USC Viterbi
School of Engineering
Department of Computer Science

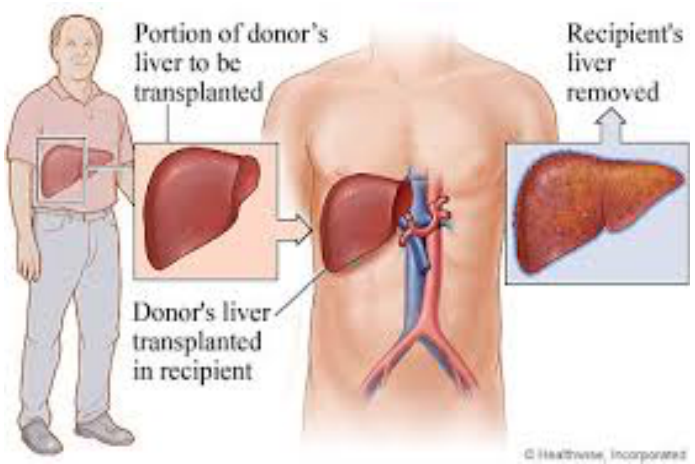
Keck School of
Medicine of USC

Multi-task Survival Analysis of Liver Transplantation using Deep Learning

Anna Farzindar, Ph.D.
and Anirudh Kashi
Viterbi School of Engineering
farzinda@usc.edu

Liver Transplantations

- Liver transplant is a lifesaving procedure for many liver-damaging conditions.
  - End stage cirrhosis, liver failure, or a cancerous lesion of the liver



Liver Transplantations

- Orthotopic liver transplantation (OLT) is the last resort for liver cirrhosis patients
  - Shortage of liver donors
  - 14,000 patients on waitlist, but only 7,000 annual OLTs
  - **3,000 patients died** while awaiting transplantation in 2015
- Must match donor and recipient features
  - Old age, sex/blood type mismatch, split grafts, and donation after cardiac death can lead to complications
  - Minimize cold ischemia time for best outcome (<24 hours)

The MELD Score

- Liver allocation policy - “Sickest first”
- Measure of sickness - MELD (Model for End-stage Liver Disease)
- Calculated using three main features:
  - Serum total bilirubin, a measure of jaundice
  - Prothrombin time, a measure of blood clotting ability
  - Creatinine, a measure of kidney function
- Higher the MELD score, more sick the patient is (Range from 1 - 40)
- Allocation also happens based on blood group.

Objective

- To select the best possible donor liver out of all the other available donor livers.
- Identify the important covariates that help in prediction (Feature importance).
- Visualizations of how much the prediction changes with change in feature values.

Datasets

- UNOS (United Network for Organ Sharing) -> Our Initial datasets
  - 59115 patients from 2002 to 2016
  - Around 150 features each (after pre-processing).
- SRTR (Scientific Registry of Transplant Recipients)
  - 87334 patients from 2002 to 2018
- Mapping of features: Using data dictionary, we manually mapped every feature in UNOS to SRTR to provide good comparison of the results.

Pre Processing

- Consider only deceased donors.
- Get records with MELD values from 12-23.
- Replacing all invalid values to NaNs
- Remove records with no label (GSTATUS, GTIME)
- Remove records with multi-organ transplantation
- Drop columns with more than 30% NaN values and records with more than 20% NaN values.
- Mean imputation for numerical columns and median imputation for categorical columns.

