Machine Learning in predicting outcomes of organ transplantation

Author: Khoa Dao

Mentor: Prof. Zhe He

Objectives / Background

Organ transplants can be life-saving for people but are also full of potential risks that physicians and transplantation teams can't predict in the future. Because every person is different from others, including their organs which have been specialized for their own body. With the advancement in the medical field, we can only reduce the risk of organ transplanting between humans. However, the reduction in risk can give a large number of potential variables that have to be considered.

With the advancements in Machine Learning models, the tasks of evaluating those big amounts of variables and processing complex relationships can be more efficiently performed by using Machine Learning models to assist physicians in decision-making and considering potential organ offers. Additionally, Machine Learning models can also be trained to predict post-organ transplant outcomes. The purpose of this poster is to compare the existing research on the topic and figure out how can they support each other. Not just Machine Learning is being researched, Deep Learning has also proven to have potential in these kinds of tasks.

Data used:

United Organ Sharing (UNOS) database

Pediatric Heart Transplant Society database

Performances

Among all of the machine learning models has been researched, artificial neural network (ANN), and deep neural network (DNN) stands out with their performances followed by other machine learning models.

In predicting mortality after Heart Transplant, the DNN model achieved an AUC of 0.72, a Brier score of 0.08, a calibration slope of 0.99, a calibration intercept of -0.01, and an IDI of 0.05. And the ANN shows the best performance for predicting outcomes at 3 years post Liver Transplant. and Pediatric Heart Transplant (HTx)(DNN is not used in these researches).

For liver transplants, the ANN model achieved an accuracy of 0.82, a sensitivity of 0.83, a specificity of 0.81, a PPV of 0.86, an NPV of 0.77, an AUC of 0.88, and an F1 score of 0.84. For Pediatric Heart Transplants, ANN had the highest accuracy and area under the receiver operating characteristic curve (AUC).

Machine Learning Algorithms and Processes:

The Machine Learning Models that are being used for research are logistic regression (LR), multiple logistic regression (MLR), support vector machine (SVM), classification and regression trees (CART), random forests (RF), support vector machine (SVM), gradient boosting machine (GBM), artificial neural network (ANN), and deep neural network (DNN).

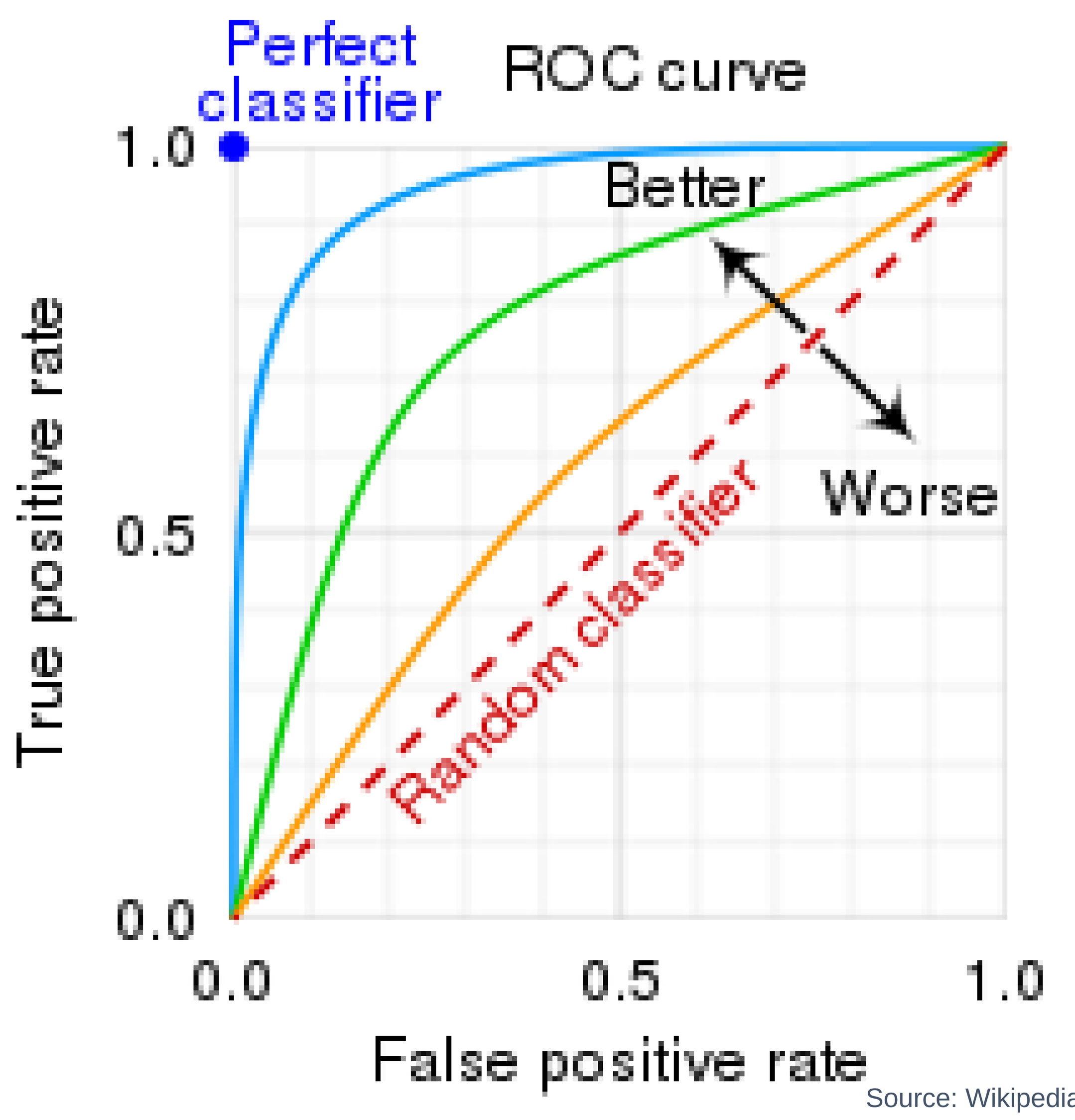
.

