CLASSIFICATION OF MELANOMA VIA MACHINE LEARNING TECHNIQUES

Nhi Hoang, Ngan Le, Tommy Pham, Jose Hernandez

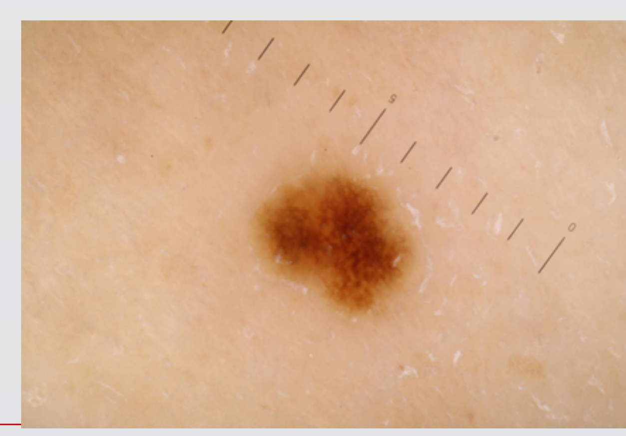
Computer Science Department (COSC 4337 – Data Science II), University of Houston

Abstract

Healthcare professionals have been relying on biopsy for diagnosis, but the application of Machine Learning approaches like Deep Learning has demonstrated its potential in solving image classification problems that are useful for diagnosis. With this paper, we will take everyone on a walk into the world of melanoma, and how Convolution Neural Network, Random Forest and Support Vector Machine perform when it comes to classifying whether the tumor is benign or malignant.

