Predicting the Survival of the Organ Transplanted Patients

Srikanth Reddy Anugu, Sai Kalyan Nomula, Lakshmi Naga Sameera Sunkara, Haribhakta Gope [1]

[1] University of Dayton, Dayton, OH, USA

This paper presents a comprehensive study on developing a predictive model aimed at forecasting survival outcomes for patients who have undergone graft transplantation. The survival of grafts post-transplantation is critical for patient recovery and long-term health. Accurately predicting these outcomes can significantly aid healthcare providers in optimizing post-operative care, tailoring individualized treatment plans, and improving overall management strategies for transplant recipients. The research leverages advanced machine learning techniques to analyze extensive datasets, providing insights into various factors influencing graft survival. By incorporating patient demographics, medical history, and post-transplantation monitoring data, the model aims to identify patterns and risk factors associated with graft failure. This predictive capability can enable early intervention, potentially reducing complications and improving the quality of life for transplant patients.A key focus of this study is the integration of data from diverse sources to enhance the model's accuracy and reliability. Rigorous validation processes are employed to ensure the robustness of the predictive model. Furthermore, the study emphasizes the importance of maintaining data confidentiality and ethical standards throughout the research process.The outcomes of this research have the potential to transform post-transplant care, offering healthcare providers a valuable tool to predict and mitigate risks, ultimately leading to better transplant success rates and improved patient outcomes.

Predictive analytics | Machine learning | Organ transplant | Survival | UNOS

Introduction

Organ transplantation is a critical medical procedure that can significantly enhance the quality of life and survival rates for patients with end-stage organ failure. Among various types of organ transplants, intestine transplants are particularly important due to the unique and complex nature of intestinal failure, which often leaves patients with few viable treatment options. The survival of transplanted grafts is paramount for patient recovery and long-term health. Accurately predicting these survival outcomes can aid healthcare providers in optimizing post-operative care and improving management strategies for transplant recipients.

The data for this study is sourced from the United Network for Organ Sharing (UNOS), a pivotal organization in the field of organ transplantation in the United States. UNOS plays a critical role in facilitating organ donation and transplantation by managing the national transplant waiting list and ensuring the equitable distribution of organs. The organization's efforts have had a profound impact on society, saving countless lives and advancing the field of transplant medicine. UNOS was established in 1984 and has since been at the forefront of transplant coordination and research. According to recent statistics, UNOS manages over 113,000 patients on the transplant waiting list annually. In 2020 alone, there were over 39,000 organ transplants performed in the United States, a testament to the organization's critical role in addressing organ failure and improving patient outcomes. The rigorous standards and policies set by UNOS ensure that organs are allocated fairly and efficiently, maximizing the chances of successful transplants.

Intestinal transplants are particularly significant due to the complexities associated with intestinal failure, which can arise from conditions such as short bowel syndrome, Crohn's disease, and intestinal atresia. For patients with severe intestinal failure, traditional treatments may be insufficient, making transplantation the most viable option to restore normal digestive function and improve quality of life. Despite advancements in surgical techniques and immunosuppressive therapies, the long-term survival of intestinal grafts remains a formidable challenge. Therefore, developing predictive models to assess the survival outcomes of transplanted grafts is crucial for enhancing patient outcomes and reducing complications.

This project aims to develop a predictive model to forecast the survival outcomes for patients who have undergone graft transplantation using advanced machine learning techniques. By analyzing comprehensive datasets provided by UNOS, the research seeks to identify key factors and patterns associated with graft survival. The ultimate goal is to provide healthcare providers with a reliable tool to predict and optimize the survival of transplanted grafts, thereby improving the overall success rates of intestinal transplants and enhancing the quality of care for transplant recipients.

A. Why is it Important?. The importance of accurately predicting graft survival outcomes in intestine transplantation cannot be overstated. Intestinal failure is a life-threatening condition that severely impacts patients' quality of life, often requiring long-term parenteral nutrition, which carries risks of severe complications, including infections and liver damage. For many patients, an intestinal transplant is the only viable treatment option that offers the potential for a return to a more normal lifestyle.

Predictive models that accurately forecast graft survival can have significant benefits:

- **Improved Patient Care:** By predicting the likely outcomes of graft survival, healthcare providers can tailor post-operative care plans to each patient's specific needs, potentially reducing complications and improving recovery times.

- **Optimized Resource Allocation:** Hospitals and transplant centers can better allocate their resources, ensuring that high-risk patients receive the intensive monitoring and care they need.

- **Informed Decision-Making:** Surgeons and patients can make more informed decisions about transplantation options, understanding the risks and likely outcomes associated with different treatment plans.

- **Enhanced Research and Development:** Insights gained from predictive modeling can drive further research into the factors affecting graft survival, leading to new therapeutic approaches and improved surgical techniques.

Ultimately, the development of a reliable predictive model for graft survival outcomes aims to enhance the overall success rates of intestinal transplants, reduce the burden on healthcare systems, and, most importantly, improve the quality of life and survival rates for transplant recipients.

Problem Statement

The survival of graft transplanted patients is a critical issue in the field of organ transplantation, directly impacting the long-term health and quality of life of recipients. Despite significant advancements in surgical techniques and immunosuppressive therapies, the prediction of graft survival remains a complex and challenging task. This project aims to address this challenge by developing a predictive model to forecast the survival outcomes of graft transplanted patients.

The primary objective is to utilize advanced machine learning techniques to analyze comprehensive datasets from the United Network for Organ Sharing (UNOS). These datasets include detailed information on patient demographics, medical history, and post-transplantation monitoring data. By identifying key factors and patterns associated with graft survival, the model will provide healthcare providers with a powerful tool to predict and optimize the survival of transplanted grafts.