1) Graft Failure prediction using limited features

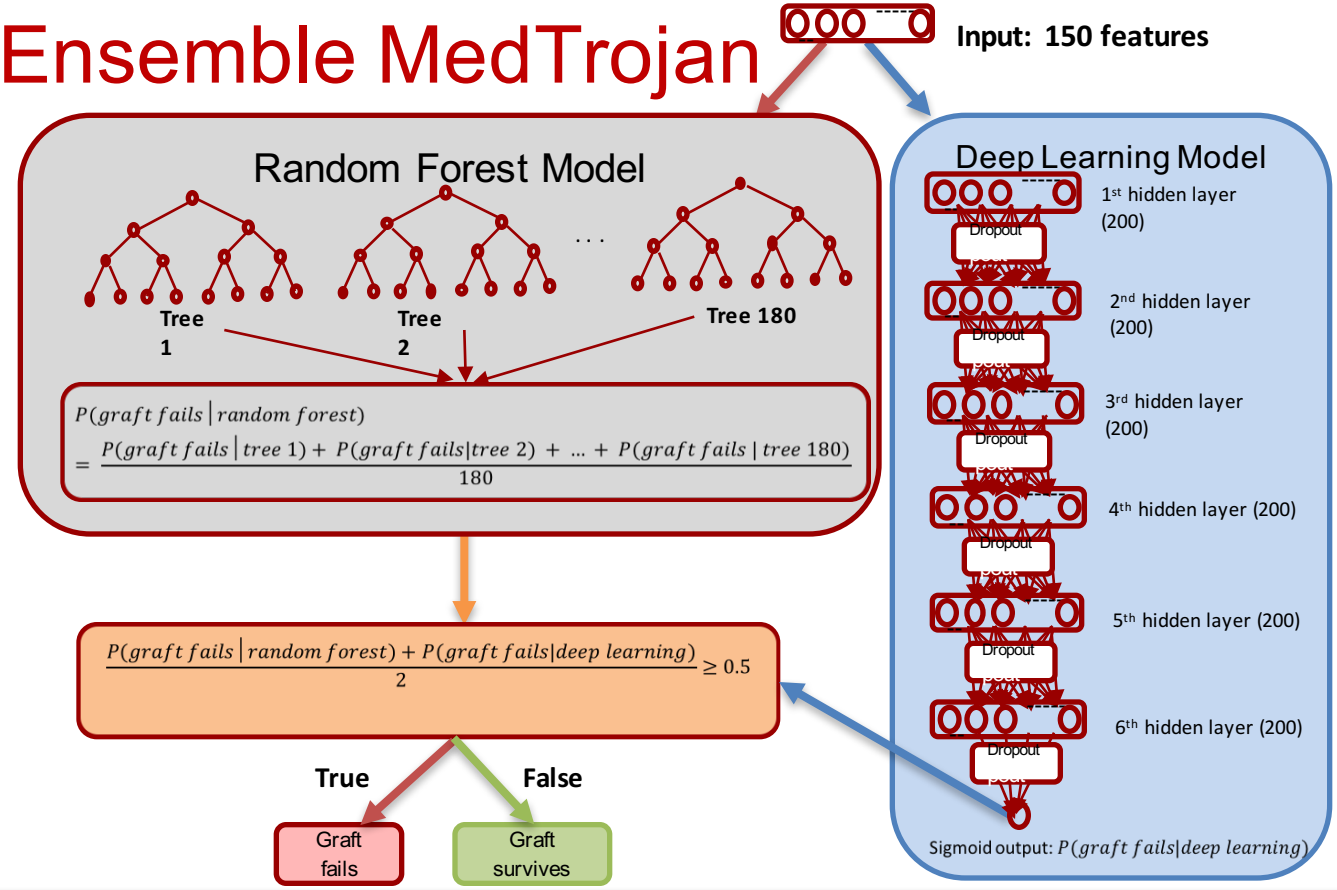
- In our Preliminary Work, we developed various models on version of the UNOS dataset with a **reduced 29 features dataset**
- Predictions for **graft failure at 1 Year Post Transplantation**
- Hoped to determine which models had the best potential to be further developed
- Models tested were Multilayer Perceptron (MLP), Random Forest, Support Vector Machine (SVM), and Deep Learning

Model	**Precision**	**Recall**	**F1**
MLP	0.14	0.40	0.20
Random Forest	0.14	0.48	**0.22**
SVM	0.10	**0.91**	0.18
Deep Learning	**0.16**	0.37	**0.22**

