MACHINE LEARNING FOR EARLY DETECTION, MANAGEMENT AND PROGNOSIS OF HYPERTENSION

Author: Suparno Datta

Abstract: Hypertension, or high blood pressure, is a major risk factor for cardiovascular disease, and early detection and management of hypertension is crucial for reducing the risk of complications. Machine and deep learning techniques have the potential to improve the prediction and management of hypertension by analyzing large amounts of patient data and identifying patterns that may not be apparent to human experts. In addition, these techniques can also be used to develop personalized treatment plans for hypertension, by analyzing patient data and identifying the most effective interventions for each individual. Overall, the use of machine and deep learning in hypertension prediction and management has the potential to improve the timeliness of hypertension diagnosis, facilitate regular monitoring of blood pressure (BP), predict future complications and optimize treatment strategies for patients. However, artificial intelligence in hypertension and cardiovascular medicine is still in its nascent stages and realizing its potential requires numerous challenges to be overcome. While there has been a surge of machine learning applications on electronic health records (EHRs) during the last decade, extracting specific disease cohorts and relevant features that can feed into the models still remain a cumbersome process as it requires deep-ranging knowledge of the underlying EHR data model, its entities, and their relationships. As a result, retrospective and fully data-driven studies for predicting the onset and complications of hypertension in large-scale real-world EHR databases are still scarce. Evidence suggests that apart from risk prediction models, continuous BP monitoring is integral to the detection, control, and treatment of hypertension. But the ability to measure BP in this way is currently limited. Fortunately, machine learning models applied to photoplethysmography (PPG) signals have shown encouraging results for non-invasive, continuous measurement of BP. Nevertheless, further evaluation is crucial to validate this approach, especially on datasets that are collected in a non-controlled environment, a setting in which hardly any previous work has been carried out. In order to address these challenges, in this thesis, I have made the following contributions: 1) designed and developed a Python framework tailored to specific needs of clinical modeling domain that enable flexible generation of modeling-ready cohorts, 2) leveraging this framework, I developed and evaluated a set of models based on machine learning (ML) and deep learning (DL) to predict the onset and complications of hypertension leveraging large-scale longitudinal EHR data, and 3) evaluated to what extent the models proposed in recent literature can estimate BP based on PPG data in a non-controlled environment, i.e., during day-to-day activities. In conclusion, I argue that appropriate tooling for cohort extraction, models that leverage large-scale longitudinal EHR data, and evaluation of these models on diverse datasets are key concerns if artificial intelligence is to achieve its full potential in diagnosis and management of hypertension.