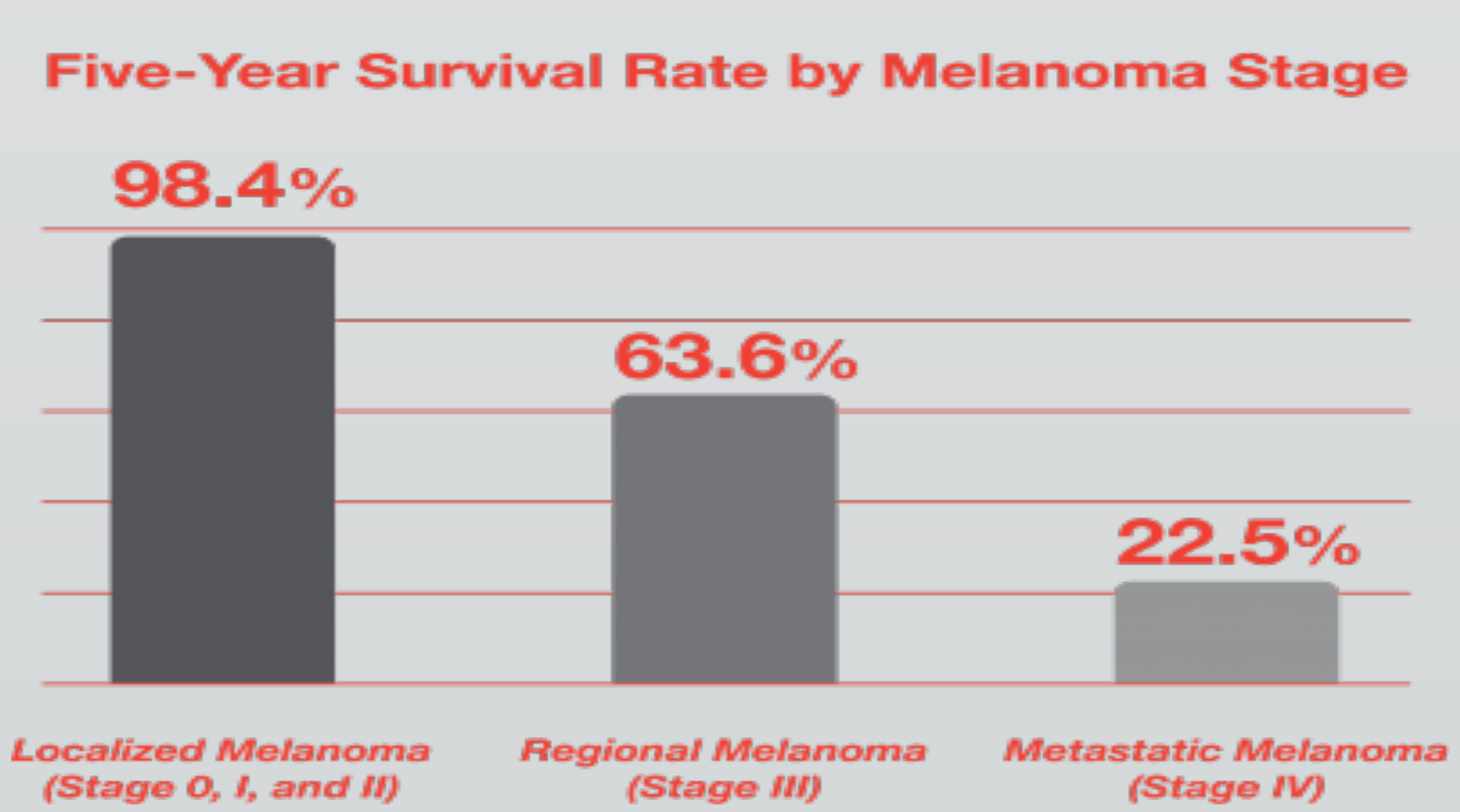
Benign

Malignant



Background

Since cancer is one of the top leading causes of death in the world, early detection and treatment is absolutely essential for maximizing one's chance at survival. Among the types of skin cancer, melanoma has the highest fatality rate due to its ability to quickly spread throughout the body. Therefore, the survival rate of melanoma increases significantly depending on how early the diagnosis and treatment is.



Methods

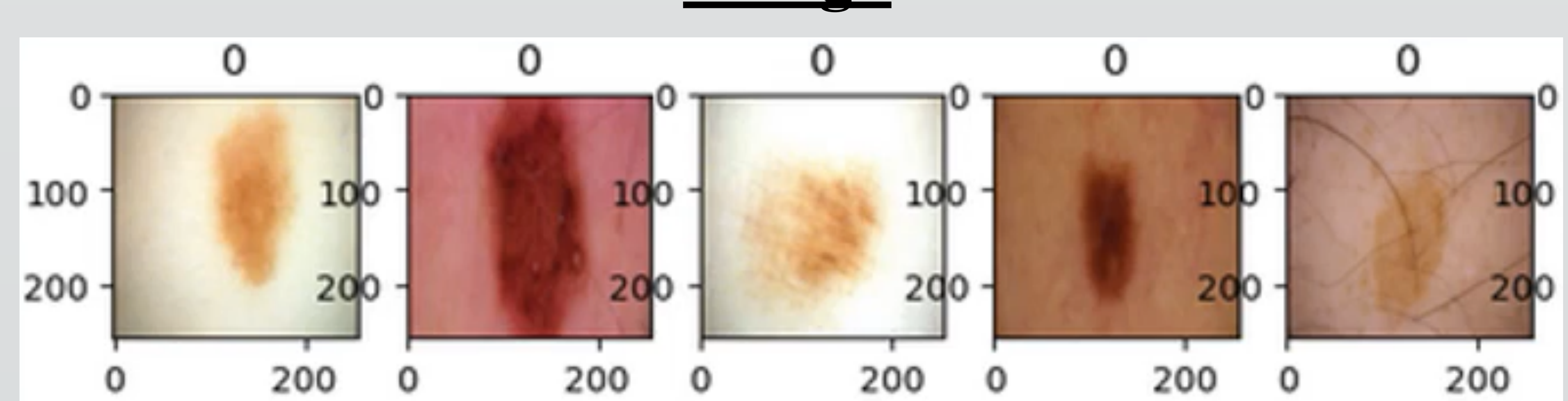
- **Random Forests:** Random Forest is a decision tree-based methodology. Its technique considers each instance separately, choosing the one with the most votes.