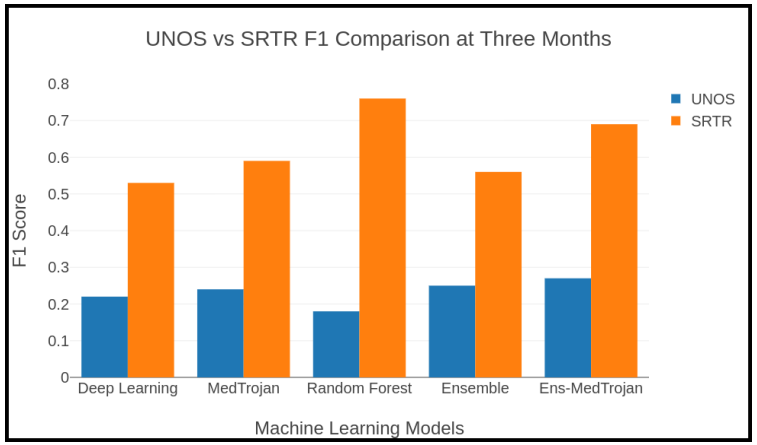
2) Graft Failure prediction using all available features

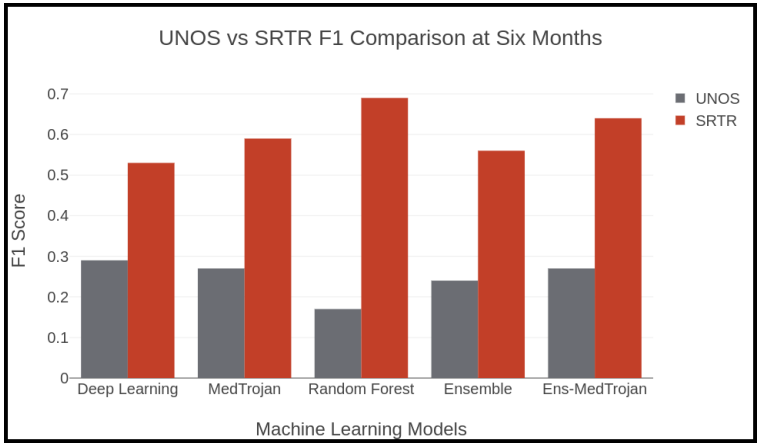
- Then we experimented using **all available features** from UNOS and SRTR dataset with Random Forest and Deep Learning models.
  - We built separate models for every duration, namely 3 months, 6 months, 1 year, 3 years **post transplantation**
  - Results of both Random Forest and Neural network was a value indicating probability of failure.
  - Final output was an ensemble result of both these models.

