B.Tech Project

on

Simulation Modelling Analyses for Kidney Transplantation Networks

(Project Number: 34)

Submitted By

Pranav Shankar Girish (2019ME10823)

Supervised By: Dr. Varun Ramamohan

Mechanical Engineering Department

Indian Institute of Technology Delhi

November 2022

Acknowledgements

The work presented in this report couldn't be completed just by the efforts of one student. I would like to thank my thesis supervisor, Prof. Varun Ramamohan for guiding me at every step. I would also like to thank the Mechanical Department at IIT Delhi for giving us students the opportunity to carry out this project. The project has been an excellent learning opportunity, and a fitting way to culminate my academic growth through my undergraduate in Mechanical Engineering. I would also like to thank my peers for their motivation throughout the semester, motivating me to work hard and sincerely. I would also like to thank my parents for their constant support. Lastly, I thank the Almighty for providing us all with these opportunities.

Abstract

In our study, we have developed a multifaceted analysis of simulated kidney transplantation networks. We have produced a computational method for predicting, at the time of a patient's registration on the kidney transplantation waitlist, whether or not the patient will receive a transplant within a pre-specified time period (e.g., one year) using classification methods. If predicted yes, we predict their time on the waitlist before receiving the transplant using regression methods. We accomplish this by simulating the patient and donor behaviour to generate data points and, then, use Machine Learning algorithms to predict the probability of a transplant and, in turn, the time to allocation. The simulation is modelled using a set of parameters, both patient-related and donor-related. Once the simulation was up and running, we trained different classifiers and regressors to the obtained data and benchmarked them against each other, to finally obtain an accurate prediction. Furthermore, we are building on this by incorporating the concept of metamodelling into this simulated network. These techniques of simulation models have the advantage of faster execution, and they can provide insight into the nature of the simulation response as a function of design and input parameters.

Contents

Nomenclature

ANN Artificial Neural Network

AV Fitsula Arteriovenous fitsula

DES Discrete-event simulation

ESRD End-stage renal disease

f1 score F-measure

KAP Kidney allocation priority

KNOS Kerala Network for Organ Sharing

MAD Mean Absolute Deviation

MAPE Mean Absolute Percentage Error

NOTTO National Organ and Tissue Transplant Organisation

OECD Organisation for Economic Co-operation and Development

PRA Panel Reactive Antibody

RMSE Root mean squared error

ROC-AUC score Receiver Operator Characteristic - Area Under Curve score

ROTTO Regional Organ and Tissue Transplant Organisation

TNOS Tamil Nadu Network for Organ Sharing

VIF Variance Inflation Factor

List of Tables

List of Figures

1 Introduction

Approximately 220,000-275,000 new end-stage renal disease (ESRD) patients require renal replacement therapy, or dialysis in India every year [1][2]. Given the growing burden of diabetes and hypertension in India, this number is likely substantially larger in 2020 due to the association of these diseases with ESRD and dialysis. Renal (kidney) transplantation has been established as a significantly more efficacious therapy in terms of improving the survival of ESRD patients when compared to dialysis. Further, it has also been found to be more cost-effective than dialysis when costs and quality of life are also considered in addition to survival [3][4]. The key issue, then, limiting access to transplantation is the substantial shortage of donated kidneys in India, with reported organ donation rates ranging between $0.34 - 0.80$ per million population, compared to 26-36 per million population in OECD countries [5].

We present in this study a combined simulation and machine learning-based method for predicting, at the time of registration of a transplant candidate (an ESRD patient on dialysis) on the transplant waitlist, whether the candidate will receive a transplant within a pre-specified period of time, and if the patient is predicted to receive a transplant, the wait time of the candidate before receiving the transplant. Here it is pertinent to note that these predictions are personalized to each patient, based on the "state" of the transplant waitlist at the time of their arrival and their clinical characteristics, and are not derived from the distribution of the waiting time estimated from the simulation. This information can be used by the transplant candidate to decide whether to seek organs from living donors, register in other transplant hospitals, etc. We demonstrate the application of this method for two Indian states; Tamil Nadu and Kerala.

The next aim of our work is to use metamodelling techniques to improve the efficiency of computational execution of the simulation, allowing for larger regions of study. The idea is to create metamodels of individual simulations and, in turn their combination, using the key features identified and network outcomes for each to simulate a pre-defined set of required results. We will compare these results with results obtained by the parent simulations, and the simulation of the states when combined.

The remainder of this report is organized as follows: in section 2, we provide a review of existing literature, mainly in the field of operations research, relevant to our study. In section 4, we provide the necessary background and methodology regarding the transplantation system being modelled and the modelling framework itself. This includes the allocation policy followed in India for kidney transplantations, the architecture used to develop the simulation model, and the machine learning techniques used for the prediction module of our work. We also describe in this section the outcomes of the simulation and the results of the prediction exercises performed for various machine learning algorithms as obtained so far. We finally summarize the work in section 5 and discuss its limitations and avenues for future research, mainly focussing on the application of metamodelling to further strengthen our analytical study.

2 Literature Review

There is a significant body of literature documenting the application of simulation and operations research methods to analyze and optimize different aspects of organ transplantation systems of multiple countries. We restrict our review of the literature to: (a) a brief overview of the literature on optimizing decision-making in transplantation (e.g., determining the optimal timing of performing transplant), (b) the use of simulation in particular to model and improve organ transplantation systems, and (c) the delay prediction literature in operations management. We focus on these areas in the literature because our study presents an approach to predict their time to transplant (i.e., delay before entering service, in operations management terminology) using a combination of discrete-event simulation and machine learning to aid in decision-making for the patient.

Many articles documenting efforts to use operations research techniques to improve decision-making in transplantation have been published; however, a majority of this work focuses on liver transplantation systems. For example, Alagoz et al. [6] optimize the timing of a liver transplant from a living donor; Sandikci et al. [7] determine how different amounts of information (none versus incomplete versus full) regarding the transplant waitlist affect organ accept/reject decisions, and Batun et al. [8] take the risk sensitivity of the waitlisted patient into account when deciding whether to accept a liver offer. We note that nearly all of these studies involve liver transplants. Our literature search yielded lesser work involving kidney transplantation in comparison to liver transplantation. Zenios et al. [9] present a Monte Carlo simulation that incorporates donor and recipient parameters with the purpose of comparing alternative allocation policies in terms of quality of life. They consider the operations of a single US organ procurement centre in their simulation. Su et al. [10] simulate the US kidney transplantation system to analyse the impact of accepting/rejecting a donated kidney on the basis of the estimated increase in quality-adjusted life years associated with the donated kidney. Davis et al. [11][12] also simulate the US kidney transplantation system as part of a larger research effort to alleviate regional disparities in kidney allocation. In recent work, Sandikci et al. [13] present a novel simulation of the US kidney transplantation system that yields substantial savings in computational expense when compared to the UNOS simulation.

We build upon our previous work in simulating the kidney transplantation system in another Indian state, Rajasthan [14]. The models simulate the generation of kidneys from the districts of each state, and subsequent transportation to the district where the transplantation to the recipient is to be performed. Using this simulation model, In this work we not only predict the delay faced by the 'customer' (i.e., the patient) before entering 'service' (i.e., receiving a transplant), but we first have to predict whether the customer will receive service at all within a pre-specified time-frame. Kidney transplantation systems can be modelled as a queueing system. Such a system would have a complex queueing discipline. Since multiple clinical parameters are used to determine a patient's position on the waitlist, and a complex algorithm is used to allocate organs, analytical approaches are likely not to be tractable, and the literature also reflects this [15]. In such situations, empirical approaches that use transplant waitlist data are likely to work better to generate delay predictions for patients registering on the transplant waitlist. Further, it has been shown that predictors that utilize system state information, such as waitlist information (e.g., queue-length-based predictors), generally outperform delay history-based predictors [16]. The majority of the empirical work on delay prediction involves using historical data regarding the queueing system under consideration and training predictors using this data. However, in situations where such data is not available, or sufficient information regarding system state data required for training an accurate predictor is not maintained in the queueing system logs, the data may be generated from a discrete-event simulation of the system. This is the case for the kidney transplantation system that we consider.

