Machine Learning in the Medical Field: Identifying Early-Stage Melanoma

Irfan Iqbal
Dept. of Computer Science
Cal Poly Pomona
Pomona, USA
iiqbal@cpp.edu

Gabriel George
Dept. of Computer Science
Cal Poly Pomona
Pomona, USA
gdgeorge@cpp.edu

Jessica Ortega
Dept. of Computer Science
Cal Poly Pomona
Pomona, USA
jeortega1@cpp.edu

Jennifer Cardenas
Dept. of Computer Science
Cal Poly Pomona
Pomona, USA
cardenas2@cpp.edu

Sebastian Chinen
Dept. of Computer Science
Cal Poly Pomona
Pomona, USA
sachinen@cpp.edu

***Abstract*—This document articulates the efficiency and positive impact of utilizing machine learning to identify and diagnose Melanoma. 7600 images from the International Skin Imaging Collaboration were used to train various pre-trained neural network models that have been fine-tuned with extra final layers added. Transfer learning was used to test multiple models efficiently to determine the best accuracy for the identification of cancerous skin segments. The end goal is to create a system that can minimize errors between human melanoma screenings and increase the availability of accurate diagnosis tools.**

I. Introduction

Melanoma is the most severe type of skin cancer that develops in the body's melanocyte cells responsible for producing melanin and can also form in several areas of the body, such as the eyes and, in some instances, inside the nose or throat [1]. When identified and treated in its earliest stages, melanoma is readily curable. While it is typically detectable by simple visual inspection, early stages may be difficult to distinguish from benign skin lesions by the naked eye. Inexpert screenings for melanoma could lead to numerous unnecessary biopsies, while the financial costs of failing to diagnose early are also considerable [2]. On the one hand, potentially lethal melanomas may be missed; on the other, many unneeded biopsies result in overdiagnosis.

Despite being the most lethal type of skin cancer, there is a 98 percent 5-year survival rate if melanoma is diagnosed while still confined to the outer layer of the skin [2]. Using computer vision and machine learning, technology can help identify melanoma while still in its early stages. Thus, our motivation for this paper is to create a machine-learning model that can differentiate if a melanoma-looking mole is benign or malignant through a single picture. Creating an accurate machine-learning model can drastically reduce unnecessary biopsies and financial costs for each screening. A successful machine-learning model can solve to issue of both overdiagnosis and increase accessibility for those who cannot get frequent screenings to prevent missed melanomas which could be potentially lethal cancers.

We will be using images from the International Skin Imaging Collaboration (ISIC) database for our data set. The database contains a plethora of mole images that are classified by their diagnosis, benign or malignant, as well as additional meta-data provided. With this data set, we aim to create a machine-learning model using supervised neural networks that can accurately perform binary classification on pictures of moles. Neural networks process data through layers of neurons meant to imitate the human brain. By inputting the image pixels and having the algorithm adjust the weights of each neuron, the model will be able to find patterns in the images and correctly identify the diagnosis of each mole.

II. Dataset to be used

The data set comes from ISIC (International Skin Imaging Collaboration). ISIC aspires to reduce melanoma mortality by providing a standard of high-quality digital dermatologic imaging that may be utilized by medical professionals and the computer science community to help diagnose melanoma at an earlier stage where the melanoma may be curable. There are currently 71670 images available of moles, and their diagnosis(benign, malignant, etc.), clinical(melanoma class, melanoma type, etc.), and technological (dermoscopic type and image type) attributes are specified as well.

The data set we will be using will only contain images and not additional metadata that is provided by the ISIC such as gender, melanoma type, and mole size. We chose a subset of 7600 images from the ISIC database. The goal of analyzing this set of images is binary classification to determine whether the moles are benign or malignant. Since the ISIC database is so large, we chose 3800 images from each class to make our dataset balanced even though the entire ISIC database is quite unbalanced with about 90% benign and 10% malignant.

III. Methodology including the machine learning model proposed

The machine learning model utilized will be a supervised-learning neural network. Supervised machine learning is a subcategory of artificial intelligence and machine learning that uses labeled data sets to train algorithms to classify/predict outcomes accurately by including inputs and correct outputs in the training data set that allow the model to learn over time [2]. The algorithm measures the model's accuracy through the loss function and adjusts until the error is minimized significantly [2]. Therefore, the method of validation for the supervised-learning neural network will be the train/test split method. ISIC's database will be split into training data that will be fed to the algorithm for training purposes and split into testing data that will be utilized to test the model's accuracy. We used a split of 75 training data nad 25 percent for testing and validation.