Ensemble MedTrojan

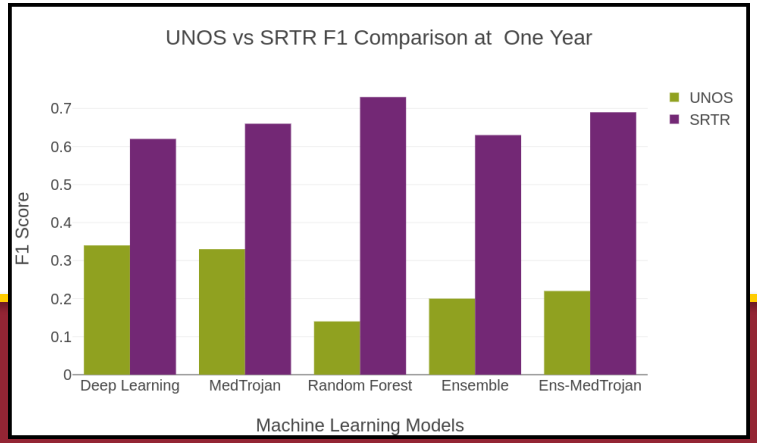


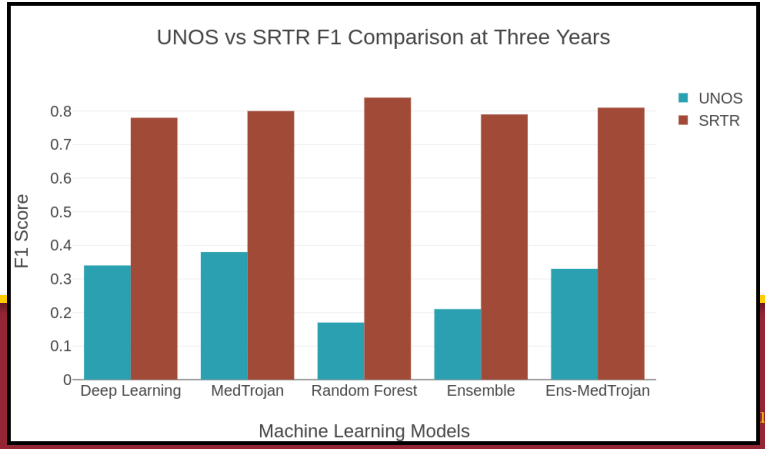
Evaluation: F1 scores and AUCROC





Note: **AUROC**s Show the similar comparison results





Limitation

- Downside to this kind of the model is that the results are not mutually exclusive. For example, if it predicts failure for 3rd month, it can predict Not Fail for 1 Year.

Survival Analysis with continuous time

- Survival Analysis is time to any event.
  - Here, the event is new liver failure.
- Major challenge – **Censoring**
  - Missing data problem in which time to event is not observed
  - Patients can enter the study at any time, leave before the desired event has occurred or the event might not happen over the course of the study at all.
  - There is the timeline of clinical trial which is important to consider.

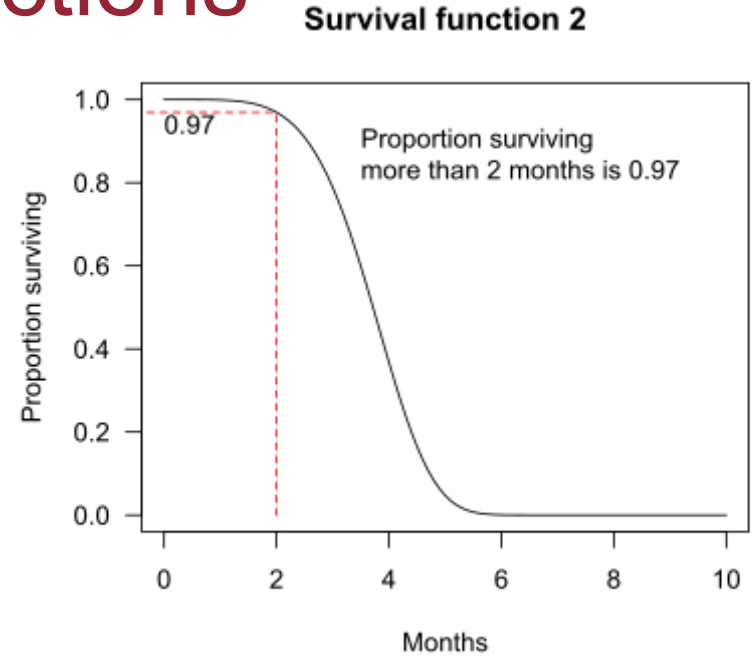
Survival and Hazard functions

- Survival function is the probability of a patient surviving longer than time t

$$S(t) = P(T \geq t)$$

- Hazard function is the instant probability that the event occurs knowing that the event did not occur before.

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T < t+dt | T \geq t)}{dt}$$



- Survival function can also expressed as a function of the hazard at all durations up to t.

Linear Cox models

$$\lambda_i(t|x_i) = \lambda_0(t).\exp(\theta.x_i)$$

- $\lambda_0(t)$ Is the baseline hazard value which is the hazard value when all the covariates are the default values.
- x_i Is the combined feature values of both the patient and the deceased donor, the ones which we are trying to find the risk.
- This model is learnt by maximizing the log-likelihood of normalized linear combination of feature values
- Limitation: For observation/characterized disease progression on existing cases
    - Not commonly used for prediction on new cases.

Linear Cox models

$$L_c(\beta) = \prod_{i:E_i=1} \frac{\exp(\hat{h}_\beta(x_i))}{\sum_{j \in R(T_i)} \exp(\hat{h}_\beta(x_j))}$$

$$l(\beta) = -\sum_{i:E_i=1} (\hat{h}_\beta(x_i) - \log \sum_{j \in R(T_i)} e^{\hat{h}_\beta(x_j)})$$

Where,

i is the patient and E_i = 1 indicates that the event happened.

