Multi-Task Learning for Post-transplant Cause of Death Analysis: A Case Study on Liver Transplant

Sirui Ding[1], Qiaoyu Tan[1], Chia-yuan Chang[1], Na Zou, PhD[1], Kai Zhang, PhD[3], Nathan R. Hoot, MD, PhD[4] Xiaoqian Jiang, PhD[3], Xia Hu, PhD[2]
[1] Texas A&M University, College station, TX, USA; [2]Rice University, Houston, TX, USA; [3]University of Texas Health Science Center, Houston, TX, USA; [4]McGovern Medical School, University of Texas Health Science Center, Houston, TX, USA.

Abstract

Organ transplant is the essential treatment method for some end-stage diseases, such as liver failure. Analyzing the post-transplant cause of death (CoD) after organ transplant provides a powerful tool for clinical decision making, including personalized treatment and organ allocation. However, traditional methods like Model for End-stage Liver Disease (MELD) score and conventional machine learning (ML) methods are limited in CoD analysis due to two major data and model-related challenges. To address this, we propose a novel framework called CoD-MTL leveraging multi-task learning to model the semantic relationships between various CoD prediction tasks jointly. Specifically, we develop a novel tree distillation strategy for multi-task learning, which combines the strength of both the tree model and multi-task learning. Experimental results are presented to show the precise and reliable CoD predictions of our framework. A case study is conducted to demonstrate the clinical importance of our method in the liver transplant.

1 Introduction

Organ transplant is a crucial therapeutic option for individuals with end-stage diseases, e.g., kidney failure [1], liver failure [2], liver cancer [3], etc. However, due to the complex surgical procedures and high risk of graft failure [4], how to allocate organs properly remains an important yet challenging problem. To increase allocation precision and effectiveness, doctors often need to consider a series of post-transplant factors, especially the cause of death (CoD) analysis [5], such as rejection, infection, cancer, and recurrent disease [6]. Accurately predicting and analyzing these CoDs before the transplant can aid doctors in making better clinical decisions regarding organ allocation [7] and precise treatment after the surgery [8]. In this work, we focus on liver transplant as a case study. Currently, the MELD score [9] is widely used as the standard medical indicator to aid doctors in making better clinical decisions. Nevertheless, MELD cannot provide a granular analysis of the aforementioned CoDs factors, since it was originally designed for the 3-month mortality prediction of liver-related diseases. While some statistical methods have been proposed, they are either intended for a limited number of predictors [10] or make strong assumptions about the input features and outcomes, such as linear relations and feature independence [11]. These limitations hinder the accurate prediction of post-transplant CoDs, necessitating the development of more advanced computational methods to support precise clinical decision-making in liver transplant.

Machine learning (ML) has recently received remarkable success in predicting transplant-related medical outcomes [12]. For example, Lau et al. employed neural networks and random forest to predict post-transplant graft failure [13]. Ding et al. developed a prediction framework based on knowledge distillation for the graft status prediction with consideration of fairness issues [14]. Despite their success, the complex nature of liver transplant makes it infeasible to apply previous ML methods directly for post-transplant CoD prediction. We identify two significant challenges from the data and model-related aspects as follows.

First, from a data perspective, a patient usually has multiple CoDs which makes the analysis a multi-label learning task. In addition, recorded CoDs (positive samples) are scarce compared to negative samples, i.e., successful transplantation or unrecorded data. As a result, there is an imbalance problem in the data, making it difficult for machine learning models to accurately predict the positive class [15]. This is because we do not have enough data to learn ML models for different CoD tasks independently. Therefore, it is infeasible to directly apply traditional multi-class learning methods [16] and existing ML methods for organ transplant [12] in the post-transplant CoDs analysis.

Second, from a modeling perspective, tree-based models like GBDT [17, 18] tend to perform better than neural net-
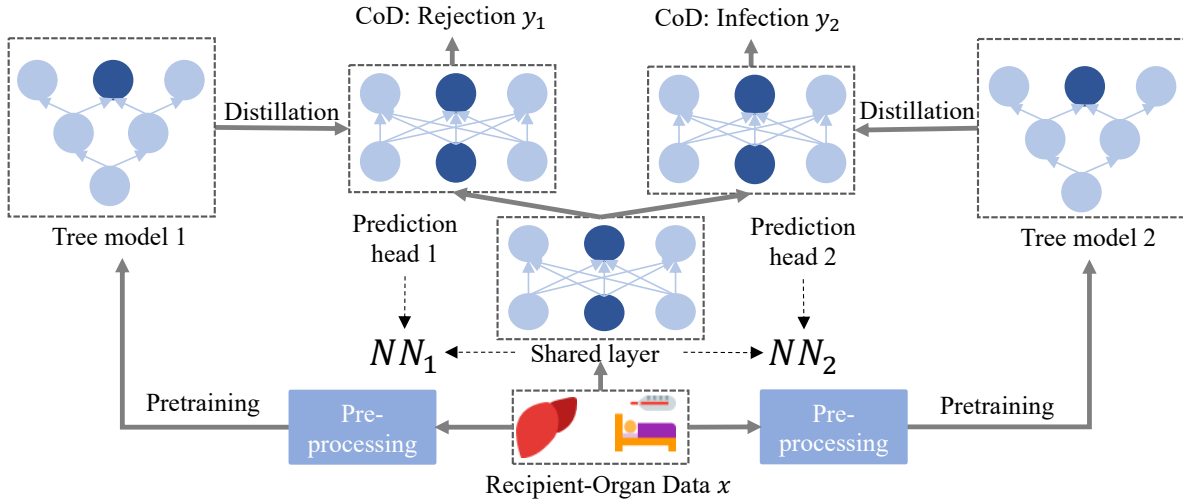
Figure 1: An overview of the CoD-MTL workflow for multiple CoDs prediction.

work (NN) based approaches [19] in the healthcare field, since the majority of organ transplantation records are EHR/tabular data [20]. We also verified this in our preliminary experiments, as shown in Table 1. Despite the relative advantages of tree-based models, they are still limited in tackling our CoD tasks, since they cannot capture the complementary correlations among different CoD tasks (a.k.a. multiple labels) [21]. Thus, there is an urgent need to devise more advanced tree-aware models that can simultaneously handle multiple prediction targets.

To tackle the above challenges, we propose a tree-distillation multitask learning framework, called **CoD-MTL**, for post-transplant CoD analysis. In this paper, we focus on the prediction of rejection and infection since they are the most common post-transplant CoDs [5]. Specifically, for challenge (1), instead of modeling the rejection and infection independently, we develop a multitask learning model [22] with a shared network layer under the CoD-MTL framework to capture their semantic correlations, since they are intrinsically associated with each other in the organ transplantation field. The shared neural networks will take advantage of the various related tasks to alleviate the unbalanced data problem in CoD analysis. For challenge (2), we design a novel tree distillation strategy in CoD-MTL to effectively transfer the advances of tree-based models into neural networks for different CoD tasks. As a result, a principled approach is obtained to integrate the capacity of multitask learning in capturing complementary information across various tasks and the power of tree-based models in modeling tabular data in an end-to-end fashion. We validate the effectiveness of our framework on the real-world liver transplant dataset. Experiment results show the CoD-MTL can accurately predict the post-transplant CoDs. The case study demonstrates the clinical importance of CoD-MTL to help doctors in organ transplant clinical decisions.

2   Data and Problem Description

Data preparation. In this work, we use a patient cohort obtained from the patients registered on the liver transplant waiting list of the Organ Procurement and Transplantation Network (OPTN) [23], consisting of a total of $8,922$ patients who underwent liver transplantation. Out of these patients, $4,160$ died due to rejection (including both acute and chronic rejection), and 3627 died due to infection after the transplant. In addition, we also randomly selected 2000 patients as negative samples who had no documented death after transplantation. In this study, we consider 102 features from both the donor organs and recipients, excluding sensitive attributes such as gender and race. The donor/organ features are divided into three categories: the donor's basic information, the donor's history of diseases, and information on the donor's death. Similarly, the patient/recipient features are categorized as the patient's basic information, history of diseases, and transplant-related laboratory tests.

Problem formulation. We are given a dataset $\{p_i, o_i, Y_i\}_{i=1}^{N}$ consisting of N patient-organ pairs. Each patient p_i (or

organ o_i) is associated with a L_p (or L_o) dimensional feature vector $\mathbf{x}_i^p \in \mathbb{R}^{L_p}$ (or $\mathbf{x}_i^o \in \mathbb{R}^{L_o}$). For each patient-organ pair (p_i, o_i), there are M possible causes of death (CoDs), denoted as $Y_i = \{y_j \in \{0, 1\}\}_{j=1}^{M}$, where $y_j = 1$ if the j-th CoD causes the death of the patient and $y_j = 0$ otherwise. The goal is to train a machine learning model that can predict the probability of each CoD for a given patient-organ pair (p_i, o_i) based on their input features $\mathbf{x}_i = \langle \mathbf{x}_i^p, \mathbf{x}_i^o \rangle$. The model should learn to predict multiple CoDs simultaneously, and the learning objective is to minimize the cross-entropy loss between the predicted probabilities and the ground-truth labels.

3 Methodology

This section will introduce the proposed post-transplant CoD prediction framework (CoD-MTL) in detail. Firstly, we will describe the pre-processing procedure for input data (Section 3.1). Then we introduce the multi-task learning framework for post-transplant CoDs prediction (Section 3.2). Finally, the proposed tree-distillation strategy for multi-task learning will be elaborated (Section 3.3).

3.1 Data pre-processing

To effectively learn from original liver transplant EHR data, we use an encoder to transform categorical features into numerical values following the standard ways of processing raw data. These numerical features are then concatenated with the original numerical features. To address any missing values, we impute all features with zero. This processed data is used as input for both the tree and multi-task learning models in the CoD-MTL framework. Additionally, to ensure a robust evaluation, the data samples are shuffled during the K-fold cross-validation stage.

3.2 Multi-task learning for multiple CoDs prediction

Immunosuppressive drugs that patients take to prevent rejection after liver transplant surgery can weaken their immune system and increase their susceptibility to infections [24]. To investigate the clinical relationship between rejection and infection prediction tasks, we adopt a multi-task learning approach that uses a shared deep learning module and customized prediction heads for different CoDs. The CoD-MTL framework is designed based on the multi-task learning paradigm [25], as illustrated in Figure 1. To predict the j-th CoD, we formulate the output as follows:

$$y_j = \text{Head}_j(\text{SharedLayer}(x_i)), \tag{1}$$

where $\text{Head}_j(\cdot)$ refers to the prediction head part for the j-th task, and $\text{SharedLayer}(\cdot)$ denotes the shared layer of multiple tasks in the multi-task learning pipeline. We will provide further details about the Head and SharedLayer modules in the following subsection.

3.3 Tree-distillation boosted multi-task learning

In this subsection, we will elaborate on the proposed tree-distillation strategy in the multi-task learning framework. Firstly, we will present the process of integrating the tree model into neural networks using knowledge distillation. Next, we will introduce a new approach to integrate the tree models into the multi-task learning framework. Then we will describe the learning process of the whole CoD-MTL framework.

Tree distillation in the neural network. Tree-based models like GBDT have shown great success across various healthcare scenarios and tabular data [26, 27]. Recently, DeepGBM [28] has been developed to combine the merits of GBDT and deep neural networks by distilling the knowledge of GBDT to deep neural networks. Despite its effectiveness, DeepGBM is designed for a single learning task and cannot model the correlations between multiple learning tasks, as shown in CoD analysis. Inspired by this, we propose to upgrade DeepGBM for multi-task learning, i.e., distilling multiple task-specific GBDT models into a multi-task deep neural network. Assume $V^{t,i}$ denotes the sparse leaf index that corresponds to the i-th patient of the training data in the t-th tree of T, we first transform the leaf outputs of one GBDT model T into a dense embedding as below:

Model type	Model	CoD: Rejection		CoD: Infection	
		AUROC	AUPRC	AUROC	AUPRC
Traditional ML (single task)	Logistic Regression	0.551+0.008	0.482+0.005	0.569+0.013	0.471+0.013
	GBDT	0.588+0.008	0.497+0.010	0.611+0.011	0.499+0.014
	Random Forest	0.583±0.016	0.504±0.009	0.608±0.009	0.506±0.020
Neural Networks (single task)	MLP	0.571+0.012	0.493+0.008	0.592+0.003	0.483+0.011
Multitask learning model	Multitask Learning	0.595+0.021	0.517+0.015	0.614+0.019	0.515+0.028
The proposed method	CoD-MTL	0.640+0.012	0.557+0.012	0.646+0.007	0.553+0.018

Table 1: Performance comparison on Two CoD Prediction Tasks

$$\mathbf{E}^i = \text{Emb}(||_{t \in T}(V^{t,i}); \theta), \tag{2}$$

where $\mathbf{E}^i$ represents the dense embedding table obtained from the embedding model $\text{Emb}(\cdot)$ with trainable parameter θ, where $\text{Emb}(\cdot)$ is a fully connected neural network. The notation $||_{t \in T}(V^{t,i})$ indicates the concatenated sparse representation across multiple trees in GBDT. To learn the embedding model, we optimize the objective function as:

$$min \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}'(\mathbf{W} \times \text{Emb}(||_{t \in T}(V^{t,i}); \theta) + \mathbf{b}, q^i), \tag{3}$$

where $\mathbf{W}$ and $\mathbf{b}$ are the parameters that map the dense embedding into the final prediction, and q^i is the corresponding leaf prediction of the i-th sample. The loss function $\mathcal{L}'$ can be chosen as the cross-entropy loss function, which is commonly used in classification tasks.

After the embedding of sparse representations from tree models' leaves, we can use this dense embedding E^i as the distillation target to further distill the tree structures into a neural network. The distilled neural network can approximate the tree model by optimizing the following objective:

$$\mathcal{L}_{distill} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\text{NN}(\mathbf{x}_i^p[\mathbb{I}^T]); \theta_{NN}), \mathbf{E}^i), \tag{4}$$

where $\text{NN}(\cdot)$ represents the distilled neural network with trainable parameters θ_{NN}, and $\mathbf{x}_i^p$ is the input feature for the i-th patient. $\mathbb{I}$ denotes the indices of the features selected from the tree model.

Integration of tree model in multi-task learning. When it comes to predicting multiple post-transplant CoDs, we propose a multi-task tree-distillation paradigm to achieve this. First, we train a GBDT model for each CoD prediction task. For the j-th CoD, we have GBDT model T_j and the distilled network $\text{NN}_j(\cdot)$ with trainable parameters θ_{NN_j}. We then develop the distilled neural network for multiple CoD tasks, as shown in Figure 1. Specifically, the distilled model NN_j for each CoD task includes a shared layer for representation learning and a task-specific prediction head for each CoD task, as shown in Formula 1. The prediction head for the j-th CoD is a simple neural network as follows.

$$y_j(x^i) = \mathbf{W}_j \times NN_j(\mathbf{x}_i[\mathbb{I}^{T_j}]); \theta_{NN_j}) + \mathbf{b}_j, \tag{5}$$

where $\mathbf{W}_j, \mathbf{b}_j$ are associated parameters to transfer the dense embedding to final predictions for the j-th task.

Learning process of CoD-MTL. To train our model, we optimize the parameters of CoD-MTL according to the following multi-task loss function.
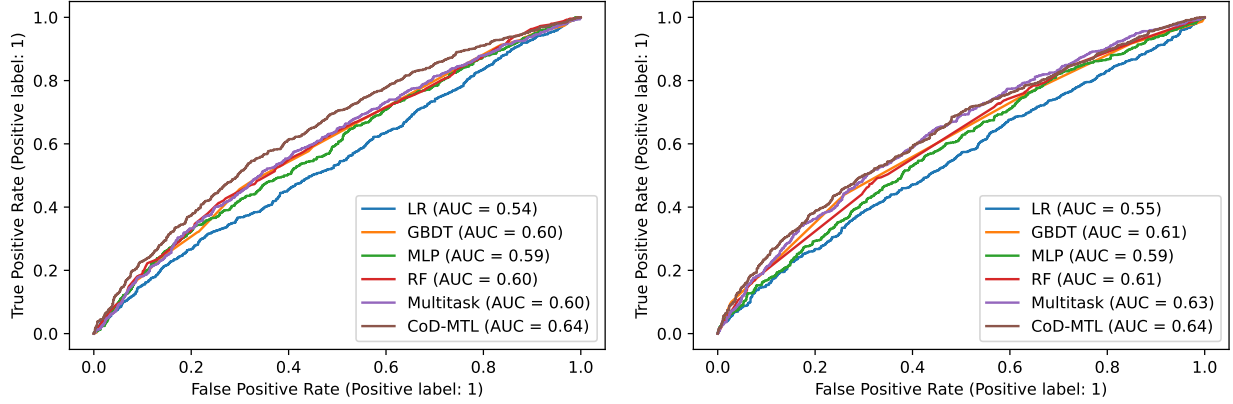
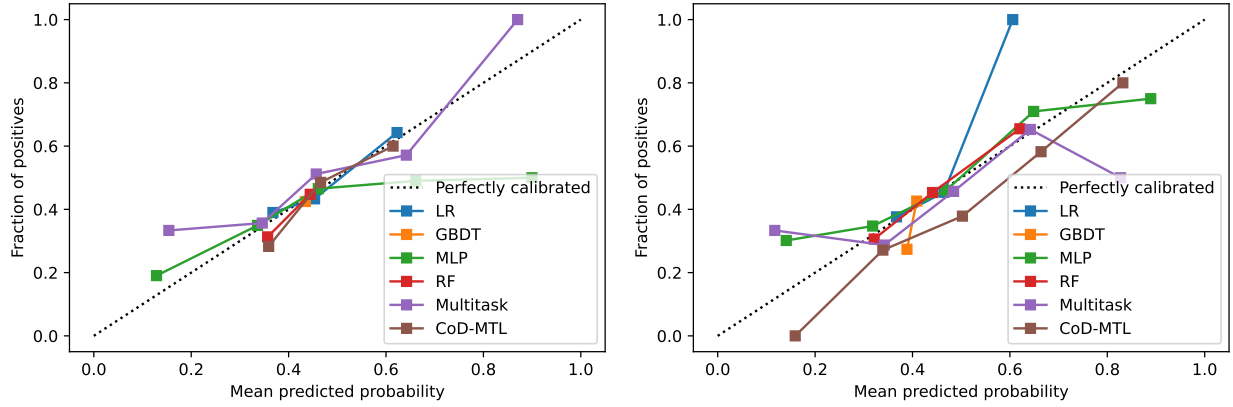Figure 2: ROC curves for rejection and infection CoDs (From left to right).

Figure 3: Calibration curves for rejection and infection CoDs (From left to right)

$$\mathcal{L}_j = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathrm{NN}_j(\mathbf{x}_i[\mathbb{I}^{T_j}]); \theta_{NN_j}), \mathbf{E}_j^i),$$

$$\mathcal{L}_{multi} = \sum_{j=1}^{M} \alpha_j(\beta_j \mathcal{L}'(y_j, y_j') + \gamma_j \mathcal{L}_j). \tag{6}$$

$\mathcal{L}_j$ is the knowledge distillation loss function for the j-th CoD task, i.e, NN_j. $\mathcal{L}_{multi}$ is the overall multi-task loss function for M CoD tasks, where α_j, β_j, and γ_j are trade-off parameters to control the importance of different terms.

4 Experiment

In this section, we provide a comprehensive evaluation of CoD-MTL from the computational and clinical perspectives by answering the following research questions (**RQ**).

- **RQ1:** Can the CoD-MTL accurately predict the rejection and infection as the CoDs? (Section 4.2)

- **RQ2:** To what extent can the Cod-MTL be considered trustworthy for predicting CoDs? (Section 4.3)

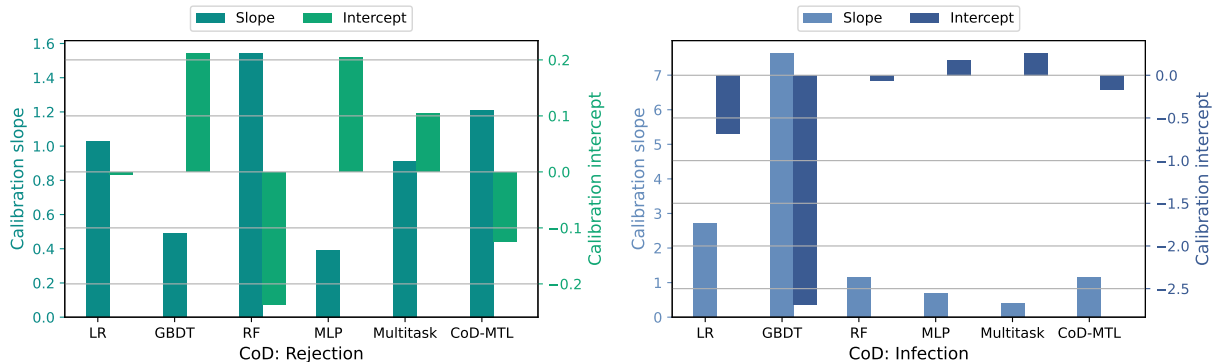- **RQ3:** How could the CoD-MTL help the doctor make the clinical decision in liver transplant? (Section 4.4)

Figure 4: Calibration performance on rejection and infection prediction tasks.

4.1 Experimental settings

Baseline methods. We choose the baseline methods from three categories which are traditional ML, neural network, and multitask learning model respectively. For traditional ML, we select three commonly used methods as baselines, which are Logistic Regression (LR) [29], Gradient Boosting Decision Tree (GBDT), and Random Forest (RF) [30]. For the neural network, we use a multi-layer perceptron (MLP) [31] as the neural network baseline model. For the multitask learning model, we use the hard parameter sharing multitask learning frameworks as the baseline method.

Evaluation metrics. To ensure a fair comparison, we adopt the K-fold cross-validation strategy to evaluate the baseline and proposed methods. The AUROC and AUPRC metrics will be computed by averaging across multiple folds to assess the prediction accuracy. Additionally, we calculate the standard deviation (STD) of AUROC/AUPRC across different folds to evaluate the model uncertainty. To evaluate the clinical significance of CoD-MTL, we will engage a clinical expert to assist us in the case study.

Implementation details. We implemented the baseline machine learning methods using scikit-learn [32] and PyTorch. The CoD-MTL framework was implemented using LightGBM [33] and PyTorch. We trained the CoD-MTL for 100 epochs using AdamW as the optimizer with a learning rate of 0.001. All the experiments were conducted on a server equipped with NVIDIA V100 GPUs and Intel Xeon CPUs. We set K to 4 for cross-validation.

4.2 Prediction performance on rejection and infection as CoDs

We present the superior performance of CoD-MTL compared to the baseline machine learning methods as shown in Table 1. Several observations can be summarized as follows:

Firstly, the tree model outperforms MLP method on the CoD prediction task. For the rejection prediction, we observe that GBDT can achieve higher AUROC and AUPRC by 3.0% and 0.81% compared to MLP. For the infection prediction task, the AUROC and AUPRC of GBDT are higher than MLP by 3.2% and 3.3% respectively. This may be due to the ability of GBDT to identify important features from the EHR and eliminate irrelevant features that are less related.

Secondly, the multitask learning model improves performance compared to the single MLP. For the rejection prediction, the multitask learning baseline outperforms the single MLP by 4.2% and 4.9% on AUROC and AUPRC, respectively. For the infection prediction, the multitask learning baseline achieves 3.7% and 6.6% higher AUROC and AUPRC than the single MLP model. Our findings demonstrate that combining two highly related tasks in the multitask learning model can boost the performance of each single task. The shared model parameters can help the model learn common knowledge for both tasks and make more precise predictions for each CoD of the patients.

Thirdly, our results demonstrate that CoD-MTL outperforms the other baseline methods by a significant margin. Specifically, we observe a maximum improvement of 16.1% and 15.6% in terms of AUROC for rejection and infection prediction, respectively. Similarly, the maximum improvement in AUPRC is 15.6% and 17.4% for rejection and
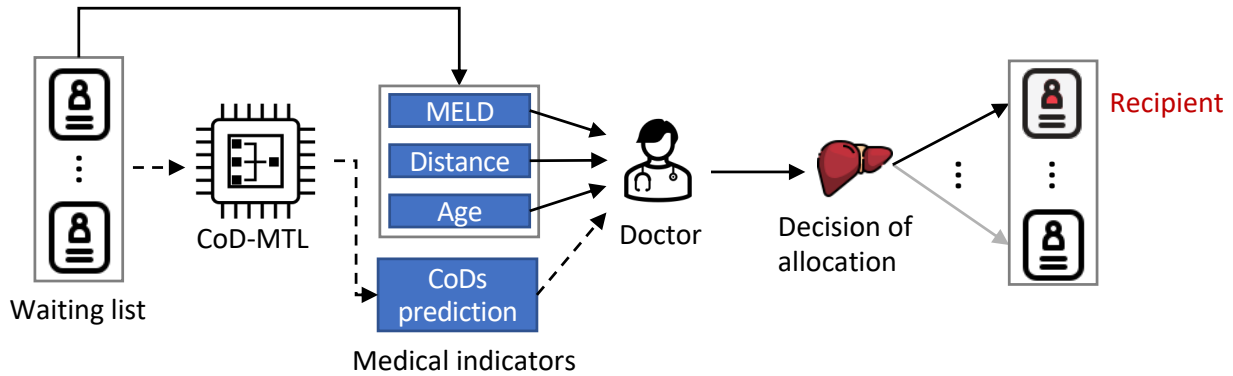
Figure 5: Illustration of how can CoD-MTL help the doctor make the clinical decisions in liver transplant.

infection prediction, respectively. These results provide strong evidence of the effectiveness of CoD-MTL in leveraging the advantages of both tree models and multitask learning. By utilizing highly related features and common knowledge between the two tasks, CoD-MTL achieves superior performance on both CoD prediction tasks.

Moreover, we performed the sensitivity analysis using the ROC curve for a single fold of data. Figure 2 shows the ROC curves for rejection and infection prediction tasks. As seen in the figure, CoD-MTL exhibits a steeper slope than other baseline methods for both tasks. This indicates that CoD-MTL has a higher sensitivity, which is crucial for accurately identifying patients at high risk of rejection and infection. Early detection of rejection and infection is critical for preventing organ failure or loss and timely medical intervention. Therefore, the superior sensitivity of CoD-MTL makes it a promising approach for liver transplant outcome prediction.

4.3 Model uncertainty analysis

Our proposed model shows moderately small uncertainty under cross-validation experiments, which is consistent with reliable machine learning models [34]. We obtain the standard deviation (std) of the AUROC and AUPRC from cross-validation to evaluate the uncertainty. In the rejection prediction task, the CoD-MTL's std of AUROC is relatively small. Although LR and GBDT have smaller std than CoD-MTL, their AUROC is too low for precise prediction, which is 13.9% lower than CoD-MTL. The CoD-MTL's std of AUPRC for rejection is fairly small. While LR reaches the smallest std of AUPRC, the AUPRC of LR is 15.6% lower than CoD-MTL to ensure an accurate prediction. The situation is analogous when it comes to infection prediction. The std of AUROC of CoD-MTL is relatively small, and the only baseline method with a smaller std is MLP. However, MLP's AUROC is 9.1% lower than CoD-MTL. MLP also has the lowest std of AUPRC, but its AUPRC is 14.5% lower than CoD-MTL, indicating the incapability of infection prediction. The small std of our model shows its stable performance across different folds of data, which makes it promising to be generalized well to unseen data and reliably used in real-world practice.

Additionally, the proposed ML model can produce reliable predictions with well-calibrated probabilities, which is crucial for clinical applications [35]. To further investigate the model uncertainty on rejection and infection prediction, we plot the calibration curves on one fold of data, as shown in Figure 3. The calibration curve of CoD-MTL in both the left and right parts of Figure 3 is close to the diagonal line, indicating that the predicted probabilities correspond to the observed fractions well. To quantitatively measure the calibration performance of the models, we calculate the calibration slope and intercept of the calibration curve in Figure 3, as shown in Figure 4. The calibration slope of CoD-MTL is close to 1 on both tasks, indicating that the predicted probabilities are well-calibrated with the true probabilities. Although LR's calibration slope is closer to 1 on the rejection prediction task, it is not well-calibrated on the infection task. Similarly, the proposed model achieves a calibration intercept near 0 on both tasks, indicating that the predicted probabilities are well-centered around the observed fractions. These observations suggest that the proposed CoD-MTL model can output reliable predictions with rather small uncertainty across different tasks, making it a promising tool for organ transplant outcome prediction.
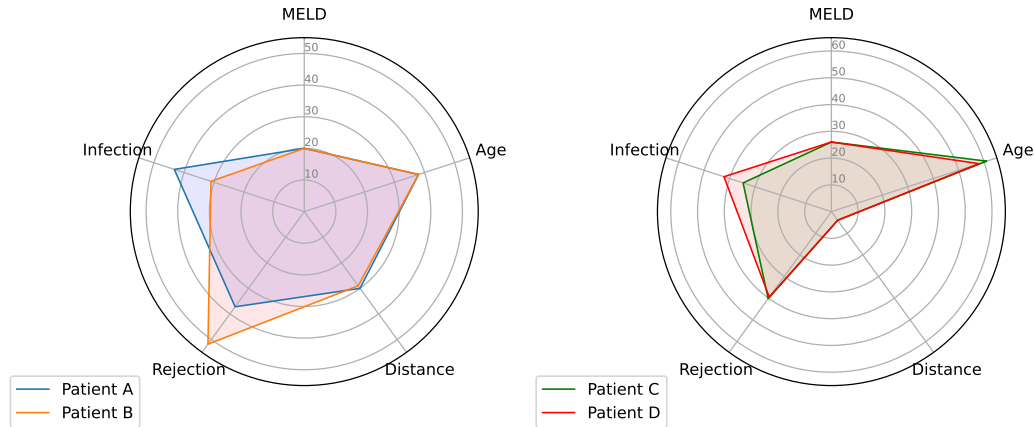
Figure 6: Two pairs of patients with similar features from the same transplant centers.

4.4 Case study

Our proposed model represents a significant improvement in clinical decision support for liver transplantation as shown in Figure 5. It takes into account a variety of factors that can affect patient outcomes, such as the likelihood of rejection or infection, to provide a more nuanced analysis of each patient's individual medical situation.

For instance, in cases where two patients from the same transplant center appear to be very similar, our model may reveal that they have different probabilities of dying from rejection or infection. We present two pairs of patients as shown in Figure 6. Patients A and B come from the same transplant center, and patients C and D come from another transplant center. We can observe that patients A and B have the same MELD score which is 20, the same age which is 38, and nearly the same distance from the donor which is 30 and 29 respectively. Our model predicts patient A with a higher probability of dying from infection and patient B with a higher probability of dying from rejection. The situation is similar for patients C and D, who share very similar characteristics related to allocation. Patient D is predicted with a higher probability of infection.

This kind of detailed analysis can be invaluable for clinicians who are looking to make more informed decisions about patient care [5]. With this level of information, doctors can develop more personalized treatment plans that are tailored to the specific needs of each patient. For example, they may choose to administer more aggressive immunosuppressant therapy to a patient who is at a higher risk of rejection, while opting for a more cautious approach for a patient who is at a lower risk. By providing doctors with more detailed and accurate information about patient outcomes, our model can help to improve the overall efficiency and effectiveness of liver transplantation. This, in turn, can lead to better outcomes for patients and more efficient use of healthcare resources.

5 Discussion of limitation

Additionally, we need to address some limitations and identify corresponding solutions for future improvement of our CoD-MTL framework. A critical constraint of our current model is its lack of interpretability, as the inability to explain predictions may impede its deployment in organ transplant scenarios. To address this gap, we will incorporate explainable AI techniques [36] into our future work to provide a human-understandable interpretation of the predictions. Another limitation is the failure to consider fairness in our current framework. Equity is an ethical goal that clinical decision support systems should aim to achieve [37, 38]. Therefore, our future work will place a strong emphasis on integrating fairness constraints within our prediction framework to ensure it is suitable for multiple outcomes in organ transplants. The proposed framework has the potential to extend beyond our focus on organ transplants and can be applied to other medical fields that use multi-task learning. For instance, the prediction of complications and the length of stay in the ICU [39, 40] may benefit from our CoD-MTL framework, subject to future refinement.

6 Conclusion

In this work, we propose a novel multi-task learning framework named CoD-MTL for the cause of death prediction in organ transplant. The key innovation lies in designing a tree-distillation strategy in multi-task learning, which serves as a bridge to combine the merits of tree-based models and multi-task deep neural networks for more accurate prediction of the transplant EHR data. Empirical results on the liver transplant cohort show the output of CoD-MTL to be accurate and reliable for the precise liver transplant. The clinical case study further demonstrates our framework can be a promising clinical decision support tool for physicians in organ transplantation-related allocation and treatment procedures. We will attach more emphasis on the explainability and fairness of the framework as a future direction.

7 Acknowledgements

References

1. Abecassis M, Bartlett ST, Collins AJ, Davis CL, Delmonico FL, Friedewald JJ, et al. Kidney transplantation as primary therapy for end-stage renal disease: a National Kidney Foundation/Kidney Disease Outcomes Quality Initiative (NKF/KDOQI™) conference. Clinical Journal of the American Society of Nephrology. 2008;3(2):471-80.

2. Kumar R, Anand U, Priyadarshi RN. Liver transplantation in acute liver failure: Dilemmas and challenges. World Journal of Transplantation. 2021;11(6):187.

3. Ravaioli M, Ercolani G, Neri F, Cescon M, Stacchini G, Del Gaudio M, et al. Liver transplantation for hepatic tumors: a systematic review. World Journal of Gastroenterology: WJG. 2014;20(18):5345.

4. Tovikkai C, Charman SC, Praseedom RK, Gimson AE, van der Meulen J. Time-varying impact of comorbidities on mortality after liver transplantation: a national cohort study using linked clinical and administrative data. BMJ open. 2015;5(5):e006971.

5. Watt KD, Pedersen RA, Kremers WK, Heimbach JK, Charlton MR. Evolution of causes and risk factors for mortality post-liver transplant: results of the NIDDK long-term follow-up study. American journal of transplantation. 2010;10(6):1420-7.

