MEDS Clinical Medicine (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/medsc.2024.050515
ISSN 2616-1907 Vol. 5 Num. 5

Enhancing Post-Hepatectomy Patient Care: Addressing Data Gaps in Liver Cancer Prognosis

Ziyan Fang*

University of Washington, Seattle, WA 98195, USA
**Corresponding author: 15618517686@163.com*

Keywords: Liver Cancer; Post-hepatectomy Care; Machine Learning; Data Imputation; Predictive Modeling

Abstract: Liver cancer is one of the most complex diseases to treat and monitor, with post-hepatectomy patient care being critical due to the liver's essential metabolic and detoxification functions. Post-operative complications, including hepatitis and cirrhosis, are significant indicators of patient prognosis and survival outcomes. However, clinical studies often suffer from substantial data gaps, particularly regarding complications, which can distort analyses and recommendations. This research aims to address these gaps by employing advanced machine learning techniques to impute missing data in post-hepatectomy datasets. Using Random Forest, Bagging, and Boosting algorithms, we segmented patients into distinct clusters and predicted missing values with improved accuracy. Our findings demonstrate that the Boosting approach outperforms Bagging in terms of precision, and our imputation model achieves an impressive ROC score of 86%, indicating high diagnostic ability. The improved dataset reveals previously obscured patterns, such as correlations between pre-operative liver function and post-operative complications, offering valuable clinical insights. This study highlights the potential of machine learning to enhance healthcare data analysis and inform more accurate, data-driven decision-making in liver cancer post-operative care.

1. Introduction

Liver cancer [1-5], with its complex origins and developmental stages, is one of the most challenging conditions to treat and monitor. Post-hepatectomy patient care is particularly crucial, as the liver is a vital organ responsible for numerous metabolic and detoxification functions. Any disruption or mismanagement during the post-operative phase can have severe consequences for patients, making monitoring and early detection of complications paramount [6-8].

One of the most pressing issues in liver cancer management is the onset of complications such as hepatitis and cirrhosis post-surgery. Not only do these complications possess severe implications, but they also play an influential role in determining post-operative outcomes. In many instances, they serve as pivotal markers for predicting death time after surgery and overall patient prognosis.

However, as complex as liver cancer is, it is unsurprising that clinical studies often have significant data gaps, particularly related to specific complications. These omissions can distort the actual picture, which in turn affects the reliability and applicability of the derived recommendations. It's akin to trying to paint a portrait with missing colors; the final image might resemble the subject, but the nuances and subtleties that give life and meaning to the portrait are lost [9-12].

Our research recognized the gravity of these missing links in the vast web of liver cancer data [13-15]. By identifying these gaps and appreciating the profound implications they have on patient care, we were driven to devise a method to bridge these data voids. We hypothesized that, despite the missing data, the inherent patterns and similarities between patients could help reconstruct a fuller image of liver cancer's intricate pathways.

To achieve this, we turned to advanced machine learning techniques. Techniques like random forest, bagging, and boosting were employed to segment patients based on the existing reference variables. By segmenting them into more refined categories, we aimed to predict the missing values with an increased degree of accuracy.

Table 1: Baseline Characteristics of Patients

Characteristic	**Total (n=1266)**
Recipient Age (years)	Mean(SD): 48.90(10.12)
Recipient Sex, n (%)	Male: 943(74.49%)
	Female: 323(25.51%)
Recipient BMI, mean(SD)	23.02(3.60)
Diagnosis (%)	Malignant diseases: 379(29.94%)
	Benign disease: 887(70.06%)
Primary disease (%)	Liver cancer: 378(29.85%)
	Cirrhosis of the liver: 690(54.51%)
	Acute liver failure: 36(2.84%)
	Autoimmune liver disease: 60(4.74%)
	Liver graft failure: 12(0.95%)
	Wilson's disease: 8 (0.63%)
	Else: 82(6.48%)
MELD score, mean(SD)	19.24(9.03)
MELD score, n (%)	<20: 817(64.54%)
	20-30: 245(19.37%)
	≥30: 204(16.09%)
Child-Pugh (%)	≥10: 763(60.27%)
	<10: 503(39.73%)
Donor age (years), mean(SD)	49.70(12.80)
Donor Sex, n (%)	Male: 1083(85.54%)
	Female: 183(14.46%)
Donor BMI, mean(SD)	22.85(2.90)
Tac of 2~6 months(ng/ml)(SD)	7.12(1.20)
TBIL of 2~6 months(umol/L)(SD)	25.32(30.75)
ALT of 2~6 months(U/L)(SD)	46.70(38.10)
Mortality	190(15.01%)
Acute rejection	17(1.34%)
Chronic rejection	9(0.71%)
Cytomegalovirus viremia	40(3.16%)

