2022 IEEE Systems and Information Engineering Design Symposium

2022 Best Paper Awards

Energy and the Environment Track

Optimizing for Water Equity in the Colorado River Basin

Hania Abboud, Erin Baker, Teagan Baiotto, Christopher Weigand, Julianne Quinn

University of Virginia

Optimization, Simulation, and Decision Analysis Track

Multi-Criteria Decision Analysis Tool for Capital Planning and Prioritization of WMATA Facilities and Assets

Latifa Al Jlayel, Kazi Asifa Ashrafi, Yumna Dahab, , Diing Manyang, Joost Santos

George Washington University

Health Track

Predicting Liver Utilization Rate and Post-Transplant Outcomes from Donor Text Narratives with natural language processing

Kristy Bell, Madeline Hennessy, Michael Henry, and Avni Malik

University of Virginia

Data Track

Golf and GameForge: Innovative Analytics for Recommender Systems

Rachel Kreitzer, Rose Dennis, Steven Wasserman, Zachary Kay, Jerry Lu, Sam Roberts, Thomas Twomey, William Scherer

University of Virginia

Systems Design Track

Developing a Dynamic Control Algorithm To Improve Ventilation Efficiency in a University Conference Room

Matthew Caruso, Jason Jabbour, Caleb Neale, Alden Summerville, Avery Walters, Arsalan Heydarian, Arthur Small, Mahsa Pahlavikhah Varnosfaderani

University of Virginia

Infrastructure and Networks Track

Design of a Prioritization Methodology for Equitable Infrastructure Planning

Rahul Dhansinghani, Ayman Ibrahim, Aditya Kannoth, Claire Miller, Lena Nguyen, Steven Pham, Reid Bailey

University of Virginia

POLICY TRACK

DEMOCRATIZING HOUSING AFFORDABILITY DATA: OPEN DATA AND DATA JOURNALISM IN CHARLOTTESVILLE, VIRGINIA

Spencer Bozsik, Xinlun Cheng, Malvika Kuncham, Evan Mitchell

University of Virginia

Predicting Liver Utilization Rate and Post-Transplant Outcomes from Donor Text Narratives with Natural Language Processing

Kristy Bell
Master of Data Science Candidate
University of Virginia
Charlottesville, VA
kab7kr@virginia.edu

Madeline Hennessy
Master of Data Science Candidate
University of Virginia
Charlottesville, VA
mlh5fy@virginia.edu

Michael Henry
Master of Data Science Candidate
University of Virginia
Charlottesville, VA
ddj9rt@virginia.edu

Avni Malik
Master of Data Science Candidate
University of Virginia
Charlottesville, VA
am7bf@virginia.edu

Abstract— **Liver transplantation is a critical, life-saving treatment option for patients with terminal liver disease. Despite an organ shortage, many donated livers are discarded for reasons such as poor organ condition and physical incompatibility with a recipient. Current clinical models for liver risk assessment only utilize tabular data and result in poor precision and recall. Critical information relevant to this decision-making is likely included in the free-text clinical notes from donor evaluations that contain pertinent medical and social history of the donor that is currently unavailable in tabular data sources.**
This article describes the development of a model using these free-text clinical notes using a variety of Natural Language Processing (NLP) and machine learning (ML) techniques to predict the outcomes of three key metrics: 1) liver utilization rate, 2) 30-day mortality rate, and 3) 1-year mortality rate. The free-text narratives were useful for predicting liver utilization, with an associated area under the curve (AUC) score of 0.81, but were not useful for predicting both mortality outcomes, with associated AUC scores of 0.53 and 0.52, for 30-day and 1-year mortality, respectively. Using a locally interpretable model-agnostic explanations (LIME) algorithm, key phrases, like "dcd" and "alcohol" were found to be associated with unutilized livers, while "brain" and "heroin" were associated with utilized livers. Based on these findings, modeling donor text narratives may substantially contribute to improved decision-making and outcomes of liver transplantation.

