Title

Imputing clinical indicators for analyzing survival times in patients with prostate cancer

Authors

Kelvin Ofori-Minta,[1] Abhijit Mandal,[1,2] and Ming-Ying Leung[1,2]

[1]Department of Mathematical Sciences, [2]Border Biomedical Research Center, The University of Texas at El Paso, El Paso, TX.

Abstract

Progression free survival (PFS), defined as the time from random assignment in a clinical trial to disease progression or death by any other cause, has seen increased interest in its adoption as a substitute for the overall survival (OS) endpoint, partly due to the ability of PFS to monitor situations where OS improvements by follow-up treatments were marginal. These numerous upvotes for PFS endpoint argue that a survival benefit is experienced when tumor progression is stunted, hence an indication of tumor stabilization. However, the clinical data freely accessible via cBioPortal for cancer genomics often come with copious instances of missing observations across several clinical indicators. After data wrangling, the final dataset contains 25 clinical indicators from 494 subjects, with 12 of the clinical indicators reporting missing values. We deploy an existing imputation method, multiple imputation by chain equation (MICE). For comparison, we use two different algorithms from mice, predictive mean matching (PMM) and random forest (RF). Checking with diagnostic visualizations suggests that both PMM and RF produce plausible imputation results and depict near consistent data distribution with the original data. With these imputed datasets we will perform variable selection to reduce the dimensions of both the clinical and genomic data for our ultimate aim of integrating them to assess the contributing factors to the recurrence of prostate cancer and patient survival using deep machine learning methods, such as random survival forest and deep survival neural network.