Failure Prediction of Heart Graft Transplantations within the first year

Ipsita Praharaj
Final Year Undergraduate (B. Arch)
Indian Institute of Technology,
Kharagpur, India
ipsita. praharaj1997 @ gmail.com

Paras Chaudhary
Final Year Undergraduate (B. Arch)
Indian Institute of Technology,
Kharagpur, India
paras7735 @ gmail.com

***Abstract*—The survival of heart transplant patients is a crucial decision to take and is mainly used for the matching procedure between a donor and a recipient. However, the disadvantages of existing methods for predicting life-expectancy of heart transplant patients are 1) these approaches rely on hand-engineering and use a small number of candidate predictors based on domain knowledge. This makes them prone to human-bias & often fail to leverage features, crucial for modeling complex functions of the input variables. 2) The resampling of the minority class (cases of failure of heart transplant) in equal proportions results in unrealistic synthetic data to train on, which affects the specificity of the classifier. In this work, we present a Machine Learning approach that performs feature selection, thus eliminating human bias and allows us to use a large number of input variables. We propose Adaptive Synthetic oversampling(ADASYN) for resampling the dataset, a data-resampling method that provides better accuracy over in-use resampling methods. Furthermore, we propose a method to score the predictors according to their sum of importance values obtained from different ML models defined as a function of the AUC and the specificity of the particular models. We demonstrate that using only 14 key features, we can obtain avg.AUC, accuracy & specificity values of 0.53, 0.58 & 0.42 respectively. While the state of the art published method uses 43 hand-engineered features and reports AUC, accuracy & specificity values of, 0.58, 0.70 and 0.38 respectively.**

Keywords—Heart Transplant, Feature elimination, Accuracy, Adaptive data resampling, Filters & Wrappers

I. Introduction

When a weakened heart cannot fill with enough blood and pump enough blood to satisfy the body's demand, this medical condition is termed as Heart failure. Within the United States, there have been 3408 heart transplants and 32 heart-lung transplants performed in 2018, nevertheless there have been over ten,700 total deceased donors, once as well as each DBD and DCD[1]. Primary graft malfunction happens once the donor heart fails and can't function. Your immune system also can reject the new heart. This rejection most commonly occurs within six months after the transplant. Over time, your new heart could fail thanks to the majorly same reasons that caused the original heart to fail.

This journal is intended to serve as an addendum to the existing literature, which focused on the heart graft survival outcome prediction for one time-period only and understand how the importance of a predictor translates into real-life.

II. Related Work

A. Survival data analysis: The Kaplan–Meier Method

The Kaplan–Meier methodology, a widespread methodology used for survival analysis. It enables us to calculate the incidence rate of events like recovery of nephritic operations, heart infarction or death by using patient data from all subjects at risk for these events. It explains how survival likelihood is calculated, survival data can be easily summarized and survival in groups is compared by using the logrank check for hypothesis testing. Also, it provides some concepts regarding the presentation of survival plots. However , it cannot provide an estimate of the size of the excellence between groups and a connected confidence interval, additionally it'll exclusively study the results of one issue at the time, and so they can not be used for applied statistical methods[2].

B. A deep learning approach

A fully connected deep dynamic neural network reads medical records, stores previous medical state history, infers current medical states and predicts future medical outcomes. At the information level, input is described as vectors and models patient health state patterns by the memory of historical records. designed on Long memory (LSTM). These models, in addition, learn expressly from models of medical interventions that amend the course of health issues and predict future medical risk. The results show improved prediction accuracy.[3]

C. Hybrid data analytic methodology

The study used for our reference developed an approach to predict the short, medium, and long outcomes for heart transplantation patients, and understand how the pattern of a predictor changes over time. They use the United Network for Organ Sharing (UNOS) dataset in their work with records of 12349 patients over thirty years. Since the datasets utilized plenty of survival cases than deaths for 1- and 5-year, it potentially going to bias the survival analysis. Hence, they use alternative methods for resampling, random below sampling (RUS) and artificial minority oversampling (SMOTE) to beat the data-imbalance issues. Their results indicate that supply regression combined with SMOTE achieves the optimum classification for the 1-, 5-,and 9-year outcome prediction, with area-under-the-curve (AUC) values of 0.624, 0.676, and 0.838, by severally exploiting 43 hand-engineered features.[4]