Keywords— Natural Language Processing (NLP); Electronic Health Records (EHR); text analytics; liver transplantation

I. Introduction

Liver transplantation is the only treatment for irreversible liver failure and cirrhosis, which affects an estimated 1-in-400 adults in the United States [1]. There is a high demand for donor livers in Western countries, with a waiting list mortality of 20-30% [2]. The shortage of available donor organs proves to be a limiting factor in liver transplantation and is only exacerbated by the number of livers that are discarded every year [3]. The exact reasons for discard are still unclear, but are generally based upon donor history and a subjective evaluation of the donor chart by the transplanting surgeon [2]. The donor chart is produced through a rigorous evaluation process managed by the United Network for Organ Sharing (UNOS), which serves as the nation's Organ Procurement and Transplantation Network (OPTN). When a potential organ donor is identified, a Deceased Donor Registration (DDR) form is filled out for initial screening by a representative from a local Organ Placement Organization (OPO) [4]. The evaluation includes physical examinations and laboratory results, which are used to populate tabular donor data. Following initial approval, medical and social history on the donor is compiled through medical records and in-depth interviews with next-of-kin, and the information is stored in DonorNet, a data management system maintained by UNOS.

An OPO medical coordinator determines the organ's suitability for transplantation based on standard or extended criteria, the latter of which includes certain risk factors such as age greater than 70, hepatitis C infection, and donation after cardiac death (DCD), which have been associated with adverse post-transplant outcomes. Preferred organ donors are young, otherwise healthy individuals having fatal brain injuries, identified via standard criteria. However, due to the scarcity of these donors, most liver transplants come from donors meeting the extended criteria [5].

The current liver allocation policy in the U.S. is primarily determined by the recipient's Model for End-Stage Liver Disease (MELD) score, a measure of the severity of liver disease [6]. Once the recipient receives an offer for the donor liver, the transplanting surgeon reviews the donor chart alongside the recipient's medical chart. Noting potential risks, the transplanting surgeon may recommend rejecting the offer due to disease transmission risk, transplantation logistics, donor history, or other reasons [7]. In such cases, the offer is passed to the next recipient according to acuity and distance until it is accepted [8]. Often donor livers will not be accepted by any recipients, at which point the liver will be discarded.

II. Background

The most common liver risk assessment index, the Donor Risk Index (DRI), highlights donor factors associated with transplant failure (re-transplantation or recipient death) [9]. These donor factors – increased age, DCD, split liver donation, African American race, shorter height, and causes of death other than trauma, stroke, or anoxia – have been verified to affect liver utilization and post-transplant outcomes; however, the DRI has been a poor predictor of liver utilization and mortality. One study that looked at predicting post-transplant mortality using the DRI, as well as the MELD score, failed to prove any significant predictive value, as their models had associated C-statistics of 0.54 and 0.66, respectively [10].

To improve upon the DRI's performance, this project incorporated unstructured clinical notes, which may contain more detailed information on the medical and social history of deceased donors. A similar study on kidney utilization demonstrated the utility of donor text narratives using an NLP approach, suggesting that a text model might capture information not in the tabular data and allow more successful prediction of liver utilization and post-transplant mortality [11].

III. Data Description

This study used data from the OPTN data system, which includes data on all donors, wait-listed candidates, and transplant recipients in the U.S. submitted by the members of the OPTN [12]. The Health Resources and Services Administration (HRSA) and the U.S. Department of Health and Human Services provide oversight to the activities of the OPTN contractor.

Unstructured text in DonorNet was concatenated from three different fields, including admission course (AC), donor highlights (DH), and social history (SH). Liver utilization and transplant outcomes were extracted from the UNOS Standard Transplant Analysis Request (STAR) file. Utilization was defined as 1 if the liver was recovered from the deceased donor and 0 if no liver was recovered. Likewise, 30-day and 1-year mortality were coded 0 if the transplant resulted in the recipient's death within 30 or 365 days of transplantation, respectively, and 1 otherwise. The complete study cohort consisted of 104,607 deceased donors from 2001 to 2021, with 17.9% being DCD (Table I). All deceased donors were included in the construction of a utilization model, with roughly 80% of livers being recovered. The mortality model cohort, composed of 75,591 transplanted donor livers, had recipient mortality rates of 2.7% and 8.1% for 30 and 365 days, respectively.