One of the ways to test and compare the Machine Learning Models is the shuffled 10-fold CV and rolling CV (Cross Validation). The shuffled 10-fold CV is a method to evaluate how well a machine learning model can predict new data. It works by dividing your data into # smaller sets, randomly shuffling them, and then using 9 sets for training and 1 set for testing. This is repeated 10 times, each time using a different set for testing. It does not have to be 10 times (5-fold CV). It can be any amount but 10 tends to be the most effective.

The other similar way that can be done is Rolling CV (Rolling Cross-Validation). Rolling cross-validation is a technique used to estimate the performance of machine learning models on time series data. It involves creating multiple folds or subsets of data that are ordered and using each fold as a validation set while using all previous folds as training sets. This way, the model can learn from past data and be tested on future data

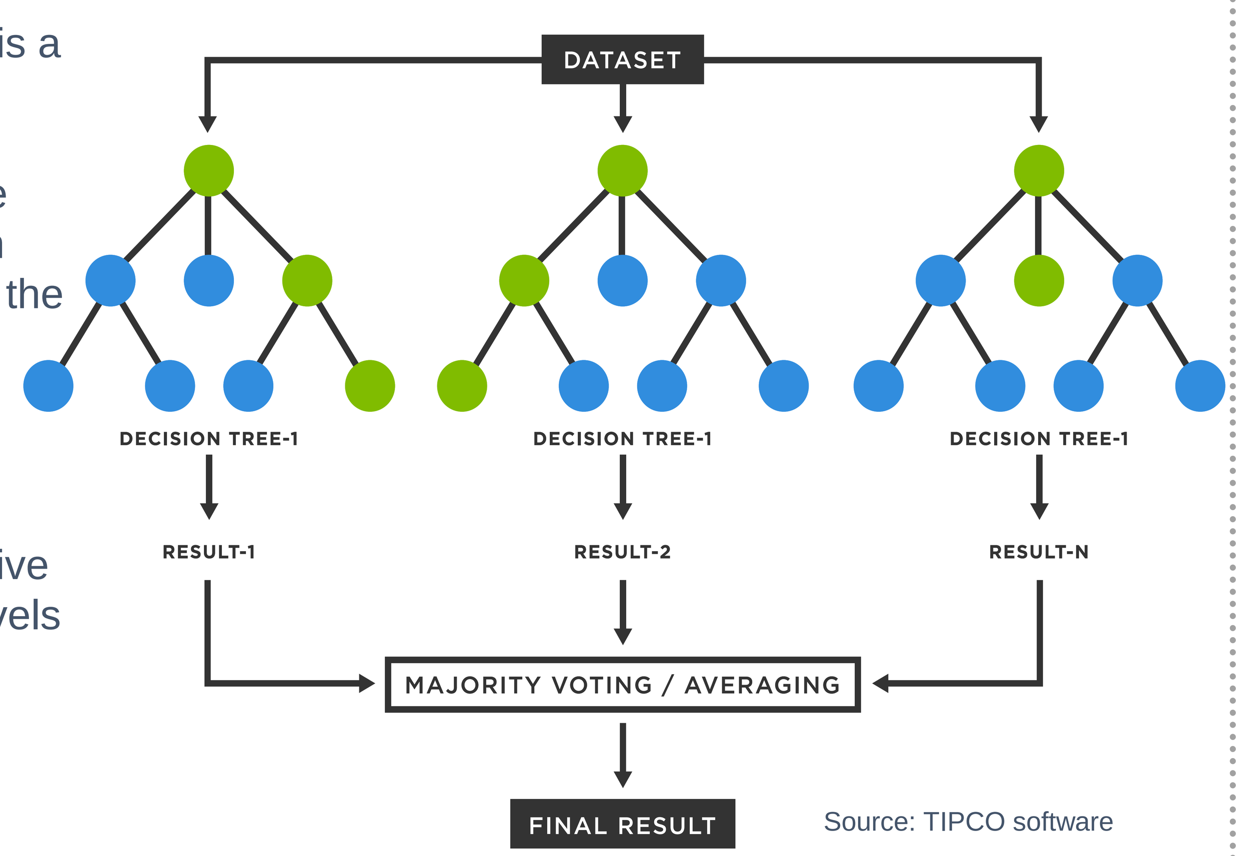
For Deep Learning (DL) models, the feature importance (SHAP) is a way for researchers to DL models. Feature importance is a way of measuring how much each feature contributes to the predictions of a machine-learning, deep-learning model. The SHAP (SHapley Additive exPlanations) approach assigns a SHAP value to each feature, which indicates how much the feature alters the prediction in comparison to the average prediction. To prioritize features and see how they affect the predictions, features can be ranked using SHAP values.

