Multimodal Approach for Early Detection of Melanoma

Ethan Fahimi (fahimi@mit.edu)
Samantha Tsang (wtstsang@mit.edu)

Fall 2023

Contents

1 Introduction

Melanoma, a malignant form of skin cancer, poses a significant global health challenge. Responsible for approximately 75% of skin cancer-related deaths, its early detection is crucial for effective treatment and improved survival rates. Traditional diagnostic methods, primarily reliant on the expertise of dermatologists, face challenges due to the subtle and often ambiguous nature of melanoma lesions. These limitations highlight the need for more advanced, accurate, and accessible diagnostic tools.

The primary objective of this project is to develop a novel, multi-modal approach to enhance the early detection of melanoma when compared to a model that is only trained on a single modality. By leveraging the synergy of machine learning techniques and the rich data available in modern medical imaging as well as detailed patient data, this project aims to assist dermatologists in making more accurate diagnoses. The focus is on utilizing a combination of image and tabular data to classify melanoma cases, thereby addressing the current gaps in early skin cancer detection methods and demonstrating the effectiveness of multi-modal models in making predictions due to the extra data that is considered.

The approach adopted in this project involves a multi-stage process integrating advanced image processing and data analysis techniques. The methodology is twofold; first, a Convolutional Neural Network (CNN) analyzes dermoscopic image data to identify visual patterns indicative of melanoma, then the embedding generated by this is then combined with patient-level contextual data, before the model undertakes the final classification. This multi-modal method is designed to harness the strengths of both visual and contextual data, offering a more comprehensive analysis than traditional single-modality approaches.

2 Data

The primary dataset for this project is sourced from the SIIM-ISIC Melanoma Classification Challenge hosted on Kaggle. This dataset is a comprehensive collection of high-quality dermoscopic images, jointly provided by the Society for Imaging Informatics in Medicine (SIIM) and the International Skin Imaging Collaboration (ISIC).

The dataset comprises a substantial number of dermatological images, each labeled for the presence or absence of melanoma. These images represent a diverse range of skin types, lesion types, and stages of melanoma, offering a realistic representation of clinical scenarios. Alongside the images, the dataset includes patient-level metadata, such as age, gender, and lesion location, which are supplementary for a more holistic analysis.

2.1 Exploratory Data Analysis

We can see from Figure 1 that the proportion of benign to malign cases is nearly identical for men and women. Figure 2 shows that the peak for benign cases is around 45, while the peak for malign is around

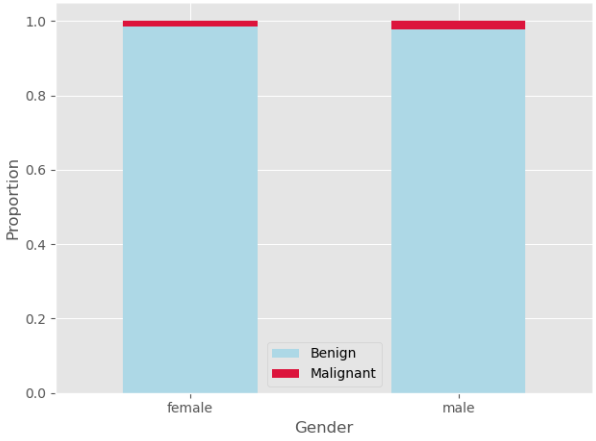

Figure 1: Proportion of melanoma cases in the dataset across different genders.

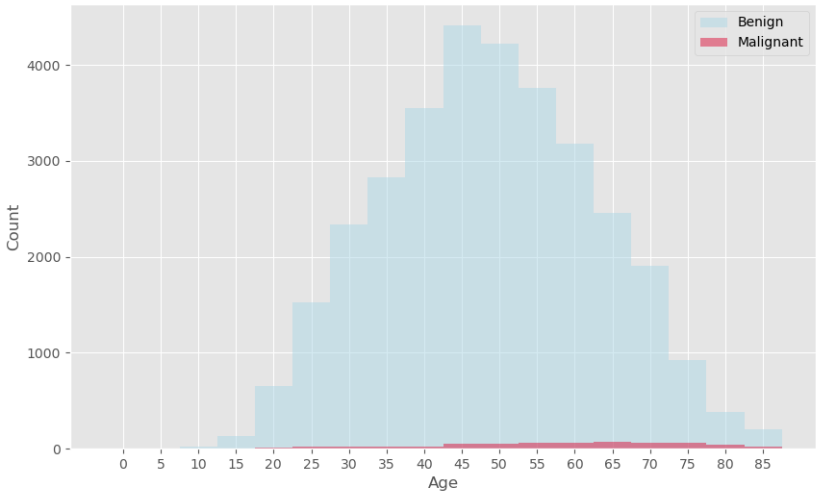

Figure 2: Overlaid bin histograms for melanoma cases in the dataset for different ages.

