Project Proposal

CPSC 8810

John Pascoe

My project will focus on predicting asthma exacerbations in patients using machine learning models. Asthma is a chronic respiratory condition characterized by airway inflammation and bronchoconstriction, leading to symptoms like wheezing, shortness of breath, chest tightness, and coughing. Predicting exacerbations can significantly improve the quality of life for asthma patients by enabling preemptive management strategies, reducing emergency hospital visits, and personalizing treatment plans. Asthma affects millions of people worldwide and is a significant cause of morbidity and healthcare utilization. Early prediction of asthma exacerbations can lead to more effective control of the condition, minimizing severe outcomes, and reducing healthcare costs. Machine learning offers the potential to identify complex patterns in data that traditional statistical methods may not capture. By leveraging EHR data, machine learning models can predict the likelihood of exacerbations with high accuracy.

The primary machine learning problem type for this project will be classification. My goal is to classify patients into groups based on their risk of experiencing asthma exacerbation within a specified future time frame. Time series analysis may also be employed to analyze temporal patterns in symptom reporting or medication use that precede exacerbations.

The dataset for this project will be derived from electronic health records of asthma patients and will include:

* Demographics: Age, gender, ethnicity
* Clinical history: Previous attacks, hospitalizations, mediction use, comorbidities
* Environmental factors: Air quality, pollen counts, weather conditions
* Patient-reported outcomes: Symptom scores, peak expiratory flow rates

Table 1: Data Characteristics

| Variable | Description | Type | Range/Value |
| --- | --- | --- | --- |
| Age | Patient’s age at the time of data entry | Continuous | 0-100 years |
| Gender | Patient’s gender | Categorical | Male, Female, other |
| Ethnicity | Patient’s self-identified ethnicity | Categorical | Various |
| Previous Exacerbations | Number of exacerbations in the past year | Continuous | 0-10+ |
| Medication Use | Types of asthma medication used | Categorical | Controller, Rescue, None, Both, other |
| Comorbidities | Presence of other chronic conditions | Categorical | Yes, No |
| Air Quality Index | Environment air quality | Continuous | 0-500+ |
| Pollen Count | Daily pollen count in patient’s area | Continuous | 0-12,000 grains per cubic meter |
| Symptom Score | Daily patient-reported symptom score | Continuous | 0-10 |

References:

Public EHR data will be utilized along with the following research papers as of this point:

Nissen, F., Quint, J. K., Wilkinson, S., Mullerova, H., Smeeth, L., & Douglas, I. J. (2017). Validation of asthma recording in electronic health records: a systematic review. *Clinical epidemiology*, *9*, 643–656. <https://doi.org/10.2147/CLEP.S143718>

Landeo-Gutierrez, J., Defante, A., Cernelc-Kohan, M., Akong, K., Rao, A., Lesser, D., Duong, T. E., Cheng, E. R. Y., Ryu, J., & Tantisira, K. (2023). Leveraging Electronic Health Records for Guideline-Based Asthma Documentation. *The journal of allergy and clinical immunology. In practice*, *11*(3), 855–862.e4. <https://doi.org/10.1016/j.jaip.2022.11.032>

Cobian, A., Abbott, M., Sood, A., Sverchkov, Y., Hanrahan, L., Guilbert, T., & Craven, M. (2020). Modeling Asthma Exacerbations from Electronic Health Records. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, *2020*, 98–107.