG16 in terms of classification of melanoma due to its capacity to collect the intricate features of the images and precisely differentiate between benign and malignant samples.

These results suggest that deep learning and transfer learning has the potential to be used in clinical settings to improve diagnostic accuracy and reduce the chance of human error.. Since this model was trained on a small dataset of 10000 images, this model would need to be adapted to a greater dataset of 1,000,000 images in order to predict skin cancer for large populations in clinical settings. Privacy concerns must also be taken into consideration with regard to using the medical data of patients.

Some ways the model can be improved is by increasing the number of images that are used for training the model, which can increase the model's accuracy and precision. The model can be further improved by optimizing the hyperparameters of the model and adding additional layers to enhance the extraction of the image's features. We can make the model identify and classify multiple types of skin cancer and not just melanoma, which can be even more applicable for the health-care industry and helpful for future innovation. The robustness of the model can also be improved by incorporating artificially noised data in the training and validation datasets.

## Conclusion

In this study, we compared a transfer learning image classifier based on the InceptionV3 and VGG16 model in Python in order to classify an image of melanoma into benign or malignant. To this end, a public dataset from Kaggle was used, containing 10000 images. We optimized both models with respect to the number of epochs and the learning rate, achieving an optimal learning rate of 0.001 and 10 epochs. Our results demonstrated that InceptionV3 outperformed VGG16 with regards to metrics, including accuracy, specificity, sensitivity, and precision. The optimal model with InceptionV3 was able to achieve an AUC of 0.9, an accuracy of 90%, a precision of 98%, and a specificity of 97% for the testing dataset. This study can potentially contribute to the future research and diagnosis for the application of deep learning to health care.

## Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

## References

Brodzicki, A., Piekarski, M., Kucharski, D., Jaworek-Korjakowska, J., & Gorgon, M. (2020). Transfer learning methods as a new approach in computer vision tasks with small datasets. *Foundations of Computing and Decision Sciences*, *45*(3), 179-193.

Cagnetta, F., Favero, A., & Wyart, M. (2023). What can be learnt with wide convolutional neural networks?.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), 832.

Goyal, M., Knackstedt, T., Yan, S., & Hassanpour, S. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in biology and medicine*, *127*, 104065.

Guy, G. P., & Ekwueme, D. U. (2011). Years of potential life lost and indirect costs of melanoma and non-melanoma skin cancer: a systematic review of the literature. *Pharmacoeconomics*, *29*, 863-874.

Javid, M. (2022). *Melanoma Skin Cancer Dataset of 10000 Images* [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/3376422

Johansen, T. H., Møllersen, K., Ortega, S., Fabelo, H., Garcia, A., Callico, G. M., & Godtliebsen, F. (2020). Recent advances in hyperspectral imaging for melanoma detection. *Wiley Interdisciplinary Reviews: Computational Statistics*, *12*(1), e1465.

Neyshabur, B., Sedghi, H., & Zhang, C. (2020). What is being transferred in transfer learning?. *Advances in neural information processing systems*, *33*, 512-523.

Painuli, D., & Bhardwaj, S. (2022). Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review. *Computers in Biology and Medicine*, *146*, 105580.

Ribani, R., & Marengoni, M. (2019, October). A survey of transfer learning for convolutional neural networks. In *2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)* (pp. 47-57). IEEE.

Safran, T., Viezel-Mathieu, A., Corban, J., Kanevsky, A., Thibaudeau, S., & Kanevsky, J. (2018). Machine learning and melanoma: the future of screening. Journal of the American Academy of Dermatology, 78(3), 620-621.

Wang, J. Y., Wang, E. B., & Swetter, S. M. (2023). What is melanoma?. *JAMA*, *329*(11), 948-948.