CA4021 BSc Data Science Formal Project Proposal

Prediction of Kidney Transplant Graft Survival using a Cluster-Then-Predict Framework

Aoife McDaid

Student ID: 19369973

aoife.mcdaid3@mail.dcu.ie

Diarmuid Brady

Student ID: 19479166

diarmuid.brady35@mail.dcu.ie

January 14, 2023

Supervisor: Tomas Ward

Kidney Transplantation is a life-changing procedure that increase a patient's lifespan for decades. Acute rejection is the primary risk factor and incurs significant costs on human life and the healthcare system. Kidney donations are in short supply, so each one must be allocated carefully. Machines learning models have been developed to predict graft survival to aid decision making in clinical practice. To date, they have not outperformed traditional statistical methods and have generally suffered from insufficient evaluation. We propose a novel strategy using a cluster-then-predict framework by utilising a large dataset of over 2500 kidney transplants containing donor and recipient characteristics and transplant factors. To support our methodology, we provide an overview of the kidney transplant domain and investigate the state-of-the-art technologies for imputation, clustering, prediction and evaluation. In our methodology, we decide to use consensus clustering along with tree-based models and gradient-boosting machines using traditional statistical methods as a baseline. For cluster analysis, we will work with domain experts to interpret and phenotype each cluster using t-SNE visualisations and relevant internal evaluation indices. For prediction, we will evaluate performance using discrimination, calibration and computational requirements.

In this research proposal, firstly, we provide the background and motivation behind carrying out this research and define a problem statement. Secondly, we perform a literature review and detail the current state-of-the-art (SOTA). Thirdly, we elucidate our proposed methodology. Additionally, we outline a plan using a Gantt chart. Finally, we summarise the key points made during the proposal.

1 Motivation and Background

Kidneys serve a vital role in our bodies performing several functions that involve filtering waste products from the blood, maintaining electrolyte balance, regulating blood pressure, and producing hormones [1]. Toxins and excess fluids build up in the body if the kidneys do not filter waste products from the blood and disrupt the chemical balance in the blood [2].

A measure of kidney function (ability to filter waste) is the estimated Glomerular Filtration Rate (eGFR). Clinicians use eGFR to test early warning signs for Chronic Kidney Disease. eGFR is measured on a scale between 0 - 120. Above 90, is considered normal, 60 - 90 is early-stage kidney disease, 15 - 60 is kidney disease, and below 15 is kidney failure or End-Stage Kidney Disease (ESKD) [3].

Patients with kidney failure must undergo renal replacement therapy (RRT). Dialysis and kidney transplantation are the two most common forms of RRT. Dialysis is a suboptimal solution as it cannot ameliorate kidney function and is associated with poor life quality and life expectancy [4]. Instead, a patient should seek out an allograft kidney transplant by enrolling on the transplant waiting list. An allograft kidney transplant is a surgery performed to replace a non-functioning kidney with a healthy kidney from a donor [5]. The donor can either be living or deceased. Living donors are usually close familial relatives like a brother, sister, son, or daughter, and deceased donors may have experienced brain death, the loss of brain function like dying on a ventilator or cardiac death, the loss of breathing and circulation, like a traumatic accident [6].

Kidney transplantation can significantly improve life quality and increase life expectancy by around 15 - 20 years on average [7]. Only 139 deceased allograft kidney transplants were performed in Ireland in 2021 despite a waiting list of 512 patients and a waiting time of 25 months on average [8].

Kidney transplants are highly valuable, but supply cannot meet demand, and effectively allocating each kidney is paramount. The severe transplant risks are hyperacute (minutes after the surgery), acute (within a year post-surgery), and chronic (after a year post-surgery) graft rejection [9]. A graft failure indicates that giving the allograft kidney to someone else would result in improved lifespan and reduced costs on screening and the procedure.

Among the factors that influence graft failure are the donor and recipient age, gender mismatch, comorbidities (diabetes, hypertension, obesity), time on dialysis (before transplant), cause of kidney failure, prior kidney transplantation, cold ischemia time, human leukocyte antigen (HLA) mismatches, eGFR, and proteinuria. Cold ischemia time measures the time a kidney transplant is stored on ice before transplantation after being removed from the donor. HLA mismatches occur when there are differences in the HLA proteins between a donor and recipient. These proteins play a role in the immune system and can cause graft rejection if there are too many mismatches. Proteinuria is the presence of excess proteins in urine which is a sign of kidney. All factors, except eGFR, are associated with an increased risk of graft rejection [10]–[12].

Polygenic risk scores are tools that are used to predict the likelihood of an individual experiencing graft loss based on their genetic makeup. Multiple genetic variants (single nucleotide polymorphisms or SNPs), rather than any single genetic mutation, are calculated to create the risk scores [13].

Kidney allocation systems developed around the world leverage statistical methods and risk scores that aid clinicians in decision-making [14]. The inclusion of post-transplant factors is a topic of uncertainty. Past research has focused solely on information available at the time of transplant because these can be used in the kidney allocation process [15]. Kidney donor risk index (KDRI) and estimated post-transplant survival (EPTS) score are two risk scores used by the Organ Procurement and Transplantation Network (OPTN) systems and have been externally validated in other countries [16]–[19]. Kaplan-Meier (KM) estimator and Cox proportional hazards (Cox P-H) model are two

statistical models used along with the risk scores for kidney patient stratification. These statistical methods are highly interpretable and are valuable in clinical practice as they rely on a reduced set of variables [15].

Outside of clinical practice, research in machine learning has attempted to capture complex relationships to improve performance but these models have not outperformed the traditional statistical methods [Ruchot et al]. Furthermore, machine learning models are associated with higher computational requirements in terms and execution time and memory usage, along with a higher risk of bias due to poor evaluation [20]–[22].

