Pre-Test Probability for Coronary Artery Disease in Patients with Chest Pain based on Machine Learning Techniques

Byoung Geol Choi, Ph.D., Seung-Woon Rha, M.D., Ph.D., Ji Young Park, M.D., Ph.D. and, Yung-Kyun Noh, Ph.D.

From: Department of Computer Science, Hanyang University, Seoul 04763, Republic of Korea (BGC, Y-KN), Cardiovascular Center, Korea University Guro Hospital, Seoul 08308, Republic of Korea (BGC, S-WR), Division of Cardiology, Nowon Eulji Medical Center, Eulji University, Seoul 01830, Republic of Korea (JYP), Cardiovascular and Research Institute, Korea University, 198, Gamasan-ro, Guro-gu, Seoul 08309, Republic of Korea (BGC), School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea (Y-KN).

Background: A correct and prompt diagnosis for coronary artery disease (CAD) is a crucial component of disease management to reduce risk of death and improve quality of life in patients with CAD. Currently, the American College of Cardiology (ACC)/American Heart Association (AHA) and the European Society of Cardiology (ESC) guidelines recommend selecting an appropriate pre-diagnosis test for an individual patient according to the CAD probability. The purpose of this study was to develop a practical pre-test probability (PTP) for obstructive CAD in patients with chest pain using machine learning (ML); also, the performance of ML-PTP for CAD is compared to the final result of coronary angiography (CAG).

Methods: We used a database from a single-center, prospective, all-comer registry designed to reflect real-world practice since 2004. All subjects underwent invasive CAG at Korea University Guro Hospital in Seoul, South Korea. We used algorithms of logistic regression, random forest (RF), supporting vector machine, and K-nearest neighbor classification for the ML models. The dataset was divided into two separate consecutive sets according to registration period to validate the ML models. ML training for PTP and internal validation used the first dataset registered between 2004 and 2012 (8,631 patients). The second dataset registered between 2013 and 2014 (1,546 patients) was used for external validation. The primary endpoint was obstructive CAD. Obstructive CAD was defined as having a stenosis diameter of greater than 70% on the quantitative CAG of the main epicardial coronary artery.

Results: We derived an ML-based model consisting of three different models according to the subject used to obtain the information, such as the patient himself (dataset 1), community first medical center (dataset 2), and doctors (dataset 3). The performance range of the ML-PTP models as the non-invasive test had C-statistics of 0.795 to 0.984 compared to the result of invasive testing via CAG in patients with chest pain. The training ML-PTP models were adjusted to have 99% sensitivity for CAD to not miss real CAD patients. In the testing dataset, the best accuracy of the ML-PTP model was 45.7% using dataset 1, 47.2% using dataset 2, and 92.8% using dataset 3 and the RF algorithm. The CAD prediction sensitivity was 99.0%, 99.0%, and 98.0%, respectively.

Conclusion: We successfully developed a high-performance model of ML-PTP for CAD, and it is expected to reduce the need for non-invasive tests in chest pain. However, since this PTP model is derived from data of a single medical center, multicenter verification is required to use it as a PTP recommended by the major American societies and the ESC.

Keywords: coronary angiography, coronary artery disease, machine learning, pre-test probability