Accurate predictions of graft survival can significantly enhance post-operative care, allowing for individualized treatment plans that reduce complications and improve patient recovery. Additionally, such predictive models can aid in the allocation of medical resources, ensuring that high-risk patients receive the necessary attention and care.

The development of this predictive model involves several critical steps:

- **Data Collection and Preprocessing:** Gathering and preprocessing the data from UNOS to ensure it is suitable for analysis. This includes handling missing values, normalizing data, and selecting relevant features.

- **Feature Selection:** Identifying the most significant variables that influence graft survival to improve the model's accuracy and performance.

- **Model Development:** Employing various machine learning algorithms to develop the predictive model, including logistic regression, decision trees, random forests, and neural networks.

- **Model Evaluation:** Assessing the model's performance using appropriate metrics such as accuracy, precision, recall, and the area under the receiver operating characteristic curve (ROC AUC).

- **Implementation and Validation:** Implementing the model in a real-world clinical setting and validating its predictions with actual patient data to ensure its reliability and effectiveness.

This project is expected to make a significant contribution to the field of organ transplantation by providing a reliable and accurate tool for predicting graft survival. The insights gained from this research will help improve patient outcomes, optimize resource allocation, and guide future research in transplantation medicine.

Dataset

Our study focuses on patients who have undergone intestinal transplantation. To facilitate our analysis, we have access to five key datasets provided by the United Network for Organ Sharing (UNOS), each related to various aspects of intestinal transplant outcomes. These datasets are crucial for developing a comprehensive predictive model for graft survival. Below is a detailed description of each dataset:

B. Intestine Malignancy. The Intestine Malignancy dataset contains detailed information on the occurrences of post-transplant malignancies among recipients. Malignancy is a significant concern in transplant patients due to the immunosuppressive therapy required to prevent graft rejection, which can increase the risk of cancer. This dataset includes:

- **Type of Malignancy:** Specific types of cancer diagnosed in post-transplant patients.

- **Time to Diagnosis:** The duration between transplantation and the diagnosis of malignancy.

- **Treatment Administered:** Details of treatments provided to patients diagnosed with malignancy.

- **Patient Outcomes:** Survival rates and health status post-treatment.

Analyzing this data helps us understand the impact of malignancies on graft survival and overall patient health.

C. PRA (Panel Reactive Antibody) Crossmatch. The PRA Crossmatch dataset includes information on antibody compatibility between the donor and the recipient. Panel Reactive Antibody (PRA) levels are used to assess the likelihood of graft rejection. A higher PRA indicates a higher level of pre-formed antibodies, which can lead to an increased risk of rejection. This dataset provides:

- **PRA Levels:** The percentage of antibodies present in the recipient's blood that react with donor antigens.
- **Crossmatch Results:** The outcomes of tests that determine compatibility between donor and recipient.
- **Rejection Episodes:** Instances of graft rejection and their severity.

This data is critical for evaluating immunological compatibility and its effect on graft survival.

D. Intestine Follow-Up. The Intestine Follow-Up dataset provides longitudinal clinical data on the health status of recipients post-transplant. This dataset includes:

- **Patient Survival:** Information on whether the patient is alive or deceased at follow-up intervals.
- **Graft Function:** Measures of how well the transplanted intestine is performing.
- **Rejection Episodes:** Details of any rejection episodes, including treatment and outcomes.
- **Infections and Complications:** Occurrences of infections and other post-transplant complications.

This data is essential for understanding long-term outcomes and identifying factors that contribute to successful graft survival.

E. Intestine Additional HLA (Human Leukocyte Antigen). The Intestine Additional HLA dataset details the finer points of HLA typing not covered during the initial matching process. HLA compatibility between donor and recipient is crucial for transplant success, as it reduces the risk of rejection and improves graft survival. This dataset includes:

- **Detailed HLA Typing:** Comprehensive HLA data, including alleles and antigen matching.
- **Compatibility Scores:** Metrics that quantify the degree of HLA compatibility between donor and recipient.
- **Impact on Outcomes:** Analysis of how HLA matching affects graft survival and patient health.

This data allows us to perform a more granular analysis of the impact of HLA compatibility on transplant outcomes.

F. Intestine Immunological Follow-Up and Discharge. The Intestine Immunological Follow-Up and Discharge dataset focuses on the immunological response of patients post-transplant and at the time of discharge. It includes:

- **Immunosuppressive Therapy:** Details of the medications and treatment regimens used to prevent rejection.
- **Immunological Markers:** Data on specific markers and indicators of the patient's immune response.
- **Patient Health Status:** Health assessments and outcomes at the time of discharge.

Understanding the immunological response is key to managing post-transplant care and improving graft survival rates.

G. Significance of the Data. The comprehensive nature of these datasets from UNOS allows us to build a robust predictive model for graft survival. By integrating data on malignancies, immunological compatibility, follow-up clinical status, detailed HLA typing, and immunological responses, we can identify critical factors influencing graft outcomes. The insights gained from this analysis will help healthcare providers make informed decisions, tailor post-operative care plans, and ultimately improve the survival and quality of life for intestinal transplant recipients.
Furthermore, this data-driven approach will enhance our understanding of the complexities involved in intestinal transplantation and guide future research in transplantation medicine. By leveraging advanced machine learning techniques, we aim to develop predictive models that can provide accurate and actionable insights, thereby contributing to better patient care and improved transplant success rates.

Our Organ: Intestine

In the context of organ transplantation, the intestine plays a critical and unique role. Intestinal failure is a severe medical condition where the intestines can no longer perform their essential functions of nutrient absorption and digestion. This failure often results from various diseases or conditions, such as short bowel syndrome, Crohn's disease, and intestinal atresia. For patients suffering from these conditions, an intestinal transplant can be a life-saving procedure.

