CP468 - Artificial Intelligence
Spring 2024
Group 4
Project Report

Member 1 Name: Adelina Javgureanu	ID: 169039718
Member 2 Name: Gracie Heer	ID: 169034824
Member 3 Name: Jamie Beatty	ID: 169027003
Member 4 Name: Micah Billington	ID: 169023902
Member 5 Name: Yusra Hassan	ID: 169024293

Submission Date: August 2, 2024

Table of Contents

All listed members have contributed, read, and approved this submission.

Member 1 Name: Adelina Javgureanu Signature:
Member 2 Name: Gracie Heer Signature:
Member 3 Name: Jamie Beatty Signature:
Member 4 Name: Micah Billington Signature:
Member 5 Name: Yusra Hassan Signature:

Abstract

This report provides a comprehensive study on Melanoma skin cancer detection using convolutional neural networks (CNNs). Melanoma is one of the most critical cases of skin cancer, so early detection and accurate diagnosis are important for better patient outcomes. This project will be divided into two phases: Design from scratch a CNN model using the TensorFlow and Keras libraries to classify skin lesions as malignant or benign. After that, implement it, train it, and finally evaluate the provided design. Further on, in the second phase, we are going to make good use of transfer learning by experimenting with three pre-trained deep learning models for fine-tuning to improve classification accuracy in Melanoma detection using the TensorFlow API and Gradio. In this regard, a custom-built CNN is compared to pre-trained models for assessing the efficiency of transfer learning in improving diagnostic accuracy. The results are analyzed by studying different metrics such as accuracy, TPR, TNR, and AUC-ROC. This paper demonstrates the deep learning potential for increasing skin cancer diagnostics and provides a concrete view into how to apply CNNs concretely in medical image analysis.

Introduction

The American Cancer Society reports that skin cancer is the most common type of cancer a person can have throughout their lifetime. Melanoma happens to be the deadliest among all its forms, as this particular skin cancer has the fastest ability to spread. It is, therefore, essential to detect it early since the five-year survival rate for stage 1 melanoma is nearly 99 percent, while this drops significantly in later stages. The Public Health Agency of Canada projected that in 2019, 7,800 Canadians would become afflicted with melanoma, whereas 1,300 would die from the disease. Early identification is key to any effective treatment, which requires a safe, cost-effective diagnostic tool that is accurate. CNNs, in identifying malignant tumors from photos, have shown a great deal of promise with an accuracy as high as 95%. Models based on color and shape can give an early diagnosis, allowing patients to seek medical attention in time for the initiation of treatment, which a doctor can do to better case management. This therefore breaks major challenges related to distance, cost, and time, improving treatment outcome.

Similar Solutions

The ISIC 2024 competition aims to create AI algorithms to detect skin cancer using 3D total body photographs (TBP). These images look like smartphone photos often used in telehealth. The goal is to make the algorithms useful in places without specialized skin care. Participants develop algorithms to identify cancerous lesions, focusing on achieving high accuracy, especially for detecting true positive cases. This competition helps underserved populations by improving early skin cancer detection and triage, leading to better patient outcomes. This competition addresses a similar solution as ours by using advanced AI techniques to detect melanoma, showing how technology can enhance early detection and treatment.

Another similar solution is the International Skin Imaging Collaboration 2017, which achieved notable results with their algorithms, reaching an area under the ROC curve of 0.87 and a specificity of 85%, surpassing the performance of sampled dermatologists[1]. Additionally, a systematic review titled "Analysis of Artificial Intelligence-Based Approaches Applied to Non-Invasive Imaging for Early Detection of Melanoma" reported a mean specificity of 85.58% and an ROC greater than 80%[2]. These findings underscore the potential of AI to enhance melanoma detection accuracy and efficiency, reflecting the objectives and methods of our project.

Methodology

The way we tackled this problem was to use three pre-trained models from keras (EfficientNet, ResNet, DenseNet) as well as a model by scratch, all specialized in image classification. We then added layers to refine the model's ability to distinguish between different moles. We included augmented photos in the training and validation data in order to improve results and minimize overfitting and finally tested our models against the test sets in order to classify photos as melanoma or non-melanoma.