65, which logically makes sense that older patients are at a higher risk of having malignant melanoma. Finally, Figure 3 illustrates a large disparity between the count of benign and malign cases for different body parts, which is perhaps indicative of the places where melanoma is most common to occur. From the above figures, we can see that the dataset presents challenges such as a severe class imbalance, with fewer instances of malignant melanoma cases compared to benign ones. This imbalance reflects the real-world prevalence of the disease but poses challenges for machine learning models in terms of learning bias and overfitting.

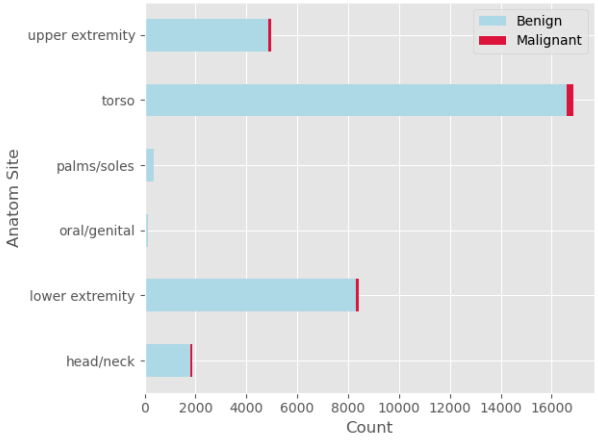

Figure 3: Distribution of melanoma cases in the dataset for different lesion locations.

3 Methods

3.1 Model Structure

The initial stage of our methodology involved comprehensive data pre-processing. Given the diverse nature of the dataset, it was essential to standardize the images and tabular data for effective analysis. This included resizing and normalizing the images to a 224 pixel by 224 pixel size to ensure uniformity in input data for the Convolutional Neural Network (CNN). For the tabular data, pre-processing involved encoding categorical variables and scaling numerical variables to avoid biases in the model due to variable scales.

To further refine our model's ability to generalize to new, unseen data, we implemented a robust data augmentation strategy for the image dataset. Data augmentation is a widely recognized technique in machine learning to increase the diversity of the training data without actually collecting new data. This technique involves applying various transformations to the original images, such as random rotations, flips (both horizontal and vertical), zooming, and shifts in brightness and contrast. These transformations simulate different angles, lighting conditions, and variations in lesions that a dermatologist might encounter in a real-world setting. By training the CNN with this augmented dataset, the model learns to recognize melanoma indicators across a broader range of image presentations, enhancing its robustness and predictive accuracy on diverse clinical images.

The core of our model is the VGG-16 architecture, a renowned deep CNN known for its effectiveness in image classification tasks. The original VGG-16 was developed by the Visual Graphics Group (VGG) at the University of Oxford and was trained on the ImageNet dataset, a large-scale dataset containing over 14 million images spanning across 1000 different classes.

We decided to implement transfer learning, leveraging the pre-trained VGG16 model to capitalize on its already learned features. A key modification is made by unfreezing the last few sets of convolutional

layers of the VGG-16. This allows these layers to be retrained on our specific dataset, making the model more attuned to the nuances of melanoma images. By retraining these layers, we expect the model to learn features that are particularly relevant to melanoma detection, enhancing its accuracy and reliability. We also modified the number of nodes in the last fully connected layers in the graph (from 4096, 4096, 1000 to 1024, 128) to prevent possible overfitting, since our dataset is comparably smaller in scale than ImageNet, and our objective is less complex. This results in embedding of length 128 representation of each image in the dataset.

Alongside image data, the model also incorporates tabular data from patients. This multi-modal approach, combining image and structured data, enriches the model's input, potentially leading to more accurate diagnoses. The integration of patient information such as age, sex, and anatomical site of the lesion can provide critical context that aids in the distinction between benign and malignant lesions. The tabular data is processing through two simple fully-connected layers with 8 nodes each. the resulting embedding to concatenated with the image embedding as the full representation of patient data.

