**Proposal 3 : Smoker Status analysis using Bio-Signals**

**Problem:**

Smoking detection in patients

**Dataset:**

Kaggle dataset (<https://www.kaggle.com/competitions/ml-olympiad-smoking/data>)

The entire dataset consists of three csv files, each of which contains over 23 bio-signal features of patients like ages, heights, weights, fasting blood sugar… and the target prediction, smoking status.

**Proposed Solution and Real world Application :**

Smoking has been extensively demonstrated to have detrimental effects on health across various dimensions, impacting nearly every organ, inducing numerous diseases, and diminishing the life expectancy of smokers overall. As of 2018, it stands as the primary contributor to avoidable illness and death worldwide, persistently undermining global health.

While evidence-based treatments for smoking cessation have been advocated and promoted, fewer than one-third of participants managed to attain abstinence. Many physicians perceive counseling for smoking cessation as ineffective and time-consuming, leading them to omit it from their daily practice routine.

Thus, there comes an urgent need for physicians to have a method to identify smokers who had a better chance of quitting. Our proposed solution is to mathematically analyze all the factors in datasets and find top-influential bio-signals to identify smoker and non-smoker .

The real world application of this solution is that it can provide a method straightforward enough for the physicians to interpret and apply. In addition, in the future, it can be a basis to develop machine learning algorithms to accurately identify smokers who are likely to quit.

**Project steps**

| Step | Estimated completion time | Person(s) in charge (among the group 19) |
| --- | --- | --- |
| 1. Extract, clean and preprocessing data | One week | (Pin-Ying, Zhiyang) |
| 2. Visualize, analyze, and investigate hidden pattern in the data | Two weeks | (Shah, Zhiyang) |
| 3. Interpret the outcomes and find the top-influential bio-factors | One weeks | (Shah, Zoom) |