In order to analyze the results from these models, researchers examine the using Area under the ROC Curve (AUC). The false positive rate (FPR) is shown against the true positive rate (TPR) at various levels on the ROC curve. The likelihood that a randomly selected positive example would be ranked higher than a randomly selected negative example is represented by the AUC. An AUC of 1 would indicate a perfect classifier, whereas an AUC of 0.5 would indicate a random classifier.



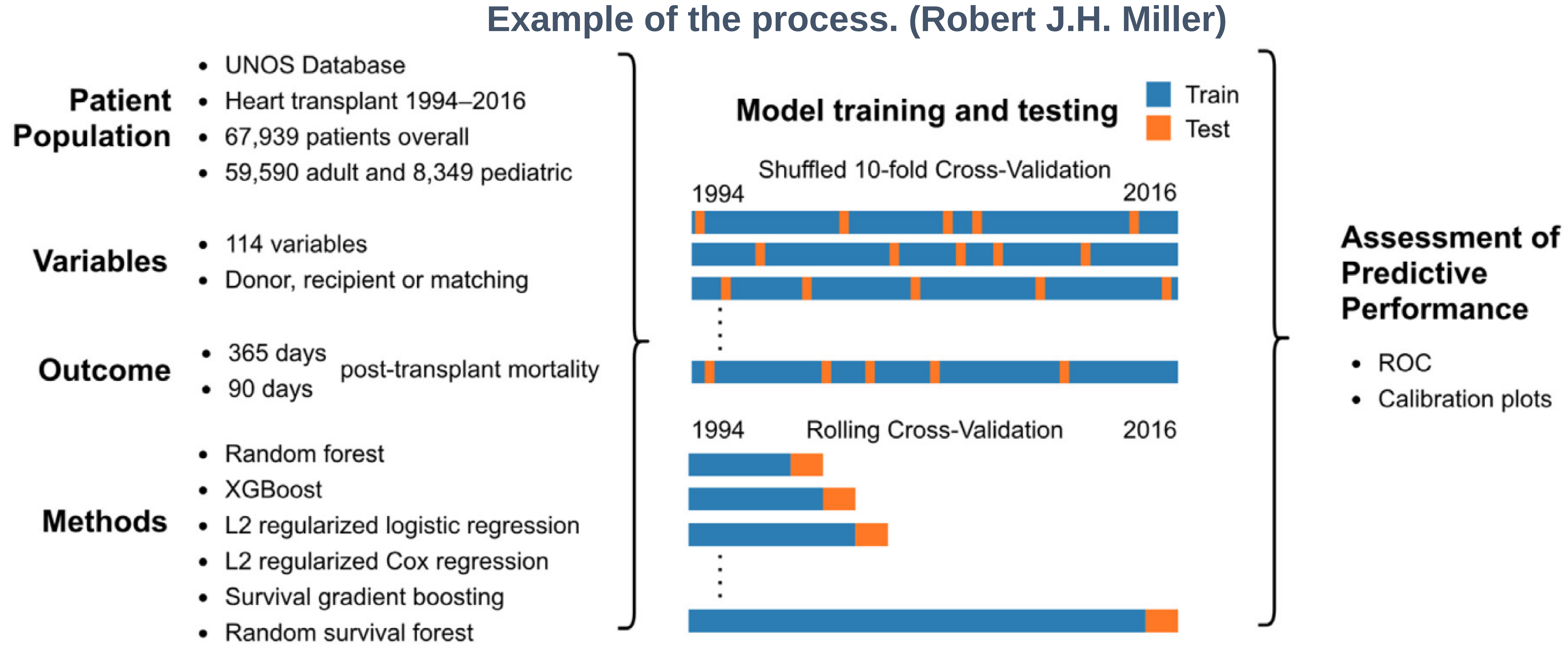
Source: Wikipedia

Random Forest Machine Learning Model



Source: TIPCO software

Example of the process. (Robert J.H. Miller)