There was also limited literature on the application of metamodelling techniques to such systems. We were introduced to such techniques through recent advances in the field and learnt the associated methods through a tutorial published along with the work [17]. The study by Fatma et al. explores the use of stochastic metamodels in primary healthcare delivery network systems [18]. They use stochastic metamodels to simplify their parent discrete-event simulation and thus reduce runtimes while still obtaining similar results. We plan on exploring this use further in our further study, but in our case, we will use metamodels to enable larger regions of study using network outcomes from a combined simulation.

A preliminary demonstration of the approach of using simulated data to make accurate predictions in cases of scare data was presented in Baldwa et al. [19]. Here, they used a discrete-event simulation of the neurosurgery ward of a large tertiary care hospital to generate data relating to the state of the neurosurgery ward at the time of arrival of the patient into the waitlist (i.e., the features for the machine learning method) to whether the patient was admitted or not within a prespecified duration. In this work, we build upon this approach by demonstrating the application of this method to the entirely different complex queueing problem represented by the kidney transplantation system. Further, Baldwa et al.[19] do not consider the characteristics of the patient in their analysis (as it was not required for their analysis); however, we consider both clinical as well as operational characteristics (i.e., the state of the waitlist, hospital in which the patient is registered, etc.) as features in this work, as both types of factors are relevant in predicting whether a patient receives a transplant. We also present a comprehensive demonstration of how this approach can be used to statistically determine the key clinical and operational characteristics influencing the waitlist outcome for the patient. An intriguing point of study here is the influence of state-localization of networks on its overall performance, which is what our proposed metamodelling approach seeks to investigate.

3 Project Objectives

Large urban public tertiary care hospitals in India typically face significantly more demand than their available capacity. This problem extends to kidney transplantations, which has been established as a treatment that allows a better quality of life and survival post-transplant than its alternatives. Here, the substantial shortage of donated kidneys in India causes an increasingly long waitlist of patients awaiting a transplant. These patients are unaware of whether they will be admitted within a certain time duration. A first step towards alleviating this problem is to provide these patients with this information at the time they present seeking admission so that they can make an informed decision about whether they should wait or seek treatment elsewhere or by other means.

The key objective of this study is to develop capabilities for end-stage renal disease patients to make better-informed treatment decisions at the time of their registration to the kidney transplantation waitlist. We strive to acheive this by developing a simulation and utilizing data generated by it to make real-time, accurate predictions regarding a patient's allocation chances. Such a study can be generalized as we are in the process demonstrating the use of simulated data to make real-time predictions. Such a method becomes vital when data is scarce or sensitive. We further aim to build on this by implementing metamodelling techniques in our simulation; to analyse the network at a higher level, combining states to obtain a more expansive view of the system for further study.

Overall, the contributions our study aims to achieve are as follows: We present a novel method, based on a combined simulation and machine learning approach, to predict waitlist outcomes for patients registering on the transplant waitlist at the time of registration. The simulation strives to provide an accurate dataset of information regarding a kidney transplantation network, which may be scarce or sensitive to obtain otherwise. The parameters taken for the simulation are based on real-world reports provided publicly by kidney transplantation organisations in the South Indian states under study, namely Kerala and Tamil Nadu. The outcomes extracted from this simulated data are multifold:

- Classification of waitlisted patients to predict whether or not they will receive a transplant in a pre-specified period.
- Regression of waitlisted patients to estimate the time taken for the allocation of a transplanted organ given that the patient will receive a transplant.
- Utilizing metamodelling techniques; exploring how to construct a high fidelity simulation of the network using metamodels of individual node (state) simulations.

Our approach for generating the above predictions is as follows. First, we develop a discrete-event simulation (DES) of the kidney transplantation system under consideration. The DES mimics the patient arrival, patient removal (through either death or transplantation), cadaveric organ generation and organ allocation processes of the transplantation system. We then use the DES to record the "state" of the transplant waitlist and the clinical characteristics of each patient at the point of their registration in the transplant waitlist. The waitlist outcomes are recorded for each such patient, thus generating a dataset suitable for training a machine learning algorithm to generate the required predictions. Note that the DES may be developed specifically for this purpose, or for the separate purpose of analysing and improving various aspects of the transplantation system under consideration. We then aim to explore further the influence of high level factors of the network by expanding the network beyond a specified state using metamodelling. This overall approach may be generalised to predict delays for any complex queueing system where adequate queue log data is not available to train an empirical predictor directly. We aim to demonstrate how our approach can be used to determine the factors influencing the waitlist outcome for the patient. We also aim to discuss how the accuracy of the simulation in representing the system affects the accuracy of delay prediction.

4 Work Progress

4.1 Allocation Methodology

The allocation of cadaveric kidneys to transplant candidates is complicated, and is determined by a variety of organ and patient characteristics, including time spent on the waitlist. ESRD patients eligible for registering on the waitlist must be of age less than 75 years at the time of registration, must have undergone regular maintenance dialysis for at least three months and must be registered in a single approved transplantation centre. Upon registration in a transplantation centre, the patient is registered in the respective state and district waitlists. The position of the patient on the waitlist is decided by a kidney allocation priority (KAP) score assigned to the patient at the time of registration. The KAP score is computed based on an algorithm provided in NOTTO's kidney allocation guidelines. (NOTTO, 2018) The kidney allocation is performed taking into account whether it was retrieved in a publicly funded hospital or a privately owned and managed hospital. If the organ is retrieved in a government hospital, patients registered in government transplant centres within the state are given higher priority for allocation; and patients registered in private hospitals are considered for allocation only if a suitable recipient is not found on the government hospital waitlist. If the kidney is retrieved from a private hospital, then patients registered in private transplant centres are first considered for allocation. Therefore, whether the patient is registered in a government or a private hospital affects their probability of transplant.

Table 1: Algorithm for determining the KAP score of a patient

S.No.	**Criteria for scoring**	**Points allocated**
1	Time on dialysis	(+1) for each month on dialysis
2	Previous immunological graft failure within3 months of transplantation.	(+3) for each graft failure
3	Age of recipient.	(+3) for less than 6 years(+2) for 6 to less than 12 years (+1) for 12 to less than 18years
4	Patient on temporary Vascular access.	
a)	With failed all AV Fistula sites.	(+2)
b)	With Failed AV Graft after all failed AVFsites.	(+4)
5	PRA (Panel Reactive Antibody).	(+0.5) for every 10% above20%

When considering patients registered in private or government transplant centres for allocation, patients registered in transplant centres within the district where the organ was retrieved are given higher priority. We also note that in the event that the kidney is retrieved from a donor aged less than 18 years, then patients aged less than 18 years are first considered for allocation. Finally, for each waitlist, donor/recipient matching is done on the basis of the blood groups of the donor and the patients. An O group (universal donor) kidney is matched to an O group recipient, and if none, it is allocated next to an A group patient, then to a B group patient, and finally to an AB group patient otherwise. A or B group kidneys are allocated to the patients with the same blood groups; else it is allocated to an AB group patient.

4.2 Simulation and Analyses

4.2.1 Simulation Model

The simulation was programmed on the Python computing platform and run on a personal system with an intel i7 10th generation processor system with 16GB RAM. A warm-up period of 12 years was used before results were collected over a period of 18 years. 20 replications were performed for collecting and reporting results. Three key events drive the execution of the simulation: the arrival of patients, removal of patients due to death, and organ arrival (which in turn drives removal of patients due to transplant). Thus a patient exits the waitlist either by receiving a transplant or due to death, implying that we do not take baulking or reneging into account in our model. Note that patient removal time due to death is determined by the time to death (estimated from the literature) assigned to the patient upon arrival in the system (a full description of model parameter estimation is provided in the following subsection). We also note here that we represent districts in each state by their district headquarters, and hence travel times between districts (for calculating organ transport times) are also calculated between the district headquarters. Overall, we have improved on the simulation model developed by Pratish Dewangan in his M.Tech Project at IIT Delhi, 2021, by incorporating an improved algorithm for removal time estimation and properly assigning probabilities of arrival at public and private hospitals based on their prevalence, among other tweaks.

