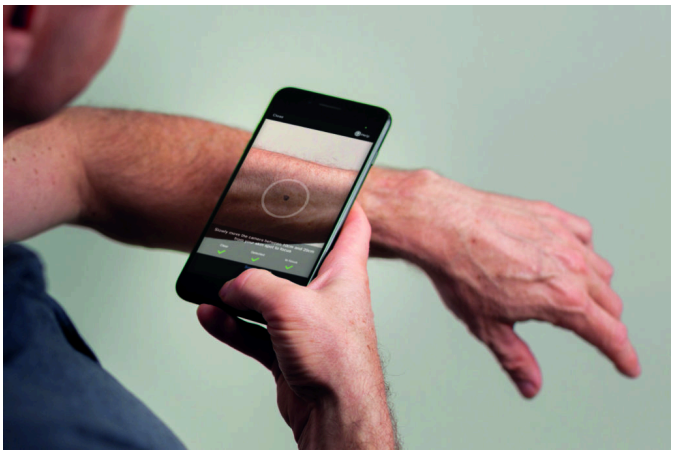
Classifying Skin Tumors from Smartphone Data

DS4002 Case Study by Grace Kang



In the United States, skin cancer is the most prevalent type of cancer, affecting one in five Americans by the age of 70. Although skin cancers are easily detectable since they develop through sudden changes in the skin, they are often detected later in the disease progression. In the past decade, the number of new melanoma cases diagnosed increased by 32%. However, when diagnosed and treated early, the 5-year survival rate for melanoma is 99%. As such, it is crucial to detect skin cancer as early as possible.

To help make diagnosis for skin cancer more accessible, the use of smartphone images and apps have grown on the rise. With the increasing popularity of machine learning, there are many existing models built for image classification. One such model is the EfficientNet model, a convolutional neural network developed by Google AI. At the intersection of machine learning skin cancer research, a common dataset used to train these models is HAM10000 – a dataset with over 10,000 dermatoscopic images. However, all the images in this dataset consist of exclusively clinical-grade images, which limits the applicability of models trained on it since real-world images typically are of lower resolution and variable quality.

As a newly budding data scientist, it's your task to explore whether you are able to successfully fine-tune an EfficientNet model so that it can classify different skin cancer and disease types from smartphone images to help lower the barrier for skin cancer diagnosis. To do this, first you will start out by exploring the original dataset to understand what the demographics of the data you're working with entails. Then, you will fine-tune two separate models: one that uses a binary classification, classifying based on whether an image is malignant or benign, and one that uses a six-classification model, classifying each skin cancer and disease type. Lastly, you will evaluate the models and present your results in a document.