6. Moreno R, Berenguer M. Post-liver transplantation medical complications. Annals of hepatology. 2006;5(2):77-85.

7. Volk ML, Goodrich N, Lai JC, Sonnenday C, Shedden K. Decision support for organ offers in liver transplantation. Liver transplantation. 2015;21(6):784-91.

8. Bhat M, Al-Busafi SA, Deschênes M, Ghali P. Care of the liver transplant patient. Canadian Journal of Gastroenterology and Hepatology. 2014;28(4):213-9.

9. Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology. 2003;124(1):91-6.

10. Gong N, Jia C, Huang H, Liu J, Huang X, Wan Q. Predictors of mortality during initial liver transplant hospitalization and investigation of causes of death. Annals of Transplantation. 2020;25:e926020-1.

11. Nitski O, Azhie A, Qazi-Arisar FA, Wang X, Ma S, Lilly L, et al. Long-term mortality risk stratification of liver transplant recipients: real-time application of deep learning algorithms on longitudinal data. The Lancet Digital Health. 2021;3(5):e295-305.

12. Gotlieb N, Azhie A, Sharma D, Spann A, Suo NJ, Tran J, et al. The promise of machine learning applications in solid organ transplantation. NPJ digital medicine. 2022;5(1):89.

13. Lau L, Kankanige Y, Rubinstein B, Jones R, Christophi C, Muralidharan V, et al. Machine-learning algorithms predict graft failure after liver transplantation. Transplantation. 2017;101(4):e125.

14. Ding S, Tang R, Zha D, Zou N, Zhang K, Jiang X, et al. Fairly predicting graft failure in liver transplant for organ assigning. In: AMIA Annual Symposium Proceedings. vol. 2022. American Medical Informatics Association; 2022. p. 415.

15. Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Computing Surveys (CSUR). 2019;52(4):1-36.
16. Li J, Liu Y, Yin R, Zhang H, Ding L, Wang W. Multi-class learning: From theory to algorithm. Advances in Neural Information Processing Systems. 2018;31.
17. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001:1189-232.
18. Sayed M, Riano D, Villar J. Predicting duration of mechanical ventilation in acute respiratory distress syndrome using supervised machine learning. Journal of Clinical Medicine. 2021;10(17):3824.
19. Shwartz-Ziv R, Armon A. Tabular data: Deep learning is not all you need. Information Fusion. 2022;81:84-90.
20. Massie AB, Kuricka L, Segev DL. Big data in organ transplantation: registries and administrative claims. American Journal of Transplantation. 2014;14(8):1723-30.
21. Zhang Z, Jung C. GBDT-MO: gradient-boosted decision trees for multiple outputs. IEEE transactions on neural networks and learning systems. 2020;32(7):3156-67.
22. Zhang Y, Yang Q. A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering. 2021;34(12):5586-609.
23. Kwong A, Kim W, Lake J, Smith J, Schladt D, Skeans M, et al. OPTN/SRTR 2018 annual data report: liver. American Journal of Transplantation. 2020;20:193-299.
24. Hernandez MDP, Martin P, Simkins J. Infectious complications after liver transplantation. Gastroenterology & hepatology. 2015;11(11):741.
25. Caruana R. Multitask learning. Springer; 1998.
26. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, et al. A machine-learning-based prediction method for hypertension outcomes based on medical data. Diagnostics. 2019;9(4):178.
27. Seto H, Oyama A, Kitora S, Toki H, Yamamoto R, Kotoku J, et al. Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data. Scientific Reports. 2022;12(1):15889.
28. Ke G, Xu Z, Zhang J, Bian J, Liu TY. DeepGBM: A deep learning framework distilled by GBDT for online prediction tasks. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2019. p. 384-94.
29. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. the Journal of machine Learning research. 2008;9:1871-4.
30. Breiman L. Random forests. Machine learning. 2001;45(1):5-32.
31. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics. 2017;22(5):1589-604.
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825-30.
33. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems. 2017;30.
34. Baumann D, Baumann K. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. Journal of cheminformatics. 2014;6(1):1-19.
35. Gawlikowski J, Tassi CRN, Ali M, Lee J, Humt M, Feng J, et al. A survey of uncertainty in deep neural networks. arXiv preprint arXiv:210703342. 2021.
36. Tjoa E, Guan C. A survey on explainable artificial intelligence (xai): Toward medical xai. IEEE transactions on neural networks and learning systems. 2020;32(11):4793-813.
37. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Annals of internal medicine. 2018;169(12):866-72.
38. Chang CY, Yuan J, Ding S, Tan Q, Zhang K, Jiang X, et al. Towards Fair Patient-Trial Matching via Patient-Criterion Level Fairness Constraint. arXiv preprint arXiv:230313790. 2023.
39. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Scientific data. 2019;6(1):96.
40. Zhang S, Li H, Tang R, Ding S, Rasmy L, Zhi D, et al. PheME: A deep ensemble framework for improving phenotype prediction from multi-modal data. arXiv preprint arXiv:230310794. 2023.