The age and demographic of the donors for both the utilization and mortality cohorts were notably similar. The average cohort ages were 40 and 39 years old, respectively, with a standard deviation of 17 years for both. The three largest demographics for both cohorts were White, Black, and Hispanic: 66%, 16%, and 14% for the utilization cohort and 64%, 18%, and 14% for the mortality cohort.

TABLE I. COHORT DESCRIPTION

Deceased Donor Group	***Count***	***Avg Age***	***DCD Donor %***
All Deceased Donors	104,607*	40.5	17.9%
Liver not recovered	21,049	42.5	56.0%
Liver recovered	83,558	40.0	8.3%
Recovered but DISCARDED (no transplant)	7,967	45.9	25.2%
Recovered and transplanted	75,591+	39.4	6.5%
Recipient death (<30 days)	2,069	38.7	6.3%
Recipient death (31-365 days)	4,066	41.6	6.7%
Recipient lived (>365 days)	69,456	39.3	6.5%

IV. Methodology

Several transformations were applied to the text in order to derive meaningful insights through ML. The key text fields (AC, DH, SH) were concatenated into one body of text, stripped of numbers and punctuation, and converted to lowercase. Next, "stop words" such as "the" and "is," were removed using the NLTK package in Python. Finally, each word in the text was stemmed with the Porter stemmer in the NLTK package.

After these pre-processing steps, all unigrams and bigrams were converted to a vector of term-frequency inverse-document-frequency (TF-IDF) values using the default Sci-Kit Learn implementation. The data was partitioned into training, validation, and testing sets. For most of the models there was a 75%/25% split for train and test sets. A separate stacking classifier was also built, and it utilized a 56.25%/18.75%/25% split for training, validation, and test sets, with the validation set being used to train the weights that were given to the predictions of the four individual models.

Logistic regression, random forest, and gradient boosting were used to predict the outcome of utilization. A stacking classifier, composed of four logistic regression models, was also fit to deal with the class imbalance of the liver utilization dataset. Each logistic regression model was trained on one fourth of the positive class and the entirety of the negative class in the training data. After hyperparameter tuning on their respective training sets, the four individual models were combined into an ensemble model and tested for performance on a validation set. In order to measure the models' AUC scores, as well the confidence intervals of those AUC scores, bootstrapping was performed using the test set. For the best performing models – as measured by specificity and sensitivity – the LIME algorithm was leveraged to reveal the top 30 most predictive unigrams and bigrams for each class (utilized liver/not utilized liver).

After similar pre-processing, the 30-day and 1-year mortality models, which had an even greater class imbalance, were trained on small subsets of the data with an even class distribution. Initial testing yielded no significant insights. Further testing using more advanced models and more data confirmed the lack of predictability seen in the down sampled models.

V. Results

A. Utilization Outcome Model Performances

Using the subset of donor text fields (AC, DH SH), liver utilization was predicted AUC scores of 0.78, 0.78, and 0.81 on a holdout dataset with a gradient boosting, random forest, and logistic regression classifiers, respectively (Fig. 1). The logistic regression model – built with an L2 penalty and trained with a balanced class weight – yielded the highest Cohen Kappa score (0.43) and specificity (0.62) on the holdout data, suggesting that it was able to recognize unutilized livers at the highest rate. The gradient boosting classifier correctly identified utilized livers at the highest rate of all models with a sensitivity of 0.97. To resolve the data imbalance and improve model performance, a stacking classifier was constructed but yielded an AUC score comparable to the logistic regression classifier.

