Machine Learning Approach to Computer Assisted Diagnosis of Skin Diseases

Haller, Daniel (School: Marvin Ridge High School)
Kulshrestha, Ashwin (School: Marvin Ridge High School)
Patel, Jainith (School: Marvin Ridge High School)

Skin cancer is the most commonly diagnosed cancer in the U.S. Although it comprises less than 1% of new diagnoses, melanoma accounts for the vast majority of skin cancer deaths. The American Cancer Society estimates that more than 9,000 people will die of the disease this year. Early detection of skin cancer is critical, but it can be difficult for patients to distinguish between a benign skin lesion and a malignant one. The purpose of this research is to develop a computer-assisted diagnostic (CAD) tool which can analyze an image of a skin lesion and provide a diagnosis. We used a machine learning approach to develop this tool. We obtained our training data from the public International Skin Imaging Collaboration (ISIC) archive of skin lesions, which contains images of the lesions, de-identified patient information, and confirmed diagnoses. Using MATLAB, we performed feature extraction on the images and encoded them based on a visual dictionary of 200 features. We then used Matlab's classification learner to develop and test classification algorithms, using stratified 5-fold cross validation to detect overfitting. Using this approach, we developed three models for use in the diagnosis of skin lesions. The first model is able to distinguish between benign and malignant lesions at an accuracy rate of 74% (AUROC=.81). The two other models distinguish between specific forms of benign and malignant lesions, respectively. This set of tools is valuable to patients without immediate access to a dermatologist, as well as to physicians seeking diagnostic assistance.

Awards Won: