Caroline Cocca

April 3, 2022

Creating AI-Enabled Systems

System Project Proposal

In my system project, I am focusing on the domain of applied machine learning in medicine. Although many models are already deployed in real-world medical settings [1], the application of machine learning to telehealth and remote patient monitoring has not yet been thoroughly explored and implemented. One issue that I believe could be addressed and improved with machine learning and telehealth is the mitigation of preventable health conditions.

Many across the country are unable to afford or access healthcare, whether due to a lack of insurance or sheer distance from any practitioners, which put them at a higher risk for having preventable conditions that were not detected or treated in time. These conditions being missed leads to many avoidable deaths– in specific, up to 40% of deaths from each of the five leading causes in the U.S. could have been prevented [2]. Remote monitoring could substantially improve the rates of preventable conditions detected and mitigated in such populations, and consequently save many lives and ease a huge burden on the healthcare system.

Specifically, I will be exploring the remote monitoring of vitals and glucose levels as a predictive indicator of future coronary heart disease (CHD). The goal of this system is to reduce time needed by doctors to review patient risk factors, as well as time needed for patients to make physical doctor visits. This time reduction for physicians would also result in more efficient and less costly monitoring, making healthcare more affordable and accessible to underserved populations.

The dataset I will utilize is an ongoing cardiovascular study based in Framingham, Massachusetts, known as the Framingham Heart Study-Cohort (FHS-Cohort.) This dataset contains features collected from participants including demographic (age, sex), behavioral (current smoker status, cigarettes per day), and medical (diabetic status, BMI, systolic and diastolic blood pressure, heart rate, etc.) risk factors. The target variable of the dataset is a boolean variable indicating whether the patient has a 10-year risk of CHD.

My system project will consist of a simple Python script outputting simulated vitals at a regular interval, which a binary classification model trained on the above dataset will listen to and process on submission. The predictions generated from this model will then be streamed to an output file. All vitals information and model outputs will be saved to a relational database. The streaming of vitals and model output will be handled with Apache Kafka. This system serves as a proof of concept for utilizing machine learning in the reception and processing of patient data for the purpose of flagging potential cases of concern for further physician review. This proof of concept will not include a user interface or mobile application for submitting/reviewing data, and will also not include the testing of bluetooth transmission from vitals measuring devices to a listening model. However, such system components would need to be explored and implemented in future work for a fully functioning system that could be easily used by patients and doctors.

References

[1] M. May, “Eight ways machine learning is assisting medicine,” Nature, 2021, https://www.nature.com/articles/s41591-020-01197-2

[2] U.S. Department of Health and Human Services, “Up to 40 percent of annual deaths from each of five leading US causes are preventable,” CDC, 2014, https://www.cdc.gov/media/releases/2014/p0501-preventable-deaths.html