Identifying Melanoma With Deep Learning

Devan Agrawal

July 2020

Inspiration

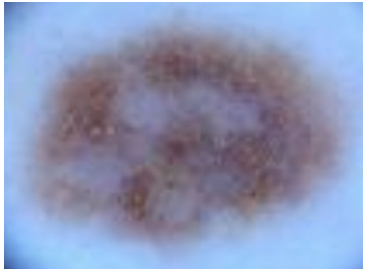
- Melanoma is the rarest type of skin cancer
- Despite this, it causes the most deaths
- Doctors aren't very accurate in identifying it, so machine learning may perform better than an average dermatologist
- Possible Application: Machine learning could provide doctors with a second opinion on melanomas
- Possible Application: Machine learning could be integrated into an app to identify malignant and benign skin lesions

Datasets

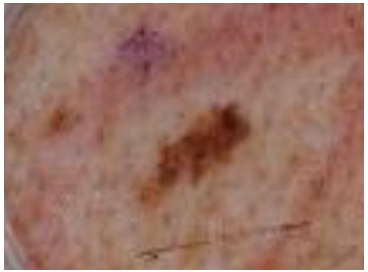
- I used a dataset from kaggle
- The model was provided with images sorted by melanoma or melanocytic nevus, a benign mole often confused with melanoma
- I used 4522 images of melanomas and the same amount for melanocytic nevi
- Each image was resized to 120x80 pixels
- Link to Data:
    - https://www.kaggle.com/andrewmvd/isic-2019

Example Images From Dataset

Melanoma:

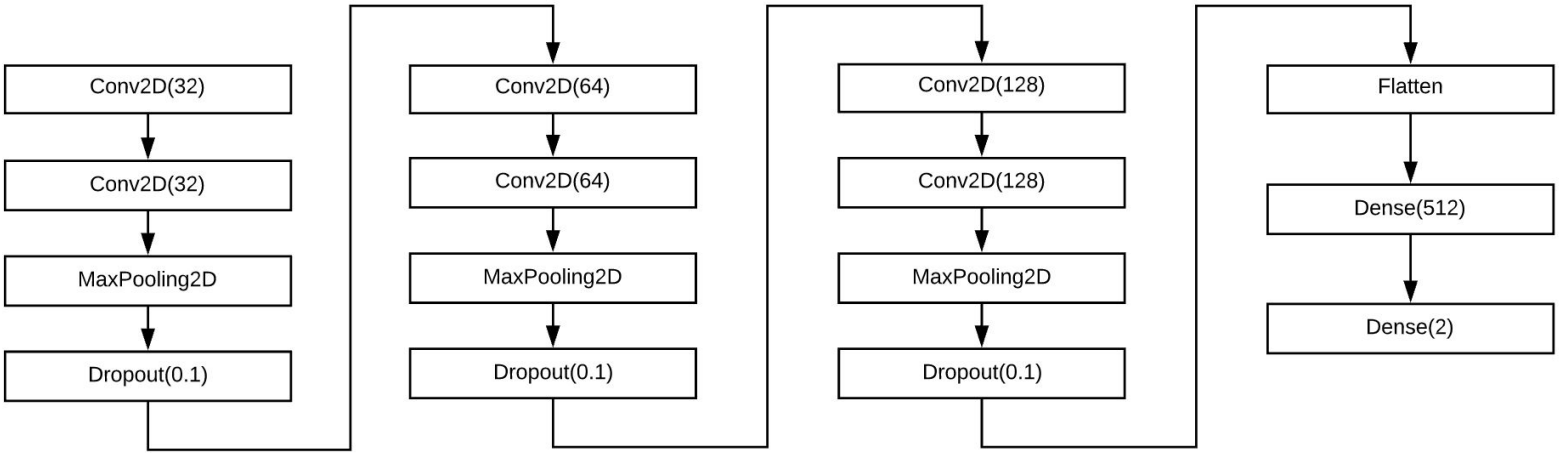




Melanocytic Nevi:

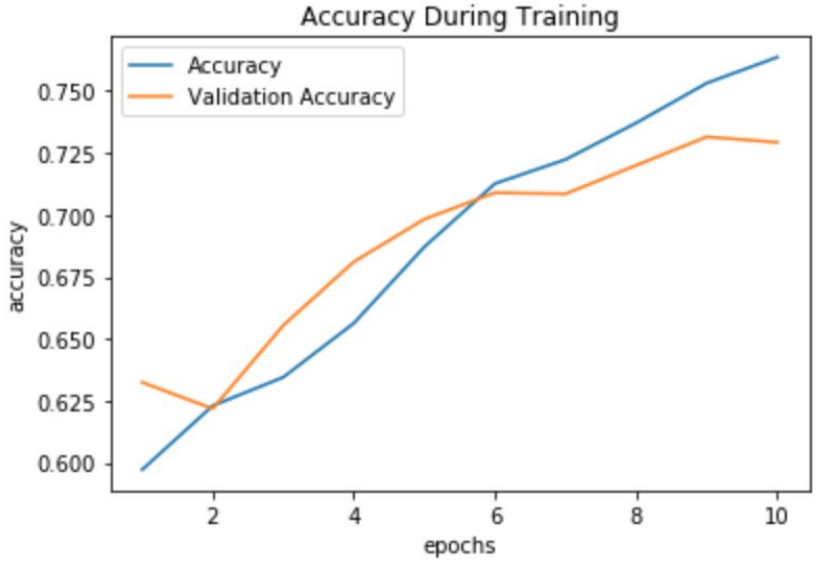


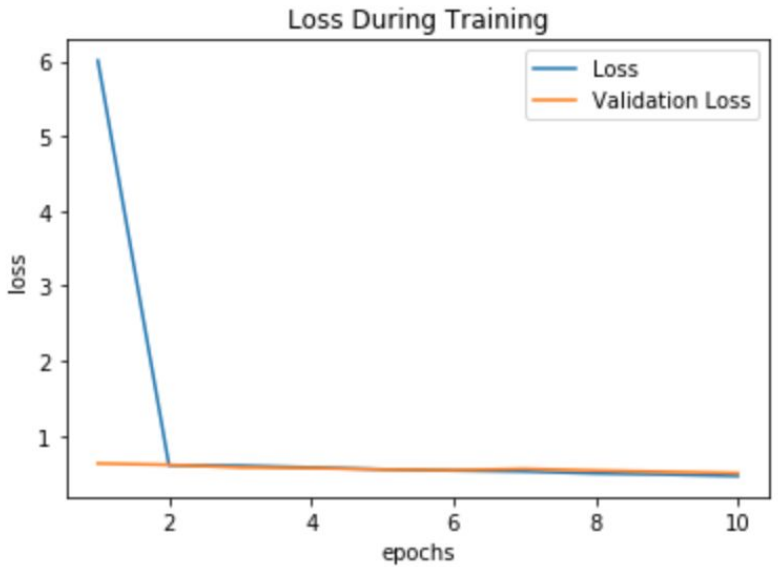
The Model

- The model was built using a 2D convolutional neural network from keras
- After experimenting to minimize overfitting, I chose this model structure:



Model Training Statistics

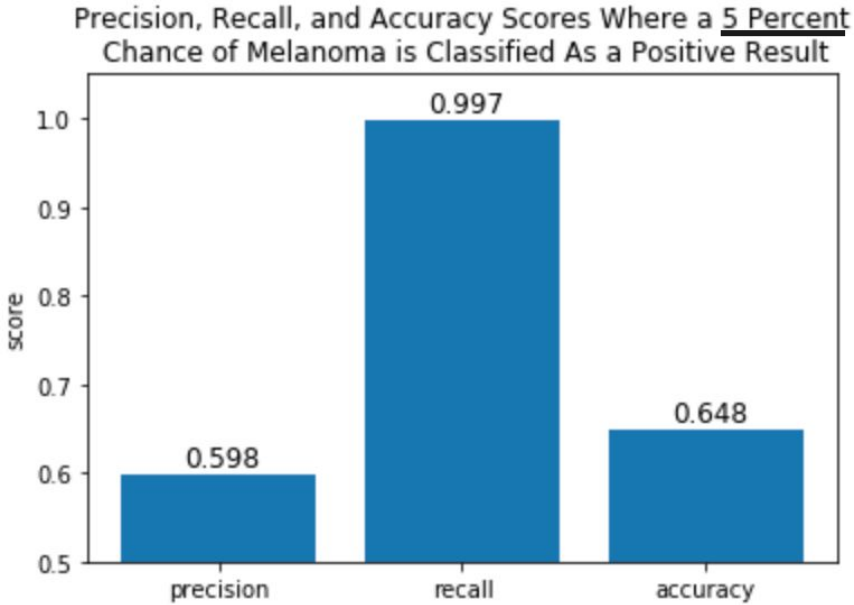


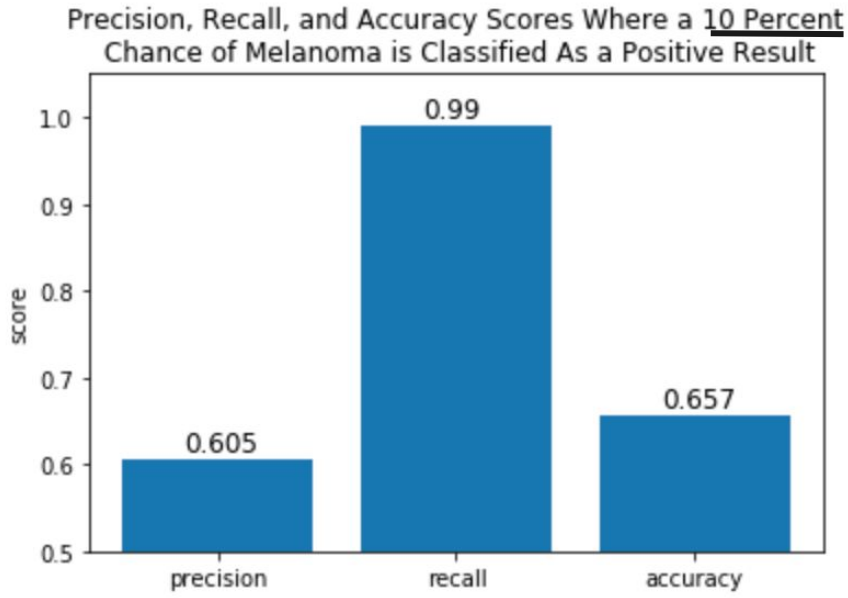


Precision, Recall, and Accuracy Scores

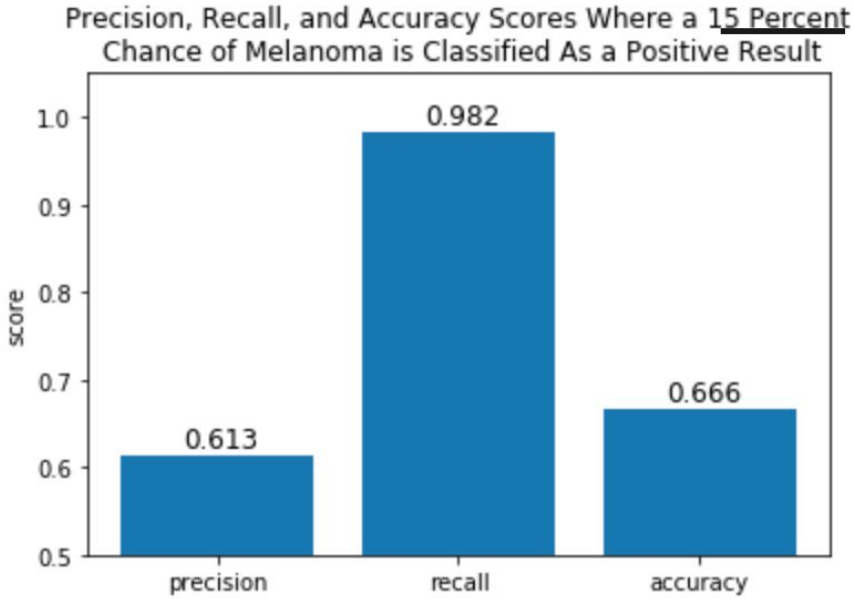
- A **precision** score gives you a number from 0 to 1 showing the chance of a predicted melanoma actually being a melanoma
- A **recall** score gives you a number from 0 to 1 showing how many actual melanomas were correctly classified
- An **accuracy** score gives you a number from 0 to 1 showing how many of all the images were correctly classified
- Since misclassifying melanomas can be fatal, while incorrectly outputting a melanoma for a benign mole simply leads to one extra doctor's appointment, we want to lean towards optimizing for recall
- Our model outputs the percent chance of a melanoma, so we can optimize based on what percent is classified as melanoma

