Mechanistic Data Science for Engineering (ME 395)

Homework # 2

Student name: Asma Ul Hosna Meem

Student ID: 3335096

Problem:

Propose a real-life problem that can be solved by the six modules of mechanistic data science (summary and six steps in one page).

Proposal: Heart failure prediction analysis

Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels. They are the leading cause of death with an estimated 17.9 million deaths globally in 2019 [1]. Heart failure is the most common among these cardiovascular diseases. Heart failure occurs when the heart muscle fails to pump enough blood that is needed by the body organs to function properly. There are many physical and behavioral factors such as, unhealthy diet and lifestyle (e.g., excessive smoking or drinking), high blood pressure, obesity etc. that can increase the risk of heart failure in an individual. Therefore, determining these risk factors can be crucial in early detection and prevention of heart failures in potential heart disease patients.

Determining which are the most critical and fatal risk factors that can cause death from heart failure and thus developing a heart failure prediction model can be accomplished through the six modules of mechanistic data science [3].

- For this project, first of all, we need to collect data on patients having heart failure. Such an open source and published dataset has been chosen for data collection where twelve (12) different features about the patients have been tabulated [2].
- This dataset includes age, gender, diabetes history, high blood pressure values and some more information about the patients. In the second step, we need to extract the most important mechanistic features from the raw dataset.
- In the third step, dimensional reduction needs to be done to reduce the amount of data for further analysis.
- After obtaining the reduced dataset, we may proceed to develop a reduced order surrogate model. Since this is a patient-information based problem, we might categorize this as a purely data driven problem with no available mathematical/scientific formulations. So, we may skip this step here.
- After that, regression analysis will be done to determine the relationship between patient survival and the important risk factors.
- Finally, with the help of this regression analysis we would be able to generate a regression model which can be further used, given the values of the important features as inputs to the

regression model, for future prediction of probability of patient survival in heart failure incidents.

References:

1. [Cardiovascular diseases (CVDs) (who.int)](#)
2. Chicco, D., Jurman, G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med Inform Decis Mak* **20,** 16 (2020). https://doi.org/10.1186/s12911-020-1023-5
3. Liu, W. K., Gan, Z., & Fleming, M. (2020). Mechanistic Data Science for Engineering Draft: An Introduction.