Description of the Data

Our dataset (found at this link) contains 10,015 jpg images of different kinds of moles including melanoma and benign lesions. The pictures comprise of 7 unique classes in total. The pictures can be of moles on different kinds of places on the body and all information including this, the mole type and the sex/age of the person is included in a metadata .csv file. The type of mole (eg. whether a picture shows melanoma or not) can be identified with its file name (without the extension) by matching it with dx_type.

Data Pre-processing

Initially, only around 11% percent of our images belonged to the melanoma class. In an attempt to create a more balanced dataset for classification, we decided to augment the existing melanoma images by applying random alterations to them such as flips, zooms, translations and rotations. After adding in the augmented melanoma images to our original dataset, our dataset increased to 27,312 images, half of which belonged to the melanoma class. We added the "Melanoma" label to each of the augmented values and created a dataframe with the image ID's and class names, which were "Melanoma" and "Non-Melanoma"

Pre-trained Models

The methodologies for using pre-trained models such as EfficientNetB0, ResNet and DenseNet121 to classify melanoma images involve several key steps, which differ from training a model from scratch in various aspects. For all pretrained models, the process begins with

importing necessary libraries and loading the pretrained base models from TensorFlow. The image data is then loaded and preprocessed, including resizing, normalizing, and augmenting the images to improve model generalization. The pretrained models' top layers are replaced with custom dense layers tailored to the binary classification task of detecting melanoma. These layers include Flatten, Dense, Dropout, and a final Dense layer with a sigmoid activation function. L2 regularization was applied to each model's dense layer in an attempt to reduce overfitting in the training. The models are compiled with an appropriate optimizer (Adam) and loss function (binary cross-entropy), and then trained using the preprocessed data. Validation and training accuracies and losses are monitored to prevent overfitting. EfficientNet, which was found to be most prone to overfitting, applies tensorflow early stopping as well.

By Scratch Model

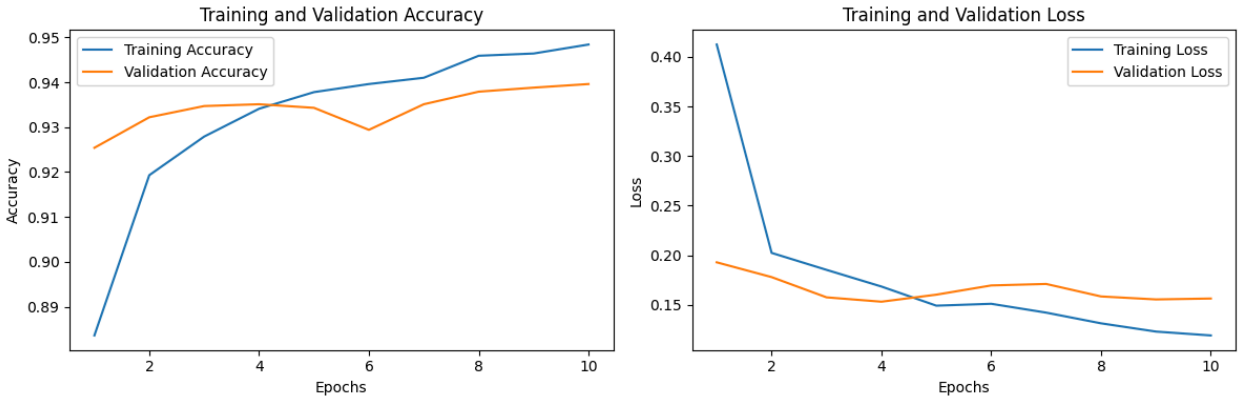
This model creates a convolutional neural network (CNN) from scratch to classify images as melanoma or non-melanoma, distinguishing itself from pretrained models by building and training the network architecture from the ground up. The process begins by importing necessary libraries, mounting Google Drive to access the dataset, and loading the image data along with corresponding labels from a metadata file. Data augmentation and preprocessing are performed using an ImageDataGenerator, splitting the data into training and validation sets. Unlike pretrained models, this approach constructs a CNN model from scratch with layers, batch normalization, activation layers, and a final dense layer for binary classification. The model is compiled with the adam optimizer and binary cross-entropy loss and trained for 10 epochs, incorporating callbacks for checkpointing, early stopping, and learning rate reduction. The trained model is then saved, evaluated on sample images and test data. This methodology emphasizes the configuration and training of a custom model architecture, requiring more data and computational resources compared to fine-tuning pretrained models but offering complete control over the model's design and training process.