Setting up the Simulation:

- Patient-related parameters: Patient-related parameters are estimated mainly using data published on the state network for organ sharing (KNOS for Kerala, TNOS for Tamil Nadu) website, census data (Census Commissioner of India 2011) and also using information obtained from the literature. The interarrival times of patients are estimated using the KNOS waitlist, fit using a chi-squared goodness of fit test and it follows an exponential distribution (p-value 0.7) with a mean of 1.382 days. Further, the removal time of a patient has been estimated as follows; Based on the KAP score of a patient, we find the percentile of this score. Using this percentile, we find the corresponding mean removal time, about which a symmetric beta distribution is defined. The final removal time value used is a random sample from this beta distribution. Lastly, The exponential distribution was found to be the best-fitting distribution to the time on dialysis at registration, with a p-value of 0.581 (test hypotheses the same as that for patient arrival rates). Per the NOTTO allocation algorithm, the KAP score of a patient is updated as the amount of time the patient spends on dialysis in the simulation increases.

- Organ-related parameters: Precise data regarding the dates of arrival of organs was not available on the organ-sharing websites. Therefore, the distribution of interarrival times for kidneys from deceased donors was assumed to be exponential, and its parameters were estimated using aggregate organ donation data published. Other parameters of the donor, such as the blood group and age, which are required to determine the allocation of the kidney, are also calculated according to the proportions of various blood groups and age ranges in the population of the entire state of Kerala. The probabilities of receiving a kidney in a public or private hospital within the district are made proportional to the number of each type of hospital in the district, instead of equal probabilities as was earlier assumed.

Note that as a rule of thumb based on the quality of data, data from patients or donors registering on the waitlist from 2016 to the end of 2018 was used to estimate model parameters. Data for patients or donors registering on the waitlist in 2019 was set aside for validating model outcomes.

Table 2: Patient-related Parameters

Parameter	Distribution	Estimate	Source
Patient inter-arrival time	Exponential	Mean= 1.382 days/patient	KNOS [20]
District origin of patient	Discrete	P(0)=0.1279, P(1)=0.0959, P(2)=0.0229, P(3)=0.0411, P(4)=0.0182, P(5)=0.1004, P(6)=0.055, P(7)=0.0637, P(8)=0.0913, P(9)=0.0501, P(10)=0.055, P(11)=0.1644, P(12)=0.0914, P(13)=0.0137	KNOS [20]
Age	Normal	Mean=49.74, SD= 7.423	KNOS [20]
Blood Group	Discrete	O = 0.458; A = 0.2385; B = 0.224; AB = 0.0795	TNOS [21]
Time on Dialysis	Exponential	Mean=260.3 days	RNOS [22]
Removal time	Beta	Mean: Beta with Mean = 40.31, SD = 26.69	
		Random sampling from Beta: a=0.66*Mean, b=1.33*Mean	Lakshminarayana et al. 2017 [23]
PRA Level	Discrete	P(PRA level = 0) = 0.650; P(1–20) = 0.056;	
		P(21–79) = 0.136;P(80–100) = 0.158	Cecka et al. (2011) [24]
Probability of a previous immunological graft failure within 3 months of transplantation	Discrete	P(yes) = 0.020; P(no) = 0.980	Abraham et al. (2009) [25]
With Failed all AV Fistula sites	Discrete	P(yes) = 0.052; P(no) = 0.948	Chandrashekar et al. (2014) [26]
With Failed AV Graft after all failed AVF sites	Discrete	P(yes) = 0.03125; P(no) = 0.96875	Chandrashekar et al. (2014) [26]

Table 3: Organ(Donor)-related parameters

Parameter	Distribution	Estimate	Source
Donor Inter-Arrival Time	Exponential	Mean=11.17 days/organ	Mohan Foundation Report [27]
Donor Blood group	Discrete	P(AB)=0.069, P(A)=0.192, P(B)=0.254, P(O)=0.485	TNOS [21]
District in which organ originates	Discrete	P(0)=0.0636, P(1)=0.0983, P(2)=0.0332, P(3)=0.0756, P(4)=0.0390, P(5)=0.0788, P(6)=0.0593, P(7)=0.0926, P(8)=0.1232, P(9)=0.0841, P(10)=0.0358, P(11)=0.0990, P(12)=0.0931, P(13)=0.0244	Census Report [28]
Number of kidney retrieval probability	Discrete	P(1)=0.777, P(2)=0.223	Statewise Organ Donors [27]

Algorithm 1: Computation of patient removal time [29]

Input: KAP score data of waitlisted patients

Output: Removal time of a patient

1 KAP scores simulated using random sampling of normal rtime distribution with removal time $mean = 40.31$ months, $SD = 26.69$ months, from the paper Lakshminarayana et al. 2017

2 Fitting this data to a beta distribution with $alpha = 0.89$ and $beta = 33.99$ (best fit out of alternatives)

3 Using the KAP score distribution obtained, we find the percentile of a given patient (A patient in the xth percentile of KAP scores would be in the $(100 - x)th$ percentile of rtimes)

4 The rtime value obtained from the percentile value on the distribution is used as the mean for the patient.

5 Using this mean, we define a symmetric beta distribution with $a = 0.67 * mean$ and $b = 1.33 * mean$. The final rtime value used is a random sample from this developed beta distribution.

Table 4: Simulation Outcomes for Kerala

Outcomes (/year)	Average	CI 95%
Average deaths	203.81	1.57
Average organs allocated	58.13	1.14
Average wait to allocation (hr/number of patient allocated)	19,175.10	215.43
Average time to transportation (min/number of patient allocated)	200.73	4.21
Allocation in government hospital	24.45	0.63
Allocation in private hsopital	33.68	0.93
Blood group A allocations	14.34	0.46
Blood group AB allocations	8.17	0.34
Blood group B allocations	14.91	0.49
Blood group O allocations	20.72	0.61
Average unallocated organs	0.35	0.09

The output parameters collected from the simulation include year-wise probabilities of receiving a transplant while on the waitlist, average organ transport time, the average time to transplant for a waitlisted patient, total number of patient deaths, number of unallocated organs and the total number of transplants in the simulation period. The key developments implemented to build on the existing model are the estimation of removal time and the assessment of probabilities of donor arrival in private or public hospitals. The former is now beta distributed with the mean dependent on the KAP score of the patient; thus, more critical patients would have a lower removal time, instead of random sampling. Also, the donor arrival rate in public and private hospitals in a city now depends on the relative number of such hospitals in the city.

