# Proposal for project in deep learning course

Netanel Indik - 311300784

Meir Nizri - 312237563

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.

During the entire course of the pandemic, one of the main problems that healthcare providers have faced is the shortage of medical resources and a proper plan to efficiently distribute them. They have been in the dark, failing to understand how much resource they can assign for their patients when even in the very next week COVID-19 curve could sway very unpredictably. In these tough times, being able to predict what kind of resource an individual might require at the time of being tested positive or even before that will be of great help to the authorities as they would be able to procure and arrange for the resources necessary to save the life of that patient.

**Our goal in this project is to build a model that, given a patient current symptom status and medical history, will predict which risk group he or she is in. Using this model healthcare providers will be able to prioritize patients effectively and thus reduce mortality rates.**

Note: We hope that through the analysis of the model we can also understand how each feature, which will be detailed below, affects the chances of mortality, and consequently understand who the populations at increased risk are.

**Data Set**

procuring patient data of COVID-19 patients containing patient-specific information regarding patient history and habits is not an easy task. This is mainly due to the regulatory security laws such as HIPAA and GDPR which makes it almost impossible for anyone to get hands-on PHI data. The data set obtained is from [Kaggle](https://www.kaggle.com/tanmoyx/covid19-patient-precondition-dataset) and was provided by the Mexican government ([link](https://www.gob.mx/salud/documentos/datos-abiertos-152127)). This data set contains a huge number of anonymized patient-related information.

We are aware that our analysis will not necessarily apply for other countries as well, as coronavirus behaves differently in different areas of the world. However, it will give a general picture that serves the purpose of the model.

Features:

1. id: The identification number of the patient.
2. sex: female or male.
3. age: of the patient.
4. patient type: hospitalized or not hospitalized.
5. entry date: The date that the patient went to the hospital.
6. date symptoms: The date that the patient started to show symptoms.
7. intubed: Intubation is a procedure that’s used when you can’t breathe on your own. Your doctor puts a tube down your throat and into your windpipe to make it easier to get air into and out of your lungs. A machine called a ventilator pumps in air with extra oxygen. Then it helps you breathe out air that’s full of carbon dioxide (CO2).
8. pneumonia: Indicates whether the patient already have air sacs inflammation or not.
9. pregnancy: Indicates whether the patient is pregnant or not.
10. diabetes: Indicates whether the patient has diabetes or not.
11. copd: Indicates whether the patient has Chronic obstructive pulmonary disease or not.
12. asthma: Indicates whether the patient has asthma or not.
13. inmsupr: Indicates whether the patient is immunosuppressed or not.
14. hypertension: Indicates whether the patient has hypertension or not.
15. cardiovascular: Indicates whether the patient has heart or blood vessels related disease.
16. renal chronic: Indicates whether the patient has chronic renal disease or not.
17. other disease: Indicates whether the patient has other disease or not.
18. obesity: Indicates whether the patient is obese or not.
19. tobacco: Indicates whether the patient is a tobacco user.
20. contact other covid: Indicates whether the patient has contacted another COVID-19 patient.
21. icu: Indicates whether the patient had been admitted to an Intensive Care Unit.
22. death: indicates whether the patient died or recovered. **This is the label we would like to predict.**