2 Problem Statement

Kidney transplants are a life-changing operation that significantly improve patient outcomes, but the supply of kidney donations does not meet the demand. The expected waiting time for a deceased donor kidney transplant is approximately 2.5 years. Each kidney is precious and irreplaceable, and patient prioritisation is a challenging process. To achieve optimal results, extensive screening must be carried out to ensure strong compatibility between the donor and recipient. Donor and recipient age, sex, comorbidities, histocompatibility (HLA mismatching), prior kidney transplant, and time on dialysis impact graft survival. All factors must be available and accurate at the time of kidney transplantation. A machine-learning model must be developed using a large dataset with over 2500 kidney transplants containing these features to predict death-censored graft failure at five years. The machine learning model must build upon previous work by devising a novel strategy for the problem. Model results must be reproducible meaning missing data handling, data wrangling, and model performance including hyperparameter tuning must all be reported. Model performance including computational requirements, execution time and memory usage are important factors to consider alongside relevant evaluation metrics.

3 Literature Review - State-Of-The-Art

Kidney transplantation is a lifesaving procedure for patients with end-stage renal disease, but graft survival remains a formidable challenge. Many factors influence the success of a transplant, including patient characteristics, donor factors, and the surgical procedure itself. Accurately predicting graft survival outcomes is important for identifying high-risk cases, optimizing patient selection and management, and improving long-term outcomes.

Over the past decades, various machine learning and statistical methods have been applied to predict kidney graft survival. These approaches include traditional statistical models, such as Cox regression and survival analysis, and more recent developments in machine learning, such as tree-based models, gradient-boosting machines, and neural networks. The results are promising, but machine learning models have not differentiated themselves from traditional statistical methods, in terms of accuracy, reliability, and overall performance.

The latest research has explored the use of consensus clustering to phenotype kidney transplants into distinct, interpretable groups with strong, differential survival outcomes [23]–[26]. These clusters can provide valuable insight to clinicians at the time of kidney donor offer and allocation. The variety of approaches presents the possibility for improvement in model performance. Unfortunately, as highlighted by several systematic reviews, multivariable prediction models often exhibit poor experiment design and evaluation [Collins et al, Wynants et al, Senanayake et al]. Unreported calibration is a key shortcoming in many of these papers. For these models to be considered in clinical practice, they must adhere to higher standards of reporting and documenting model development, hyperparameter tuning and performance.

In this section, we aim to explore the use of advanced machine learning and statistical techniques for graft survival prediction. Firstly, we examine the latest imputation methods. Secondly, we investigate the state-of-the-art (SOTA) clustering techniques. Thirdly, we review the current highest-performing machine learning models. Finally, we assess the up-to-date evaluation metrics used for each task.

Barry et al. outlined a methodology for performing a state-of-the-art (SOTA) literature review involving six steps [paper]. We implemented this methodology and began by defining our initial research question as "what is the current SOTA for kidney transplant survival prediction?" For our search strategy, we entered the query "kidney" AND "transplant" AND "survival" AND "prediction" into the PubMed search engine [27].

3.1 Imputation Methods

The quality and quantity of data are highly impactful factors in the success of a data science project. One strategy for handling missing data is a complete case analysis, which involves removing all observations with null values. This is not recommended if missing data proportions are greater than 5%. Missing data imputation is a whole area of research that implements statistical and machine learning methods to maximize the value of a dataset.

Multiple Imputations using Chained Equations is a statistical method used to handle missing data. It imputes values by analyzing the relationships between variables and making a prediction. Li et al used this method to handle incomplete traffic accident data and produced impressive results [28].

There has been a surge in machine learning-based imputation methods since 2016 [29]. K-Nearest Neighbours (K-NN) Imputer is an imputation method that leverages the K-NN model to predict missing data for both continuous and categorical data [30]. Juna et al used K-NN Imputer effectively to aid the prediction of water quality resulting in an accuracy of 99& [31].

3.2 Advanced Clustering Techniques

There has been a growing increase in research focusing on phenotyping recipients and donors of kidney transplants. A person's phenotype refers to the set of observable characteristics resulting from the interaction of their genetic makeup with the environment [32].

Consensus clustering is a method for combining multiple clustering results on the same data set into a single consensus clustering. It is a way to improve the stability and robustness of a clustering result by aggregating the results of multiple clustering. Consensus clustering algorithms work by generating multiple clusterings of the data using different clustering algorithms or different parameter settings and then combining these clusterings into a single consensus clustering. There are several ways to combine multiple clusters into a consensus cluster, including voting methods and probabilistic approaches. Consensus clustering is often used in situations where it is difficult to determine the correct number of clusters or the appropriate clustering algorithm for a given data set. It can also be used to improve the reliability of a clustering result by aggregating the results of multiple clusters, which can be particularly useful when the data is noisy or has a complex structure [33].

Thongprayoon et al. published three papers in 2022 focusing on discovering the distinct phenotypes of kidney recipients for their respective groups; limited functional status recipients, black recipients, and morbidly obese recipients all using consensus clustering [23]–[25].