- **Support Vector Machines:** SVMs are generally useful to make binary classifications by drawing a decision boundary between the two classes in a multidimensional space. It will then generate the optimal hyperplane by maximizing the margin.

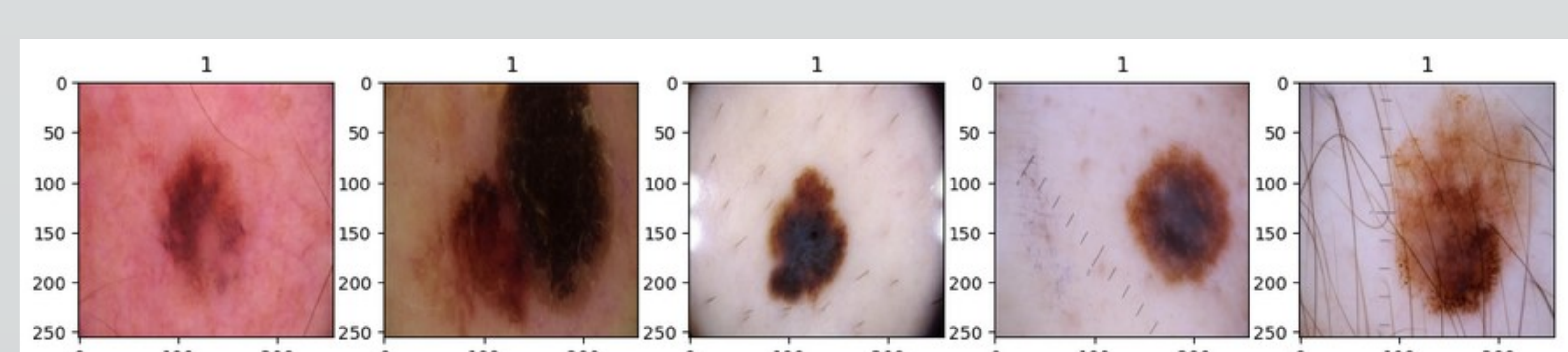
- **Convolution Neural Network**: The convolution operation will perform feature extraction on our images along with max pooling to withhold useful features. The feature map is flattened into a one-dimensional array and ran through a multilayer perceptron.

  + *Resampling*: dataset has less than 2% malignant cases, we will randomly pull the equivalent amount of benign cases for a more balanced dataset
  + *Regularization*: technique to battle overfitting (30% dropout)

Benign



Malignant



Results

We created models for our three methods using the complete dataset, which resulted in high accuracy rates for the training and validation sets. Since the training dataset is mostly comprised of benign class images and our test set consists of only benign class images, this may not create a truly useful model. Adding regularization to our model did not alter these results.

Resampling with a subset of our data to create a balanced class set resulted in a lower training and testing accuracy rates on our CNN model but this creates a more useful model.

Model	Accuracy Score
Random forest:	1
Random forest (with regularization):	0.999
Support Vector Machine:	0.987
CNN:	0.981
CNN (with regularization):	0.981
CNN (with resampling):	0.662
CNN (with resampling and regulariation):	0.599

Conclusion

In conclusion, we implemented a comparative analysis of CNN, RF, and SVM in order to determine the best model for the purpose of classification of melanoma. Based on the result, we believe that the CNN model is the most suitable solution to our research problem. The problem is the imbalance data set yielding high accuracy model, but it can lead to poor average precision during machine learning classification. For further analysis, there are many ways in which our project could have been improved. We can address the challenges with methods like data augmentation, designing new loss functions, transfer learning, lightweight CNN.

Future Direction

It can be expected that AI has the potential to play an active role in a paradigm shift in skin cancer diagnosis in the near future. The application of deep learning in melanoma detection can speed up diagnosis, buying time for medical treatment. Recently, with the success of deep learning in medical image analysis, several researchers have applied deep learning methods for skin cancer classification in an end-to-end manner and achieved satisfactory results.

Acknowledgments

American Cancer Society
https://www.cancer.org/

SIIM-ISIC Melanoma Classification
https://www.kaggle.com/competitions/siim-isic-melanoma-classification