H. Importance of the Intestine. The intestine is vital for maintaining the body's nutritional balance and overall health. Its primary functions include the digestion of food, absorption of nutrients, and the protection of the body from harmful pathogens through its extensive immune system. When the intestine fails, patients may require long-term parenteral nutrition (PN), which is intravenous feeding that bypasses the gastrointestinal tract. While PN can sustain life, it comes with significant risks, including infections, liver damage, and reduced quality of life.

I. Challenges in Intestinal Transplantation. Intestinal transplantation is considered one of the most complex and challenging types of organ transplants. The complexities arise from several factors:

- **High Risk of Rejection:** The intestine has a high density of immune cells, making it particularly prone to rejection compared to other transplanted organs.

- **Infection Susceptibility:** Due to the immunosuppressive therapy required to prevent rejection, intestinal transplant patients are at a heightened risk of infections, which can be severe and life-threatening.

- **Complex Post-Transplant Care:** Post-operative care for intestinal transplant patients is intricate, involving close monitoring for signs of rejection, infection, and other complications.

J. Why Focus on Intestinal Transplantation?. Given the significant challenges and complexities associated with intestinal transplantation, it is crucial to develop predictive models that can accurately forecast the survival outcomes for these patients. The ability to predict graft survival and identify potential complications early can vastly improve patient care and outcomes. Here's why focusing on the intestine is particularly important in this context:

- **Enhancing Patient Outcomes:** Accurate predictions of graft survival can help healthcare providers tailor post-operative care to the specific needs of intestinal transplant recipients, reducing the incidence of complications and improving overall outcomes.

- **Optimizing Resource Allocation:** By identifying patients at higher risk of complications, hospitals can allocate resources more effectively, ensuring that these patients receive the necessary intensive care and monitoring.

- **Improving Quality of Life:** Successful intestinal transplants can significantly improve the quality of life for patients who otherwise would rely on long-term parenteral nutrition, which carries its own risks and complications.

- **Advancing Medical Research:** Insights gained from analyzing intestinal transplant outcomes can drive further research into improving transplant techniques, immunosuppressive therapies, and patient management strategies.

K. Significance in Predictive Modeling. Intestinal transplantation represents a high-stakes area within organ transplant medicine where predictive modeling can have a substantial impact. By leveraging detailed datasets and advanced machine learning techniques, we aim to develop robust models that can provide reliable survival predictions. These models can assist clinicians in making informed decisions, tailoring treatments, and ultimately improving the survival rates and quality of life for intestinal transplant patients.

Data Preprocessing

Effective data preprocessing is crucial for building a robust and accurate predictive model. This section outlines the comprehensive steps taken to preprocess the datasets provided by the United Network for Organ Sharing (UNOS) to ensure data quality and suitability for analysis.

Data Integration. The first step involved merging the five key datasets—Intestine Malignancy, PRA Crossmatch, Intestine Follow-Up, Intestine Additional HLA, and Intestine Immunological Follow-Up and Discharge—using the common identifier `TRR_ID_CODE`. This unique identifier allowed us to combine information from different aspects of intestinal transplant outcomes into a unified dataset, providing a holistic view of each patient's post-transplant journey.

Handling Missing Values. Data completeness is essential for accurate model training and prediction. We employed a multi-step approach to handle missing values in the dataset:

Column Removal. First, columns with more than 70% missing values were removed from the dataset. This threshold was chosen to ensure that the remaining data was sufficiently complete for reliable analysis without losing critical information due to excessive missingness.

MCAR Test. For the remaining columns, we conducted a Missing Completely At Random (MCAR) test to determine the randomness of the missing values. The MCAR test helps identify whether the missing data points are randomly distributed or if there is an underlying pattern that needs to be addressed.

KNN Imputation. After the MCAR test indicated the presence of non-random missing values, we employed the K-Nearest Neighbors (KNN) imputation method to fill in the missing data. KNN imputation replaces missing values by considering the 'k' nearest neighbors in the dataset, using their values to estimate the missing data points. This method is effective in preserving the relationships between variables and maintaining the integrity of the dataset.

Encoding Categorical Variables. To prepare the dataset for machine learning algorithms, it was necessary to convert categorical variables into numerical format. We applied one-hot encoding to the categorical columns, which creates binary columns for each category level, representing the presence or absence of that category. This transformation allows algorithms to process categorical data effectively without introducing bias.

Feature Scaling. Feature scaling was performed to normalize the range of independent variables. This step is critical because machine learning algorithms often perform better when the numerical features are scaled to a similar range. We used standard scaling, which transforms the data to have a mean of zero and a standard deviation of one, ensuring that each feature contributes equally to the model's performance.

Outlier Detection and Treatment. Outliers can skew the results of a predictive model and lead to inaccurate predictions. We conducted outlier detection using statistical methods such as Z-score analysis and the Interquartile Range (IQR). Identified outliers were either removed or treated depending on their impact on the dataset and the overall analysis.

Data Splitting. Finally, the processed dataset was split into training and testing sets. The training set was used to train the machine learning model, while the testing set was reserved for evaluating the model's performance. We maintained an 80:20 split ratio to ensure sufficient data for both training and validation.
By meticulously preprocessing the data, we aimed to enhance the quality and reliability of the predictive model, ensuring robust and accurate survival predictions for graft transplanted patients. The following sections will detail the methodologies and results of the predictive modeling process.

Feature Selection

Feature selection is a crucial step in the development of a predictive model, as it involves identifying and selecting the most relevant variables that contribute to the accuracy and performance of the model. By focusing on these key features, we can improve the model's predictive power and reduce overfitting. This section outlines the comprehensive steps taken for feature selection in our study.