Jalakam demonstrated how consensus clustering using Latent Class Analysis (LCA) and using KAMILA (KAy-means for MIxed LArge data) and Gaussian Mixture Models as base clusters produced superior results to using either base clustering method alone. Jalakam demonstrated how these three methods were superior to other popular clustering methods such as k-prototypes, k-modes, and Majority Voting. LCA is a statistical technique and a form of clustering analysis that works with categorical data and is used to identify subgroups, or latent classes, within a population that shares similar characteristics or traits [34]. KAMILA is a semi-parametric unsupervised clustering technique that is based on k-means that caters to mixed-type data and balances the contribution of categorical and continuous variables [35]. Gaussian Mixture Model (GMM) is a statistical algorithm for clustering, it is a probabilistic approach for grouping similar data points together by modelling the data as a mixture of multiple Gaussian distributions, each distribution represents a cluster, the mean and covariance of each component define the shape and location of the cluster. The Expectation-Maximization (EM) algorithm is commonly used to estimate the GMM parameters and assign each data point to a cluster [36]. He leveraged LCA to phenotype over 25,000 transplant patients producing three clusters with strong and differential clinical characteristics and associations with graft survival outcomes. Recipient and donor age, end-stage renal disease cause (diabetes, hypertension, and glomerulonephritis), and type of event (death, graft failure) are among the characteristics found to have statistically significant differences across each cluster.

3.3 Predictive Models

Multiple studies have shown [18] the effectiveness of using the Cox proportional hazards model for investigating the relationship between the kidney transplant outcome and one or more predictor variables. Multiple studies have found that models such as random survival forests and artificial neural networks have a better model performance [37]. Random survival forests (RSF) is a machine learning ensemble method used for survival data and artificial neural networks (ANN) is a machine learning method that uses activation functions to deliver outputs from inputs received. Three artificial neural network models - DeepSurv, DeepHit, and recurrent neural networks (RNN) - were compared to the Cox proportional hazards model. Using the C-index metric, they found that Deepsurv, DeepHit, and RNN had better discrimination (0.650, 0.659 and 0.659 respectively) than the Cox model (0.646). DeepSurv is a Cox proportional hazards deep neural network survival method that has been used for a personalized treatment recommender system. It successfully models relationships between patient characteristics (age etc.) and their risk of failure [38]. Another study [39] compared the predictions from Cox proportional hazards model with artificial neural networks for 5-year graft survival. They found that ANNs were more accurate and sensitive than the Cox regression model. The predictive accuracy was 88% for ANNs and 72% for the Cox model. Decision trees have been used previously to predict graft survival in kidney transplant recipients [40]. A decision tree is a hierarchical machine learning model which is used to make predictions based on how certain 'questions' are answered. Comparing the survival decision tree model to the conventional decision tree model showed that the survival decision tree model had a better c-index of 0.80 while

the conventional decision tree model had a c-index of 0.71.

Goel MK, Khanna P, Kishore J [41] state that the Kaplien-Meier estimate is one the most effective models to be used to measure the survival time of subjects after treatment. It calculates the probability that a patient will survive up to time t. When used with the log-rank test it can compare survival between groups [42]. However, they cannot be used for multivariate analysis (analyzing data involving more than one type of measurement). Cox proportional hazards model can be used for this purpose. Both approaches use parametric and non-parametric statistical techniques (e.g., dependent and independent variables) [43].

A study consisting of kidney transplant recipients from Europe and the United States obtained accurate discrimination with a C index of 0.8 [44]. The study found eight prognostic factors to be independently associated with allograft failure. These factors were made up of functional (e.g eGFR), histological (e.g. Interstitial fibrosis), and immunological (e.g. HLA) factors and were combined into a risk prediction score for kidney allograft failure called iBox.

Research has discovered that longer waiting times on dialysis negatively impact post-transplant graft and patient survival [45]. It has been shown that waiting 6 - 12 months for kidney transplantation increases the mortality rate after transplantation by 21% while waiting over 48 months results in a 72% increase in mortality risk. Using Cox proportional hazard model it has been discovered that multiple variables are associated with 1-year mortality post-transplant [44]. These variables include increasing age and history of select pre-transplant medical comorbidities e.g. history of diabetes.

Kidney Transplant Morbidity Index (KTMI) calculates the impact that a patient having one or more diseases diagnosed pretransplant has on the transplant outcome. Studies have shown that the risk of kidney transplant failure increases as the KTMI scores increase. This tells us that diseases such as heart disease, high blood pressure, etc. are independent predictors of higher graft failure [46].

3.4 Evaluation

In this project, imputation, clustering, and prediction are three tasks that will be evaluated. Multivariable Imputation of Chained Equations (MICE) is the SOTA for imputing missing data in the preprocessing stages of machine learning tasks. Evaluating the quality of imputation involves comparing the distribution of a given variable before and after imputation and determining whether there are statistically significant differences using the Kolmogorov-Smirnov (KS) test. The KS test involves calculating the p-value for the difference between two distributions. If the p-value is less than 0.05 then there is a significant difference and the imputation method has not accurately imputed the missing data [paper].

For clustering analysis, there are several state-of-the-art approaches to evaluating cluster quality. A high-quality cluster is characterized by its ability to simultaneously group observations most similar to each other and separate them from all other observations. To do this When working with mixed data types, a unifying distance measure is required to integrate both numerical and categorical data before calculating any cluster evaluation metrics. J. C. Gower proposed such a distance in his seminal paper "A General Coefficient of Similarity and its Properties". The similarity measure, namely Gower's Distance, considers numerical, binary, and multiple categorical data when constructing the distance matrix. Furthermore, internal evaluation indices are calculated using this distance matrix. The Calinski-Harabraz index measures both the compactness and separateness of clusters.