Results

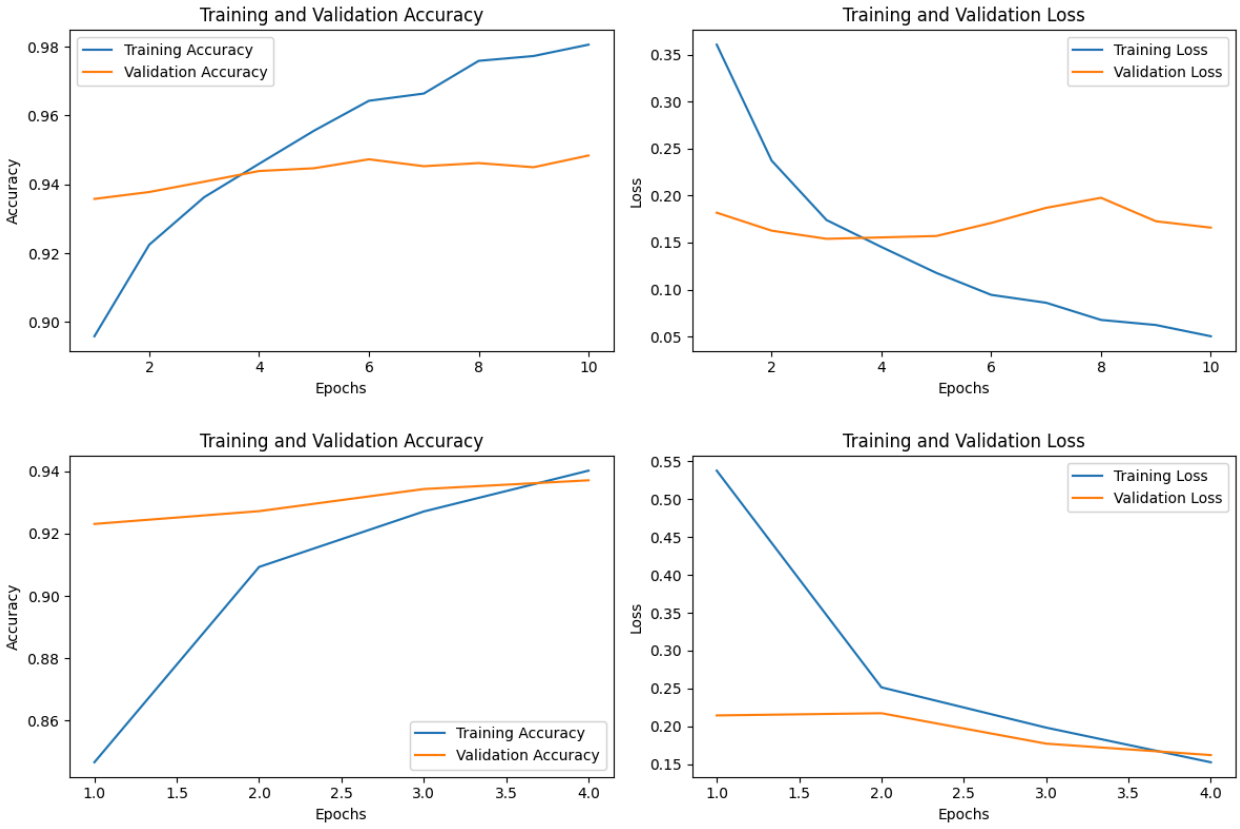
Training Curves

These graphs show how the accuracy of the model increases over time and the loss decreases.