III. Motivation

Steady rate of growth in the gap between supply and demand of donated healthy hearts lead to longer waiting times and thereby leaving many to die while on the waiting lists. The human bias found in the statistical models, due the hand-engineering features and multiple base assumptions, prevents them from getting newer insights into the data. Thus data-mining models should be preferred over the former, as they are feature intensive and predict with performance measures that are very strong to replace the former techniques used. Resampling the dataset, could prove to be very precarious in terms of healthcare analytics because it generates synthetic data points that don't exist in the real world. Thus, there is a need to deploy newer techniques of data synthesis that turn out to be very close to the real-life cases. Feature selection is crucial when it comes to scalability and viability, the study used for reference makes a prediction based on well tabulated data from an authentic organisation, using 43 variables. This might not be possible in many healthcare organisations across the world currently. Hence, obtaining similar results with lesser number of features is crucial. Generating a strong grasp of knowledge (in the context of heart transplantation research) about predictive factors model to the outcome of a heart transplant for the 1 year period is more crucial for the following reasons:

1. The mortality due to natural causes has been found to increase in older patients(>9 years)
2. Graft rejection most commonly occurs within six months after the transplant.
3. Cuts down the major monetary loss at smaller gaps during transplant procedure, which has several distinct cost components.

IV. Objective

The main objectives of this study are to develop a data-driven method:

1. To replace the large feature set with major predictors to enhance the understanding of failure of a transplant for the 1 year period beforehand.
2. To obtain similar results using the smaller feature set so that it can be easier to implement across many healthcare systems.
3. To generate new insights with correlation heat maps of these features obtained

We have organized our pipeline based on research on data-driven methods to predict post-transplantation graft survival. These methods are mentioned below:

1. Simulation and operations research.i.e. Synthesizing data point addition for solving class imbalance.
2. Filter methods to remove redundancy
3. Conventional statistics.i.e. Logistic regression(LR)
4. Data analytic approaches.i.e. Using 2 popular data analytic models (artificial neural networks(ANN), and decision trees(CRT)).

V. Methodology

A. Data preparation

The dataset used for this study has been acquired from the United Network for Organ Sharing, UNOS, which is a "private, non-profit organization that manages the United States's organ transplant system under contract with the federal government". In our study we have used the cleaned and censored data from the previous literature, the data is censored. For convenience, we represent all the Dead/Live grafts within a given time frame under one variable (C_i). The variable gstatus indicates that the patient is still alive.

The decision (C_i) can be represented mathematically as:

$$C_i = (Yes, \forall gstatus = 0 | gtime \leq i * 365 days) OR (No) \quad (1)$$

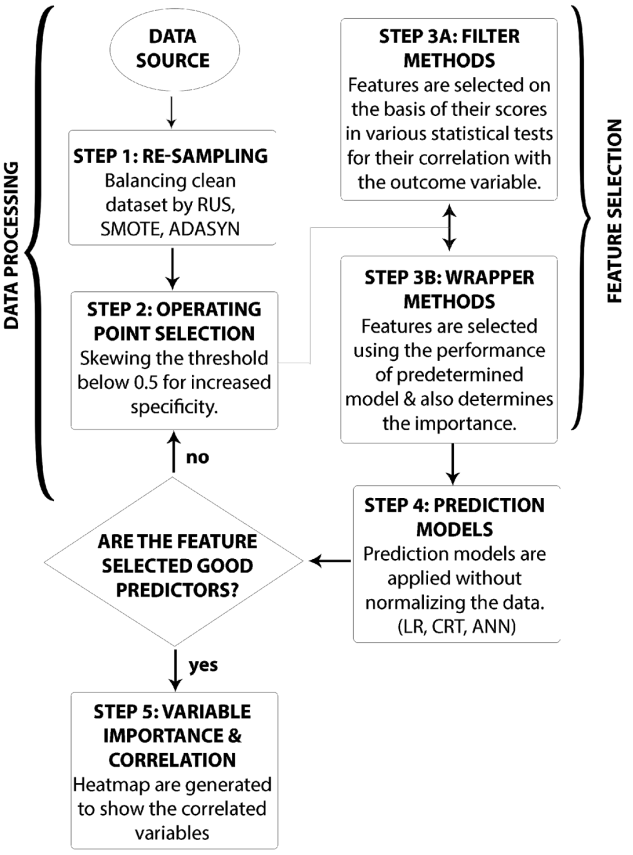

Fig. 1. The flowchart above illustrates our research workflow processes.