For visualizing the clusters, t-SNE (t-Distributed Stochastic Neighbor Embedding) is a very popular and common method. t-SNE is a machine learning algorithm that is used for dimensionality reduction. It is particularly well-suited for visualizing high-dimensional datasets because it maps the data points to a low-dimensional space in a way that preserves the structure of the original data. This makes it easier to understand and interpret the data, as it reduces the complexity of the data by condensing it into fewer dimensions. t-SNE works by minimizing the divergence between the distribution of the high-dimensional data and the distribution of the low-dimensional data. This is achieved by minimizing the Kullback-Leibler divergence, which is a measure of the difference between two probability distributions. The t-SNE algorithm has several hyperparameters that can

be adjusted to influence the way that the data is embedded in the low-dimensional space. One of the main benefits of t-SNE is that it can reveal patterns in the data that are not immediately visible when the data is visualized in higher dimensions. It is often used in fields such as data visualization, machine learning, and natural language processing [47]. Jalakam used t-SNE to evaluate cluster compactness and separateness [26].

Previous studies have evaluated predictive models by looking at the discrimination and calibration [18], [37]. It is harmful if these scores are not evaluated [48]. The discrimination of the model refers to the model's ability to correctly distinguish the outcomes [49]. The calibration of a model refers to the agreement between the predictions made by the model and the observed outcomes [50].

4 Methodology

We propose to carry out our project in the following steps:

- Data Cleaning and Processing
- Data Modelling - Clustering and Phenotyping
- Data Modelling - Prediction
- Data Insights - Key Learnings

For each step, we will be working on a High-Performance Cluster from King's College London. We will work on Jupyter notebooks using python and commit our work to Gitlab. In Python, we will use the pandas library for data cleaning and manipulation, seaborn and matplotlib libraries for data visualization, and the sci-kit-learn library for imputation, clustering, prediction, and evaluation. Our proposed consensus clustering algorithm KAMILA is only available in R. Therefore, for this step we will work in an R environment on Jupyter Lab and then return to working in a python environment.

4.1 Datasets

Our data consists of genotype and phenotype data relating to both donor and recipient. The genotype has been used to create five polygenic risk scores which are discussed below. The phenotype data is stored within files containing different datasets. Each dataset contains different variables relating to the donor and recipients including a unique id for each patient. For example, the iGene file contains variables describing the donor and recipient's age, sex, date of transplant, HLA mismatch between the donor and recipient, etc. The UKIRTC0 file contains variables describing the date of graft failure, acute kidney rejection at 3 and 12 months, dialysis status at the time of transplant, etc. The UKIRTC1 file contains data relating to the graft survival times in days, and serum creatinine levels at 3 months, 1 year, 2 years, 3 years, 4 years, and 5 years post-transplant. Serum creatinine measures the amount of creatinine in the blood. This is a small indication of the type of variables that are stored in the datasets we are investigating.

4.2 Data Cleaning and Processing

During data cleaning, we will assess the quality of the data and define a set of rules to abide by to ensure we maintain a high standard when carrying out each phase of the data science pipeline. Key steps involve:

- Variable name alignment: we will be following the python PEP8 style guide meaning that variable names will be lowercase and underscore separated (variable_name) and DataFrame column names will be camel case (ColumnName). All column names must refer to the same measure and share the same variable name.
- Conversion of data types: in the dataset, there will be object, numeric and date-time data types, and each variable will be set as the correct data type.
- Tidy up each data frame: Wickham et al coined the term tidy data as a dataset in the appropriate form to perform machine learning techniques on. Each row represents an observation and each column represents a variable.
- Missing data handling: If all missing data is below 5%, we will use a complete case analysis where any null rows will be dropped. Based on initial data inspection, there is missing data in key variables like HLA mismatches and cold ischemia time. Therefore, rows with greater than 20% missing data will be dropped, and rows with less than 20% will be imputed. We will implement K-NN Imputer
- Feature transformation: with the aid of domain experts, we will create new features. We will group common diseases, conditions, and disorders to generate more differential features.

4.3 Data Modelling - Clustering

We will break down clustering analysis into separate steps

- Cluster generation
- Cluster evaluation
- Cluster interpretation

4.3.1 Cluster Generation

We will implement the KAMILA clustering algorithm when producing the base clusters. Using the labels from the base clusters we will generate the consensus clusters. Consensus clustering will be executed using the Latent Class Analysis.

4.3.2 Cluster Evaluation

We will evaluate the base and consensus clusters using internal evaluation indices and dimensionality reduction visualizations. The Calinski-Harabasz (CH) index will be used as an internal evaluation index. This is an increasing score, the higher the better. t-SNE plots will be used to visually inspect the clusters based on their compactness and separation.

4.3.3 Cluster Interpretation

We will work alongside domain experts to evaluate the cluster with the highest CH index. By leveraging their domain expertise, we gain valuable insight that can drive further feature selection and model selection for the prediction task.

4.4 Feature Importance and Feature Selection

Before running the predictive models on our data we will implement feature importance and feature selection. Feature importance will give us a better understanding of our data and help reduce dimensionality. To carry out feature importance analysis we will use the Gini index to improve our model's performance. We will then carry out feature selection on our data. To do this we will use a wrapper method such as the Backward Elimination approach. This approach removes the least important feature each turn which improves the model's performance.

4.5 Feature Engineering

We plan to include Polygenic risk scores (PRS) in our clustering algorithm. Polygenic risk scores estimate the effect of many genetic variants on an individual's disease status. These scores have been pre-calculated using each patient's genomic data found in the datasets and as a result, five PRS have been produced.

4.6 Predictive Models

Once we have evaluated the clusters, we plan to run separate prediction models on each of these clusters produced. We plan to assess many different models on each of the clusters to obtain the most accurate prediction scores. After researching models best suited for the prediction of this type [22], we plan to run and evaluate the following models on our clusters; Random Survival Forest, XGBoost, and Bayesian Belief networks. XGBoost is a gradient-boosted decision tree machine learning model and the Bayesian Belief network is a probabilistic graphical model that represents dependencies between random variables [51], [52]. We plan to use Harrell's Concordance Index to assess the models and the Receiver Operating characteristic (ROC) curve to compare the performances.