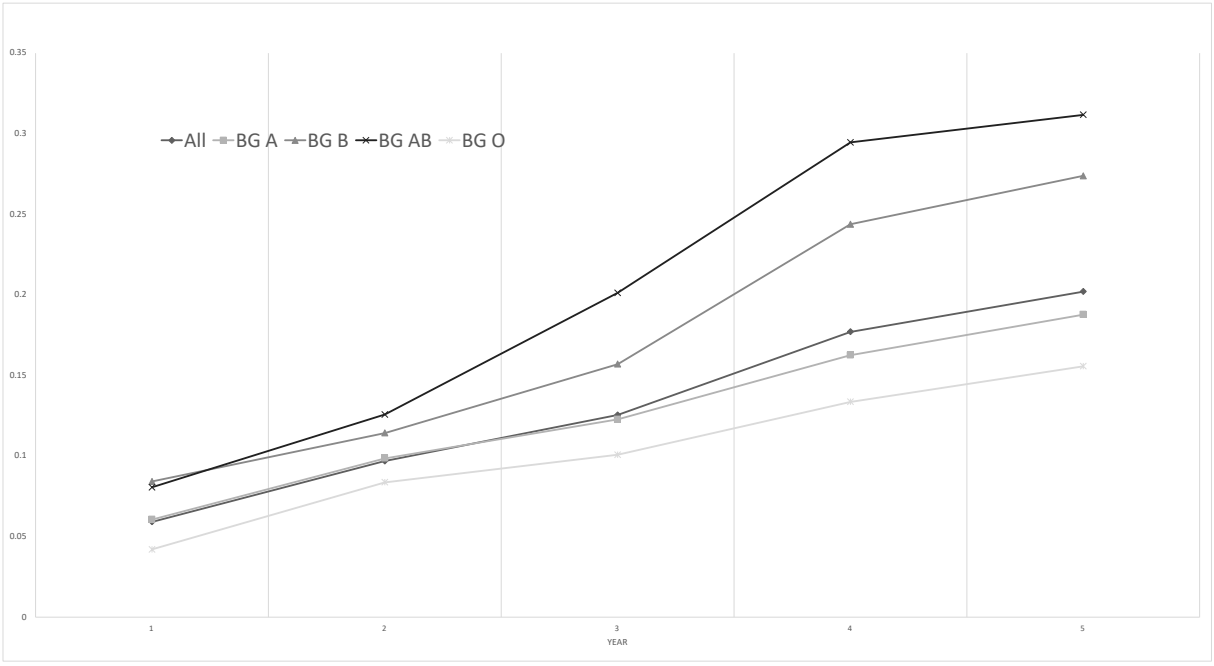

Figure 1: Relative Probabilities of receiving a transplant based on blood group

The results portrayed were obtained by running 20 replications of the simulation for a 30-year period, including a 12-year warmup. The probabilities of transplant are calculated as follows: patients arriving in each year are tracked separately, and the proportion of these patients receiving a transplant at the end of each subsequent year is updated. Average organ transport time is defined as the average time required to transport an organ from the retrieval location to its destination (a transplant centre). The average time to transplant is calculated only for patients who received a transplant during the steady state simulation period. Both the probabilities of receiving a transplant and the average time to transplant are calculated separately for different blood groups and the type of transplant hospital in which patients are registered in order to quantify disparities in transplantation outcomes on the basis of these characteristics. The death count is calculated by counting those removed from the waitlist without receiving a transplant.

Table 5: Validation of Simulation Outcomes

Parameters	Actual (KNOS, 2019)	Simulation Estimate (95% CI)	Upper limit	Lower Limit
Number of patients registered in 2019	264	261.61	265.90	257.32
patient in BG A	82	61.56	63.71	59.40
patient in BG AB	13	21.22	22.37	20.08
patient in BG B	51	59.43	61.29	57.56
patient in BG O	118	119.41	122.60	116.22
Organ arrived in 2019	32	57.61	60.45	54.77
Donor arrived in 219 (Mohan foundation)	19	32.70	34.32	31.09

Validation of the simulation model is a challenge, given the limited data available regarding kidney transplantation outcomes in the Indian context. We validated our simulation results against the actual values for 2019 (as discussed earlier, 2019 data was kept aside for validation purposes).

4.2.2 Classification Model

The first component of the prediction model is the prediction of whether or not a patient on the waitlist will receive a transplant. The computational experiments performed to consolidate our results include the overall waitlist outcome in terms of allocation probabilities, year-wise and blood-group-wise classification, and further feature selection using VIF and logistic regression methods. Several machine learning methods were benchmarked to find the method that provided the best accuracy given the nature of the input dataset. The methods benchmarked included standard vector classifiers, decision tree methods such as bagging and random forests, and artificial neural networks. The model was fed the simulation outcome, which represented the waitlist data over the simulation period, as input data. The model is trained using features which would be known to a patient at the time of registering to the waitlist, these include their age, city, hospital, blood group, KAP score, and waitlist data. The input data was split into training and test sets with a 75/25 split ratio, and scaled using MinMaxScaler, part of the scikit-learn package. The input dataset consisted of 929 successful allocations out of 3945 patients. Thus it is necessary to balance the dataset, for which we use the Synthetic Minority Oversampling Technique. The classification is done using the status features of the input dataset as targets, and randomized features such as ID and removal time are removed.

Before training the model, the hyperparameters were tuned using GridSearchCV. The obtained ranges for the same are given below. To measure the performance of the developed models, we use the ROC-AUC (Receiver Operator Characteristic - Area Under Curve) and f1 scores. The ROC is a probability curve that plots the true positive rate against the false positive rate at various threshold values. The AUC is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. It is preferred over simply using the classifier error rate as it calibrates the trade-off between sensitivity and specificity at the best-chosen threshold. [30]. The f1 score is the harmonic mean of precision and recall. Precision is the number of correct positive predictions relative to total positive predictions, while recall is the correct positive predictions relative to total actual positives. This score provides us with an insight into both the type I (false-positive) and type II (false-negative) errors in the model. [31].

The hyperparameters for each machine learning model and the corresponding ranges considered during tuning, along with the final used values, are provided in Table 6. Multiple trials have been performed by changing the considered range to make sure that the obtained value is not just a local optima for that corresponding range.

The input dataset features that we are using for classification are given along with their characteristics in Table 7. We use different types of features to train our model; operational data, clinical data, and waitlist data.

The results of classification when the target is the status of transplant are given in Table 8. The table shows the mean and standard deviation of AUC score of the model along with the precision, recall, and f1 scores for both cases where a transplant is received (target is 1) and a transplant is not received (target is 0). The mean and standard deviations are taken by iterating the model over multiple random states, so as to present a more consolidated result. From these results, we see that the classifiers, especially the decision tree ensembles (bagging and random forest) and gradient boost techniques, are performing well in classifying the data. It is pertinent to note that we're achieving over 80% precision and recall for successful transplant cases.

Table 6: Hyperparameter Tuning performed for classification

Model	Hyperparameters	Used	Range Considered
Logistic Regression	C	100	[0.1,1,10,100,200,500,1000]
	penalty	l1	[l1,l2]
	solver	newton-cg	[‘newton-cg’, ‘lbfgs’, ‘liblinear’]
SVM Classifier	C	100	[0.1,1,10,100,200,500,1000]
	degree	1	[1,2,3,4]
	kernel	poly	[‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’]
	gamma	scale	[’scale’]
Bagging Classifier	n_estimators	1000	[50,100,200,500,1000,1500,2000]
Random Forest Classifier	max_features	sqrt	[‘sqrt’, ‘log2’]
	n_estimators	500	[50,100,200,500,1000,1500,2000]
Gradient Boosting Classifier	learning_rate	0.1	[0.001,0.1,0.3,0.5]
	n_estimators	100	[50,100,200,500,1000]
Artifical Neural Network	hidden layers	2	[1,2,3]
	Neurons_Trial	30	[10,20,30,40,50,100]
	Optimizer_Trial	adam	[adam,rmsprop]
	batch_size	10	[10,20,30]
	epochs	20	[10,20,30]

We further drill down by capturing results for successful predictions for receiving a transplant in 5 and 2 years, and by categorizing performance into the different blood groups. The results of the 5th-year status experiment in Table 9 are fairly similar to the overall status results, reflecting a low chance of receiving a transplant in time (before removal) if not received in 5 years. We see that the precision and recall values for the case of 2nd-year status in Table 10 cross 90% for successful transplant cases (target is 1). We also observe that the AUC score for this case exceeds 95%. This is particularly encouraging as patients can be confident in their predictions for at least a 2-year time scale. The results of classification for the different blood groups in Table 11 reflect the influence of blood group on allocation policy, where the AUC score of blood group AB is the highest but with the lowest precision, while blood group O has the lowest AUC score with larger precision. This can be attributed to the difference in sample size for the blood groups and the differences in compatibility between patient and donor blood groups.

The input data used for classification is generated by the simulation model as a wait-list dataset. This data is generated in each replication, and we use the dataset from the final replication for our prediction module. To make sure the results from different replications are statistically similar, the results obtained when running classification on different replications of the same simulation run is shown in Table 12.