Removal of Direct Death-Related Columns. To ensure that the model's predictions are based on pre-transplant and peri-transplant factors rather than outcomes, we removed columns directly related to the patient's death. These columns, if included, could introduce bias and reduce the model's generalizability. The columns removed include:

- **COD (Cause of Death):** This column indicates the specific cause of the patient's death. Including it would directly influence the survival prediction, as it is an outcome variable rather than a predictor.

- **DTIME (Death Time):** This column records the exact time of the patient's death. Similar to the cause of death, this is an outcome measure and does not serve as a predictor for the model.

Removal of Irrelevant Columns. In addition to death-related columns, we also removed columns that do not add any value to the predictive model. These columns include identifiers and dates that are not informative for the model's purpose. The columns removed include:

- **DATE:** Columns representing specific dates (e.g., transplant date) that do not directly influence survival predictions.

- **ID:** Unique identifiers (e.g., patient ID, transplant ID) that are essential for data management but not useful for prediction.

- **YEAR:** Year columns that do not provide specific information relevant to the patient's medical condition or treatment.

Selection of Relevant Features. After removing irrelevant and outcome-related columns, the next step involved selecting the most relevant features that contribute to the prediction of graft survival. We employed several techniques to identify these key features:

Feature Importance from Models. Using machine learning algorithms such as Random Forest and Gradient Boosting, we evaluated the importance of various features. These algorithms provide feature importance scores, indicating how much each feature contributes to the prediction. Features with higher importance scores were prioritized for inclusion in the model.

Domain Knowledge. Knowledge from the field of transplant medicine was also utilized to select features. Clinically significant variables known to impact graft survival, such as immunosuppressive therapy details, patient demographics, and pre-transplant health status, were included based on their relevance and impact.

Target Variable

In the context of our predictive model for graft transplantation outcomes, the target variable is a critical component as it defines the outcome we aim to predict. For our study, the target variable is `PSTATUS`, which indicates whether the person is alive or dead following the transplantation. This binary variable is essential for training the model to distinguish between successful and unsuccessful transplant outcomes.

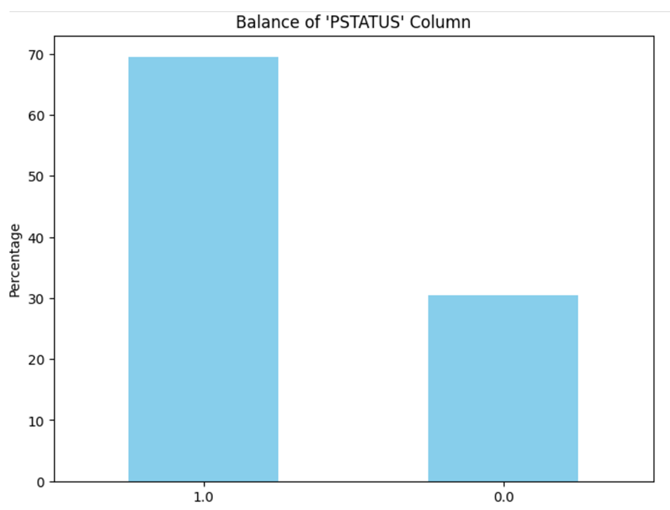

Fig. 1. Imbalanced Target Variable

PSTATUS: Indicator of Survival. The `PSTATUS` variable is binary, with two possible values:

- **1 (Alive):** Indicates that the patient is alive post-transplant.

- **0 (Dead):** Indicates that the patient has passed away post-transplant.

By focusing on this variable, we aim to predict the survival of patients who have undergone graft transplantation, enabling healthcare providers to identify high-risk patients and tailor post-operative care accordingly.

Imbalanced Target Variable. One of the significant challenges we face in predictive modeling is dealing with imbalanced target variables. In our dataset, the `PSTATUS` variable is imbalanced, meaning there are significantly more instances of one class compared to the other. This imbalance can lead to several issues:

- **Model Bias:** The predictive model may become biased towards the majority class, as it will learn to predict the more frequent outcome more accurately at the expense of the minority class.

- **Poor Performance on Minority Class:** The model may perform poorly on the minority class, leading to high false-negative rates, which is critical in medical predictions where identifying at-risk patients is paramount.

Addressing Imbalanced Data. To mitigate the effects of an imbalanced target variable, we implemented several strategies using specific Python packages:

Resampling Techniques.

- **Oversampling the Minority Class:** We used the `SMOTE` (Synthetic Minority Over-sampling Technique) package to generate synthetic samples of the minority class. SMOTE works by creating new synthetic instances that are similar to existing minority instances, effectively balancing the class distribution.

- **Undersampling the Majority Class:** We employed the `RandomSampler` package to randomly select a subset of the majority class. This technique reduces the number of majority class instances to match the size of the minority class, ensuring an equal representation of both classes.

By combining these resampling techniques, we were able to balance the dataset effectively, thus enhancing the model's ability to learn from both classes without bias.

Balanced Target Variable. After applying the resampling techniques, we achieved a balanced target variable distribution. The following figure illustrates the balanced `PSTATUS` variable:

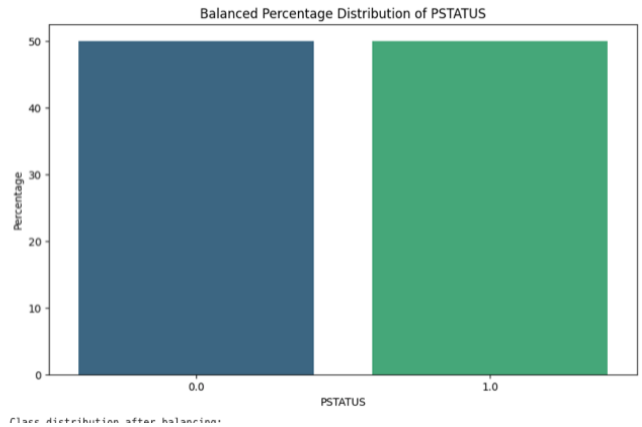

Fig. 2. Balanced Target Variable

This balanced dataset ensures that our predictive model can accurately identify and predict the outcomes for both survival and non-survival cases, leading to more reliable and generalizable results.
By carefully addressing the imbalance in the `PSTATUS` variable, we aim to develop a predictive model that accurately identifies patients at risk, thereby enhancing post-transplant care and improving patient outcomes.