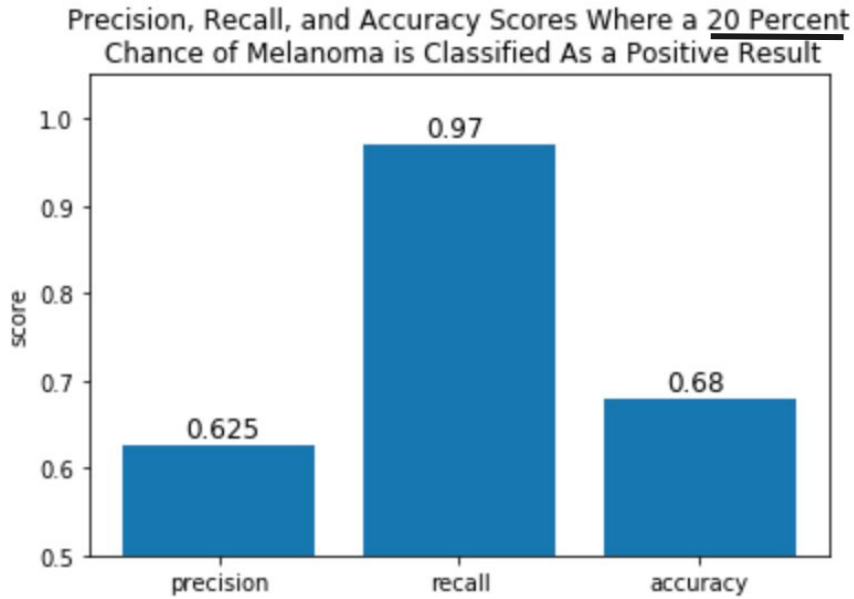
Melanoma CNN Scores (5%, 10%)



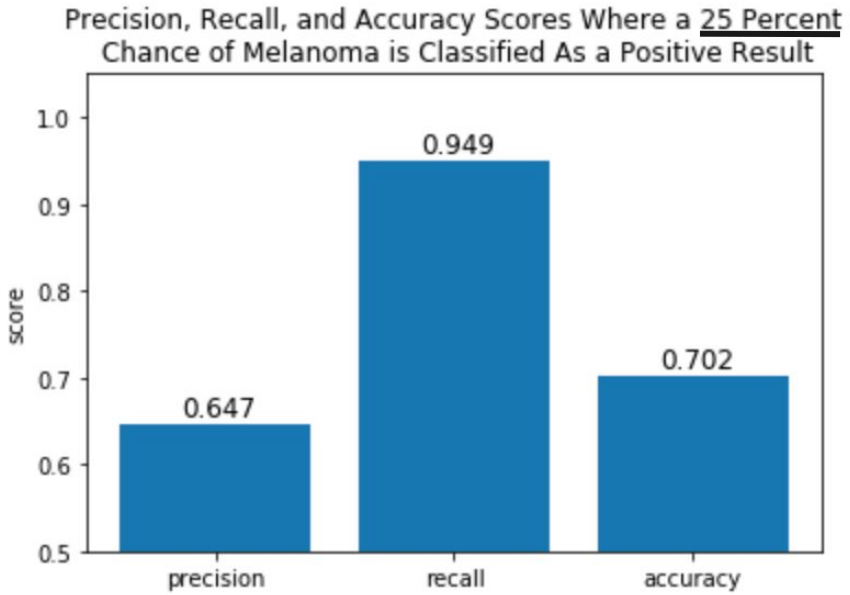


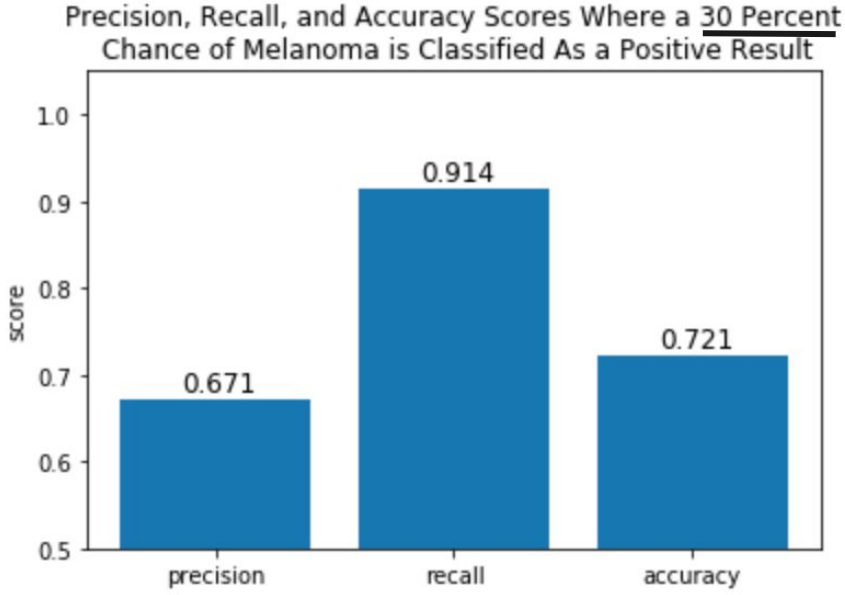
Melanoma CNN Scores (15%, 20%)