$R(t) = \{i : T \geq t\}$ Indicates the set of all patients still at risk at time t

Non-linear methods

- Non-linear methods mainly include shallow and deep neural networks
- Deep Survival is extension of Cox model, using deep learning to model just the logarithm of hazard value.
- Exponentiate the output to get actual hazard value.

Our Multi-task deep survival analysis model

- We train our model to simultaneously learn **two tasks**:
  - Hazard value
  - Exact time when the liver might fail
- **Hazard value** is predicted as a real number.
- **Time** is modeled as multiple sigmoid units, where entire time frame is split equally among these sigmoid units
- Input to model: features from dataset for donor-recipient couples

First Loss functions: Hazard value

Cox's partial likelihood loss combined with Efron's Approximation to handle ties (Ref. Bradley Efron 1977).

$$l(\beta) = -\sum_{i:E_i=1}(\hat{h}_\beta(x_i) - \log\sum_{j\in R(T_i)} e^{\hat{h}_\beta(x_j)})$$

$$l_1(s^{(1)}) = -\sum_j \left(\sum_{i\epsilon H_j} \log s_j^{(1)} - \sum_{l=0}^{m-1} \log\left(\sum_{i:Y_i\geq t_j} s_i^{(1)} - \frac{l}{m}\sum_{i\epsilon H_j} s_i^{(1)} \right) \right)$$

where t_j denotes unique times, H_j the set of indices i such that $Y_i = t_j$ and $C_i = 1$

The output of first layer is a value for proportional to the hazard value.

Second Loss functions: Timeline

Combining a ranking loss with isotonic regression and ranking loss (Ref. Melon et al. 2012)

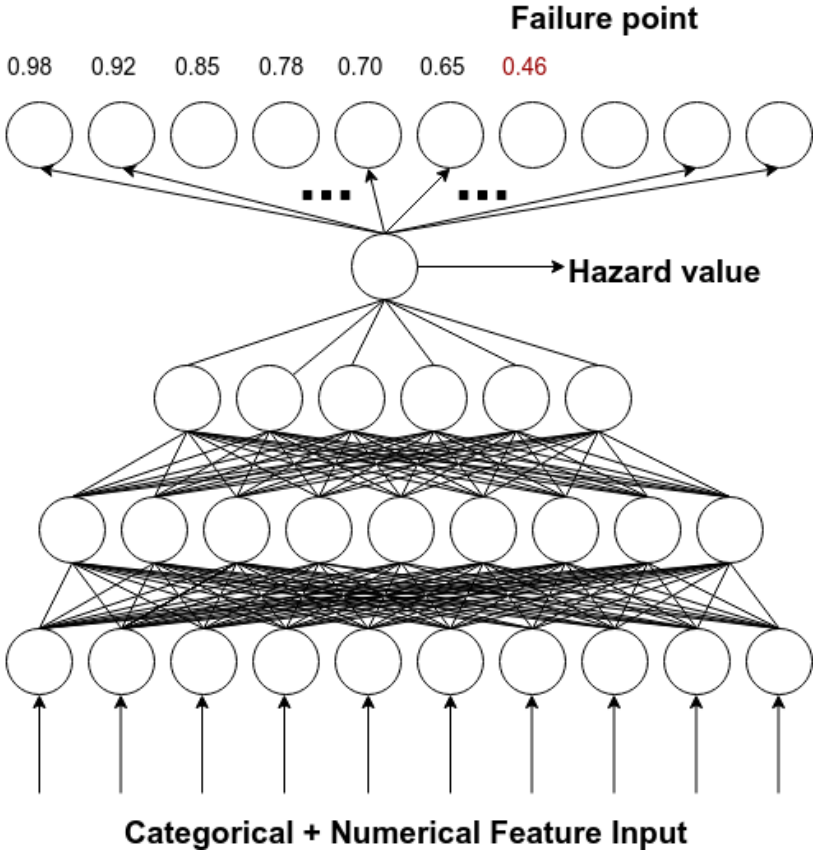
$$l_2(w) = \sum_{Acc(i,j)} l\left(s^{(2)}(x_j; w) - s^{(2)}(x_i; w), 1\right) y_j(1 - y_i)$$

where Acc(i, j) selects acceptable pairs, that is: i not censored, and at time of i's event, j is not censored