Feature Importance

Feature importance is a crucial aspect of building predictive models, as it helps in identifying which variables significantly influence the model's predictions. Understanding feature importance can lead to better model interpretation, improved performance, and more informed decision-making.

Importance of Feature Importance. By evaluating feature importance, we can:

- **Improve Model Performance:** Identify and focus on the most influential features, leading to more accurate and robust models.
- **Enhance Interpretability:** Understand which features drive the predictions, providing insights into the underlying data patterns and relationships.
- **Reduce Overfitting:** Simplify the model by removing irrelevant or less important features, thus reducing the risk of overfitting.
- **Aid in Feature Selection:** Guide the feature selection process, ensuring that only the most relevant features are included in the final model.

BORUTASHAP: A Unique Approach. For our analysis, we used the BORUTASHAP method to determine the feature importances. BORUTASHAP combines the strengths of the Boruta algorithm and SHapley Additive exPlanations (SHAP) values. Here's why BORUTASHAP is unique and powerful:

- **Boruta Algorithm:** Boruta is a feature selection algorithm that identifies all relevant features by iteratively comparing the importance of real features with that of randomly permuted features (shadows). It is robust and ensures that no relevant feature is overlooked.

- **SHAP Values:** SHAP values provide a unified measure of feature importance by explaining the contribution of each feature to the model's predictions. They offer consistency and interpretability, making it easier to understand the impact of each feature.

By integrating these two methods, BORUTASHAP offers a comprehensive and reliable approach to feature importance analysis, ensuring that all relevant features are identified and their contributions are well understood.

Feature Importance Analysis.

Consistent Proportions with Stratified Splits. To ensure robustness and consistency, we performed feature importance analysis across five stratified splits of the dataset. This approach maintains the same proportion of the target variable (alive or dead) in each split, providing reliable and generalizable results.

Feature Importance Plots. We plotted the feature importances for each split and then averaged them to highlight the significant variables for our model. Below are the average feature importance plot for all five splits:

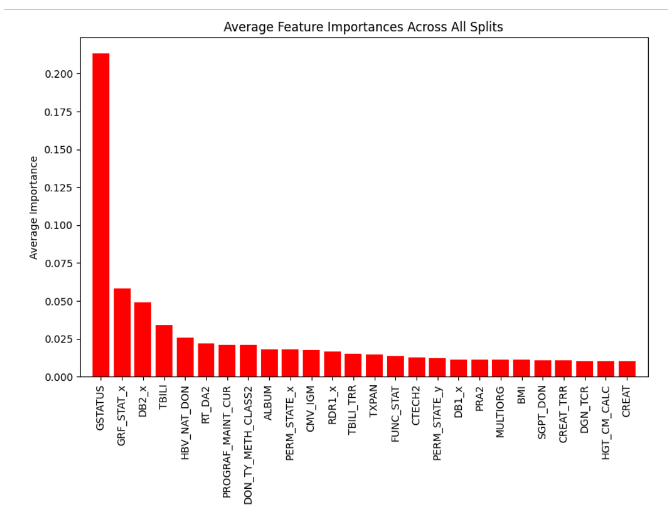

Fig. 3. Average Feature Importances Across All Splits

Key Findings. From our feature importance analysis, we identified several key features that significantly influence the model's predictions. Notably, the **GSTATUS (Graft Status)** feature stands out as the most significant predictor, with an average importance of over 0.20, significantly higher than other features. This indicates its crucial role in the model's predictions and underscores the importance of graft status in determining patient survival.
By leveraging BORUTASHAP, we were able to gain a deep understanding of the feature importances, ensuring that our model focuses on the most relevant variables. This comprehensive analysis not only improves model accuracy but also enhances interpretability, providing valuable insights into the factors that influence graft survival.

Final Features

For the final feature selection, we focused on identifying features with an importance score greater than 0.01. This threshold ensured that only the most relevant and significant features were included in the final dataset. The final dataset contains 27 features and 1 target variable, carefully chosen to enhance the predictive power of our model.

Selected Features. The following table lists the final features used in our analysis, along with their descriptions:

Feature Name	Description
GSTATUS	Graft status of the patient
GRF_STAT_x	Graft function status
DB2_x	Diabetes status (x)
TBILI	Total bilirubin levels
HBV_NAT_DON	Hepatitis B Virus NAT result for the donor
RT_DIA2	Renal transplant diabetes status (2)
PROGRAF_CUR	Current use of Prograf as maintenance
DON_CLASS2	Donor HLA typing method (Class 2)
ALBUM	Albumin levels
PERM_STATE_x	Permanent state (x)
CMV_IGM	Cytomegalovirus IgM status
RD1_x	Renal donor status (x)
TBILI_TRR	Total bilirubin at transplant
TXPAN	Transplant pancreas status
FUNC_STAT	Functional status of the patient
CTEC2E	Cytomegalovirus status (2)
PERM_STATE_y	Permanent state (y)
DB1_x	Diabetes status (1)
PRA2	Panel reactive antibody (2)
MULTIORG	Multi-organ transplant status
BMI	Body Mass Index
SGPT_DON	Serum glutamic-pyruvic transaminase levels in donor
CREAT_TRR	Creatinine levels at transplant
DGN_TCR	Diagnosis at transplant
HGT_CM_CALC	Height in cm (calculated)
CREAT	Creatinine levels
PSTATUS	Patient status (Alive/Dead)