Another approach we plan on looking into is running a predictive model on the data as a whole. This will differ from the previous method as instead of running different predictive models on each

cluster produced, we will run a predictive model on the data as a whole. Researchers [53] have previously used the Cox Proportional Hazards model for this approach however we plan to use machine learning models to do this. We plan to investigate multiple models as mentioned above; XGBoost, Random Survival Forest, and Bayesian Belief networks. We will use the ROC curve analysis for this comparison and Harrell's Concordance index and integrated Brier score for the evaluation of the models.

4.7 Predictive Evaluation Metrics

We plan to compare the two approaches; running a predictive model on the data as a whole and running multiple predictive models on the clusters. We will compare the discrimination using the ROC curve analysis which is a graph that is constructed by plotting the true positive rate against the false positive rate [54] and by calculating and comparing the Harrell's Concordance Index for both approaches. We will finally look at the calibration of the predictive models. A graphical assessment of calibration is built with predictions on the x-axis and the outcome on the y-axis. Ideal model predictions should be on the 45-degree line [55]. We will evaluate the calibration between the two approaches by looking at the different plots. We plan to further evaluate the models using the Brier skill score which reflects the skill gained in the model compared with a model with no skill [54]

To further evaluate our approaches we will compare them to a baseline produced using the Cox proportional hazards statistical model by looking at each model's discrimination and calibration.

5 Project Plan

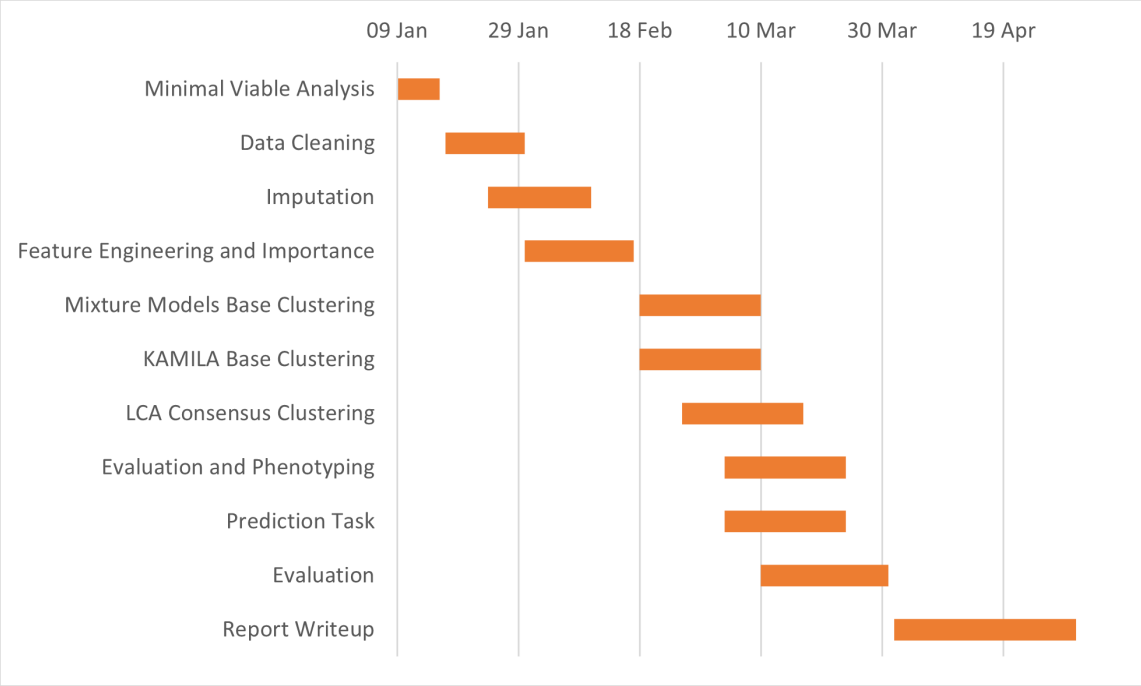

Figure 1: Gantt Chart

Figure 1 shows our Gantt Chart which demonstrates our breakdown of tasks and the timeline for those tasks to be completed.

As you can see we have divided our project into eleven main tasks which are to be carried out from 09/01/23 to the project's due date. The order of these tasks is vital for the smooth running of our project. We will begin by carrying out a Minimal Viable Analysis approach. This entails having a complete understanding of the project as a whole and gathering insights to enhance our understanding. We will then carry out cleaning on our datasets. This is a vital task that will result in suitable data for our machine-learning models. As you can see above, we plan to have this task completed by the end of January. After beginning cleaning our data we will begin the imputation stage and following this task we will carry out feature engineering and importance. These three tasks can be carried out with a slight overlap as they are involved in the preparation of our data. These tasks must be completed before executing the tasks following.

Next, we will carry out clustering models on our data. The chart above illustrates the three clustering tasks to be completed. First, the base clusters will be created using Mixture Models and KAMILA. These will be carried out within the same time frame between 18/02/23 and 10/03/23. After creating the base clusters, we will begin consensus clustering using LCA. We will then evaluate these clusters and phenotyping.

During the cluster evaluation and phenotyping stage, we will begin running our predictive models on these clusters and the data as a whole. By the end of March, we plan to have carried out our evaluation of these predictive models. Finally, our timeline shows that once this evaluation is complete, 1/04/23, we will begin our report writeup which we will continue to work on until the project deadline.

