Paper ID: 1570940882

Paper Title: Using Machine Learning to Predict Heart Failure: Evaluating Model Performance on Clinical Data

Authors: Peeraya Wetchasit, Patararin Phaisakamas and Suphornthip Pongsuwan (Satriwithaya School, Thailand); Piyaon Sapphaphab and Oraya Rinthon (Satriwithaya School, Bangkok, Thailand); Suejit Pechprasarn (Rangsit University, Thailand)

Email: 49463@satriwit.ac.th

Abstract

Heart Failure or cardiovascular diseases (CVDs) are the number one cause of death globally, taking an estimated 17.9 million lives each year - that is 50,000 people a day or one death every 1.7 seconds, which accounts for 31% of all deaths worldwide. Heart failure is a common event caused by CVDs and this study uses a publicly available clinical dataset of 303 patients UCI Machine Learning Repository, containing 13 variables that can be used to predict a possible heart disease. The data is preprocessed and curated to verify its objectivity and accuracy before using principal component analysis and machine learning (ML) models. The various machine learning models are trained, tested, and compared, including trees, linear discriminant, quadratic discriminant, logistic regression, naive Bayes, support vector machine (SVM), K-nearest neighbor (KNN), ensemble, neural network, and kernel. The best model that can achieve the highest accuracy is Linear SVM with an accuracy of 84.3%. The principal component analysis method is then used to identify critical components and create an accurate model with fewer parameters. The Medium Gaussian SVM provides the best cross-validation classification performance and only requires nine predictors.