Table 1. Final Features Used in the Analysis

Detailed Summary of Final Features. The final feature selection process was driven by the need to identify the most impactful variables for predicting graft survival. Below is a detailed summary of the selected features and their significance:

- **GSTATUS and GRF_STAT_x:** These features represent the graft status and function status of the patient, which are critical indicators of the transplant's success and the patient's overall health post-transplant.
- **DB2_x and DB1_x:** These variables capture the diabetes status of the patient at different points, as diabetes can significantly impact transplant outcomes and patient recovery.
- **TBILI and TBILI_TRR:** Total bilirubin levels, both at transplant and overall, are important markers of liver function and overall metabolic health.
- **HBV_NAT_DON and SGPT_DON:** These features provide information on the donor's health, particularly regarding Hepatitis B and liver enzyme levels, which can influence graft acceptance and function.

- **RT_DIA2 and RD1_x:** Renal transplant status and donor status are vital for understanding the kidney function and health status of the patient, impacting long-term survival.
- **PROGRAF_CUR:** Current use of Prograf, an immunosuppressive drug, is critical for preventing graft rejection and ensuring patient stability.
- **DON_CLASS2:** This feature indicates the donor HLA typing method, which is crucial for assessing immunological compatibility.
- **ALBUM and CREAT:** Albumin and creatinine levels are essential indicators of kidney and liver function, affecting the patient's metabolic health.
- **PERM_STATE_x and PERM_STATE_y:** Permanent state features provide longitudinal information about the patient's health status, offering insights into their stability over time.
- **CMV_IGM and CTEC2E:** Cytomegalovirus status is important for understanding the patient's immunological challenges post-transplant.
- **TXPAN:** Transplant pancreas status can influence metabolic stability and graft function.
- **FUNC_STAT:** Functional status provides a comprehensive view of the patient's overall health and ability to perform daily activities.
- **PRA2:** Panel reactive antibody status is a key immunological metric, indicating the likelihood of graft rejection.
- **MULTIORG:** Multi-organ transplant status helps assess the complexity and risk associated with the patient's transplant procedure.
- **BMI:** Body Mass Index is a general health indicator, impacting the patient's recovery and long-term health.
- **CREAT_TRR:** Creatinine levels at transplant are crucial for assessing kidney function at the time of the procedure.
- **DGN_TCR:** Diagnosis at transplant provides context for the patient's medical history and underlying conditions.
- **HGT_CM_CALC:** Height, calculated in centimeters, is used for normalizing other health metrics.
- **PSTATUS:** The target variable, indicating whether the patient is alive or dead post-transplant, is the primary outcome we aim to predict.

By selecting these 27 features, we ensure that our model focuses on the most relevant and significant variables, leading to more accurate and reliable predictions. The next sections will detail the methodologies and results of our predictive modeling process, highlighting the impact of these final features on model performance.

Models

In this project, we employed five different machine learning models to predict the survival outcomes of graft transplanted patients. Each model has unique strengths and characteristics that contribute to its predictive power. By using a combination of models, we aimed to leverage the strengths of each to achieve the best possible predictive performance. The models used are Logistic Regression, Random Forest, CatBoost, XGBoost, and LightGBM.

1. Logistic Regression. Logistic Regression is a widely used statistical method for binary classification problems. It models the probability that a given input point belongs to a certain class. The logistic function, also known as the sigmoid function, is used to map predicted values to probabilities. Key characteristics include:

- **Interpretability:** Coefficients can be interpreted as the impact of the corresponding feature on the probability of the target outcome.
- **Simplicity:** Easy to implement and computationally efficient, making it suitable for large datasets.
- **Baseline Model:** Often used as a baseline model to compare with more complex models.

2. Random Forest. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Key characteristics include:

- **Robustness:** Reduces overfitting by averaging multiple trees, leading to improved generalization.
- **Feature Importance:** Provides insights into the importance of different features in the prediction process.
- **Handling Non-Linearity:** Capable of modeling complex non-linear relationships between features and the target variable.

3. CatBoost. CatBoost is a gradient boosting algorithm that is particularly effective with categorical data. It handles categorical features internally, reducing the need for extensive preprocessing. Key characteristics include:

- **Handling Categorical Data:** Efficiently processes categorical features without the need for one-hot encoding.
- **Performance:** Often provides superior performance compared to other boosting algorithms, especially with categorical data.
- **Robustness:** Less prone to overfitting and more stable in terms of predictions.

4. XGBoost. XGBoost (Extreme Gradient Boosting) is an optimized gradient boosting library designed to be highly efficient and flexible. It implements machine learning algorithms under the Gradient Boosting framework. Key characteristics include:

- **Speed and Performance:** Highly efficient in terms of computation and memory usage, often leading to faster training times.
- **Regularization:** Includes L1 and L2 regularization, which helps prevent overfitting and improves model generalization.
- **Flexibility:** Supports a wide range of customizations and hyperparameters, allowing fine-tuning for optimal performance.

5. LightGBM. LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework that uses tree-based learning algorithms. It is designed to be distributed and efficient, making it capable of handling large-scale data. Key characteristics include:

- **Efficiency:** Faster training speed and higher efficiency, with lower memory usage compared to other gradient boosting algorithms.
- **Scalability:** Capable of handling large datasets and high-dimensional data efficiently.
- **Accuracy:** Often provides high accuracy and robust performance, particularly with large and complex datasets.

Implementation

The implementation of our predictive model for graft transplantation outcomes involved several key steps, including data preparation, model selection, cross-validation, and performance evaluation. Below is a detailed explanation of the implementation plan used in this project.