References

[1] NKF, *National Kidney Foundation - Kidney Guide*. [Online]. Available: https://www.nhs.uk/Livewell/Kidneyhealth/Documents/kidney%20guide.pdf.

[2] M. Clinic, *Acute Rejection - Symptoms and Causes*. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/kidney-failure/symptoms-causes/syc-20369048.

[3] NKF, *What is the Difference Between sCr, eGFR, ACR, and BUN?* en-US, Mar. 2018. [Online]. Available: https://www.kidney.org/newsletter/what-difference-between-scr-egfr-acr-and-bun (visited on 11/21/2022).

[4] NHS, *Dialysis*, Sep. 2021. [Online]. Available: https://www.nhs.uk/conditions/dialysis/.

[5] J. Hopkins, *Kidney Transplant*, en, Aug. 2021. [Online]. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/kidney-transplant (visited on 01/08/2023).

[6] N. Miller, "Cardiac death vs. brain death," en-US, *Nursing made Incredibly Easy*, vol. 13, no. 2, p. 44, Apr. 2015, ISSN: 1544-5186. DOI: 10.1097/01.NME.0000460361.65798.bc. [Online]. Available: https://journals.lww.com/nursingmadeincrediblyeasy/fulltext/2015/03000/cardiac_death_vs__brain_death.10.aspx (visited on 01/08/2023).

[7] Beaumont, *Guide to Kidney Transplantation*, 2022. [Online]. Available: http://www.beaumont.ie/kidneycentre-forpatients-kidneytransplant.

[8] HSE, *Organ Donation Transplant Ireland Annual Report*, Dec. 2021. [Online]. Available: https://www.hse.ie/eng/about/who/acute-hospitals-division/organ-donation-transplant-ireland/publications/odti-annual-report-2021-final-14-53.pdf.

[9] Beaumont, *Transplant Complications*, 2022. [Online]. Available: http://www.beaumont.ie/kidneycentre-forpatients-aguide-complications.

[10] G. Naderi, A. Azadfar, S. R. Yahyazadeh, F. Khatami, and S. M. K. Aghamir, "Impact of the donor-recipient gender matching on the graft survival from live donors," *BMC Nephrology*, vol. 21, no. 1, p. 5, Jan. 2020, ISSN: 1471-2369. DOI: 10.1186/s12882-019-1670-x. [Online]. Available: https://doi.org/10.1186/s12882-019-1670-x (visited on 01/07/2023).

[11] C. Legendre, G. Canaud, and F. Martinez, "Factors influencing long-term outcome after kidney transplantation," eng, *Transplant International: Official Journal of the European Society for Organ Transplantation*, vol. 27, no. 1, pp. 19–27, Jan. 2014, ISSN: 1432-2277. DOI: 10.1111/tri.12217.

[12] J. Y. C. Y. Yang and M. M. Sarwal, *Transplant genetics and genomics — Nature Reviews Genetics*, Mar. 2017. [Online]. Available: https://www.nature.com/articles/nrg.2017.12 (visited on 11/14/2022).

[13] *Polygenic Risk Scores — CDC*, en-us, Sep. 2022. [Online]. Available: https://www.cdc.gov/genomics/disease/polygenic.htm (visited on 01/08/2023).

[14] OPTN, *OPTN Policies Effective as of Jan 5 2023 [eGFR]*, Jan. 2023. [Online]. Available: https://optn.transplant.hrsa.gov/media/eavh5bf3/optn_policies.pdf.

[15] S. Riley, Q. Zhang, W.-Y. Tse, A. Connor, and Y. Wei, "Using Information Available at the Time of Donor Offer to Predict Kidney Transplant Survival Outcomes: A Systematic Review of Prediction Models," *Transplant International*, vol. 35, p. 10 397, Jun. 2022. DOI: 10.3389/ti.2022.10397.

[16] P. Rao, D. Schaubel, M. Guidinger, *et al.*, "A Comprehensive Risk Quantification Score for Deceased Donor Kidneys: The Kidney Donor Risk Index," *Transplantation*, vol. 88, pp. 231–6, Aug. 2009. DOI: 10.1097/TP.0b013e3181ac620b.

[17] A. Young, G. Knoll, E. McArthur, *et al.*, "Is the Kidney Donor Risk Index a Useful Tool in Non-US Patients?" *Canadian Journal of Kidney Health and Disease*, vol. 5, p. 205 435 811 879 114, Jul. 2018. DOI: 10.1177/2054358118791148.

[18] P. A. Clayton, S. P. McDonald, J. J. Snyder, N. Salkowski, and S. J. Chadban, "External Validation of the Estimated Posttransplant Survival Score for Allocation of Deceased Donor Kidneys in the United States," en, *American Journal of Transplantation*, vol. 14, no. 8, pp. 1922–1926, 2014, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajt.12761, ISSN: 1600-6143. DOI: 10.1111/ajt.12761. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/ajt.12761 (visited on 11/23/2022).

[19] P. Clayton, K. Dansie, M. Sypek, *et al.*, "External validation of the US and UK kidney donor risk indices for deceased donor kidney transplant survival in the Australian and New Zealand population," *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, vol. 34, Jun. 2019. DOI: 10.1093/ndt/gfz090.

[20] G. Collins, J. De Groot, S. Dutton, *et al.*, "External validation of multivariable prediction models: A systematic review of methodological conduct and reporting," *BMC medical research methodology*, vol. 14, p. 40, Mar. 2014. DOI: 10.1186/1471-2288-14-40.

[21] L. Wynants, B. Van Calster, M. Bonten, *et al.*, "Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal," *BMJ*, vol. 369, p. m1328, Apr. 2020. DOI: 10.1136/bmj.m1328.