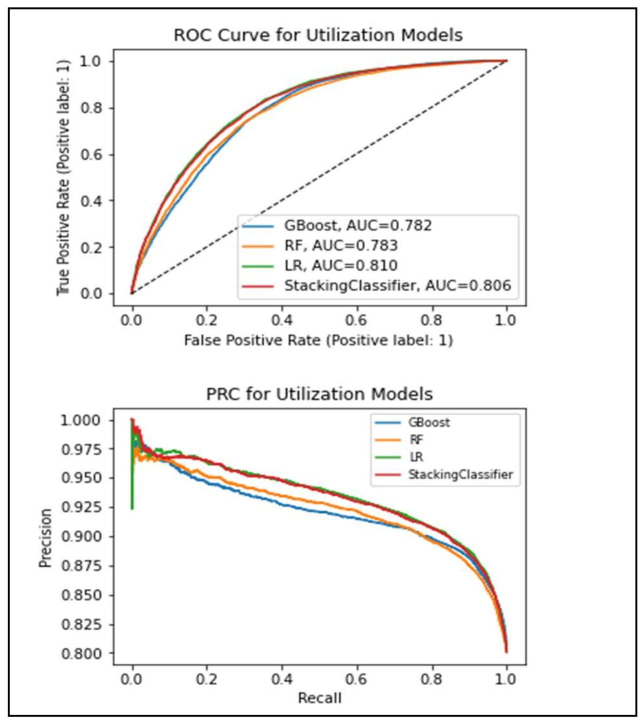

Fig. 1. Receiver operating characteristics (ROC) curve and Precision Recall Curve (PRC) colored by supervised methods built for predicting utilization of deceased donor livers.

To examine the correlation of predictions for the two models with the highest specificity and sensitivity, a scatter plot was produced. The color indicates the true class label (Fig. 2). There appears to exist a non-linear relationship between prediction probabilities on the test set. The Spearman's rank correlation coefficient was 0.74 between the predictions of the two models.

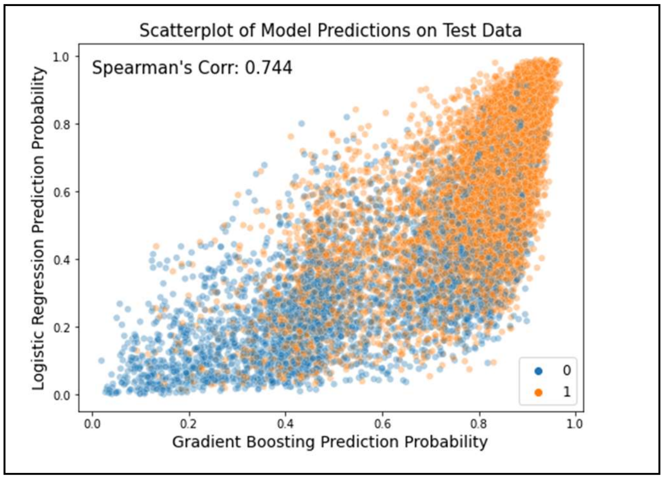

Fig. 2. A scatterplot reveals the non-linear relationship between the prediction probabilities on the test set for the gradient boosting and logistic regression models colored by true class label (1: utilized, 0: unutilized). Predictions are correlated with a Spearman's Rank Order correlation coefficient of 0.744.

B. Top Predictive Words

The results from the LIME algorithm for the logistic regression classifier and the gradient boosting classifier are visualized in Fig. 3. Many of the top predictive words in the logistic regression model were also highly predictive in the gradient boosting model. It is evident from both LIME outputs that the donor cause of death as well as history of substance abuse are key factors in determining whether a liver is utilized. The top predictive words for the utilized liver class include "marijuana," "drug," "smoke," "heroin," as well as words suggesting brain death ("brain," "bd"). Notably, words identifying cardiovascular death ("dcd," "respiratori," "cpr"), or a history of alcohol abuse ("vodka," "alcohol," "etoh") had high predictive influence towards the unutilized liver class. Other, more general words and abbreviations, including "death," "liver," "pt," "donor," "increas," and "neg", were also flagged as predictive words via the LIME algorithm.

