Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques

In study, dimensionality is a big problem because categorical representation of drugs. SMILES data are described with binary data that reaches 2048 dimension for study. However, this number of dimensions may mislead the training process and increase complexity in both time & space. To prevent that, applying multiple feature selection techniques will reduce dimensions and decrease complexity of process.

In article, prediction of cardiovascular diseases which is the most common diseases, is addressed. By early identification of it, measurement can be taken before serious problems. In project, feature selection techniques LASSO and Relief were used. As training algorithm, multiple methods are proposed to provide ensemble learning which gives powerful results compared to single applied methods.