Neural networks process training data through layers of nodes meant to mimic the inter-connectivity of a human brain. Each node is composed of inputs, a bias/threshold, weights, and an output [2]. Through supervised learning, a neural network learns a mapping function and, through gradient descent, adjusts based on the loss function [2]. The goal is for the model's cost function to be near zero to ensure the model's accuracy is high in obtaining the correct answer [2].

Through the utilization of a supervised-learning neural network, complex patterns of pixels in mole images can be recognized that correlate to the diagnosis of early-stage melanoma. ISIC's database was made available to medical professionals so that they can view the mole images and recognize the patterns in early-stage melanoma. The goal is to utilize ISIC's database and teach an algorithm to replicate the inter-connectivity of a human brain and reduce the human error factor of melanoma assessments and reduce the wait time between appointments of getting a mole assessed for melanoma so that melanoma may be diagnosed in an early stage and can be readily treatable.

In order to achieve this, the process of transfer learning was utilized. Transfer learning is a machine learning technique that involves using knowledge gained from a previous task with a large dataset that is incorporated into a similar task in order to improve the performance of the new task as the model utilizes the previously learned information. To train a model to classify the various moles and distinguish between the malignant and benign variety, several pre-trained models that had already been trained on large image-based datasets were used. The pre-trained models were then respectively fine-tuned to fit the melanoma dataset, which was significantly smaller and contained fewer classes than the original dataset. By using a pre-trained model as a starting point, transfer learning significantly reduced the amount of data and time required to train a new model while increasing the accuracy. It also allows us to test multiple models and variables to efficiently find which model best suits our task.

By leveraging pre-trained models such as ResNet50 and MobileNetV2, the model can easily identify edges in images and learn specific layers to identify shapes and other features that are important for classifying malignant or benign moles.

Since neural networks may be subject to over-fitting, which occurs when the model is able to correctly classify your dataset, however, it is inaccurate when tested on a different dataset. This is due to the way the model learns the data set's patterns instead of a universal pattern that could be applied to a different data set.

One of the workarounds that was utilized was data augmentation through image prepossessing. The training data was modified through normalization, reshaping, and augmentation. Firstly, the 7600 images that were used for training and validation were subject to normalization which divide all pixel values by 255 to convert them to range 0 to 1. This made it easier to compare each one to the other. The images were also reshaped so that the shape or size would be uniform across all images. While doing this it is important to preserve the content and relative proportions of its features. Reshaping images to a standard size and shape ensures that the input data to the machine learning model is consistent. Each image will subsequently contain the same number of pixels, channels, and aspect ratio which leads to an increase in accuracy. Another way that over-fitting is avoided is by changing the images' colors, brightness, and contrast. This improves the robustness of the model to handle variations in the input data, as it ensures that the model is not over-fitting to specific color or brightness levels in the training set. Similarly, the images were also flipped in orientation and rotated to create variation in data. The aggregate effect of these augmentation strategies improved the accuracy and applicability of the model.

To determine which pre-trained model would yield the best accuracy, multiple models were tested along with pre-processing and image augmentation before training. We used the TensorFlow Keras library to import the pre-trained models ResNet50, Xception, MobileNetV2, and InceptionV3. Any additional layers to be added to the models were also imported from the Keras library. Before training the models, ImageDataGenerator and flow_from_dataframe were used to pre-process images. We chose to re-size all images to 224 x 224 pixels in the process to allow for faster computation while still maintaining a good amount of detail. The batch size chosen was 32. Images were shuffled since they were initially loaded in order of all benign and then malignant.

The Images for the ResNet50 model were pre-processed with the preprocess_input preprocessing_function from the Keras applications library for ResNet50. The additional special layers added for this model were two dense layers with 128 neurons and the rectified linear unit function as well as a final output dense layer with the softmax activation function. The Xception model was tested with custom pre-processing. The pre-processing included normalization as mentioned above, horizontal and vertical shifting of twenty percent, rotation range of 20 degrees, and allowing for random horizontal flipping. The additional layers included were slightly different from those added to the ResNet50 model. For this model, a

GlobalAveragePooling layer was added followed by a dense layer, a dropout layer, and another dense layer. The dense layers had 200 and 170 neurons respectively with the Exponential Linear Unit activation function. The Dropout Layer had a rate of 0.4 or 40 percent. The MobileNetV2 and InceptionV3 models had the same pre-processing and final layers as the Xception model.

