[FA-0037] [Basic]

Are serum creatinine and ejection fraction enough to model decision for patient survival prediction: An Investigation for Machine Learning Solution Generation

Ali Syed Imran[1], **Lee Sungyoung**[1]

[1]*Kyung Hee University (Global Campus), Department of Computer Science and Engineering*

Objective: Machine learning (ML) has revolutionized the healthcare domain. As computing and storage resources have become less expensive and readily available along with the availability of huge among of data, intelligent automated techniques grounded in empirical research have become more prevalent in healthcare applications. Patient screening for potential health problems is yet another area that is suitable for the automation using latest machine learning techniques such as deep learning and ensemble models. Although these techniques provide amazingly accurate results on a wide variety of tasks, the underlying dynamics remain opaque to the decision-makers e.g. physicians, who are ultimately responsible for providing the final recommendation. Therefore, it is of utmost importance that the underlying logic of the black box models is explained to the human experts so that they can synthesize the empirical and theory-based insights to advance medical knowledge.

Methods: A feature attribute technique is investigated for explaining the behavior of a set of ML models. The model is trained on heart failure patients' data where the objective is to predict the death event i.e. if the patient has a high risk of dying during the follow-up period. Previous research has identified serum creatinine and ejection fraction as the two most important features to model the patient survival prediction. Through feature attribution, we have investigated whether the black box models also give higher weightage to the aforementioned two features or not?

Results: The experimental design consists of 5 ML models i.e., Gradient Boosting Trees (BGT), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and Deep Learning (DL). Table 1. provides information on the top-5 features for each model, along with the model's F-measure, bold values indicate important features selected by the model as in the previous studies.

Conclusions: This research investigated whether two important features as identified by previous research, serum creatinine and ejection fraction, are enough for ML modeling? Empirically it is demonstrated that although one or both of these features are selected by ML models, their importance varies from one model to another. Therefore, in modeling patient survival prediction, one should consider more features for better reliability of the prediction.

Acknowledgement: This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-0-01629) supervised by the IITP(Institute for Information & communications Technology Promotion)", and No.2017-0-00655(Lean UX core technology and platform for any digital artifacts UX evaluation) and IITP-2020-0-01489 (2020 Grand ICT Research Center)and 2022-0-00078,(Explainable Logical Reasoning for Medical Knowledge Generation) by Institute for Information & communications Technology Promotion(IITP)

Keyword: Serum Creatinine , Ejection Fraction , Machine Learning