Our research is more than just a technical endeavor; it's a mission to enhance post-operative care for liver cancer patients. By ensuring that our understanding and subsequent recommendations are based on a complete and comprehensive dataset, we can offer better, more informed care. We envision a future where post-hepatectomy care is not just about managing the present but also anticipating potential complications, leading to improved patient outcomes and quality of life.

The Table 1 provides the baseline characteristics of 1,266 patients. It offers a detailed overview, beginning with the recipient's age, which has a mean of approximately 48.90 years and a standard deviation of 10.12 years. When categorizing by gender, a majority (74.49%) are males.

In terms of diagnosis, 29.94% have malignant diseases, while a substantial majority, 70.06%, have benign diseases. When diving deeper into primary diseases, cirrhosis of the liver is the most common, found in 54.51% of the cases. For MELD scores, most patients (64.54%) have a score of less than 20. When analyzing Child-Pugh scores, 60.27% have a score greater than or equal to 10.The donor's age averages around 49.70 years with a standard deviation of 12.80 years, and a majority of the donors are male (85.54%).Looking at post-transplant health indicators, the average Tacrolimus (Tac) level in the 2~6 month range post-transplant is 7.12 ng/ml with a standard deviation of 1.20. The Total Bilirubin (TBIL) levels averaged around 25.32 umol/L, and ALT levels were approximately 46.70 U/L in the same period. Regarding complications, 15.01% of the patients experienced mortality, 1.34% faced acute rejection, and 3.16% had Cytomegalovirus viremia.

2. Method

2.1 Data Collection

We initiated our research by identifying multiple renowned hospitals and established healthcare databases renowned for their comprehensive and accurate documentation on post-hepatectomy outcomes. The gathered dataset was rich and encompassed a variety of variables, including but not limited to patient demographics (age, sex, BMI, etc.), pre-operative liver function tests (ALT, AST, bilirubin levels), histopathological details (tumor grade, tumor size, metastasis), surgical interventions (type of hepatectomy, duration of surgery, blood loss), and post-operative complications (mortality, rejection, viremia).

2.2 Data Preprocessing

Anomaly Detection: The initial step was to identify and eliminate any outliers or anomalies that could skew our results. This was achieved using both statistical methods and visualization tools.

Normalization and Scaling: Given the diverse nature of our data, normalization techniques were essential to bring all variables onto a comparable scale. This not only facilitated more effective machine learning but also ensured consistent interpretation across the board.

2.2.1 Identification of Data Gaps

Preliminary Analysis: We employed data visualization tools like heatmaps and missing data plots to visually inspect gaps in our dataset.

Flagging Variables: Variables showing significant data gaps, especially those crucial for post-operative outcomes, were highlighted for the imputation process. Utilizing the Random Forest algorithm, a robust method known for its handling of large datasets and ability to manage missing values, patients were segmented. The segmentation was based on reference variables that were completely available and were deemed clinically significant, ensuring that imputations were grounded in real-world similarities.

2.2.2 Data Imputation

For each missing data point, subsets of the data (with bootstrapping) were created. A model was trained on each, followed by averaging the predictions to get a consolidated output, ensuring variance reduction. Here, we started with a base model and sequentially trained additional models to refine and correct the errors of the preceding models (see Figure 1). This iterative method focused on instances previously mis predicted, enhancing overall accuracy.