Table 7: Input feature characteristics for classification model

Feature	age	city_index	bloodgroup	KAP score	hosp_name	hosp_type
Mean	50.36	8.22	0.91	10.11	1.66	0.90
STD	7.41	3.36	0.99	8.80	2.17	0.29
Min	24	1	0	1	0	0
Max	75	12	3	76	9	1
Feature	PRA	PRA type	AVG	AVF	time_on_dailysis	PIGF
Mean	21.70	3.41	0.03	0.05	9.00	0.02
STD	34.84	0.92	0.17	0.22	8.64	0.14
Min	0	1	0	0	1	0
Max	100	4	1	1	76	1
Feature	position on waitlist	patients aboove this patient	A patients above	B patients above	O patients above	AB patients above
Mean	614.66	613.66	137.27	130.93	299.83	45.63
STD	115.21	115.21	28.88	25.86	56.68	9.44
Min	2	1	0	0	1	0
Max	748	747	183	169	387	62
Feature	total patients on waitlist	total A patients	total B patients	total O patients	total AB patients	
Mean	696.08	156.71	148.39	339.37	51.61	
STD	27.52	13.89	10.97	17.88	4.38	
Min	623	126	121	303	39	
Max	752	187	170	387	62	

Table 8: Classification results for status of transplant

Mean (STD)	LogR	SVM	Bagging	Random Forest	ANN	Gradient Boost
Mean AUC Score	0.91 (0.01)	0.91 (0.01)	0.9 (0.009)	0.9 (0.008)	0.9 (0.018)	0.91 (0.011)
Mean Precision 0	0.97 (0.007)	0.97 (0.006)	0.96 (0.007)	0.96 (0.005)	0.97 (0.018)	0.96 (0.007)
Mean Recall 0	0.91 (0.013)	0.91 (0.012)	0.93 (0.011)	0.94 (0.007)	0.92 (0.031)	0.94 (0.011)
Mean f1 score 0	0.94 (0.007)	0.94 (0.007)	0.95 (0.006)	0.95 (0.005)	0.94 (0.011)	0.95 (0.007)
Mean Precision 1	0.76 (0.03)	0.76 (0.028)	0.8 (0.028)	0.82 (0.021)	0.77 (0.063)	0.81 (0.03)
Mean Recall 1	0.91 (0.019)	0.91 (0.015)	0.87 (0.016)	0.87 (0.012)	0.89 (0.057)	0.87 (0.017)
Mean f1 score 1	0.83 (0.018)	0.83 (0.018)	0.83 (0.016)	0.84 (0.013)	0.82 (0.021)	0.84 (0.019)

Table 9: Classification results for 5th year status of transplant

Mean (STD)	LogR	SVM	Bagging DT	Random Forest	ANN	Gradient Boost
Mean AUC Score	0.91 (0.01)	0.91 (0.01)	0.9 (0.009)	0.9 (0.008)	0.91 (0.011)	0.91 (0.011)
Mean Precision 0	0.97 (0.007)	0.97 (0.006)	0.96 (0.007)	0.96 (0.005)	0.97 (0.008)	0.96 (0.007)
Mean Recall 0	0.91 (0.013)	0.91 (0.012)	0.93 (0.011)	0.94 (0.007)	0.9 (0.035)	0.94 (0.011)
Mean f1 score 0	0.94 (0.007)	0.94 (0.007)	0.95 (0.006)	0.95 (0.005)	0.93 (0.017)	0.95 (0.007)
Mean Precision 1	0.76 (0.03)	0.76 (0.028)	0.8 (0.028)	0.82 (0.021)	0.74 (0.06)	0.81 (0.03)
Mean Recall 1	0.91 (0.019)	0.91 (0.015)	0.87 (0.016)	0.87 (0.012)	0.92 (0.024)	0.87 (0.017)
Mean f1 score 1	0.83 (0.018)	0.83 (0.018)	0.83 (0.016)	0.84 (0.013)	0.82 (0.031)	0.84 (0.019)

Table 10: Classification results for 2nd year status of transplant

Mean (STD)	LogR	SVM	Bagging DT	Random Forest	ANN	Gradient Boost
Mean AUC Score	0.96 (0.008)	0.96 (0.012)	0.95 (0.011)	0.95 (0.011)	0.96 (0.012)	0.95 (0.016)
Mean Precision 0	0.99 (0.003)	0.99 (0.003)	0.99 (0.003)	0.99 (0.003)	0.99 (0.003)	0.99 (0.004)
Mean Recall 0	0.97 (0.007)	0.97 (0.01)	0.99 (0.004)	0.99 (0.002)	0.97 (0.011)	0.99 (0.004)
Mean f1 score 0	0.98 (0.003)	0.98 (0.005)	0.99 (0.002)	0.99 (0.002)	0.98 (0.006)	0.99 (0.003)
Mean Precision 1	0.8 (0.032)	0.81 (0.049)	0.92 (0.026)	0.95 (0.015)	0.8 (0.061)	0.93 (0.028)
Mean Recall 1	0.95 (0.02)	0.95 (0.026)	0.9 (0.024)	0.91 (0.022)	0.95 (0.021)	0.91 (0.031)
Mean f1 score 1	0.87 (0.017)	0.87 (0.029)	0.91 (0.015)	0.93 (0.015)	0.87 (0.037)	0.92 (0.022)

Table 11: Classification Results using Gradient Boost for different blood groups

Mean (STD)	Blood Group O:	Blood Group A:	Blood Group B:	Blood Group AB:
Mean AUC Score	0.84 (0.022)	0.91 (0.026)	0.9 (0.017)	0.93 (0.014)
Mean Precision 0	0.87 (0.03)	0.96 (0.016)	0.95 (0.013)	0.98 (0.006)
Mean Recall 0	0.87 (0.039)	0.95 (0.016)	0.94 (0.015)	0.94 (0.014)
Mean f1 score 0	0.87 (0.029)	0.95 (0.013)	0.95 (0.009)	0.96 (0.007)
Mean Precision 1	0.81 (0.032)	0.87 (0.036)	0.83 (0.043)	0.79 (0.044)
Mean Recall 1	0.81 (0.05)	0.88 (0.046)	0.86 (0.033)	0.91 (0.025)
Mean f1 score 1	0.81 (0.028)	0.87 (0.034)	0.84 (0.026)	0.85 (0.027)

Table 12: Results from different sampled simulation replications

Replication	1	2	3	Combined
Mean AUC Score	0.9 (0.01)	0.92 (0.01)	0.91 (0.007)	0.91 (0.005)
0 Mean Precision	0.96 (0.007)	0.96 (0.007)	0.96 (0.007)	0.96 (0.004)
0 Mean Recall	0.94 (0.009)	0.95 (0.004)	0.93 (0.013)	0.93 (0.005)
0 Mean f1 score	0.95 (0.006)	0.95 (0.004)	0.95 (0.005)	0.95 (0.002)
1 Mean Precision	0.82 (0.023)	0.84 (0.013)	0.81 (0.029)	0.81 (0.015)
1 Mean Recall	0.86 (0.018)	0.89 (0.019)	0.88 (0.02)	0.88 (0.012)
1 Mean f1 score	0.84 (0.016)	0.86 (0.011)	0.84 (0.012)	0.84 (0.007)

Table 13: Feature Selection

Feature	Coefficient Estimate	VIF
age	0.006764	31.521921
city_index	-0.07769	7.0050044
bloodgroup	0.850925	1.8328548
kap_score	0.390441	10171.34
hosp_name	-0.0337	1.6401899
hosp_type	-6.86927	9.0357345
pra	-0.00036	174.86633
avg	0.034474	26.62936
av_fitsula	0.085149	13.499979
time_on_dailysis	-0.06209	8796.0121
pigf	-0.04439	10.669061
pra type	0.04552	27.723559

Table 14: Gradient Boosting Classification Results after feature selection

Mean (STD)	VIF-based selection	LR-based selection	Combined
Mean AUC Score	0.91 (0.011)	0.91 (0.012)	0.91 (0.01)
Mean Precision 0	0.96 (0.006)	0.96 (0.007)	0.96 (0.006)
Mean Recall 0	0.94 (0.009)	0.94 (0.01)	0.94 (0.011)
Mean f1 score 0	0.95 (0.006)	0.95 (0.007)	0.95 (0.007)
Mean Precision 1	0.82 (0.026)	0.82 (0.027)	0.81 (0.03)
Mean Recall 1	0.89 (0.018)	0.88 (0.018)	0.88 (0.014)
Mean f1 score 1	0.85 (0.018)	0.84 (0.019)	0.84 (0.018)

Building on the results obtained, we performed feature selection to further consolidate our results. We used VIF and logistic regression technqiues for the same. VIF (Variable Inflation Factors) is the most commonly used method to detect multicollinearity. VIF determines the strength of the correlation between the independent variables. It is predicted by taking a variable and regressing it against every other variable. To calculate VIF, we calculate the R^2 value for the features. Using this method, we saw very high VIF values for age and KAP score features (out of the operational features used). The logistic regression method fits the dataset using a set of weights for each feature. These weights or coefficients determine the redundancies in the features, and are those features whose coefficients come out as very small or zero. Using this method, we saw very low coefficients for age and hospital name features (out of the operational features used). The VIF values and coefficient estimates are presented in Table 12.

