Abstract

Cardiovascular diseases continue to be the leading cause of death worldwide. The term *personalized medicine* was first used in the late 90s after decades of research had established its scientific basis. The Council of Advisors on Science and Technology of the USA provided the following definition: "Personalized medicine is the practice of tailoring medical care to the unique needs of individual patients". It refers to the ability to divide people into groups based on differences in response to a given disease or reaction to a given treatment, rather than the actual production of drugs or medical devices customized for each patient.

Using personalized medicine, a physician may be able to treat a patient more effectively, safely, and for a shorter period, while also saving money. In recent years, the field of personalized cardiovascular medicine has seen a significant increase in scientific publications. Soon, personalized medicine will have an impact on the entire healthcare system.

In this context, *cardiovascular risk assessment* is therefore essential for many of the current treatment protocols. Risk estimates are also used to project the magnitude of future cardiovascular disease mortality and morbidity at the population level and in particular subgroups.

Novel techniques are required to process and comprehend the large and complex datasets required by personalized medicine diagnostic approaches. In this context, the goal of machine learning, a subfield of *artificial intelligence*, is to find intricate patterns in data that can be utilized for sophisticated exploratory data analysis, advanced prediction, or classification of previously undiscovered data.

However, a significant problem is that many machine learning model findings are difficult to understand and interpret, particularly as the input data and the models

themselves become more complex. Current research in machine learning focuses on trying to reduce the *black box* aspect of machine learning models and increase their interpretability (which will be essential for their acceptance in clinical decision-making).

Therefore, the main results and findings of this Ph.D. work are a review of the current risk prediction tools used in cardiology compared to the innovative machine learning approaches, which often improve the predictions. This analysis led to two different studies: one methodological, i.e., the development of a mathematical model that can calculate the maximum performance achievable on a given dataset (in terms of accuracy), and one application of machine learning techniques to the prediction of all-cause death, recurrent acute myocardial infarction and major bleeding in patients with acute coronary syndrome (0.92, 0.81, 0.86 as mean AUC over the external validation). This approach also made it possible to study the relationship between clinical variables and disease outcome, providing information of interest on the rationale learned by the machine learning model. To then improve the temporal information of the above-discussed predictor, time-to-event prediction models were also applied to the same clinical context and outcomes (0.77, 0.69, 0.67 as mean C-index). The last study of the present Ph.D. thesis is a methodological work in which an explainable, reliable, and usable decision support system was developed to address the black box issue of machine learning in cardiology. The method was tested over the same clinical scenario of acute coronary syndrome, and it can provide a personalized risk prediction, feature importance, and confidence score.

In summary, the present work is presented as an innovation in the field of clinical decision support systems in cardiology, following the dictates of personalized medicine and risk prediction approaches by exploiting the computational potential of data-driven models.