SAIS 2020 workshop

Medical Diagnosis with Machine Learning

Saga Helgadottir[1], Stefano Romeo[2] and Giovanni Volpe[1]

[1]Department of Physics, University of Gothenburg, SE-41296 Gothenburg, Sweden

[2]The Wallenberg Laboratory for Cardiovascular and Metabolic Research, Sahlgrenska University Hospital, SE-413 45 Gothenburg, Sweden

saga.helgadottir@physics.gu.se

In healthcare the first and most important step to treat a patient is to get a correct diagnosis of the disease or condition the patient is suffering from. Medical doctors use many tools and tests to assist them in the diagnostic process. The methods used can however often be time consuming, expensive, limited in availability and sometimes even not so reliable. This is a general problem and we believe that in many cases the diagnostic process could be improved using novel methods based on machine learning. In particular we have implemented a virtual diagnosis of a genetic disease called Familial hypercholesterolemia (FH).

FH is the most common genetic disorder of lipid metabolism, with a prevalence of 1/250. FH is characterized by lifetime elevated low-density lipoprotein cholesterol (LDL-C) levels resulting in premature cardiovascular disease. The gold standard for the diagnosis of FH is genetic diagnosis with identification of pathogenic variants in heterozygosity in genes involved in LDL-C catabolism. Although genetic diagnosis has become very effective its implementation requires specialized expertise and is thus only available in selected university hospitals. Consequently, clinical scores are often employed as less expensive, but also less accurate, alternatives to genetic diagnosis. One of the most used is the Dutch Lipid Score, which is calculated by adding the scores given to a patient's family history, the clinical history of the patient, their untreated LDL-C levels, and physical examination. The final score is used to categorize the patients into four classes given the probability that a patient has the disease, from unlikely to definite.

We have recently published a study where we assessed the use of machine-learning algorithms to implement a virtual genetic test with improved performance as compared to the commonly employed clinical diagnosis by the Dutch Lipid Score. We have used two machine-learning algorithms to identify patients with a positive genetic test for FH mutation, namely, a neural network and a classification tree. The reason for using these two techniques is that while neural networks generally perform better, they are a so-called 'black-box' method and thus their diagnosis cannot be easily explained or interpreted. In contrast, classification trees are simpler algorithms that do not perform as well but are transparent and easily interpreted so that they can provide simple diagnostic rules. Both algorithms were

trained in the same way using LDL-C/age, triglyceride (TG)/LDL-C, and high-density lipoprotein cholesterol (HDL-C) from patients from Gothenburg, Sweden as input variables. We tested the algorithms internally on another set of patients from Gothenburg as well as externally on patients in Milan, Italy. By evaluating their area under the receiver operating characteristic (AUROC) curves, we found that both machine-learning algorithms performed better (AUROC 0.79 (CT) and 0.83 (NN) on the Gothenburg internal test, and 0.70 (CT) and 0.76 (NN) on the Milan external test) than the clinical Dutch Lipid Score (AUROC 0.68 and 0.64 on the Gothenburg and Milan tests, respectively) in predicting carriers of FH-causative mutations.

Mentor sessions

I'm in my fourth year of PhD studies at the Department of Physics at the University of Gothenburg, meaning that I have roughly one year left to figure out in what direction I want to take my career. In the first week of my PhD my supervisor told me that there was no way of knowing how my studies would develop. He was right indeed, I started my PhD as an experimental physicist, dealing with optics and culturing bacteria, but now I have abandoned the laboratory to focus on applications of machine learning, especially in the biomedical field. I have been very lucky to have had the chance to collaborate with researches abroad and in the medical field but as my studies are coming to an end, I wish to have the opportunity to get into contact with the AI community in Sweden.