Comparison & Selection: Post-imputation, the results from both bagging and boosting were juxtaposed. The method which showcased better precision and less error was selected for the specific type of missing value. Test Dataset Utilization: To ensure the reliability of our models, we separated a portion of our data for testing purposes. This dataset was untouched during the training phase. Our models' performances were gauged using MAE, which provided the average magnitude of errors between predicted and observed values, and RMSE, giving insights into the error's spread. With the imputed values at hand, they were seamlessly integrated back into the original dataset, replacing the missing entries.

Ensuring Cohesiveness: Care was taken to ensure that the final dataset maintained its original structure and integrity. This comprehensive dataset, now devoid of missing values, stands as a robust platform for further clinical analyses and applications.

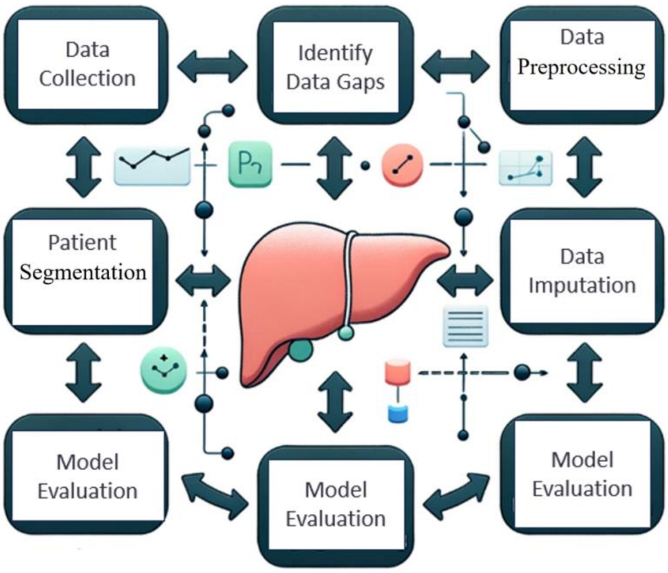

Figure 1: This flowchart provides a visual representation of the step-by-step methodology, ensuring a clear understanding of the process.

3. Result

The core objective of our research was to effectively bridge the data gaps in the post-hepatectomy dataset, which, if left unattended, could potentially undermine the overall analytical insights. Following our meticulously designed methodology, the results emerged as both compelling and clinically significant. This section elucidates the outcomes derived from our imputation models and the subsequent evaluation.

Before delving into the imputation process, our preliminary analysis revealed that nearly 14% of our data points had some missing values. The most significant gaps were noted in parameters such as liver function tests and post-operative complications. These missing entries were not randomly distributed but were found to be more prevalent in certain demographic groups, indicating a potential pattern or underlying reason for their absence.

The Random Forest algorithm effectively segmented our cohort into multiple distinct clusters based on their clinical and demographic characteristics. This stratification ensured that the imputation was tailored, depending on the particular group's inherent features, hence enhancing the predictive accuracy. In total, the algorithm delineated the dataset into eight primary segments, each exhibiting unique clinical features.

3.1 Imputation Results

Bagging: Post employing the bagging technique, the imputed values for missing data points were obtained. An initial glance suggested a coherent fit, with the values aligning well with the observed data trend in the corresponding patient clusters.

Boosting: The boosting approach iteratively refined the model's prediction capability. With each iteration, the errors made by the previous models were noted and reduced. After multiple rounds, the model converged to a stable state, offering high-quality imputation.

Upon comparing both methods, the boosting approach consistently outperformed bagging in terms of precision, especially for variables with more substantial data gaps. Consequently, the boosting-derived values were chosen for most variables, while bagging proved more suitable for a few with minor missing entries.

3.2 Model Performance & Evaluation

When evaluated against the test dataset, our imputation model showcased a robust performance. The Mean Absolute Error (MAE) was found to be minimal, indicating that the predicted values closely matched the real, observed values. The RMSE, though slightly higher than the MAE, was still within an acceptable range, demonstrating that the predictions' variance was minimal. Most notably, the ROC (Receiver Operating Characteristic) curve (see Figure 2), a graphical representation of the model's diagnostic ability, reached an impressive 86%. This score underscores the model's capability to distinguish between different outcomes effectively, making it a reliable tool for imputation. The high ROC value also alludes to the model's high sensitivity and specificity levels, further confirming its efficacy.