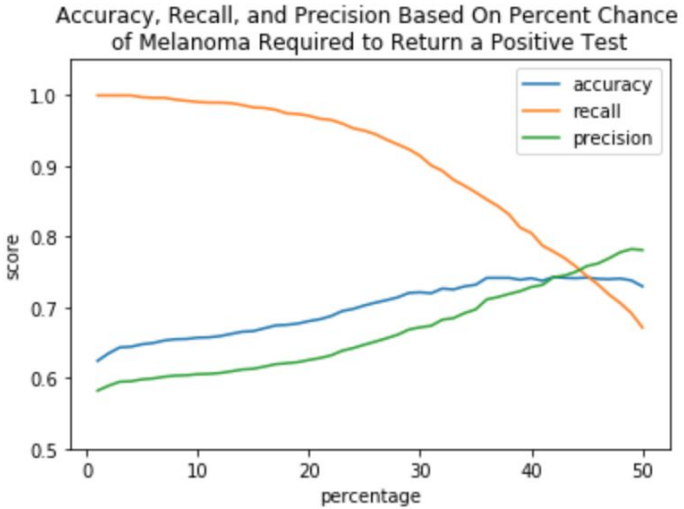
Melanoma CNN Scores (25%, 30%)





Line Graph Showing the Relationship of Precision, Recall, and Accuracy Scores

Unfortunately, there is no way to get a high precision and recall score at the same time. This graph shows the trade-offs between making sure the model catches as many melanomas as possible, which can be fatal when left untreated, without causing too many unnecessary biopsies, which take time and can be expensive depending on the patient's insurance.



Possible Next Steps to Improve Model

- Run more optimization tests to get the best possible model structure
- Find a larger dataset to train on
- Cross-test the model (on kaggle, it can be difficult to tell which dataset came from which original source, so cross-testing can be a little difficult)
- Test how the model performs on different image sizes
- See how the model performs on “amelanotic melanomas”, which don’t show the normal patterns of melanoma (doctors struggle with this too, so the accuracy of this isn’t as important for application, though it would be very useful if the model could be extremely accurate on such melanomas. Finding a dataset for this could be difficult.)
- Add the use of metadata such as age, sex, and location of skin lesion to improve the model
- Create a model which is effective with multiple images each taken a week apart, since melanomas evolve quickly (finding a dataset for this could be difficult)

Possible Application For Doctors

Application: A doctor can get an instant second opinion on possible melanomas they aren't confident about. Based on their confidence in their diagnosis, they can choose the percent chance which should be classified as melanoma based on the bar graphs and line graphs shown before. Alternatively, they can get the percent chance themselves, and use the graphs to decide whether or not to biopsy.

Possible Next Steps To Implement An App Which Works on Melanomas and Other Skin Lesions

- Train on other common skin cancers and benign images (unlike an application for doctors, which mainly struggle with melanomas vs. melanocytic nevi, an application for the general public would require the ability to work on many types of skin lesions. I decided not to implement this because of the discrepancy of the amount of data in different skin lesion categories, as well as the fact that the urgency of the removal of some skin lesions wasn't perfectly clear
- Cross-test on other datasets
- Application: create an app which tells you whether or not to visit a dermatologist, as well as the urgency of a visit based on an image and possible metadata submitted by a user. If the model is unsure of the urgency of the visit, have the user take another picture in a week. Keep doing this until a decision is reached, or until 3-4 weeks pass. If 3-4 weeks pass, visit a dermatologist.