[22] S. Senanayake, N. White, N. Graves, H. Healy, K. Baboolal, and S. Kularatna, "Machine learning in predicting graft failure following kidney transplantation: A systematic review of published predictive models," *International Journal of Medical Informatics*, p. 103 957, Aug. 2019. DOI: 10.1016/j.ijmedinf.2019.103957.

[23] C. Thongprayoon, C. C. Jadlowiec, W. Kaewput, *et al.*, "Distinct Phenotypes of Kidney Transplant Recipients in the United States with Limited Functional Status as Identified through Machine Learning Consensus Clustering," eng, *Journal of Personalized Medicine*, vol. 12, no. 6, p. 859, May 2022, ISSN: 2075-4426. DOI: 10.3390/jpm12060859.

[24] C. Thongprayoon, S. Mao, C. Jadlowiec, *et al.*, "Machine Learning Consensus Clustering of Morbidly Obese Kidney Transplant Recipients in the United States," *Journal of Clinical Medicine*, vol. 11, p. 3288, Jun. 2022. DOI: 10.3390/jcm11123288.

[25] C. Thongprayoon, P. Vaitla, C. C. Jadlowiec, *et al.*, "Use of Machine Learning Consensus Clustering to Identify Distinct Subtypes of Black Kidney Transplant Recipients and Associated Outcomes," eng, *JAMA surgery*, vol. 157, no. 7, e221286, Jul. 2022, ISSN: 2168-6262. DOI: 10.1001/jamasurg.2022.1286.

[26] K. K. Jalakam, "Multi-Modal Consensus Clustering to Identify Phenotypes of Kidney Transplant Donors and Recipients and Their Association With Survival," en, Nov. 2022, Accepted: 2022-11-21T13:19:35Z. [Online]. Available: https://DalSpace.library.dal.ca//handle/10222/82083 (visited on 01/07/2023).

[27] E. S. Barry, J. Merkebu, and L. Varpio, "State-of-the-art literature review methodology: A six-step approach for knowledge synthesis," en, *Perspectives on Medical Education*, vol. 11, no. 5, pp. 281–288, Oct. 2022, ISSN: 2212-277X. DOI: 10.1007/s40037-022-00725-9. [Online]. Available: https://doi.org/10.1007/s40037-022-00725-9 (visited on 12/18/2022).

[28] L. Li, J. Zhang, Y. Wang, and B. Ran, "Multiple imputation for incomplete traffic accident data using chained equations," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, ISSN: 2153-0017, Oct. 2017, pp. 1–5. DOI: 10.1109/ITSC.2017.8317639.

[29] M. Alabadla, F. Sidi, I. Ishak, *et al.*, "Systematic Review of Using Machine Learning in Imputing Missing Values," *IEEE Access*, vol. 10, pp. 44 483–44 502, 2022, Conference Name: IEEE Access, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3160841.

[30] *Python — Imputation using the KNNimputer()*, en-us, Section: Machine Learning, Aug. 2020. [Online]. Available: https://www.geeksforgeeks.org/python-imputation-using-the-knnimputer/ (visited on 01/09/2023).

[31] A. Juna, M. Umer, S. Sadiq, *et al.*, "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron," en, *Water*, vol. 14, no. 17, p. 2592, Jan. 2022, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2073-4441. DOI: 10.3390/w14172592. [Online]. Available: https://www.mdpi.com/2073-4441/14/17/2592 (visited on 01/09/2023).

[32] "What is XGBoost?" NVIDIA Data Science Glossary. (), [Online]. Available: https://www.nvidia.com/en-us/glossary/data-science/xgboost/ (visited on 01/09/2023).

[33] H. G. Ayad and M. S. Kamel, "On voting-based consensus of cluster ensembles," en, *Pattern Recognition*, vol. 43, no. 5, pp. 1943–1953, May 2010, ISSN: 0031-3203. DOI: 10.1016/j.patcog.2009.11.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320309004312 (visited on 01/09/2023).

[34] V. Karnowski, "Latent Class Analysis," in Nov. 2017. DOI: 10.1002/9781118901731.iecrm0130.

[35] A. Foss, M. Markatou, B. Ray, and A. Heching, "A semiparametric method for clustering mixed data," en, *Machine Learning*, vol. 105, no. 3, pp. 419–458, Dec. 2016, ISSN: 1573-0565. DOI: 10.1007/s10994-016-5575-7. [Online]. Available: https://doi.org/10.1007/s10994-016-5575-7 (visited on 01/09/2023).

[36] D. Reynolds, "Gaussian Mixture Models," en, in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds., Boston, MA: Springer US, 2009, pp. 659–663, ISBN: 978-0-387-73003-5. DOI: 10.1007/978-0-387-73003-5_196. [Online]. Available: https://doi.org/10.1007/978-0-387-73003-5_196 (visited on 01/10/2023).

[37] F.-X. Paquette, A. Ghassemi, O. Bukhtiyarova, *et al.*, "Machine learning support for decision making in kidney transplantation: Step-by-step development of a technological solution (Preprint)," *JMIR Medical Informatics*, vol. 10, Oct. 2021. DOI: 10.2196/34554.

[38] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: Personalized treatment recommender system using a cox proportional hazards deep neural network," *BMC Medical Research Methodology*, vol. 18, no. 1, p. 24, Feb. 26, 2018, ISSN: 1471-2288. DOI: 10.1186/s12874-018-0482-1. [Online]. Available: https://doi.org/10.1186/s12874-018-0482-1 (visited on 12/24/2022).

