WTUM-19-2022 - Building responsibleML models for prioritizing patients awaiting for transplantation

Hoang Thien Ly

March 11, 2022

1 Introduction

In the scope of the course Machine Learning Workshop, we would like to perform the analysis and investigate the possibility of using ML models in the field of medicine. In detailed, the States currently adopts Sickest-First Policy, which means for the patients having end-stage disease (e.g: end-stage liver disease) and requiring transplantation a new organ, there is a formula used to assess the severity of patients, afterwards, place them on the waiting list. The sicker the patient is, the higher place he/she is on the waiting list, and when compatible donated organs from deceased donors, sooner that patient will get transplanted.

In our study, we will place our emphasis for liver transplant. To assess the severity of end-stage liver patients, physicians in the US are using MELD-Score formula:

$$MELD = 3.78 \times \ln\left[Bili(mg/dL)\right] + 11.2 \times \ln\left[INR\right] + 9.57 \times \ln\left[Creati(mg/dL)\right] + 6.43$$

where:

- Bilirubin: how well liver clears substance "bile"

- INR: how well liver makes proteins needed for blood to clot

- Creatinine: how well kidneys work

Thanks to the use of MELD-Score, the instant fruition came to hopitalization system in the US. "The MELD-based allocation system was immediately successful, leading to first ever reduction in the number of waiting list candidates and a 15 % reduction of mortality among those on waiting list", mentioned in Freeman, R., Wiesner, R., Edwards, E., Harper, A., Merion, R., Wolfe, R.: Results of the first year of the new liver allocation plan. Liver Transplant, 10, 7-15 (2004).

But the problem locates at the point, the current system of assessing severity in the US is so naive and easily manipulated by doctors. In addition, the log-transformed values of Bilirubin, INR, Creatinine at 1 can be problematic, as a large percentage waiting list candidates possess those lab test below this threshold. And from a very simple model, what we can expect from, correlation between MELD and outcome may not be equally strong for all patients, and MELD may not accurately reflect the severity of their conditions.

2 ML Models

Dataset: Organ Procurement and Transplantation Network (OPTN) Standard Transplant Analysis and Research (STAR) dataset.

Target: Probability of patients dying or becoming unsuitable for transplant within 3 months.

Metric: out-of-sample area under the curve (AUC).

There are some noted work related to this study of proposing some groundbreaking ML models:

- Optimal classification tree [1](Bertsimas, D., Kung, J., Trichakis, N., Wang, Y., Hirose, R., Vagefi, P.: Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation.Am. J. Transplantat. 19, 1109–1118 (2018)).

- Logistic Regression and gradient-boosting ensembles with decision trees [2](Byrd J., Balakrishnan S., Jiang X., Lipton Z.C. (2021) Predicting Mortality in Liver Transplant Candidates. In: Shaban-Nejad A., Michalowski M., Buckeridge D.L. (eds) Explainable AI in Healthcare and Medicine. Studies in Computational Intelligence, vol 914. Springer, Cham. https://doi.org/10.1007/978-3-030-53352-6$_3$1)

The outcomes from those study are slightly better in term of higher AUC than the traditional MELD-Score model. But the question turns out to be, in the phase of deployment those models, is the model trust-worthy? Is the model being fair between ethnicity or gender? Do the important features taken after first model training cycle go alongside with domain knowledge from medical doctors? For single new instance, will the implemented model perform well?

With the hope of replying those questions, we will try to reproduce the results from their study then apply XAI Methods to partially answers those questions.

3 About the focal project

Partial Dependence Profile, Break Down method, Permutation-based variable important and Shapley values will be calculated to provide explanations for first-cycle of training models. Technical requirements:

1. Programming language: R

2. Necessary packages: xgboost, LightGBM, catboost, ranger, caret, forester, DALEX (for explaining models) and fairmodels (arXiv:2104.00507).

3. Member: Hoang Thien Ly

4. Report in RMarkdown + Presentation slides

As in any new projects, the model performances can turn out to be lower than expectation. But the main aim of this project is to build models responsible and to have applicable models in the field of medicine, which requires transparent and explanations to making even smallest decisions related to human life.

4 References:

- Bertsimas, D., Kung, J., Trichakis, N., Wang, Y., Hirose, R., Vagefi, P.: Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation.Am. J.

Transplantat. 19, 1109–1118 (2018)

- Byrd J., Balakrishnan S., Jiang X., Lipton Z.C. (2021): Predicting Mortality in Liver Transplant Candidates. In Shaban-Nejad A., Michalowski M., Buckeridge D.L. (eds) Explainable AI in Healthcare and Medicine. Studies in Computational Intelligence, vol 914. Springer, Cham.

- Waiting List, MELD Score, Liver Transplants — Dr. Robert S.Brown

- Liver Transplant Waitlist, Part 1— UCLA Transplantation Services

- UNOS.org