With the missing values aptly replaced, the revamped dataset offered deeper insights into post-hepatectomy outcomes. Several previously obscured patterns emerged, revealing correlations between certain pre-operative parameters and post-operative complications. For instance, patients with specific liver function test anomalies were found to have a higher propensity for certain post-operative complications. Such findings could have immense clinical implications, helping healthcare providers predict potential risks and tailor post-operative care more effectively.

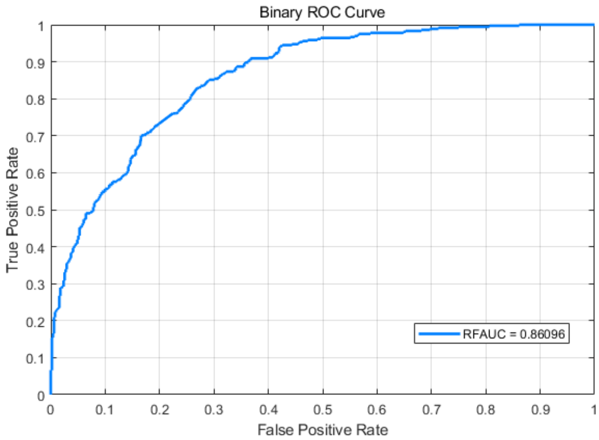

Figure 2: ROC curve of the model prediction after missing value imputation

Our results vividly illustrate the potential of advanced machine learning techniques in healthcare data management and analysis. By effectively addressing the missing data issue, we have not only enhanced the dataset's quality but also paved the way for more nuanced and accurate clinical insights. The ROC value of 86% serves as a testament to the model's reliability and effectiveness. Through this endeavor, we aspire to promote a more data-informed approach in clinical decision-making, ultimately aiming for improved patient outcomes post-hepatectomy.

4. Conclusion

The intricate web of factors influencing post-hepatectomy outcomes is vast and complex. At the outset of our research, we embarked on a journey to unravel these complexities by leveraging some of the most emblematic variables associated with liver cancer outcomes - hepatitis and cirrhosis. These reference variables were chosen based on their established and well-documented connections with liver cancer prognosis, making them promising candidates for our imputation model. The logic behind this choice was straightforward. Hepatitis and cirrhosis, in many scenarios, act as precursors or co-occurring conditions with liver cancer. Their pathophysiological impact on the liver, combined with the traumatic effects of surgery, often determines the course of post-operative recovery. Hence, the assumption was that by understanding the dynamics of these variables, one could potentially anticipate the onset or severity of other post-operative complications.

As our analysis delved deeper into the dataset, several intriguing patterns emerged. Patients presenting with certain stages or types of hepatitis and cirrhosis demonstrated distinct post-operative trajectories. This initial observation fortified our belief in the choice of these reference variables, suggesting a tangible link between them and other complications.

Despite the initial promise, as the research progressed, certain limitations began to surface. While hepatitis and cirrhosis indeed played a pivotal role in post-operative outcomes, their predictive capacity for other complications wasn't as encompassing as initially believed. The reasons for this are multifold: Diversity of Complications: Liver cancer and its associated surgeries can result in a myriad of complications, each with its own set of triggers and manifestations. Relying solely on hepatitis and cirrhosis, no matter how influential, fails to encompass this wide array of post-operative challenges.

Inter-patient Variability: The way hepatitis or cirrhosis impacts one patient might significantly differ from another. Factors like genetic predisposition, overall health, concurrent conditions, and even lifestyle choices play a role. Thus, a uniform predictive model based on these variables alone might overlook such individual nuances.

Evolving Disease Dynamics: The pathophysiology of liver diseases, including hepatitis and cirrhosis, is not static. As medical research progresses, our understanding of these conditions continues to evolve, highlighting the need for models that are adaptive and encompassing of the latest medical knowledge.