[39] A. Akl, A. M. Ismail, and M. Ghoneim, "Prediction of graft survival of living-donor kidney transplantation: Nomograms or artificial neural networks?" *Transplantation*, vol. 86, no. 10, p. 1401, Nov. 27, 2008, ISSN: 0041-1337. DOI: 10.1097/TP.0b013e31818b221f. [Online]. Available: https://journals.lww.com/transplantjournal/Fulltext/2008/11270/Prediction_of_Graft_Survival_of_Living_Donor.12.aspx (visited on 12/24/2022).

[40] K. D. Yoo, J. Noh, H. Lee, *et al.*, "A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: A multicenter cohort study," *Scientific Reports*, vol. 7, no. 1, p. 8904, Aug. 21, 2017, Number: 1 Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/s41598-017-08008-8. [Online]. Available: https://www.nature.com/articles/s41598-017-08008-8 (visited on 12/24/2022).

[41] M. K. Goel, P. Khanna, and J. Kishore, "Understanding survival analysis: Kaplan-meier estimate," *International Journal of Ayurveda Research*, vol. 1, no. 4, pp. 274–278, 2010, ISSN: 0974-7788. DOI: 10.4103/0974-7788.76794. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/ (visited on 12/24/2022).

[42] K. J. Jager, P. C. van Dijk, C. Zoccali, and F. W. Dekker, "The analysis of survival data: The kaplan–meier method," *Kidney International*, vol. 74, no. 5, pp. 560–565, Sep. 1, 2008, ISSN: 0085-2538. DOI: 10.1038/ki.2008.217. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0085253815533681 (visited on 12/24/2022).

[43] W. N. Dudley 1, R. Wickham 2, and N. COOMBS 3, "An introduction to survival statistics: Kaplan-meier analysis," *Journal of the Advanced Practitioner in Oncology*, vol. 7, no. 1, pp. 91–100, 2016, ISSN: 2150-0878. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5045282/ (visited on 12/24/2022).

[44] A. Loupy, O. Aubert, B. J. Orandi, *et al.*, "Prediction system for risk of allograft loss in patients receiving kidney transplants: International derivation and validation study," en, *BMJ*, vol. 366, p. l4923, Sep. 2019, Publisher: British Medical Journal Publishing Group Section: Research, ISSN: 0959-8138, 1756-1833. DOI: 10.1136/bmj.l4923. [Online]. Available: https://www.bmj.com/content/366/bmj.l4923 (visited on 01/07/2023).

[45] H.-U. Meier-Kriesche, F. K. Port, A. O. Ojo, *et al.*, "Effect of waiting time on renal transplant outcome," *Kidney International*, vol. 58, no. 3, pp. 1311–1317, Sep. 1, 2000, ISSN: 0085-2538. DOI: 10.1046/j.1523-1755.2000.00287.x. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0085253815472229 (visited on 12/24/2022).

[46] D. Pieloch, V. Dombrovskiy, A. J. Osband, *et al.*, "The kidney transplant morbidity index (KTMI): A simple prognostic tool to help determine outcome risk in kidney transplant candidates," *Progress in Transplantation*, vol. 25, no. 1, pp. 70–76, Mar. 2015, ISSN: 1526-9248, 2164-6708. DOI: 10.7182/pit2015462. [Online]. Available: http://journals.sagepub.com/doi/10.7182/pit2015462 (visited on 12/24/2022).

[47] L. van der Maaten and G. Hinton, *Visualizing data using t-SNE*, Nov. 2008. (visited on 01/09/2023).

[48] B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, and E. W. Steyerberg, "Calibration: The Achilles heel of predictive analytics," *BMC Medicine*, vol. 17, p. 230, Dec. 2019, ISSN: 1741-7015. DOI: 10.1186/s12916-019-1466-7. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6912996/ (visited on 01/09/2023).

[49] R. B. D'Agostino and B.-H. Nam, "Evaluation of the performance of survival analysis models: Discrimination and calibration measures," in *Handbook of Statistics*, ser. Advances in Survival Analysis, vol. 23, Elsevier, Jan. 1, 2003, pp. 1–25. DOI: 10.1016/S0169-7161(03)23001-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169716103230017 (visited on 12/27/2022).

[50] W. Luo, D. Phung, T. Tran, *et al.*, "Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view," *Journal of Medical Internet Research*, vol. 18, no. 12, e5870, Dec. 16, 2016, Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada. DOI: 10.2196/jmir.5870. [Online]. Available: https://www.jmir.org/2016/12/e323 (visited on 12/27/2022).

[51] MCP, *What is Phenotyping? – The Department of Molecular & Comparative Pathobiology*, en-US. [Online]. Available: https://mcp.bs.jhmi.edu/what-is-phenotyping/ (visited on 01/09/2023).

[52] A. Güney. "Introduction to bayesian belief networks," Medium. (Nov. 30, 2019), [Online]. Available: https://towardsdatascience.com/introduction-to-bayesian-belief-networks-c012e3f59f1b (visited on 01/09/2023).

[53] "Machine learning to predict transplant outcomes: Helpful or hype? a national cohort study," DOI: 10.1111/tri.13695. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/tri.13695 (visited on 12/24/2022).

[54] "Performance metrics for the comparative analysis of clinical risk prediction models employing machine learning — circulation: Cardiovascular quality and outcomes." (), [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/CIRCOUTCOMES.120.007526 (visited on 12/27/2022).

[55] E. W. Steyerberg, A. J. Vickers, N. R. Cook, *et al.*, "Assessing the performance of prediction models: A framework for some traditional and novel measures," *Epidemiology (Cambridge, Mass.)*, vol. 21, no. 1, pp. 128–138, Jan. 2010, ISSN: 1044-3983. DOI: 10.1097/EDE.0b013e3181c30fb2. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3575184/ (visited on 12/27/2022).