Thus, age and KAP score are removed according to the VIF methods, while age and hospital name are removed according to the Logistic Regression method. We also combine the two methods' results by experimenting removing age, hospital name, and KAP score. We trained the model again using the updated feature set according to the VIF result (removing age and KAP score features), the Logistic Regression result (removing age and hospital name), and combining both (removing age, KAP score, and hospital name). The results of this classification are presented in Table 13, and we see a minimal change in results before and after feature selection.

4.2.3 Regression Model

The second component of the prediction model is the prediction of the time after which a patient (who was predicted to receive a transplant) will receive the transplant. We used the same simulated dataset we used for classification, but this time restricted the dataset to those cases where a transplant was successful. Here, the time to allocation was our target for the same feature set. Several machine learning methods were benchmarked to find the method that provided the best accuracy given the nature of the input dataset. Preprocessing of the input data and balancing was followed by hyperparameter tuning of the model, similar to that done in classification. The methods we benchmarked included standard vector regressors, decision tree methods such as bagging and random forests, and artificial neural networks. We saw that the bagging method was giving the best results compared to the others, without the presence of any negative predictions. Once a model is trained, we use the RMSE (Root mean squared error), MAD (Mean Absolute Deviation), and MAPE (Mean Absolute Percentage Error) values calculated to compare the relative performance of the different methods.

We made multiple attempts to model the regression framework better, but have been limited by the stochastic nature of the dataset. Given this, we attempted to create a tolerance range to understand at what tolerance we get reasonable results. The results for the same represent a 76% accuracy for bagging (since this method gave non-negative values with minimum error) with 50% tolerance. We also performed outlier detection; what we were able to identify was that the worst-performing prediction cases were predominantly due to disproportionately fast transplants (in all such cases, the target is much less than the prediction). On removing these, the MAPE score (for bagging decision tree ensemble) decreases to 22%, making this a considerable improvement in results.

Table 15: Regression results for time to allocation

Total	SVM	Bagging	Random Forest	ANN
Root Mean Squared Error (RMSE)	193.29	193.62	208.61	856.4
Mean Absolute Percentage Error (MAPE)	80.31	89.01	190.39	91.49

Figure 2: Tolerance range improving the accuracy of regression

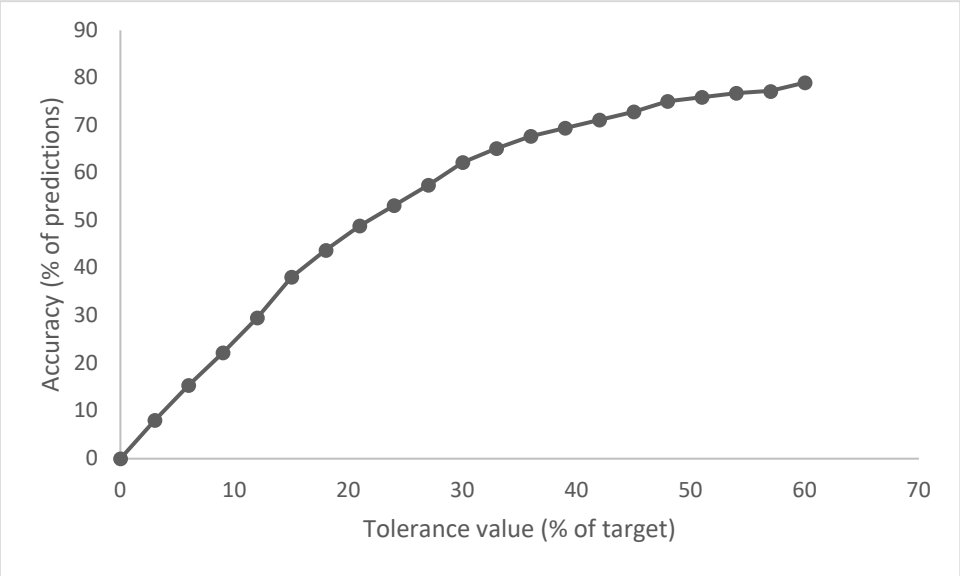

Table 16: Blood-group-wise and total scores after outlier removal

bloodgroup	MAPE	MAD
A	19.83022589	121.7792026
AB	29.86263632	186.0555242
B	18.0491531	160.6457215
O	18.53271187	162.5481762
Grand Total	22.04149807	154.4090442

4.2.4 Metamodelling Efforts

The next step in our study is to utilize metamodelling techniques to increase the scope of study with limited computational power. The use of such methods allows for a larger region of study through faster execution of the simulation with a better understanding of the network outcomes of the complex system. Following the published tutorial on metamodelling, the concepts behind this novel method became familiar.

Metamodels generally have three characteristics that give them advantage over the simulation model for certain purposes. A metamodel f generally has explicit form, deterministic output, and, once fitted, is computationally inexpensive to evaluate. Metamodels, once fitted, can be used as a proxy, to evaluate instead of making (computationally expensive and stochastic) simulation runs. Further, because of their explicit form, they can be used in many computationally intensive operations, such as optimization.

There are multiple steps required to incorporate metamodelling techniques in our simulation model. First, we need to select the relevant design parameters and outputs for our metamodels. Then, we need to scale the chosen parameters, generally done between -1 and 1. Once this is done, we need to choose a metamodel type basis our study; linear regression metamodel or a gaussian process metamodel. Out of the two, the former provides a more detailed output and is easier to fit, whereas the latter provides more flexibility. A third type of metamodel also exists, artificial neural network (ANN) metamodels, however the insights obtained from such metamodels are seen to be highly restricted by the selected codification scheme. While insight is possible from either linear regression coefficients or spatial correlation parameters, ANN coefficients generally do not provide insight on the impact of independent variables on simulation output.

Table 17: Validation of the combined simulation using individual (state) simulations

Values expressed as "Mean (STD)"	Kerala		Tamil Nadu	
Yearly average of results	Individual	Combined	Individual	Combined
Number of organs allocated	59.22 (8.11)	57.91 (6.97)	265.5 (40.30)	277.98 (15.67)
Number of deaths	209.22 (16.60)	204.00 (7.34)	1246.50 (10.60)	1252.54 (25.28)
Wait to allocation (days)	1090.19 (63.45)	1081.50 (30.23)	1238.26 (10.50)	1268.01 (18.36)
Time to transportation (minutes)	134.30 (46.65)	130.25 (33.27)	227.84 (6.24)	159.73 (9.10)
Allocation in government hospital	29.11 (6.47)	28.09 (3.64)	119 (9.89)	138.17 (10.31)
Allocation in private hospital	30.11 (6.89)	29.81 (5.14)	146.5 (30.40)	139.81 (9.54)
Blood group A allocations	15.44 (3.71)	14.46 (2.91)	64 (11.31)	61.06 (3.90)
Blood group AB allocations	9.11 (3.18)	8.06 (1.61)	23.5 (3.53)	29.00 (5.03)
Blood group B allocations	15.78 (4.37)	16.54 (2.27)	85 (16.97)	95.74 (7.76)
Blood group O allocations	18.89 (4.74)	18.85 (3.98)	93 (8.48)	92.19 (7.33)
Average unallocated organs	0.44 (1.10)	0.02 (0.08)	0 (0)	0 (0)

In our study, the use of metamodels becomes relevant when we try to model multiple states together in a single simulation to understand inter-state characteristics. We created a combined simulation of Kerala and Tamil Nadu using the current allocation policy. The policy states that if a suitable patient is not found in the state of the donor, then the regional waitlist (the waitlist with patients from both states) is considered. Using the allocation algorithm as stated in the allocation methodology, we developed the combined simulation. The results of the simulation were cross-checked with the individual state simulations to validate that they were in check.

