Team Members: Ahmad Abdalmageed, Mohamed Almotasem

# Final Project Phase 1 Proposal

## Problem

The highest prevalence of HCV infection is present in Egypt, with 92.5% of patients infected with genotype 4, 3.6% patients with genotype 1, 3.2% patients with multiple genotypes, and < 1% patients with other genotypes. Grading the Virus has an impact on the treatment given to the patient this is where Machine Learning Steps in.

## Dataset

Recorded by Ain Shams University, this Data provides records of patients who underwent treatment dosages for HCV about 18 months. Data consists of 29 Attributes and 1385 instances. The Target Variable is a Discretized grade of the patient's infection degree.

## Proposed Methodology

1. **Data Preprocessing**: Exploring the data for potential problems, cleaning and preparing
2. **Exploratory Data Analysis**: Exploring the Data for Insightful Characteristics
3. **Prototyping**: Instantiate a basic Model from potential ML Algorithms and find best Candidates
4. **Feature Engineering and Fine Tunning**: Given the Data Exploration find the best correlating features and tune the model to it's best performance.