Machine learning methods in real-world studies of cardiovascular disease

March 24 2023

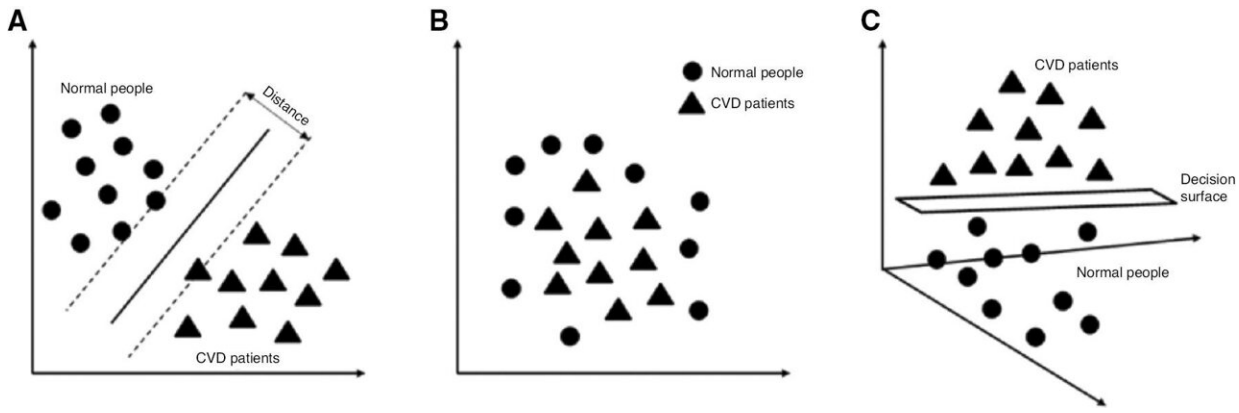

Illustration of the Support Vector Machine (SVM) Algorithm. The Black Circles and Triangles Indicate Unaffected Individuals and Patients with CVD, Respectively. A: Normal people and CVD patients are linearly separable. B: Normal people and CVD patients are nonlinearly separable. C: Normal people and CVD patients are mapped into high-dimensional space and separated by a decision surface. Credit: *Cardiovascular Innovations and Applications* (2023). DOI: 10.15212/CVIA.2023.0011

Cardiovascular disease (CVD) is one of the leading causes of death worldwide, and answers are urgently needed regarding many aspects, particularly risk identification and prognosis prediction. Real-world studies with large numbers of observations provide an important basis for CVD research but are constrained by high dimensionality, and missing or unstructured data.

Machine learning (ML) methods, including a variety of supervised and unsupervised algorithms, are useful for data governance, and are effective for high dimensional data analysis and imputation in real-world studies. This article reviews the theory, strengths and limitations, and applications of several commonly used ML methods in the CVD field, to provide a reference for further application.

This article introduces the origin, purpose, theory, advantages and limitations, and applications of multiple commonly used ML algorithms, including hierarchical and k-means clustering, principal component analysis, random forest, support vector machine, and neural networks. An example uses a random forest on the Systolic Blood Pressure Intervention Trial (SPRINT) data to demonstrate the process and main results of ML application in CVD.

ML methods are effective tools for producing real-world evidence to support clinical decisions and meet clinical needs. This review explains the principles of multiple ML methods in plain language, to provide a reference for further application. Future research is warranted to develop accurate ensemble learning methods for wide application in the medical field.

The study is published in the journal *Cardiovascular Innovations and Applications*.

More information: Jiawei Zhou et al, Machine Learning Methods in Real-World Studies of Cardiovascular Disease, *Cardiovascular Innovations and Applications* (2023). DOI: 10.15212/CVIA.2023.0011