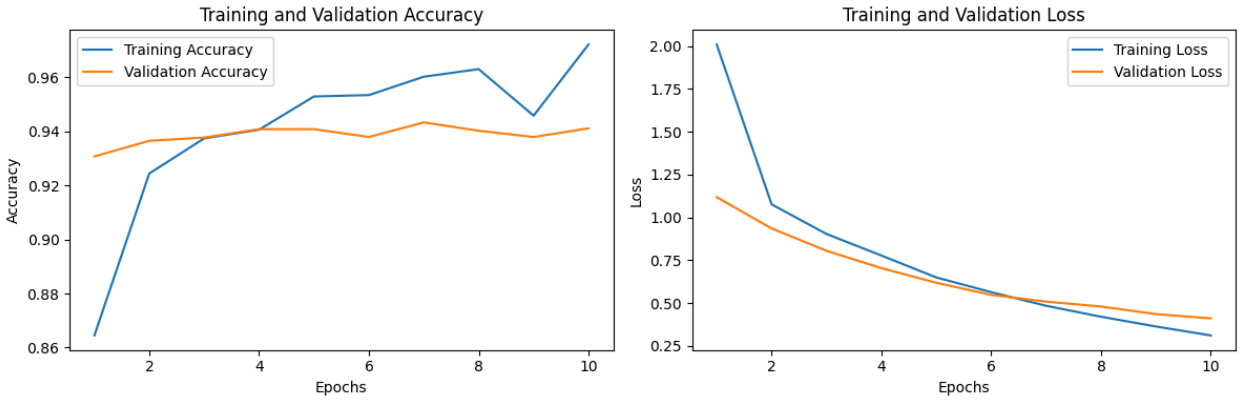
DenseNet: From the first graph, we can see that the validation accuracy actually has a slight dip at 6 epochs, however it is important that the model continues training as we see the accuracy improves over time. On the second graph we see that although the training loss is consistently decreasing, the validation loss begins to increase after the 4th epoch, so in order to maximize accuracy and minimize loss, we continue to train until the 10th epoch.



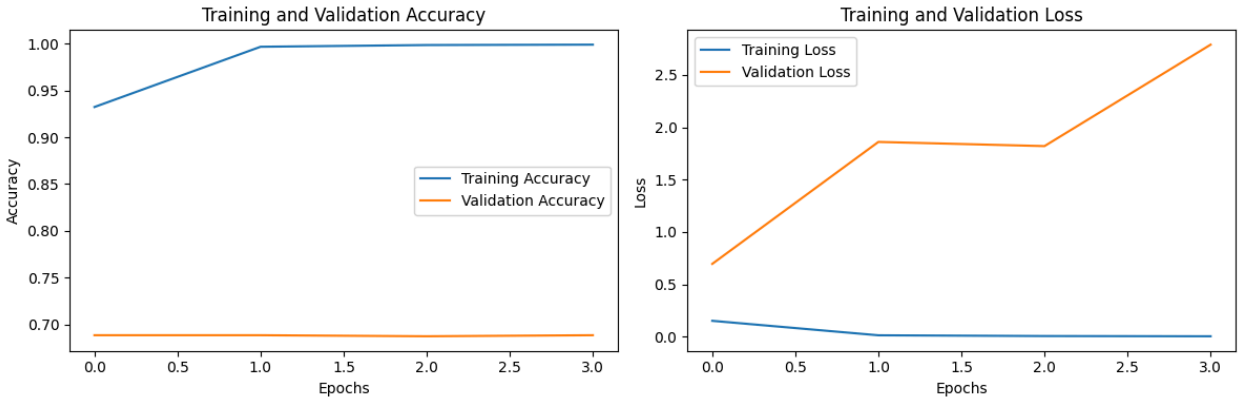
EfficientNet: This model, similar to the one above, has much better training accuracy and loss than validation. For this reason, we attempted early stopping to stop the overfitting, however it failed. So the first set of graphs shown here show that accuracy in validation is barely improving, while its loss is simply unpredictable. In the second set of graphs however, we see that by strategically lowering the number of epochs, we can receive better results where both accuracies are increasing together and both losses are decreasing together.