Now that the requisite infrastructure has been developed, namely the combined simulations and the validated individual simulations, we can work on preparing metamodels to test the improvement in computational ability. This poses the next phase of our project.

4.2.5 Analytical Framework

The classifier predicts allocation status of a patient on the waitlist as a function of the system state S. The classifier essentially sets a threshold probability p_{th} using the training set, and a probability estimate for a given patient p_{pj} (predicted for a patient j). If $p_{pj} \geq p_{th}$, the allotment is classified as yes (the patient will receive a transplant), else it is classified as no (the patient will not receive a transplant). For each patient, there would also be an actual probability of receiving a tranplant, p_{aj}.The error in prediction (E_{pred}) can thus be defined as:

$$E_{pred} = \begin{cases} |p_{aj} - p_{pj}|, & \text{if } (p_{aj} - p_{th})(p_{pj} - p_{th}) \leq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Thus we require:$P[(p_{aj} - p_{th})(p_{pj} - p_{th}) < 0]$

We can write $p_{pj} = p_{aj} + \delta s + \delta m$, signifying that the predicted probability is the actual probability plus any error induced by the simulation (δs) and the machine learning model (δm). We can also write the sum of errors as the total error (δt). Thus, the expression reduces to:

$$(p_{aj} - p_{th})(p_{pj} - p_{th}) = (p_{aj} - p_{th})(p_{aj} + \delta s + \delta m - p_{th}) \tag{2}$$

$$= p_{aj}^2 + p_{at}\delta t - 2p_{aj}p_{th} - p_{th}\delta t \tag{3}$$

$$= (p_{aj} - p_t)^2 + \delta t(p_{aj} - p_{th} < 0) \tag{4}$$

$$\text{Thus, we have the condition that: } \begin{cases} \delta t(p_{aj} - p_{th}) & \text{and,} \\ \delta t > (p_{aj} - p_{th}) \end{cases} \tag{5}$$

Qualitatively this means that the total error should be of opposite sign and of greater absolute value as that of the difference of the actual probability and the threshold.

Note that for the regressor, the error is straightforward and is given by $T_{pj} - T_{aj}$ where T_{pj} is the predicted time to allocation and T_{aj} is the actual time to allocation. The remaining framework is synonymous to both regressors and classifiers, hence we present the derivation considering the classifier alone:

Consider the predicted probability for a patient j in a perfect simulation to be p_{cj}.

$$\text{Total Error} = p_{aj} - p_{pj} = p_{aj} - p_{cj} + p_{cj} - p_{pj}$$

Now $p_{aj} - p_{cj}$ can be thought of as the error due to the simulation.
Thus,

$$\text{total error: } \sum_{tot(i)} = \text{simulation error} + \text{classification error.}$$

Here, a point to note is that the training error during the estimation of p_{th} has not been considered separately. Now, assuming the variance of the system state S_j from the perfect simulation state is $\triangle S_j$:

$$p_{pj} = f(S_j + \triangle S_j)$$

Doing a first order Taylor series approximation,

$$p_{pj} \approx f(S_j) + \triangle f(S_j)^T \triangle S_j$$
$$p_{pj} = p_{cj} + \triangle f(S_j)^T \triangle S_j$$
$$\epsilon_{sim(j)} \approx \triangle f(S_j)^T \triangle S_j$$
$$\text{Thus, } \epsilon_{tot(j)} \approx \triangle f(S_j)^T \triangle S_j + \epsilon_{classifier(j)}$$

Now $\triangle S_j$ can be thought of as a random vector of x random variables representing simulation errors in estimating individual state variables. Thus,

$$\epsilon_{tot(j)} \approx \sum_{i=1}^{n} \frac{\partial f}{\partial S_{j(i)}} \triangle S_{j(i)} + \epsilon_{classifier(j)}$$

Now, $\frac{\partial f}{\partial S_{j(i)}}$ can be approximated by the average sensitivity of the classifier to changes in the state variable. This can be in particular be estimated for the expression . Now $\frac{\partial f}{\partial S_j(i)}$ can be approximated by $\frac{\partial f}{\partial S_{(i)}}$ in general. Thus,

$$\epsilon_{tot(j)} \approx \frac{\partial f}{\partial S_{(i)}} \triangle S_i + \epsilon_{classifier(j)}$$

Taking expectations,

$$\text{E}[\epsilon_{tot(j)}] = \sum_{i=1}^{n} \frac{\partial f}{\partial S_{(i)}} E[\triangle S_{j(i)}] + \epsilon_{classifier(j)}$$

Thus, if $\frac{\partial f}{\partial S_{(i)}}$ for all i is taken as given, then $\text{E}[\epsilon_{tot(j)}]$ is particularly dependent on $\text{E}[\triangle S_{j(i)}]$. Now for a validated simulation where performance outcomes as estimated by the simulation are found to be statistically similar to the observed data, one can argue that $\text{E}[\triangle S_{j(i)}] \approx 0$. Thus for validated simulation

$$\text{E}[\epsilon_{tot(j)}] \approx \text{E}[\epsilon_{classifier}]$$

4.2.6 Results

The simulation was seen to be able to synthesize data effectively based on the input parameters provided. The results of the simulation show that it is in line with the expected values, as seen in the validation process. The changes made to the simulation improved its ability to properly model the real-world data:

- The updation of the removal time estimation algorithm resulted in a more realistic distribution of allocations, wherein patients with high KAP scores have a low time to allocation or are not allocated (since their time before removal would be low) whereas those with low KAP scores see higher times to allocation.
- The proper assignment of arrival probabilities to private and public hospitals means that now the majority of patients arrive in private hospitals while kidney retreival happens in both types of hospitals. This results in a much smaller government waitlist and thus a high probability of allocation in public hospitals.

The simulation results show a varying probability of receiving a transplant based on the patient's blood group. This is due to the allocation policy of kidneys based on blood groups, and the relative prevalence of each blood group in donors and patients. AB blood group sees the largest probability, mainly down to the relative low numbers of AB patients and donors, similar to how O blood group patients have the lowest probability.

In the classification model, we were able to achieve an AUC (area-under-curve) score greater than 0.9, signifying an accuracy of more than 90%, by tweaking the hyperparameters of the classifiers. The influence of blood group on allocation policy is reflected further in the classification outcomes, where the AUC score of blood group AB is the highest but with lowest precision, while blood group O has the lowest AUC score with a larger precision. This can be attributed to the difference in sample size for the blood groups and the existing differences in the simulation outcomes. The drill down on 2nd year status of transplant shows an AUC score of over 95% and is particularly encouraging as it means that patients can be confident in their predictions for at least a 2-year time scale. Performing feature selection resulted in a minimal improvement in results, which enabled us to be confident in our initial feature set.

In the regression model, we were able to achieve an accuracy of more than 75%, by tweaking the hyperparameters of the regressors and defining a tolerance level of half of the target time value on either side of the predicted value. We also found outliers in our output set, and the predominant nature of these outliers were that they were cases where the patient receives a transplant disproportionately fast as a result of the allocation compatibilities. Removing these outliers caused the MAPE score to fall to 22%, citing a considerable improvement in the performance of our model.