Finally the concatenated embedding goes through three fully connected layers with a sigmoid activation at the last layer to predict the probability of Melanoma for each sample.

The architecture of the neural network can be seen in Figure 4.

3.2 Model Training

Due to the size of each image and the quantity of the data, it was necessary to create a data generator, which loads the data as it is called to train the model as opposed to loading the entire set of images simultaneously. Although there are multiple pre-written data generator classes available, the multi-input aspect of our project required us to create our own version of a data generator, which ensures the tabular data and the corresponding image for that patient (identified by image ID) is loaded in as one input set. In addition, the image augmentation is completed in this step (only for the training data), and the training data is shuffled at the end of each training epoch to ensure the model does not memorize the training data based on its sequence. The batch size has an impact on the training of the model, so it was set to 64 images per batch, in order to maintain a relatively large amount of data being processed at the same time; however a more powerful computer that can handle larger batch sizes to train this model could then result in an improvement in the models performance.

In the training configuration of the melanoma detection model, a weighted loss function is utilized to counteract class imbalance, enhancing sensitivity to less frequent classes. This approach is crucial for accurate diagnosis in medical image classification. The model employs the Adam optimizer, known for efficiently handling large datasets and adapting learning rates, an ideal choice given the typically noisy and sparse gradients in medical imaging. The model is then trained for 15 epochs, and was able to finish training within 15 hours on a GPU.

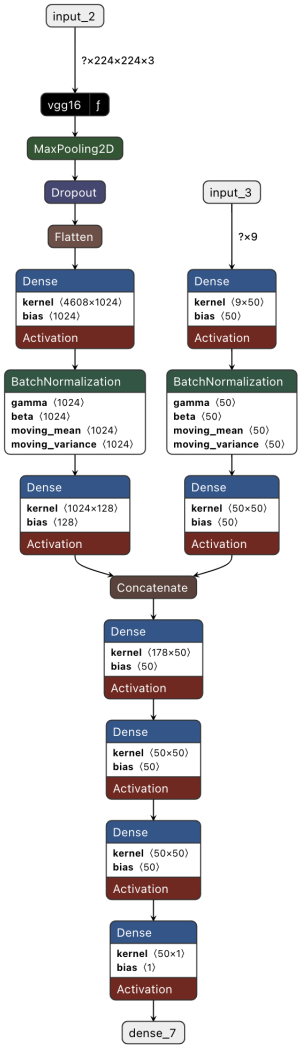

Figure 4: Neural Network Architecture

4 Results & Conclusions

To benchmark the results of the multi-modal approach, we also implemented a baseline model, which has the same structure as the image classification portion of our model, but removes the integration of the tabular data. In this baseline approach, the model simply makes a prediction after each image is transformed in the 128-dimension vector.

The table below compares the performance of the two approach in terms of area under curve (AUC) to capture each model's true performance with the data imbalance taken into account.

Model	Accuracy	AUC	F1 Score	Training Time (per epoch)
Baseline (CNN only)	0.61	0.79	0.74	1 hour
Multi Modal Approach	0.68	0.84	0.79	1 hour

Table 1: Comparison of results

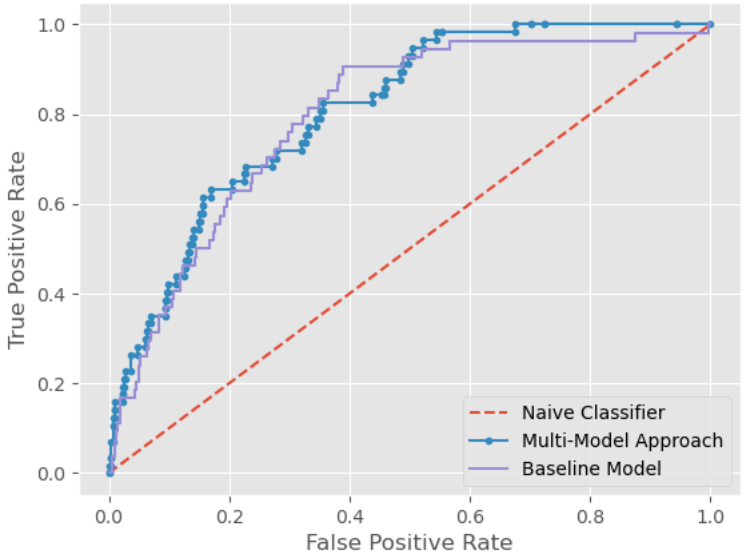

Figure 5: ROC curve comparison of baseline and multi-modal approach

The multi-modal approach was able to achieve an overall AUC of 0.84, 5% higher than the baseline model that just utilized the image data. From Figure 5 below, we also see that the multi-modal approach outperforms the baseline almost every threshold in terms of balancing true positive rate and false positive rate. These results are promising, considering that the tabular data only included simple information, such as gender, age, and which part of the body the image corresponded to, and was already able to create a significant edge against the model without this information.