B. Data Sampling Methods

The dataset employed in this study contains a lot of survival cases in distinction to failure cases. If the information set is unbalanced the model is going to be biased. To label whether or not a dataset is unbalanced we tend to check the count of the dependent class values the quantitative relation ought to be 10:1 for the information set to be thought of as a balanced knowledge set. Re-sampling techniques are accustomed to solve this issue.

a) Random under-sampling (RUS): It is a simple under-sampling technique to under-sample the majority class randomly and uniformly.[5]

b) Superiority of over-sampling method (SMOTE): It finds the n-nearest neighbors within the minority category for every of the samples within the category . Then it

produces a line between the the neighbors associate generates random points on the lines[6]

c) Adaptive Synthetic oversampling (ADASYN): The idea is to use a density distribution as a criterion to mechanically decide the amount of artificial samples that require to be generated for every minority information example. It's an improved version of SMOTE for two reasons: 1) Reducing the bias introduced by the category imbalance. 2) Adaptively shifting the decision boundary toward the rare examples.[7]

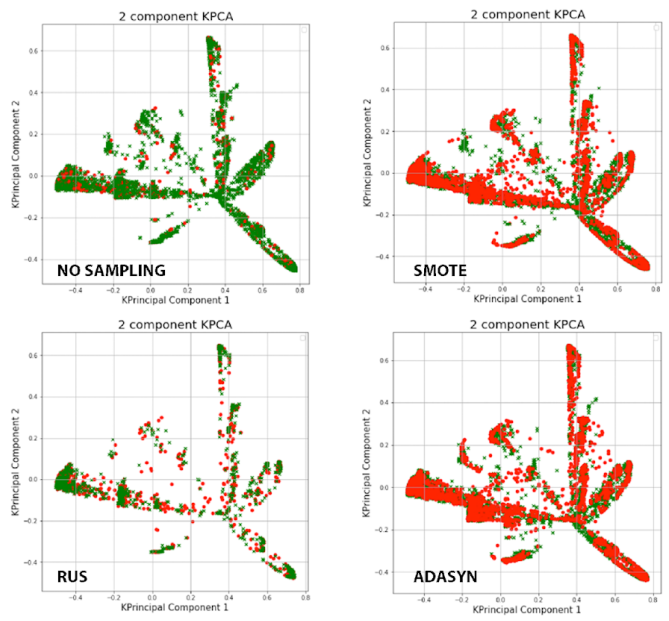

Fig. 2. Visualization of Data Sampling methods (L-R clockwise: No sampling, SMOTE, ADASYN, RUS)

C. Operating Point Selection

To ensure that the prediction strongly penalizes the latter error, we introduce this alteration in the threshold point of classification. This alteration is done based on the feature subset after feature selection algorithms. Operating point is decreased below 0.5 for classification to ensure that the classifier classifies the minority class (i.e. gstatus=0, failed heart graft),with more accuracy.

D. Feature Selection using a filter method, FCBF (Fast Correlation Based Filter)

Feature selection algorithms divide into 2 broad classes, the filter or wrapper model. Within the filter technique, features are selected on the basis of their scores in numerous statistical tests for their correlation with the result variable.[8]

Total Importance(V_i) in model(M_k) =

$$\frac{\sum_{i=1} AUC(M_k) * Specificity(M_k)}{frequency(V_i)} \qquad (2)$$

Where (V_i) belongs to feature space(122 variables) and (M_k) denotes any predictive model(LR, ANN, CRT) applied to the feature space. In our case, 0 denotes the failed grafts, therefore we focus on specificity to ensure

E. Feature importance calculation using wrapper method

The wrapper model needs one predetermined learning algorithmic rule in feature choice and uses its performance to gauge and verify that features are chosen. We have to use wrappers to gauge the importance of the variables. The importance is calculated based on the product of the AUC values of the model, the importance came from the wrapper model and the Specificity of that model.