L(·, ·) is some convex loss function, here the L2 distance

-s(2) is output from the second layer

Model details



Metrics: C - index

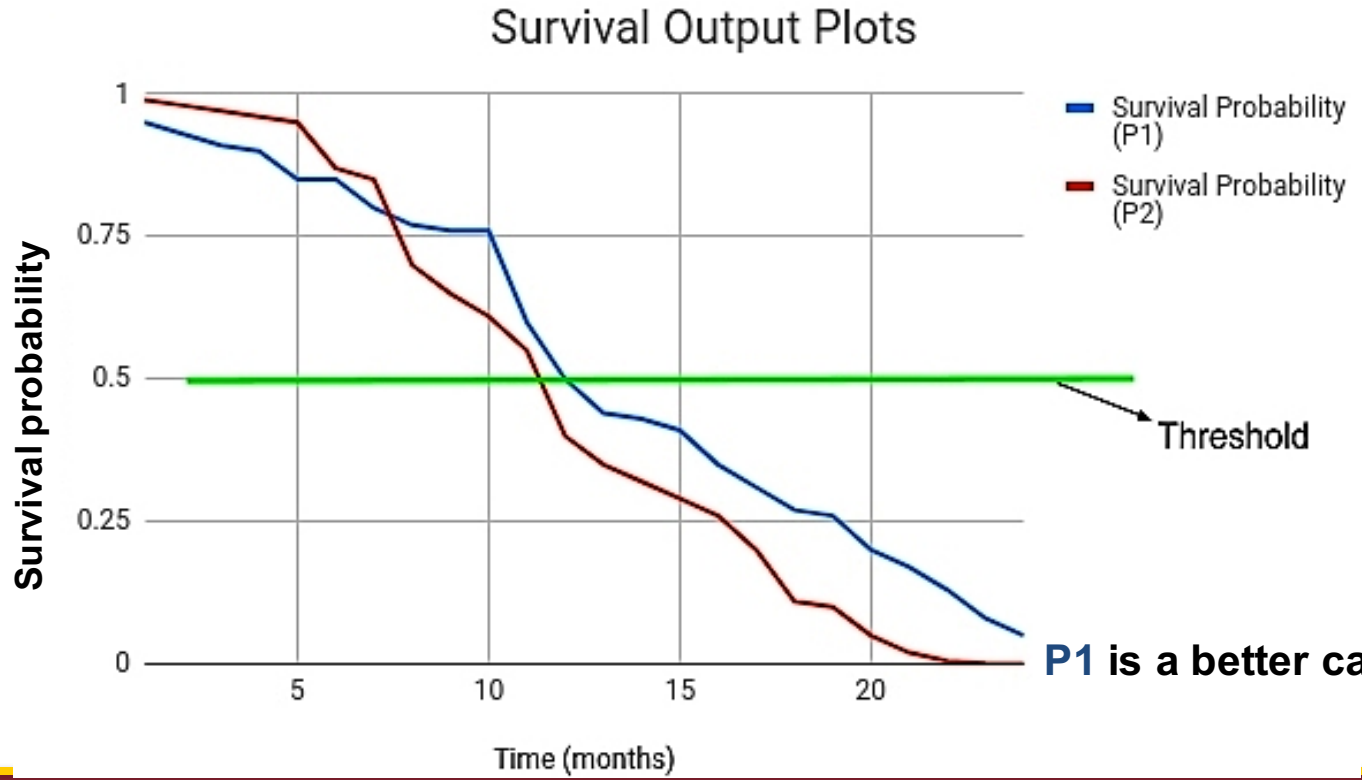
- For Survival Analysis we use a metric called C-index (Concordance index)
- Measures how good the ranking system is by finding the probability of non-inverted pairs.
- For example, if (T1, E1), (T2, E2), ..., (Tj, Ej) are the event times and occurrences in our dataset, C-index starts by counting the number of pairs which are correctly ordered by the model.
- C-index is the ratio of this value by the total number of admissible pairs.
- A pair (Ti, Ei), (Tj, Ej) are considered admissible if i is not censored and during i's event, j is still under risk.

