Predicting Patients with Liver Disease

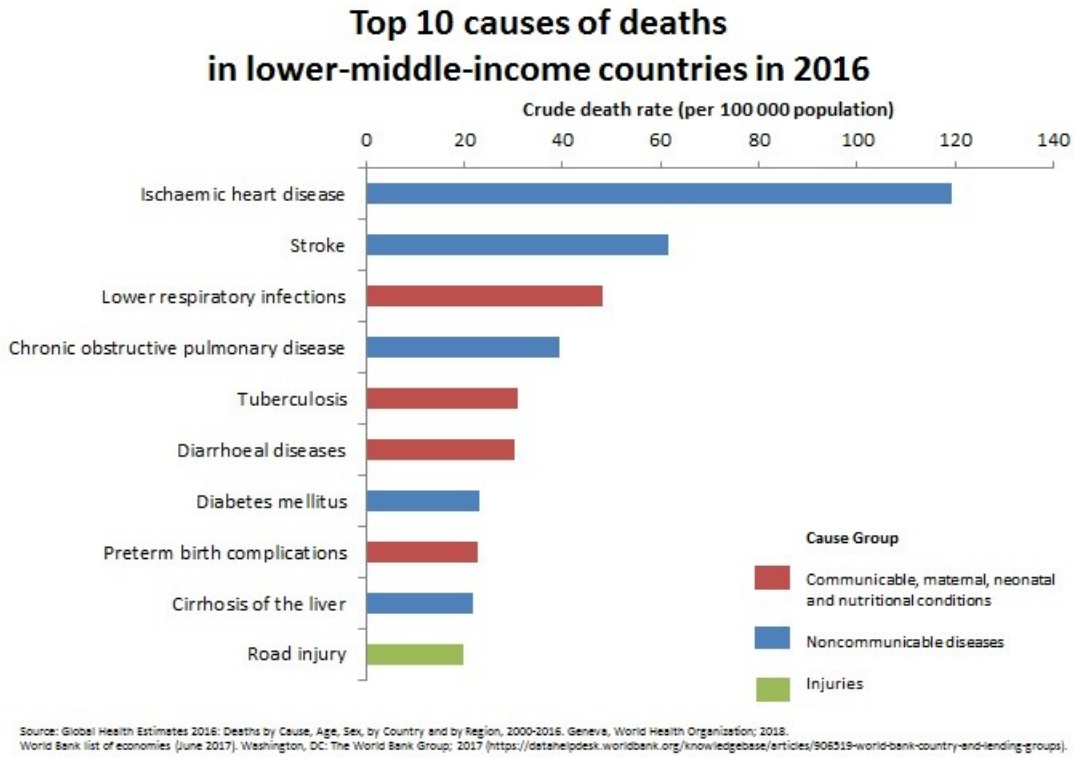
Capstone Project 1 Proposal, Springboard

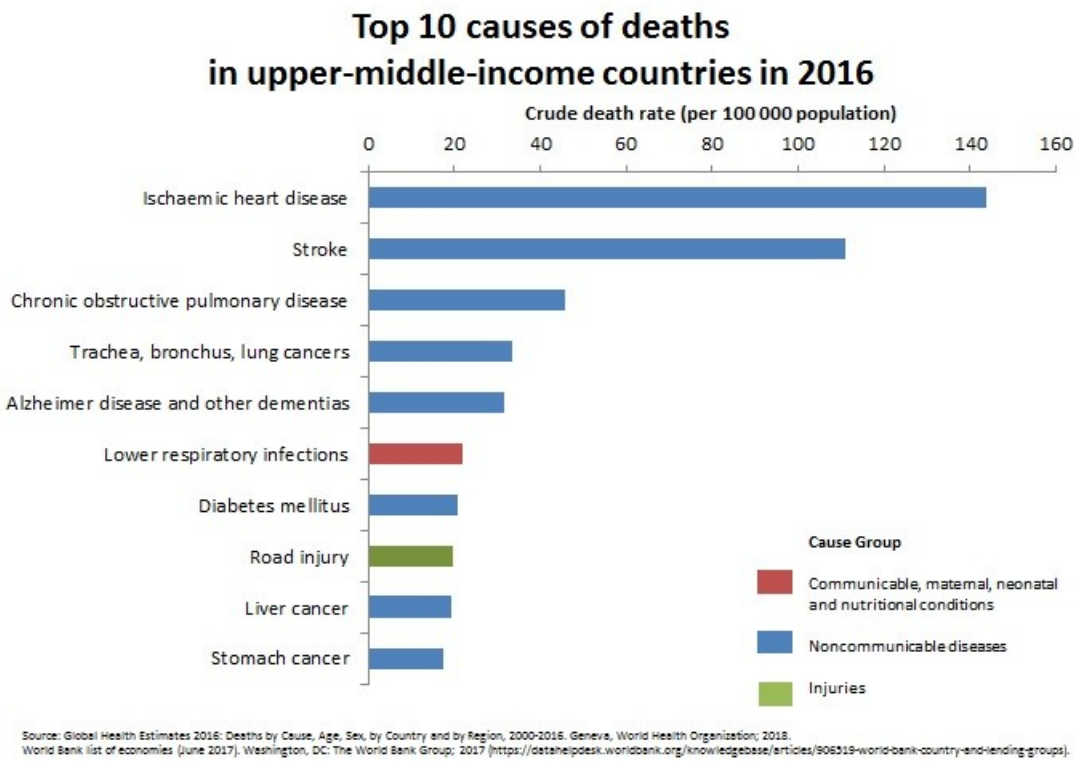
Predicting whether a patient has liver disease or not based on set of records can significantly reduce burden on doctors in an effort to correctly identify liver disease. By applying predictive machine learning algorithms against the patients dataset, we can solve the problem of identifying such patients. In this first capstone project, I am going to use machine learning models and a dataset that contains 416 liver patient records and 167 non liver patient records collected from North East of Andhra Pradesh, India.

The dataset for patient records will be taken from UC Irvine (UCI) ML repository: (*Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science)*. The current dataset contains 10 columns with patient data and a resulting 'outcome' column with the patient liver diagnosis (where 1 means patient has a liver disease, 2 means no disease). Thus, this is a binary classification problem. The techniques used in the project can be applied to other datasets with liver patient records.

It is important to note that the dataset is relatively clean and easy to use. However, the dataset will still require some data cleansing, data wrangling and possibly feature engineering. The exploratory data analyses and statistical learning will be applied to find more insights from data. Based on the above mentioned details regarding the data, several relevant machine learning algorithms (such as logistic regression, random forest and others) will be applied for predicting a liver patient diagnosis.

Overall, liver disease caused death rate is among top 10 death causes in the lower-middle income and upper-middle income countries in the world. Thus, having an accurate model for predicting patient live disease on early stages based on their records can significantly improve the diagnosis and help in early disease preventive cares. Below are the charts provided by World Health Organization (WHO):





(Source: Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, World Health Organization; 2018.)