Each model was trained with the binary_crossentropy loss function and the adam optimizer with a learning rate of 0.001. We ran 10 epochs of each model which ran at different speeds but were generally about 5-6 hours when testing multiple variations of each one. The Xception model took the longest at almost 6 hours while the MobileNetV2 was surprisingly fast at only about an hour and a half.

IV. Results

After training the models, we can get the validation accuracy by running and evaluating them using test data. Below are the test accuracies for each model.

Model	Test Accuracy
ResNet50	79.52%
Xception	85.84%
MobileNetV2	82.00%
InceptionV3	82.53%

Figure 1 shows the Accuracy and Loss graphs for ResNet50 as it ran through the 10 epochs. Figure 2 depicts the confusion matrix after evaluating the model on the testing data.

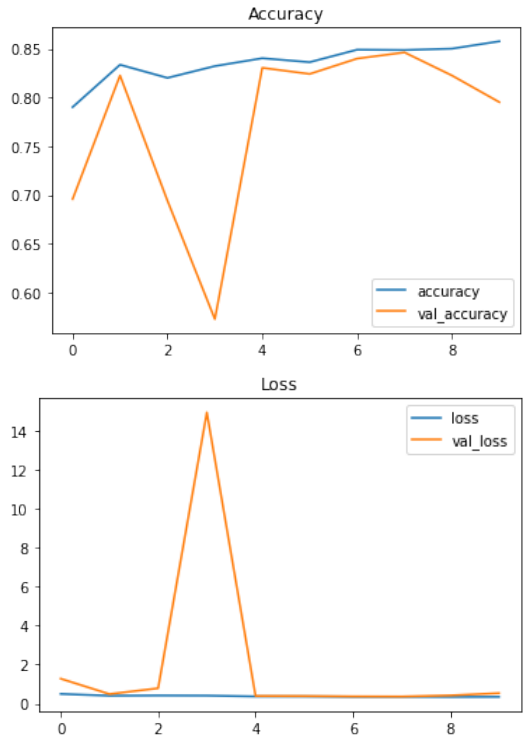

Fig. 1. Accuracy and Loss for ResNet50

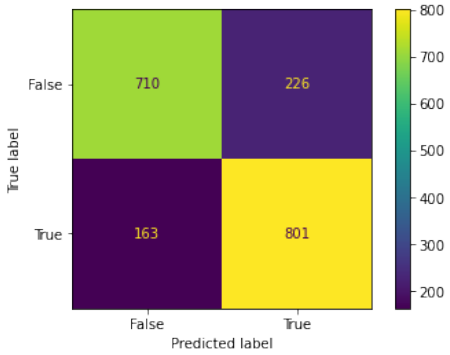

Fig. 2. Confusion Matrix for ResNet50

Figure 3 shows the Accuracy and Loss graphs for Xception as it ran through the 10 epochs. Figure 4 depicts the confusion matrix after evaluating the model on the testing data.

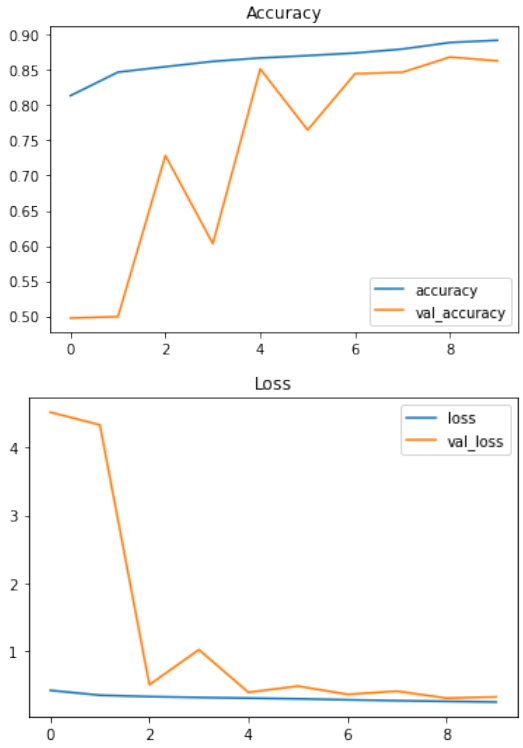

Fig. 3. Accuracy and Loss for Xception

Figure 5 shows the Accuracy and Loss graphs for MobileNetV2 as it ran through the 10 epochs. Figure 6 depicts the confusion matrix after evaluating the model on the testing data.

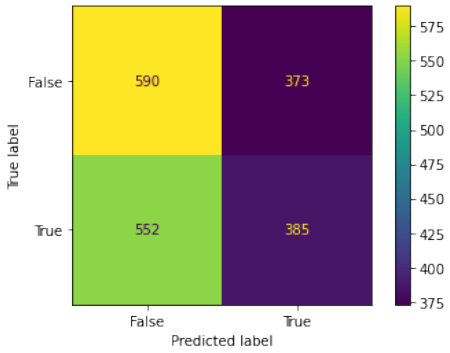

Fig. 4. Confusion Matrix for Xception

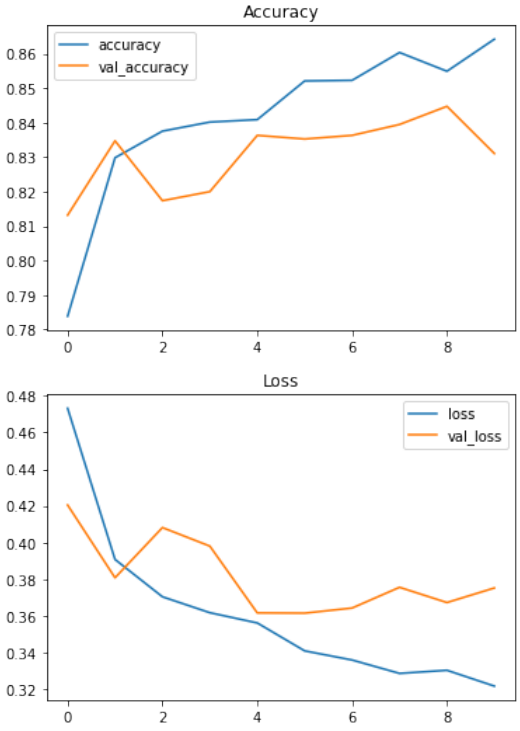

Fig. 5. Accuracy and Loss for MobileNetV2

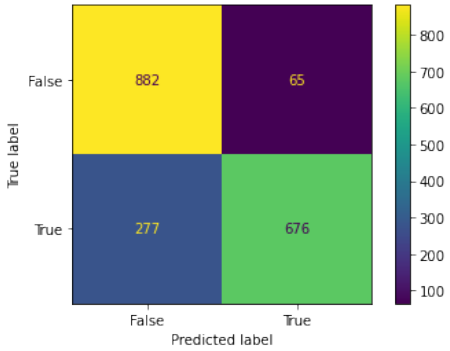

Fig. 6. Confusion Matrix for MobileNetV2

Figure 7 shows the Accuracy and Loss graphs for MobileNetV2 as it ran through the 10 epochs. Figure 8 depicts the confusion matrix after evaluating the model on the testing data.

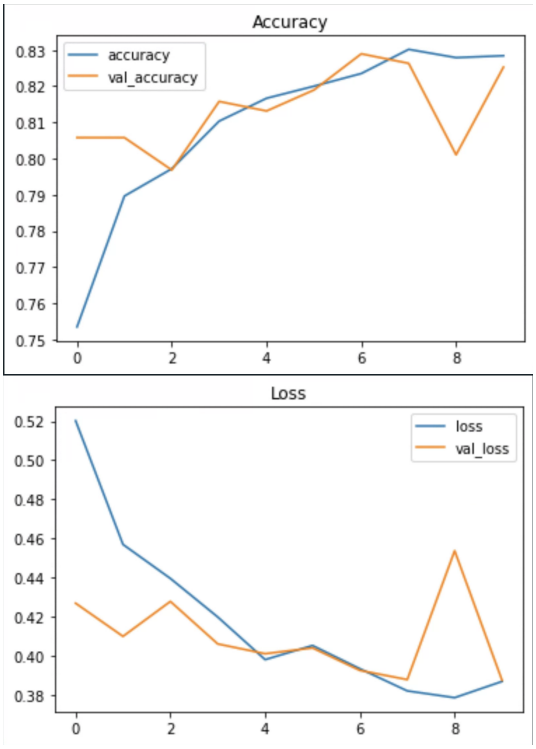

Fig. 7. Accuracy and Loss for InceptionV3

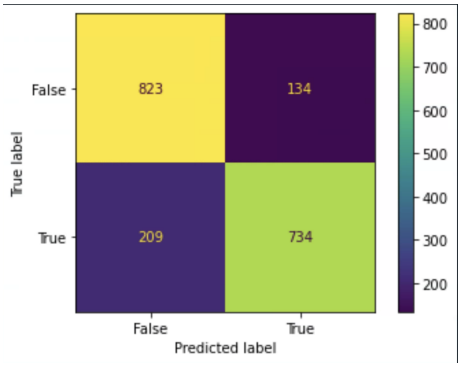

Fig. 8. Confusion Matrix for InceptionV3

Our results show that the model with the best accuracy after 10 epochs was the Xception pre-trained model with the GlobalAveragePooling, Dropout, and Dense layers added.