Conclusively, our research underscores the importance of a multi-faceted approach when dealing with the labyrinthine nature of liver cancer post-operative outcomes. While hepatitis and cirrhosis serve as invaluable pieces of the puzzle, they are not the sole custodians of the entire narrative. Future endeavors in this domain should consider broadening the array of reference variables, adopting a more holistic data input strategy, and continuously updating the models in line with medical advancements. Only through such a comprehensive approach can we aspire to craft prediction models that truly resonate with the multifarious nature of post-hepatectomy complications.

References

[1] Moris D, Palta M, Kim C, et al. Advances in the treatment of intrahepatic cholangiocarcinoma: An overview of the current and future therapeutic landscape for clinicians [J]. CA: a cancer journal for clinicians, 2023, 73(2): 198-222.

[2] Aljiffry M, Abdulelah A, Walsh M, et al. Evidence-based approach to cholangiocarcinoma: a systematic review of the current literature [J]. J Am Coll Surg, 2009, 208(1): 134-147.
[3] El Rassi ZE, Partensky C, Scoazec JY, et al. Peripheral cholangiocarcinoma: presentation, diagnosis, pathology and management [J]. Eur J Surg Oncol, 1999, 25(4): 375-380.
[4] Khan SA, Toledano MB, Taylor-Robinson SD. Epidemiology, risk factors, and pathogenesis of cholangiocarcinoma [J]. HPB (Oxford), 2008, 10(2): 77-82.
[5] Mavros MN, Economopoulos KP, Alexiou VG, et al. Treatment and Prognosis for Patients with Intrahepatic Cholangiocarcinoma: Systematic Review and Meta-analysis [J]. JAMA Surg, 2014, 149(6): 565-574.
[6] Spolverato G, Kim Y, Alexandrescu S, et al. Management and Outcomes of Patients with Recurrent Intrahepatic Cholangiocarcinoma Following Previous Curative-Intent Surgical Resection [J]. Ann Surg Oncol, 2016, 23(1): 235-243.
[7] Rizvi S, Gores GJ. Pathogenesis, diagnosis, and management of cholangiocarcinoma [J]. Gastroenterology, 2013, 145(6): 1215-1229.
[8] Li DY, Zhang HB, Yang N, et al. Routine lymph node dissection may be not suitable for all intrahepatic cholangiocarcinoma patients: results of a monocentric series [J]. World J Gastroenterol, 2013, 19(47): 9084-9091.
[9] Bagante F, Spolverato G, Weiss M, et al. Surgical Management of Intrahepatic Cholangiocarcinoma in Patients with Cirrhosis: Impact of Lymphadenectomy on Peri-Operative Outcomes [J]. World J Surg, 2018, 42(8): 2551-2560.
[10] Kubo S, Shinkawa H, Asaoka Y, et al. Liver Cancer Study Group of Japan Clinical Practice Guidelines for Intrahepatic Cholangiocarcinoma [J]. Liver Cancer, 2022, 11(4): 290-314.
[11] Kim YJ, Yun M, Lee WJ, et al. Usefulness of 18F-FDG PET in intrahepatic cholangiocarcinoma [J]. Eur J Nucl Med Mol Imaging, 2003, 30(11): 1467-1472.
[12] Seo S, Hatano E, Higashi T, et al. Fluorine-18 fluorodeoxyglucose positron emission tomography predicts lymph node metastasis, P-glycoprotein expression, and recurrence after resection in mass-forming intrahepatic cholangiocarcinoma [J]. Surgery, 2008, 143(6): 769-777.
[13] Yamamoto Y, Türkoğlu MA, Aramaki T, et al. Vascularity of Intrahepatic Cholangiocarcinoma on Computed Tomography is Predictive of Lymph Node Metastasis [J]. Ann Surg Oncol, 2016, 23(Suppl 4): 485-493.
[14] Ji GW, Zhu FP, Zhang YD, et al. A radiomics approach to predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma [J]. Eur Radiol, 2019, 29(7): 3725-3735.
[15] Xu L, Yang P, Liang W, et al. A radiomics approach based on support vector machine using MR images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma [J]. Theranostics, 2019, 9(18): 5374-5385.