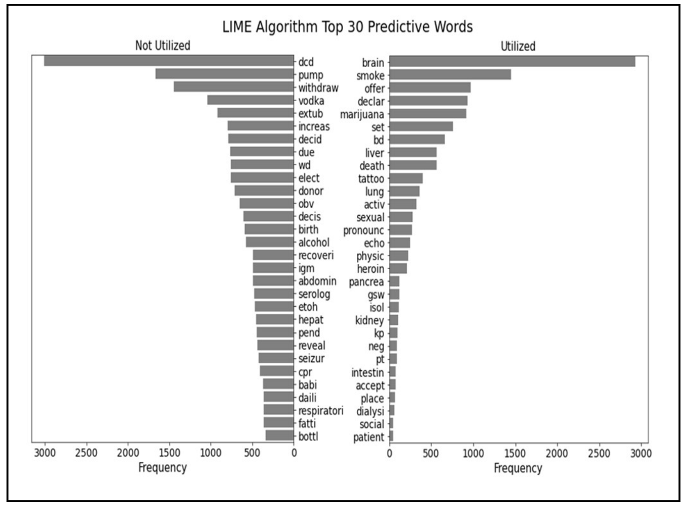

Fig. 3. Top 30 predictive words for the logistic regression model in predicting non-utilized livers (left), and the gradient boosted model in predicting utilized livers (right) based on the LIME algorithm.

C. Mortality Outcome Model Performances

The NLP models for both mortality outcomes (30-day and 1-year) performed poorly. Initial models trained on down sampled data with an even class distribution yielded the highest AUC scores of any models built to predict mortality outcomes. The highest AUC scores of this initial testing were 0.53 and 0.52 for 30-day and 1-year mortality outcomes, respectively. Further testing with more advanced methods did not yield any improvements in AUC score.

VI. Discussion

The results confirm the efficacy of donor narratives in predicting liver utilization but show their limited predictive ability for short term (30-day) and long term (1-year) mortality. The best NLP model trained for predicting liver utilization used a logistic regression model and had a test AUC score of 0.81. Of the models built, the logistic regression model had a specificity of 0.62 and the gradient boosting model had a specificity of 0.97. Both models were able to fairly accurately predict one of the two classes, suggesting that they identified important features of both outcomes. However, the predictions on the same test of the two models had a Spearman's rank correlation coefficient of only 0.74, with model predictions agreeing on only 84% of test observations classified with a 0.5 threshold.

The limited predictability of the mortality outcomes may stem from the fundamental relationship between utilization and mortality: high-risk livers are likely not being recovered from deceased donors. Any donor information which may lead to increased mortality likely was flagged prior to organ recovery from the deceased donor. In further analyses, it may be necessary to include recipient information to build a more accurate mortality model.

The results of the LIME algorithm identified unigrams associated with DCD and brain death as predictive of the unutilized and utilized classes, respectively, aligning with results of other studies. Clinically, DCD leads to warm ischemic damage of organs (including the liver), causing a higher risk of mortality with transplantation [13-14]. Conversely, brain dead donors typically yield higher quality organs [15]. Similarly, "marijuana" and "heroin," among the top predictive words for the utilized liver class, reflect the policies that have been adopted to alleviate the organ shortage. For example, livers infected with hepatitis C—a common consequence of using intravenous drugs, like heroin—may be transplanted to hepatitis C positive recipients [16]. This suggests that the results of the algorithm performed well in identifying predictive unigrams and bigrams.

The LIME algorithm identified unigrams associated with alcohol use as being predictive of unutilized livers. While alcohol abuse is a leading cause of liver cirrhosis, alcohol abuse has not been proven to have a negative effect on post-transplant outcomes [17-18]. These results may indicate potential biases during offer acceptance, stemming from the belief that alcohol abuse negatively impacts liver quality, and further investigation should be done to understand these effects.