The combined simulation developed for Tamil Nadu and Kerala gives us valuable insights into the network outcomes of the system, and further experimentation with organ and patient arrival rates and different allocation policies can help us further explore these effects. This simulation however is very computationally complex. This gives rise to the need for metamodelling techniques to effectively study this larger region. Metamodelling is thus our next plan of action. We have prepared the requisite infrastructure for its incorporation and familiarized ourselves with the underlying concepts.

5 Conclusion

The work presented in this report is the first step toward modelling, analyzing and optimizing the organ transplantation system in India. As the transplantation infrastructure in India develops further, the need for such a model to analyse and optimize logistical aspects of the transplantation system will be felt more acutely. However, it is also evident from the outcomes generated by the model that it is imperative to consolidate and expand public awareness programs to increase the organ donation rate in the country so that the average time to transplant is reduced and the number of deaths while on the waitlist is reduced. At a larger level, this work also demonstrates the use of synthetic data for prediction, which can prove to be vital in situations where data is scarce or sensitive.

We have thus successfully developed a simulation model to generate data on organ allocations, connecting patients to donors and conducting transplantations to obtain the overarching statistics relevant to real-world outcomes. Using this simulated data, we perform predictions which can accurately provide patients with vital information about their allocation probabilities and time to allocation, if applicable, at the time of their registration on the waitlist. This is greatly useful to help such patients with end-stage renal disease to make better-informed decisions about their treatment.

It is pertinent to note that we aim to publish our work done in real-time prediction (classification and regression models). To this end, we performed extensive experimentation and validation efforts to build a comprehensive study. These efforts led to improved results and deeper insights but were at the cost of progress on the metamodelling front. The next steps of the project carried forward in the next semester would thus be chiefly to implement metamodelling techniques to develop this study further. The plan is to integrate the simulations prepared for Kerala and Tamil Nadu effectively in order to obtain higher-level insights. This combined simulation, when simplified in the form of metamodels of the individual states or the combined simulation itself, promises greatly reduced computational time and, in turn, enables researchers to study larger regions of such complex systems using a high-fidelity simulation.

References

[1] Vivekanand Jha. "Current status of end-stage renal disease care in India and Pakistan". In: *Kidney International Supplements* 3.2 (2013), pp. 157–160.

[2] Gopesh Modi and Vivekanand Jha. "Incidence of ESRD in India". In: *Kidney International* 79.5 (2011), p. 573.

[3] Kirsten Howard et al. "The cost-effectiveness of increasing kidney transplantation and home-based dialysis". In: *Nephrology* 14.1 (2009), pp. 123–132.

[4] Diego Rosselli, Juan-David Rueda, Carlos Eduardo Diaz, et al. "Cost-effectiveness of kidney transplantation compared with chronic dialysis in end-stage renal disease". In: *Saudi journal of kidney diseases and transplantation* 26.4 (2015), p. 733.

[5] ORGAN India. "A study of the deceased organ donation environment in Delhi/NCR". In: *An initiative of the Parashar Foundation in partnership with MOHAN Foundation. Outline India* (2014).

[6] Oguzhan Alagoz et al. "The optimal timing of living-donor liver transplantation". In: *Management Science* 50.10 (2004), pp. 1420–1430.

[7] Burhaneddin Sandıkçı et al. "Alleviating the patient's price of privacy through a partially observable waiting list". In: *Management Science* 59.8 (2013), pp. 1836–1854.

[8] Sakine Batun et al. "Optimal liver acceptance for risk-sensitive patients". In: *Service Science* 10.3 (2018), pp. 320–333.

[9] Stefanos A Zenios, Lawrence M Wein, and Glenn M Chertow. "Evidence-based organ allocation". In: *The American journal of medicine* 107.1 (1999), pp. 52–61.

[10] Xuanming Su, Stefanos A Zenios, and Glenn M Chertow. "Incorporating recipient choice in kidney transplantation". In: *Journal of the American society of Nephrology* 15.6 (2004), pp. 1656–1663.

[11] Ashley Davis et al. "Characteristics of a simulation model of the national kidney transplantation system". In: *2013 Winter Simulations Conference (WSC)*. IEEE. 2013, pp. 2320–2329.

[12] Ashley E Davis et al. "Improving geographic equity in kidney transplantation using alternative kidney sharing and optimization modeling". In: *Medical Decision Making* 35.6 (2015), pp. 797–807.

[13] Burhaneddin Sandıkçı, Sait Tunç, and Bekir Tanrıover. "A new simulation model for kidney transplantation in the United States". In: *2019 Winter Simulation Conference (WSC)*. IEEE. 2019, pp. 1079–1090.

[14] Mohd Shoaib et al. "A discrete-event simulation model of the kidney transplantation system in Rajasthan, India". In: *Health Systems* 11.1 (2022), pp. 30–47.

[15] Rouba Ibrahim. "Sharing delay information in service systems: a literature survey". In: *Queueing Systems* 89.1 (2018), pp. 49–79.

[16] Rouba Ibrahim and Ward Whitt. "Wait-time predictors for customer service systems with time-varying demand and capacity". In: *Operations research* 59.5 (2011), pp. 1106–1118.

[17] Russell R Barton. "Tutorial: simulation metamodeling". In: *2015 Winter Simulation Conference (WSC)*. IEEE. 2015, pp. 1765–1779.

[18] Najiya Fatma et al. "Primary healthcare delivery network simulation using stochastic metamodels". In: *2020 Winter Simulation Conference (WSC)*. IEEE. 2020, pp. 818–829.

[19] Vaibhav Baldwa et al. "A combined simulation and machine learning approach for real-time delay prediction for waitlisted neurosurgery candidates". In: *2020 Winter Simulation Conference (WSC)*. IEEE. 2020, pp. 956–967.

[20] Kerala Network for Organ Sharing (KNOS). *Kerala Common Waitlist For Kidney Data*. `http://knos.org.in/Forms/patient/Waitlist_Kidney.aspx`. [Online; accessed 2022]. 2022.

[21] Tamil Nadu Network for Organ Sharing (KNOS). *Report-2013*. `https://tnos.org/pdf/report.pdf`. [Online; accessed 2022]. 2013.

[22] Regional Network for Organ Sharing (KNOS). *RNOS*. `https://www.rnos.org/faqtnos.aspx`. [Online; accessed 2022]. 2022.

[23] GR Lakshminarayana et al. "Hemodialysis outcomes and practice patterns in end-stage renal disease: Experience from a Tertiary Care Hospital in Kerala". In: *Indian journal of nephrology* 27.1 (2017), p. 51.

[24] JM Cecka et al. "Calculated PRA: initial results show benefits for sensitized patients and a reduction in positive crossmatches". In: *American Journal of Transplantation* 11.4 (2011), pp. 719–724.

[25] Georgi Abraham et al. "Evolution of renal transplantation in India over the last four decades". In: *NDT plus* 3.2 (2010), pp. 203–207.

[26] A Chandrashekar, S Ramakrishnan, and D Rangarajan. "Survival analysis of patients on maintenance hemodialysis". In: *Indian journal of nephrology* 24.4 (2014), p. 206.

[27] Mohan Foundation Report. *List of transplantation center available.* `https://www.mohanfoundation.org/transplant-centres/hospitals.asp?page=2&startpage=1&Category=Kidney&State=Kerala`. [Online; accessed 2022]. 2018.

[28] Census Commissioner of India. *Census of India3.* `http://censusindia.gov.in/` `http://census2011.co.in/`. [Online; accessed 2022]. 2011.

[29] William J Mcbride and Charles W Mcclelland. "PERT and the beta distribution". In: *IEEE Transactions on Engineering Management* 4 (1967), pp. 166–169.

[30] Peter A Flach. "ROC analysis". In: *Encyclopedia of machine learning and data mining.* Springer, 2016, pp. 1–8.

[31] Zachary Chase Lipton, Charles Elkan, and Balakrishnan Narayanaswamy. "Thresholding classifiers to maximize F1 score". In: *arXiv preprint arXiv:1402.1892* (2014).