Data Preparation. The first step was to prepare the data by defining the features and the target variable. The dataset was resampled to address the imbalance in the target variable, ensuring that the model would not be biased towards the majority class. The features included various medical and demographic variables, while the target variable was `PSTATUS`, indicating whether the patient was alive or dead post-transplant.

Cross-Validation. To ensure that the model's performance is robust and generalizes well to unseen data, we used k-fold cross-validation. This technique involves splitting the data into k folds and using each fold as a test set while the remaining folds are used for training. In this project, we used 5-fold cross-validation (k=5). This approach helps in assessing the model's performance across different subsets of the data, providing a comprehensive evaluation.

Model Selection. We selected five machine learning models for this analysis, each offering unique strengths:

- **Logistic Regression:** A simple yet effective linear model for binary classification, providing interpretable results and serving as a baseline.
- **Random Forest:** An ensemble learning method that constructs multiple decision trees and aggregates their predictions, known for its robustness and ability to handle non-linear relationships.
- **CatBoost:** A gradient boosting algorithm that handles categorical features internally and often provides superior performance with categorical data.
- **XGBoost:** An optimized gradient boosting library that offers high efficiency and flexibility, known for its regularization capabilities and speed.
- **LightGBM:** A gradient boosting framework designed for efficiency and scalability, capable of handling large datasets with high accuracy.

Evaluation

Evaluation metrics are crucial in assessing the performance of machine learning models. They provide quantitative measures to compare different models and select the best one for the task. For this analysis, we chose the following evaluation metrics: ROC AUC Score, F1 Score, Precision, Recall, and Accuracy. Each metric offers unique insights into the model's performance, particularly in the context of an imbalanced dataset where the target variable (MALIG) has a skewed distribution.

ROC AUC Score. The ROC AUC Score (Receiver Operating Characteristic Area Under the Curve) measures the model's ability to distinguish between classes. It plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The AUC represents the degree of separability achieved by the model.

$$\text{AUC} = \int_0^1 TPR(FPR^{-1}(x))\,dx \qquad \textbf{(1)}$$

where:

- TPR: True Positive Rate (Sensitivity) = $\frac{TP}{TP+FN}$
- FPR: False Positive Rate = $\frac{FP}{FP+TN}$
- TP: True Positives
- FP: False Positives
- TN: True Negatives
- FN: False Negatives

The ROC AUC score ranges from 0 to 1, with a score of 1 indicating perfect classification and 0.5 indicating random guessing.

F1 Score. The F1 Score is the harmonic mean of Precision and Recall. It provides a single metric that balances the trade-off between precision and recall, making it particularly useful for imbalanced datasets where focusing on a single metric can be misleading.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \textbf{(2)}$$

where:

- Precision = $\frac{TP}{TP+FP}$
- Recall = $\frac{TP}{TP+FN}$

The F1 Score ranges from 0 to 1, with a higher score indicating better model performance.

Precision. Precision, also known as the Positive Predictive Value, measures the proportion of positive predictions that are actually correct. It is an important metric when the cost of false positives is high.

$$\text{Precision} = \frac{TP}{TP+FP} \quad \textbf{(3)}$$

A higher precision indicates fewer false positive predictions.

Recall. Recall, also known as Sensitivity or True Positive Rate, measures the proportion of actual positives that are correctly identified by the model. It is crucial when the cost of false negatives is high, as in medical diagnosis scenarios.

$$\text{Recall} = \frac{TP}{TP+FN} \quad \textbf{(4)}$$

A higher recall indicates fewer false negative predictions.

Accuracy. Accuracy measures the proportion of all predictions that are correct. While it provides a broad measure of model performance, it can be misleading for imbalanced datasets, as it does not differentiate between the types of errors (false positives and false negatives).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \textbf{(5)}$$

In summary, each of these metrics provides different insights into the model's performance. The ROC AUC score helps assess the model's ability to distinguish between classes, the F1 Score balances precision and recall, precision measures the accuracy of positive predictions, recall assesses the model's ability to identify all positive instances, and accuracy provides an overall correctness measure. By considering all these metrics, we can comprehensively evaluate the models and select the best one for predicting malignancy in patients.

Results

The results of our predictive model for graft transplantation outcomes are summarized in the table below. We evaluated the performance of five different machine learning models using a variety of metrics to ensure a comprehensive assessment.

Metric	Logistic	Random Forest	CatBoost	XGBoost	LightGBM
ROC AUC	0.86	0.86	0.89	0.88	0.87
F1	0.87	0.88	0.90	0.89	0.88
Precision	0.89	0.86	0.92	0.88	0.90
Recall	0.86	0.90	0.90	0.90	0.86
Accuracy	0.86	0.87	0.88	0.88	0.87

Table 2. Model Performance Comparison

Explanation of Results. The performance metrics used to evaluate the models include ROC AUC, F1 score, Precision, Recall, and Accuracy. Each metric provides unique insights into the model's performance:

- **ROC AUC (Receiver Operating Characteristic Area Under the Curve):** This metric evaluates the model's ability to distinguish between positive and negative classes. A higher ROC AUC score indicates better performance. CatBoost achieved the highest ROC AUC score of 0.89, indicating its superior ability to differentiate between the survival outcomes of patients.
- **F1 Score:** The F1 score is the harmonic mean of precision and recall, providing a balance between the two. CatBoost achieved the highest F1 score of 0.90, suggesting it performs well in both precision and recall.
- **Precision:** Precision measures the accuracy of positive predictions. CatBoost achieved the highest precision of 0.92, indicating it had fewer false positives compared to the other models.
- **Recall:** Recall measures the ability to capture all positive instances. Random Forest, CatBoost, and XGBoost all achieved a high recall score of 0.90, indicating their effectiveness in identifying positive cases.
- **Accuracy:** Accuracy measures the proportion of correctly predicted instances among the total instances. CatBoost and XGBoost both achieved an accuracy of 0.88, demonstrating their overall effectiveness in predicting the survival outcomes.