Notably, the DRI, which is based on tabular data, does not highlight alcohol use as a significant factor in liver risk assessment [9]. One possible explanation for this is that the DDR form only includes one yes/no question for heavy alcohol use, described as 2+ drinks per day, which is perhaps insufficient information to factor into the assessment [4]. In fact, only four yes/no questions in the DDR form pertain to drug use, which is most likely insufficient to provide a full context of a donor's history of drug abuse. Much like with alcohol use, there is often a perceived increased risk of negative post-transplant outcomes in donors with a history of drug abuse, especially cocaine [19]. Additional data collection on drug usage in the DDR form would generate more informative tabular data and allow researchers to more easily examine the effect of drug use in liver donors without having to examine the unstructured text fields used in this study.

VII. Conclusion

The analysis from this study determined several key areas that need to be focused on when assessing the viability of liver donation. The first was the cause of death. DCD donors were found to be less likely to have their livers utilized. Donors that had suffered brain death, on the other hand, were associated with utilized livers. While this discovery is not new, it does confirm that the methodology from this project was comparable to the existing clinical models.

The other key area of interest was substance abuse. There was a strong association with alcohol abuse and a liver not being utilized, despite inconclusive linkages between alcohol abuse and adverse post-transplantation outcomes. This may indicate a potential bias in the transplantation process that needs to be examined further.

Furthermore, the DDR form provides little space to describe the drug use of the donor, simplifying the data into yes/no fields rather than including fields to indicate specific amounts. With the analysis suggesting associations between certain drugs and liver utilization, donor registration forms should be revised to capture more information about a donor's substance abuse history.

Beyond providing insights into the transplantation process, the project was able to prove the application of a novel NLP based method. Further exploration of the data with different methods could yield even better results. The dataset did not provide enough data to outperform the simpler ML methods, but as more data is accrued, an RNN based approach that uses word embeddings could outperform traditional ML methods that were used in this study.

Acknowledgment

We would like to thank the University of Virginia School of Data Science, and the United Network for Organ Sharing (UNOS) for data access and contribution to this research. In particular, we would like to acknowledge our capstone advisor, Dr. John Van Horn, as well as our the UNOS Data Science team, including Andrew Placona, Kip McCharen, and Brittany Shean. Their expertise, guidance, and support made this study possible.

References

[1] "Cirrhosis," *National Institute of Diabetes and Digestive and Kidney Diseases*. [Online]. Available: https://www.niddk.nih.gov/health-information/liver-disease/cirrhosis. [Accessed: 01-Apr-2022].

[2] H. Mergental, R. W. Laing, A. J. Kirkham, M. T. Perera, Y. L. Boteon, J. Attard, D. Barton, S. Curbishley, M. Wilkhu, D. A. Neil, S. G. Hübscher, P. Muiesan, J. R. Isaac, K. J. Roberts, M. Abradelo, A. Schlegel, J. Ferguson, H. Cilliers, J. Bion, D. H. Adams, C. Morris, P. J. Friend, C. Yap, S. C. Afford, and D. F. Mirza, "Transplantation of discarded livers following viability testing with normothermic machine perfusion," *Nature Communications*, vol. 11, no. 1, Jun. 2020. [Accessed: 07-Apr-2022].

[3] S. Farkas, C. Hackl, and H. J. Schlitt, "Overview of the indications and contraindications for liver transplantation," *Cold Spring Harbor Perspectives in Medicine*, vol. 4, no. 5, 2014.

[4] "Deceased donor registration worksheet - UNOS," *UNOS Data Collection*. [Online]. Available: https://www.unos.org/wp-content/uploads/unos/Deceased_Donor_Registration.pdf. [Accessed: 01-Apr-2022]. [Accessed: 01-Apr-2022].

[5] S. J. Cotler, "Liver transplantation in adults: Deceased donor evaluation and selection," *UpToDate*, 19-Apr-2021. [Online]. Available: https://www.uptodate.com/contents/liver-transplantation-in-adults-deceased-donor-evaluation-and-selection#H17. [Accessed: 01-Apr-2022].

[6] P. Martin, A. DiMartini, S. Feng, R. Brown, and M. Fallon, "Evaluation for liver transplantation in adults: 2013 practice guideline by the American Association for the study of liver diseases and the American Society of Transplantation," *Hepatology*, vol. 59, no. 3, pp. 1144–1165, 2014.