We can see from the accuracy and the F1 score of our model that although our model is able to achieve a high AUC score, and significantly reduce the number of false positives compared to a naive model, we are not able to achieve the same level of performance in terms of accuracy and F1 score. More specifically, due to the balancing of class weights in the loss function, the high penalization of misclassifying a minority class sample over a majority class sample likely incentivized the model to more likely predict positive class over negative class.

To combat such issues, threshold tuning can be implemented. The current results shown reflects the "optimal threshold" in terms of true positive rate and false positive rate, ie. the threshold that gives the best G-means, $\sqrt{\text{true positive rate} * (1 - \text{false positive rate})}$, and gives a false positive rate of 0.33% (out of all samples without Melanoma, how many are we predicting incorrectly). This threshold can be altered depending on the trade off that between the false positive rate and false negative rate that healthcare providers feel will best satisfy their overall objective.

Our model's enhanced performance over the baseline model underscores the value of integrating additional modalities of data, affirming its efficacy in providing a more accurate and comprehensive approach to melanoma detection.

5 Next Steps

The promising results obtained from our multi-modal approach in melanoma detection underscore its potential in enhancing diagnostic accuracy. However, there remain several avenues for further research and refinement of our model to optimize its performance and clinical applicability.

5.1 Fine Tuning Class Weights

A key challenge in our current model is balancing sensitivity and specificity, especially given the class imbalance inherent in our dataset. Our model currently employs a "balanced" formula for the weighted loss function to counteract this imbalance, where each respective part of the loss function is multiplied by the inverse proportion of the number of samples in that class. Given the severe imbalance in the dataset, the weights has a ratio set to 1:250 (majority:minority). However, this choice of class weights requires further exploration. We propose a series of experiments with varying class weight ratios to identify the balance that maximizes the model's diagnostic accuracy while minimizing false positives and negatives. This step is critical to ensure that our model can be reliably used in clinical settings where the cost of misdiagnosis is high.

5.2 Incorporation of More Granular Tabular Data

The current model leverages basic patient data such as age, gender, and lesion location. To enhance the model's predictive power, we aim to incorporate more granular patient data. This could include metrics such as patient medical history, duration of lesion, and patient-reported symptoms. By integrating these additional dimensions of data, we can provide the model with a more holistic view of each case, potentially improving its ability to distinguish between benign and malignant lesions.

5.3 Threshold Optimization for Clinical Utility

The current model operates on an "optimal threshold" based on the trade-off between true positive rate and false positive rate. However, this threshold might not align with clinical priorities, where the cost of a false negative (missing a melanoma case) is typically higher than that of a false positive. We propose a detailed analysis to adjust this threshold, considering the clinical implications of false negatives and false positives. In actual implementation, this adjustment would be made in close collaboration with clinical partners to ensure that the threshold aligns with real-world clinical needs and priorities.

In conclusion, while our current model demonstrates significant promise in the early detection of melanoma, these next steps are crucial for refining the model's accuracy, usability, and acceptance in clinical settings. Through careful consideration of these aspects, we aim to develop a tool that not only enhances diagnostic capabilities but also integrates seamlessly into the clinical workflow, ultimately improving patient outcomes in melanoma care.

References

[1] Cassimiro, Gabriel"Transfer Learning with VGG16 and Keras" Towards Data Science. URL: https://towardsdatascience.com/transfer-learning-with-vgg16-and-keras-50ea161580b4

[2] "Everything you need to know about VGG16" Medium. URL: https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918

[3] Kang, Chanseok "Multiple Inputs in Keras" GitHub. URL: https://goodboychan.github.io/python/datacamp/tensorflow-keras/deep_learning/2020/07/28/02-Multiple-Inputs-in-keras.html

[4] "SIIM-ISIC Melanoma Classification" Kaggle. URL: https://www.kaggle.com/competitions/siim-isic-melanoma-classification/data?select=train.csv