Interpretation of Results. The results indicate that all five models performed well, with CatBoost consistently achieving the highest scores across multiple metrics. This suggests that CatBoost is particularly well-suited for this predictive task, likely due to its ability to handle categorical features effectively and its robustness against overfitting.

- **Logistic Regression:** While Logistic Regression provided a good baseline performance with an ROC AUC of 0.86 and an accuracy of 0.86, it was outperformed by the more complex models.
- **Random Forest:** Random Forest showed strong performance, particularly in recall, with a score of 0.90, indicating its ability to capture positive cases effectively.

- **CatBoost:** CatBoost stood out as the best-performing model with the highest scores in ROC AUC, F1, and Precision, making it the most reliable model for predicting graft survival outcomes.

- **XGBoost:** XGBoost also demonstrated strong performance with an ROC AUC of 0.88 and high scores across other metrics, indicating its robustness and efficiency.

- **LightGBM:** LightGBM performed well with an ROC AUC of 0.87 and balanced scores across metrics, showcasing its ability to handle large datasets efficiently.

Previous Works

In comparing our work with previous studies, we refer to the study titled "A Multi-Step Precision Pathway for Predicting Allograft Survival in Heterogeneous Cohorts of Kidney Transplant Recipients" by Zhang et al. This study developed and validated the P-Cube model for predicting graft survival, demonstrating a high concordance index (C-index) across various cohorts [1].

Comparison of Results. The study by Zhang et al. used classical regression methods to predict graft survival and achieved an accuracy of 80%. In contrast, our analysis employed five different machine learning models, each evaluated using multiple performance metrics including ROC AUC, F1 score, Precision, Recall, and Accuracy.

Key Observations.

- **Diverse Model Approaches:** Unlike the classical regression methods used by Zhang et al., our analysis utilized five different machine learning models, each contributing unique strengths to the predictive process. This multi-model approach enhances the robustness and reliability of our predictions.

- **Higher Performance Metrics:** All our models achieved performance metrics greater than 80% across various evaluations. Notably, the CatBoost model achieved the highest ROC AUC score of 0.89, indicating superior predictive capability compared to the 80

- **Comprehensive Evaluation:** Our evaluation included a comprehensive set of metrics such as ROC AUC, F1 score, Precision, Recall, and Accuracy. This holistic evaluation provides a more detailed understanding of model performance and its effectiveness in predicting graft survival outcomes.

Conclusion and Future Work

Conclusion. In this study, we developed and validated a predictive model for graft survival outcomes in patients who have undergone intestinal transplantation. Utilizing data provided by the United Network for Organ Sharing (UNOS), we employed five machine learning models: Logistic Regression, Random Forest, CatBoost, XGBoost, and LightGBM. Each model was evaluated using multiple performance metrics, including ROC AUC, F1 score, Precision, Recall, and Accuracy, to ensure a comprehensive assessment.
The CatBoost model emerged as the best-performing model, achieving the highest scores across several metrics:

- **ROC AUC:** CatBoost achieved the highest ROC AUC score of 0.89, indicating its superior ability to distinguish between positive and negative classes.

- **F1 Score:** With an F1 score of 0.90, CatBoost demonstrated a balanced trade-off between precision and recall.

- **Precision:** The model achieved a precision score of 0.92, reflecting its accuracy in identifying positive cases.

- **Recall:** A recall score of 0.90 indicates the model's effectiveness in capturing all positive instances.

- **Accuracy:** CatBoost achieved an overall accuracy of 0.88, confirming its robustness and reliability.

These results highlight the effectiveness of using advanced machine learning models for predicting graft survival outcomes. The comprehensive evaluation metrics provide valuable insights into the strengths of each model, with CatBoost standing out as the most reliable for this specific predictive task.

Future Work. While the current study has demonstrated significant advancements in predicting graft survival outcomes, there are several areas for future research and improvement:

- **Integration of Additional Data Sources:** Future research could integrate additional data sources, such as genomic data, lifestyle factors, and detailed clinical histories, to enhance the predictive power of the models.

- **Real-Time Predictive Analytics:** Developing a real-time predictive analytics tool that can be used by healthcare providers during patient consultations could significantly improve decision-making and patient outcomes.

- **Longitudinal Studies:** Conducting longitudinal studies to track patients over a more extended period post-transplantation could provide deeper insights into the factors influencing long-term graft survival.

- **Model Interpretability:** Enhancing the interpretability of the models, particularly for complex algorithms like CatBoost and XGBoost, can help clinicians understand the reasoning behind predictions and increase trust in the models.

- **Personalized Medicine:** Developing personalized predictive models tailored to individual patient characteristics could further improve accuracy and patient care.
- **Clinical Trials:** Conducting clinical trials to validate the predictive models in real-world settings and understand their impact on clinical practice.
- **Ethical and Bias Considerations:** Addressing ethical considerations and potential biases in the data and models to ensure fairness and equity in predictive analytics.

The findings from this study provide a solid foundation for future research aimed at improving graft survival predictions. By continuing to refine and expand these predictive models, we can contribute to better clinical outcomes and enhance the overall quality of care for patients undergoing graft transplantation.

References

[1] Zhang, Y., Deng, D., Muller, S., Wong, G., Yang, J. Y. H. (2020). A Multi-Step Precision Pathway for Predicting Allograft Survival in Heterogeneous Cohorts of Kidney Transplant Recipients. *Frontiers in Medicine*. https://www.frontiersin.org/articles/10.3389/fmed.2020.00604/full