[7] "Updated list of organ refusal reasons will go into effect on Dec. 2," *UNOS*, 23-Nov-2021. [Online]. Available: https://unos.org/news/updated-list-of-organ-refusal-reasons-will-go-into-effect-on-dec-2/. [Accessed: 01-Apr-2022].

[8] "Liver policy: Distribution," *UNOS*, 21-Dec-2021. [Online]. Available: https://unos.org/policy/liver/distribution/. [Accessed: 01-Apr-2022].

[9] S. Feng, N. P. Goodrich, J. L. Bragg-Gresham, D. M. Dykstra, J. D. Punch, M. A. DebRoy, S. M. Greenstein, and R. M. Merion, "Characteristics associated with liver graft failure: The concept of a donor risk index," *American Journal of Transplantation*, vol. 6, no. 4, pp. 783–790, 2006. [Accessed: 01-Apr-2022].

[10] A. Flores and S. K. Asrani, "The donor risk index: A Decade of Experience," *Liver Transplantation*, vol. 23, no. 9, pp. 1216–1225, 2017. [Accessed: 01-Apr-2022].

[11] A. M. Placona, C. Martinez, H. McGehee, B. Carrico, D. K. Klassen, and D. Stewart, "Can donor narratives yield insights? A natural language processing proof of concept to facilitate kidney allocation," *American Journal of Transplantation*, vol. 20, no. 4, pp. 1095–1104, 2019. [Accessed: 01-Apr-2022].

[12] "Organ Procurement and Transplantation Network," *OPTN*. [Online]. Available: https://optn.transplant.hrsa.gov/data/. [Accessed: 07-Apr-2022].

[13] E. S. Orman, A. S. Barritt, S. B. Wheeler, and P. H. Hayashi, "Declining liver utilization for transplantation in the United States and the impact of donation after Cardiac Death," *Liver Transplantation*, vol. 19, no. 1, pp. 59–68, 2012. [Accessed: 01-Apr-2022].

[14] D. Monbaliu, J. Pirenne, and D. Talbot, "Liver transplantation using donation after cardiac death donors," *Journal of Hepatology*, vol. 56, no. 2, pp. 474–485, 2012. [Accessed: 01-Apr-2022].

[15] L. Kumar, "Brain death and care of the organ donor," *Journal of Anaesthesiology Clinical Pharmacology*, vol. 32, no. 2, p. 146, 2016. [Accessed: 01-Apr-2022].

[16] J. M. Jones, I. Kracalik, M. E. Levi, J. S. Bowman, J. J. Berger, D. Bixler, K. Buchacz, A. Moorman, J. T. Brooks, and S. V. Basavaraju, "Assessing solid organ donors and monitoring transplant recipients for human immunodeficiency virus, hepatitis B virus, and hepatitis C virus infection — U.S. Public Health Service Guideline, 2020," *MMWR. Recommendations and Reports*, vol. 69, no. 4, pp. 1–16, 2020. [Accessed: 01-Apr-2022].

[17] R. S. Mangus, C. A. Kubal, J. A. Fridell, J. M. Pena, E. M. Frost, and A. Joseph Tector, "Alcohol abuse in deceased liver donors: Impact on post-transplant outcomes," *Liver International*, vol. 35, no. 1, pp. 171–175, 2014. [Accessed: 01-Apr-2022].

[18] S. S. Arora, K. S. Chahal, P. Axley, D. E. Eckhoff, Y.-F. Kuo, A. K. Singal, R. Wiesner, and S. K. Satapathy, "Heavy alcohol use among deceased donors does not impact recipient outcomes after Liver Transplantation," *Transplantation*, vol. 102, no. 11, pp. 1864–1869, 2018. [Accessed: 01-Apr-2022].

[19] T. Komokata, S. Nishida, S. Ganz, D. M. Levi, T. Fukumori, and A. G. Tzakis, "The impact of donor cocaine use on the outcome of adult liver transplantation," Clinical Transplantation, vol. 20, no. 3, pp. 295–300, 2006. [Accessed: 01-Apr-2022].