ResNet: This model has a dip in its training accuracy at 9 epochs, despite pretty consistently increasing validation accuracy. The training and validation accuracy by the 10th epoch are both over 90%. Its loss curves both decrease towards 0 with more epochs, however, the validation loss becomes slower in its decrease. By the 10th epoch, both are less than 0.5.

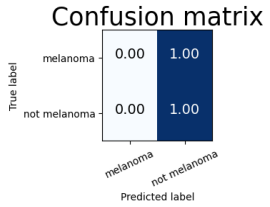
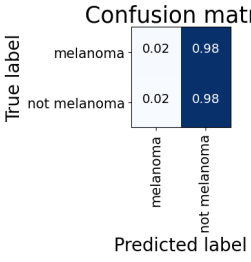
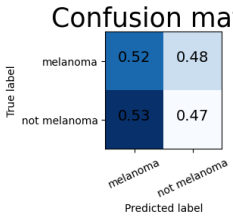


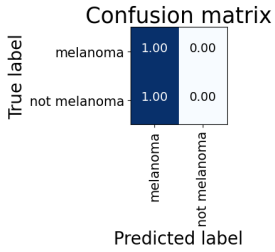
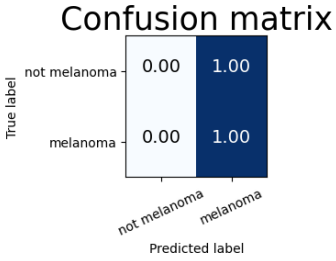
Model from scratch: In this model we can see the effect of stopping the epochs when validation accuracy is not improving. The model has stayed largely constant, this could, however, be also from the very fact that there are only four epochs. It is interesting to note that the validation loss is the only statistic with much change, but it is increasing. Since the training accuracy is much greater than validation, without much change, it would be beneficial to reduce the amount of layers in the model, as it is immediately overfitting.



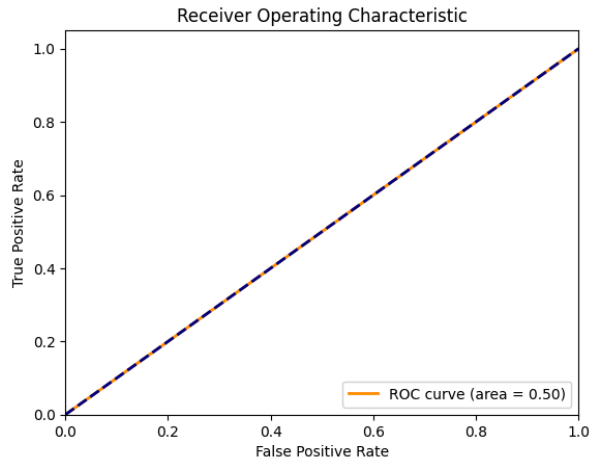
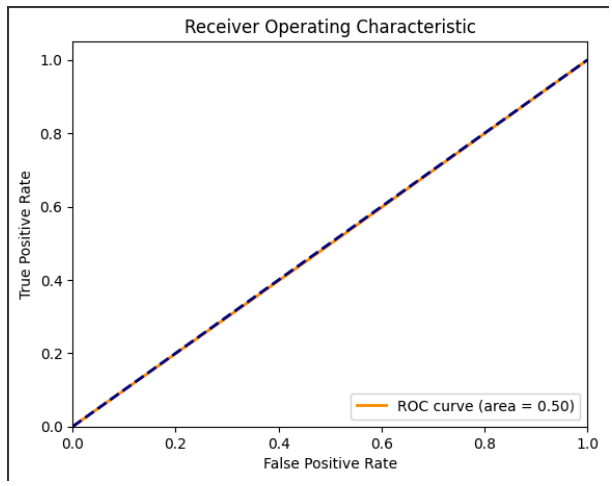
Quality Measures