Total Importance(V_i) in model(M_k) =

$$\frac{\sum_{i=1}(Importance(V_i|M_k) * AUC(M_k) * Specificity(M_k))}{frequency(V_i)} \qquad (3)$$

An unique feature subset is generated by fusing the most important features from both these feature selection techniques and the prediction is done in this feature space.

F. Target variable(C_i) prediction using the select features by Data analytic models

For this study, we apply two popular data analytic models (ANN, and CRT) and a conventional statistical method (logistic regression). We selected these three models due to: a) superior performance in several transplantation papers[9], and b) their superior performance in our preliminary analysis.

VI. Results

Three data analytic models (CRT, ANN and LR) were applied to predict the effect of the existing variables on the response/outcome variable gstatus.

1. The table 1 shows the importance of variables, calculated using eqn (3) from the Feature selection step.The UNOS variables are defined here[10].
2. The table 2 shows the results from Wrapper methods after being fed with select features set of 14 variables
3. The table 3 shows ADASYN vs SMOTE performance in the models
4. The table 4 illustrates changes in specificity by varying Operating point
5. The table 5 shows the correlated variables among the selected feature subset.

TABLE 1 : Selecting features

Variables	Cuml. Imp.	Variables	Cuml. Imp.
ECMO_TRR	0.727	AMIS	0.484
VENTILATOR_TRR	0.660	PULM_INF_DON	0.483
CREAT_TRR	0.623	INOTROP_VASO_ SYS_TRR	0.482
WGT_KG_DON_CALC	0.583	DRMAT	0.482
AGE_MATCH_LEVEL	0.574	IMPL_DEFIBRIL	0.481
DIAL_PRIOR_TX	0.558	TATTOOS	0.479
NUM_PREV_TX	0.553	HEP_C_ANTI_DON	0.425

TABLE 2: Prediction models

Modified Result	Model	Auc	Acc.	Recall	Specificity	Variables
1-year Survival 10 Cross Validation Threshold=0.4 W/ Selected 14 Features	CRT.NO	0.514	0.792	0.878	0.150	DISTANCE, ECMO_TRR, DAYS_STATI
	CRT.ADA	0.517	0.780	0.862	0.172	TBILI_DON, CREAT_TRR
	CRT.S	0.522	0.787	0.869	0.176	BUN_DON, CREAT_TRR
	CRT.RUS	0.536	0.531	0.530	0.542	DISTANCE
	ANN.NO	0.512	0.866	0.975	0.048	DISTANCE, ECMO_TRR, DAYS_STATI
	ANN.ADA	0.550	0.576	0.584	0.516	TBILI_DON, CREAT_TRR
	ANN.S	0.516	0.426	0.399	0.632	BUN_DON, CREAT_TRR
	ANN.RUS	0.516	0.426	0.399	0.632	DISTANCE
	LR.NO	0.531	0.865	0.969	0.093	DISTANCE, ECMO_TRR, DAYS_STATI
	LR.ADA	0.556	0.342	0.276	0.837	BUN_DON, CREAT_TRR
	LR.S	0.560	0.342	0.275	0.372	BUN_DON, CREAT_TRR
	LR.RUS	0.555	0.316	0.242	0.868	DISTANCE, ECMO_TRR
Avg. Specificity					0.42	
Avg. .Auc					0.59	

*NO: No sampling, S: Smote

TABLE 3: SMOTE vs ADASYN

SMOTE				
	AUC	ACC	RECALL	SPECIFICITY
Avg.	**0.537**	**0.500**	**0.489**	**0.585**
ADASYN				
	AUC	ACC	RECALL	SPECIFICITY
Avg.	**0.541**	**0.566**	**0.574**	**0.508**

TABLE 4: Operation point variation

Operating Point	Auc.Avg	Specificity .Avg
0.5	0.529	0.287
0.4	0.537	0.484

TABLE 5: Variables Correlation

VARIABLE 1	VARIABLE 2	CORRELATION
CORONARY_ANGIO	AGE_DON	0.605
HLAMAT	DRMAT	0.635
HLAMAT	AMIS	-0.535
BUN_DON	HLAMAT	-0.535

VII. Observations

There are several interesting observations that can be made from table 1, 2, 3, 4 and 5.