Evaluation results

Dataset	Method	C-index
UNOS	Single Loss Function Double Loss Function	0.62 0.57
SRTR	Single Loss Function **Double Loss Function**	0.76 **0.82**

C-index values for survival analysis of 0.5 is considered random, It is acceptable if it is in between 0.6 and 0.7 and excellent above 0.7.

Survival values of two donor-recipient pair



P1 is a better candidate over P2

Conclusion

- Develop modern model for prediction of liver graft futility using DL and machine learning
- UNOS and SRTR Datasets
- First experience ML: Prediction for time intervals (3M, 6M, 1Y and 3Y)
- Then: Prediction of survival analysis for continuous time and handling censored data
- And multi tasks using Deep survival : Multi-task because we are learning two things together in the same model: Hazard value and Time.

References

Bambha, K., and Kamath, P. S. 2013. Model for end-stage liver disease (meld). *UpToDate. Waltham, MA: UpToDate*.

Cotler, S. J. 2015. Liver transplantation: donor selection.

Efron, B. 1977. The efficiency of cox's likelihood func- tion for censored data. *Journal of the American Statistical Association* 72(359):557–565.

Harrell Jr, F. E.; Califf, R. M.; Pryor, D. B.; Lee, K. L.; Rosati, R. A.; et al. 1982. Evaluating the yield of medi- cal tests. *Jama* 247(18):2543–2546.

Kim, W.; Lake, J.; Smith, J.; Skeans, M.; Schladt, D.; Ed- wards, E.; Harper, A.; Wainright, J.; Snyder, J.; Israni, A.; et al. 2017. Optn/srtr 2015 annual data report: liver. *Ameri- can Journal of Transplantation* 17:174–251.

Lau, L.; Kankanige, Y.; Rubinstein, B.; Jones, R.; Christophi, C.; Muralidharan, V.; and Bailey, J. 2017. Machine-learning algorithms predict graft failure after liver transplantation. *Transplantation* 101(4):e125–e132.

Luck, M.; Sylvain, T.; Cardinal, H.; Lodi, A.; and Bengio, Y. 2017. Deep learning for patient-specific kidney graft survival analysis. *arXiv preprint arXiv:1705.10245*.

Menon, A. K.; Jiang, X. J.; Vembu, S.; Elkan, C.; and Ohno- Machado, L. 2012. Predicting accurate probabilities with a ranking loss. In *Proceedings of the... International Con- ference on Machine Learning. International Conference on Machine Learning*, volume 2012, 703. NIH Public Access.

Merion, R. M.; Schaubel, D. E.; Dykstra, D. M.; Freeman, R. B.; Port, F. K.; and Wolfe, R. A. 2005. The survival benefit of liver transplantation. *American Journal of Trans- plantation* 5(2):307–313.

Raji, C., and Chandra, S. V. 2016. Artificial neural networks in prediction of patient survival after liver transplantation. *J. Health. Med. Inform* 7:1.

Schlansky, B.; Chen, Y.; Scott, D.; Austin, D.; and Naugler, W. 2014. Waiting time predicts survival after liver transplan- tation for hepatocellular carcinoma: a cohort study using the united network for organ sharing registry. *Liver transplanta- tion : official publication of the American Association for the Study of Liver Diseases and the International Liver Trans- plantation Society* 20(9):1045–1056.

Schlegel, A.; Linecker, M.; Kron, P.; Gyo ̈ri, G.; Oliveira, M. D.; Mu ̈llhaupt, B.; Clavien, P.; and Dutkowski, P. 2016. Risk assessment in high- and low-meld liver transplantation. *American journal of transplantation : official journal of the American Society of Transplantation and the American So- ciety of Transplant Surgeons* 17(4):1050–1063.

Wiesner, R.; Edwards, E.; Freeman, R.; Harper, A.; Kim, R.; Kamath, P.; Kremers, W.; Lake, J.; Howard, T.; Merion, R. M.; et al. 2003. Model for end-stage liver disease (meld) and allocation of donor livers. *Gastroenterology* 124(1):91– 96.