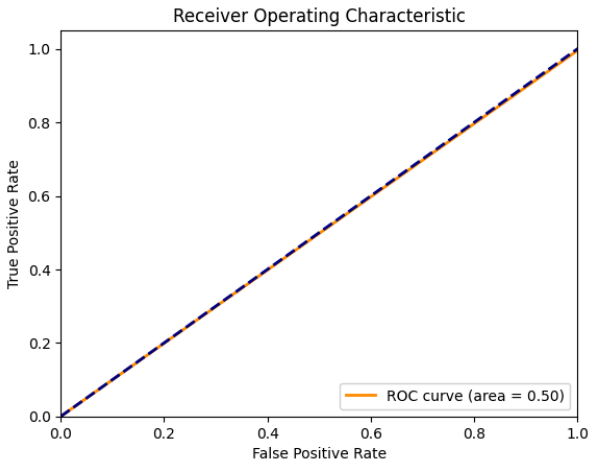
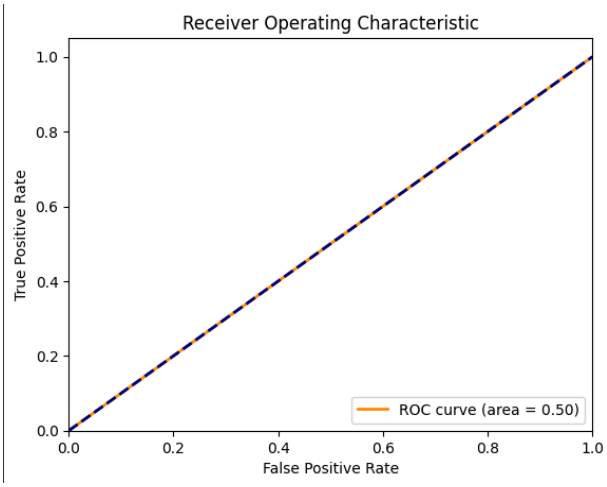
True positives, true negatives, false positives and false negatives can be found below in the confusion matrix of each model. We can see from the results that densenet and efficientnet seem to be the best at minimizing error while ResNet and the Scratch Model have extreme predictions.

ResNet	DenseNet	EfficientNet
		

Scratch	Ensemble
	

AUC & AUROC

ResNet	EfficientNet
	

DenseNet	Scratch
	

All the models showed the same results for the AUROC and AUC curves -an area of 50. This is expected from the confusion matrices. Other than efficient net this is overfitting, in its case it would benefit from additional epochs. The other models would improve with more data, and trimming the model.

Conclusion

In conclusion, our models demonstrated robust training accuracy, although this did not translate to equally strong test accuracy. Throughout our project, we encountered challenges with overfitting during testing and validation phases, an issue we would aim to fix with additional time and resources. Further exploration into identifying the specific types of moles our model commonly misclassified as melanoma would enable targeted improvements. Improvements could be implemented by trimming the models layers, using more or less epochs, using different augmented data, increasing the randomness, obtaining more data samples, and other modeling techniques, such as were used in the ISIC competitions. Looking ahead, our melanoma classification model may prove useful for medical pre-screening applications, potentially serving as a tool for earlier detection of melanoma. The model could find applications in dermatological clinics as well as telemedicine platforms focused on skin health monitoring and assessment.

Works Cited

[1]Author links open overlay panelMichael A. Marchetti MD a, a, b, c, d, e, … Codella, N. (2019). Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the International Skin Imaging Collaboration 2017. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0190962219323734?casa_token=-VtHm8QvJ8cAAAAA%3AUwqA2q44erkKsOVQtCrqfl-z8CiQXFP7zJ-ka5lGL4tQO9iVTKy5NpwUMOiro8-osTbFomuvzbsg

[2]Patel, R. H., Foltz, E. A., Witkowski, A., & Ludzik, J. (2023). Analysis of artificial intelligence-based approaches applied to non-invasive imaging for early detection of melanoma: A systematic review. Retrieved from https://www.mdpi.com/2072-6694/15/19/4694

[3]Common moles, dysplastic nevi, and risk of melanoma. NCI. (n.d.). https://www.cancer.gov/types/skin/moles-fact-sheet#:~:text=In%20advanced%20melanoma%2C%20the%20texture,itchy%2C%20tender%2C%20or%20painful

[4]International Skin Imaging Collaboration (ISIC) melanoma project. (2021). Retrieved https://dermatology-research.centre.uq.edu.au/project/international-skin-imaging-collaboration-isic-melanoma-project