V. Related Work

A method for classifying melanoma images into two classes, malignant and benign, through a deep learning model and transfer learning utilizing the MobileNetV2 network as the base model was proposed due to its lightweight network[5].

In the research, 4 data sets were utilized, including a data set from the ISIC-archive repository. The ratio of training data to testing data was 70:30. The number of epochs used was 20 and the number of images per batch was 32[5]. In the training process, an Adaptive Momentum (Adam) optimizer was used and binary-cross-entropy was used as the cost function[5]. The utilization of the MobileNetV2 network as the base model, with a proposed head model of a global pooling layer followed by two fully-connected layers, led to a high accuracy of 85 percent for the ISIC-Archive data set [5].

Another method for classifying skin cancer was proposed with the utilization of the transfer learning nets ResNet50 and Xception [6]. The ratio of training data to testing data was 80:20[6]. The data set consisted of 10015 dermatoscopic images from the HAM10000 data set [6]. Both models were trained with 10 epochs and an image batch size of 32[6]. In the training process, an Adam optimizer with a learning rate of 0.001 was used and categorical-cross-entropy was used as the loss function [6]. In the results of the study, the ResNet50 model reached an accuracy of 77 percent while the Xception model reached an accuracy of 90.48 percent [6].

VI. Conclusion

Our approach to training a machine learning model that could identify cancerous moles was to use transfer learning to implement pre-trained models trained on extremely large databases while only modifying the final layers to fit our needs. Using this method, we were able to use a much smaller dataset and train different models a lot more efficiently. By testing multiple pre-trained models, we can determine which model would be best suited for further training with more epochs and more data. From the testing done, we can conclude that the Xception model would be the best suited for our goal of binary classification for melanoma.

While we kept the hyper-parameters and pre-processing mostly the same across the models tested, there are many variables that could have been changed to improve accuracy and determine the best model for our task. Further pre-processing for more variation in test data may have helped improve accuracy. We could have tested changing the final layers added to the pre-trained models as well as the number of neurons and activation functions for each layer. Different optimizers and learning rates could also have been tested when training each model.

More epochs and test data would almost certainly have improved accuracy at the cost of time and efficiency. Rather than change all the variables mentioned and try to fine-tune a single model, we wanted to experiment with how varied the results would be in different pre-trained models. From the results, we can further tweak the Xception model to produce the best accuracy with further training.

References

[1] P. Gottfrois, "Detect mole cancer with your smartphone using Deep Learning," InTech, May 22, 2018. https://medium.com/intech-conseil-expertise/detect-mole-cancer-with-your-smartphone-using-deep-learning-8afad1efde8a (accessed Mar. 13, 2023).

[2] "ISIC Archive," www.isic-archive.com. https://www.isic-archive.com (accessed Mar. 14, 2023).

[3] Mayo Clinic, "Melanoma - symptoms and causes," Mayo Clinic, 2019. https://www.mayoclinic.org/diseases-conditions/melanoma/symptoms-causes/syc-20374884 (accessed Mar. 13, 2023).

[4] IBM, "What is Supervised Learning? — IBM," www.ibm.com. https://www.ibm.com/topics/supervised-learning (accessed Mar. 15, 2023).

[5] R. Indraswari, R. Rokhana, and W. Herulambang, "Melanoma image classification based on MobileNetV2 network," Procedia Computer Science, vol. 197, pp. 198–207, 2022, doi: https://doi.org/10.1016/j.procs.2021.12.132.(accessed May 5, 2023).

[6] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep Learning-Based Transfer Learning for Classification of Skin Cancer," Sensors, vol. 21, no. 23, p. 8142, Dec. 2021, doi: https://doi.org/10.3390/s21238142.(accessed May 6, 2023).

VII. Supplementary Materials

LaTeX Files: *https://www.overleaf.com/read/pmhydgmwcrwx*
Source Files: *https://github.com/irfan61802/Melanoma-Detection*