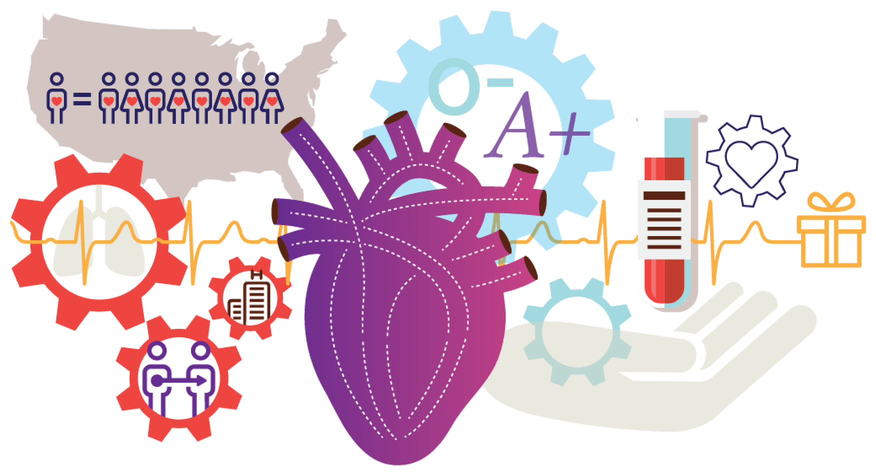
Limitations

One of the main limitations of this field of research is the availability of the data for training Machine Learning and Deep Learning models. Most of the current research uses the UNOS registry database and reported the limited amount of data that can be used to train the models. It is believed that with more data to train the models, the models will give more accurate predictions. All models showed a decline in performance over time due to temporal shifts in patient and donor characteristics and selection criteria in the research. Additionally, when models give out predictions, sometimes the researchers can't rationale the reason why it has given the prediction such as risk score. Thus, physicians can't reliably use it for certain patients' unique circumstances.

Conclusions

The results of these studies/researches show that both Machine Learning and Deep Learning models have fair predictive utility for the data that is being tested and trained on but the sensitivity for them still needs improvements. Notably, their prediction performance will be limited by temporal shifts in patient and donor selection. It is believed that with more data the result can be improved.

The author of these researches concluded that these predictions can help physicians and patients make informed decisions about HTx candidacy and post-transplant management, identify high-risk patients who may benefit from early intervention and optimization strategies to reduce their hospital stay and improve their survival, and need to be updated regularly to account for changes in practice patterns and outcomes over time.



Source: UNOS

References

1. Gupta, D., Bansal, N., Jaeger, B. C., Cantor, R. C., Koehl, D., Kimbro, A. K., Castleberry, C. D., Pophal, S. G., Asante-Korang, A., Schowengerdt, K., Kirklin, J. K., & Sutcliffe, D. L. (2022). Prolonged hospital length of stay after pediatric heart transplantation: A machine learning and logistic regression predictive model from the Pediatric Heart Transplant Society. The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation, 41(9), 1248–1257. https://doi.org/10.1016/j.healun.2022.05.016
2. Killian, M. O., Payrovnaziri, S. N., Gupta, D., Desai, D., & He, Z. (2021). Machine learning-based prediction of health outcomes in pediatric organ transplantation recipients. *JAMIA open*, *4*(1), ooab008. https://doi.org/10.1093/jamiaopen/ooab008
3. Miller, R., Tumin, D., Cooper, J., Hayes, D., Jr, & Tobias, J. D. (2019). Prediction of mortality following pediatric heart transplant using machine learning algorithms. *Pediatric transplantation*, *23*(3), e13360. https://doi.org/10.1111/petr.13360
4. Miller, R. J. H., Sabovčik, F., Cauwenberghs, N., Vens, C., Khush, K. K., Heidenreich, P. A., Haddad, F., & Kuznetsova, T. (2022). Temporal shift and predictive performance of machine learning for heart transplant outcomes. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation*, *41*(7), 928–936. https://doi.org/10.1016/j.healun.2022.03.019

Acknowledgements: Dr. Zhe He, Associate Professor