1. By changing the threshold point to 0.4 from 0.5, true negatives detection increases
2. The filter method used, FCBF (Fast Correlation Based Filter), provides comparable AUC values and specificity values.
3. ADASYN outperforms SMOTE for the feature set
4. From the correlation table, we can infer that, HLA-DR matching shows Improved Graft Survival
5. The role of HLA-A antigen loci mismatch is an important factor to be ensured before transplant.

VIII. Discussions

We have shown that comparable results can be obtained by feature engineering and correlated variables have also shed light on the inter- disciplinary fields of development. We demonstrated the use of only 14 key variables to obtain AUC, accuracy and specificity values as, 0.532, 0.612 and 0.484, whereas the reference paper showed AUC, accuracy & specificity values of, 0.582, 0.70 and 0.380 respectively.

In the context of transplantation research, after improving upon the previous results, there are several opportunities

1. To make a web-API where the user/ practitioners see the outcomes of these probabilities?
2. To improve upon the synthetic data examples used in our prediction models by engaging GAN's to mimic real data?
3. Suggest new variables to take into consideration while data sampling of the patients

To conclude, it should be noted that the analysis demonstrated in this paper can provide new prospective studies that can test hypotheses based on the recommended features for this prediction model.

Acknowledgment

We would also like to thank Prof Ram Babu Roy, Saurabh Singh, Phd Student at Rajendra Mishra School of Engineering Entrepreneurship, for his offered help in our execution and rectification of the models used for prediction.

Appendix A. Research Paper used for reference

Research Paper used for reference to this article can be found online at
https://www.sciencedirect.com/science/article/pii/S0167923616301816

IX. References

[1] Inci, Ilhan. "Donors after Cardiocirculatory Death and Lung Transplantation." Journal of Thoracic Disease, vol. 9, no. 8, Aug. 2017, pp. 2660–69.

[2] Jager, Kitty J., et al. "The Analysis of Survival Data: The Kaplan-Meier Method." Kidney International, vol. 74, no. 5, Sept. 2008, pp. 560–65.

[3] Pham, Trang, et al. "Predicting Healthcare Trajectories from Medical Records: A Deep Learning Approach." Journal of Biomedical Informatics, vol. 69, May 2017, pp. 218–29

[4] Dag, A., Oztekin, A., Yucel, A., Bulur, S., & Megahed, F. (2017). Predicting heart transplantation outcomes through data analytics. Decision Support Systems, 94, 42-52.

[5] T, E., & M, A. (2016). Classification of Imbalance Data using Tomek Link(T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method. Journal Of Informatics & Data Mining, 1(2)

[6] V. Chawla, N. (2002). SMOTE: Synthetic Minority Over-sampling Technique [PDF]. Journal of Artificial Intelligence Research 16 (2002) 321–357.

[7] Haibo He, & Garcia, E. (2009). Learning from Imbalanced Data. IEEE Transactions On Knowledge And Data Engineering, 21(9), 1263-1284

[8] Qinbao Song, Jingjie Ni, & Guangtao Wang. (2013). A Fast Clustering-Based Feature Subset Selection Algorithm for

High-Dimensional Data. IEEE Transactions On Knowledge And Data Engineering, 25(1), 1-14

[9] Dreiseitl, Stephan, and Lucila Ohno-Machado. “Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review.” Journal of Biomedical Informatics, vol. 35, no. 5, Oct. 2002, pp. 352–59

[10] SAF Data Dictionary. (2020). Retrieved 4 August 2020, from https://www.srtr.org/requesting-srtr-data/saf-data-dictionary/.