Estimating the quality and success of kidney transplantation surgeries

Ali Rafieepouralavialavijeh

Department of Management Sciences, Faculty of Engineering, University of Waterloo
Email: arafieep@uwaterloo.ca

Clinical datasets are commonly restricted in size, and access to them is difficult and expensive in many cases, thus using novel methodologies such as ML of particular interest and application. Renal transplantation has become the treatment of choice for many patients with end-stage renal disease. Given the recent progress in data analysis areas, optimizing transplantation outcomes has become of paramount importance to increase graft survival time and the patients' quality of life. Therefore, a comprehensive understanding of the mechanisms before and after transplantation for both the host and the graft would make the way for an improved prognosis. In this report, a valuable set of data collected from a hospital in Tehran, Iran, were examined and compared with multiple similar articles in the literature regarding the methodologies used and the outcomes. Since Iran faces constraints in the medical equipment compared to other nations in the world, some comparisons were not possible, and the study parameters were partially different from the approach others had. This report has examined common methodologies in the literature such as Support Vector Machines, Neural Networks, and Decision Forests. It would recommend a ranking of features to facilitate and improve further analysis.

Keywords: Renal Transplantation, Machine Learning, Survival Analysis, Feature Selection, SVM, ANN, Decision Forests

1. INTRODUCTION

Artificial intelligence and, in particular, Machine learning has been increasingly employed recently to devise predictive models to diagnose various disease situations. By the nature of the kidney transplant population, which is heterogeneous, making good predictions of graft outcomes is extremely challenging. Several kidney graft outcome prediction models have been introduced using machine learning and are available in the literature for review. However, this report aims at evaluating and predicting the quality and success of kidney transplantation surgeries with a data set collected from a hospital in Tehran, Iran.

Various ML methods are compared in predicting the same outcome. Extensive efforts have been put into finding the minimal, most effective feature-set to facilitate the process and provide the practitioners' ability to prioritize the features in the case; not all clinical non-clinical features were available. This will allow our approach to be applicable in various situations where not all the most effective features are available. This data set comes with significant differences to the common metrics used in the literature, making this study of particular significance in improving the renal transplantation outcomes and the quality of life of the patients. Besides, extensive analysis has been performed on understanding the most effective features through the time intervals after the transplantation, which adds to the importance of our efforts.

The remainder of this report is organized as follows. First, a brief review is done to show the place of our approach in comparison to the published articles in the literature. Second, We focus on understanding the data set, the features, and the preprocessing required before making our models. Third, models are built, experiments are explained in detail, and the results are depicted via tables and images. Finally, we analyze the outcomes and propose a future direction. You can see high-level report progress in Figure. 1

2. LITERATURE REVIEW

Over the recent years, the increase in the prevalence of Chronic Kidney Disease (CKD) and end-stage kidney disease has ended in the growing demand for kidney replacement therapy [Wan+19]. Between the available kidney replacement therapy methodologies, kidney transplantation has shown significant superiority and survival rates [Kar+03]. The ability to forecast graft failure among various patients is crucial in organ allocation systems to minimize the number of patients returning to an already-burdened waiting list [Bro+12]. Models that can precisely predict graft failure following

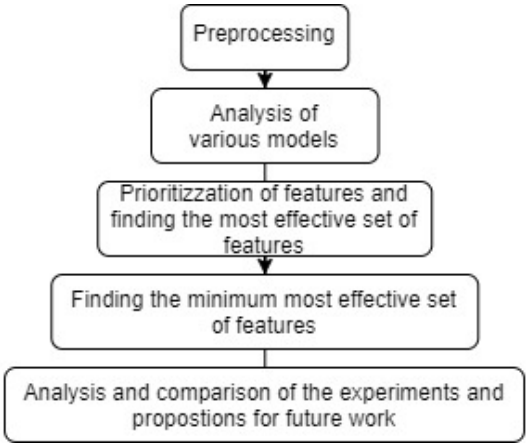

FIGURE 1. Report progress diagram

transplant may therefore help inform medical decision-making. Studies of several predictive models based on published regression methods have resulted in different predictive power. Therefore, alternative modeling approaches are motivated to tackle this challenge (for example, papers, please refer to [Moo+11], and [Fou+10].

Machine learning (ML) is an arsenal of methods whose theoretical structure may improve predictive performance over conventional statistical modeling. As a result, ML has evolved over recent decades and is already widely applied in many areas, including medical diagnose [Pat+09]. In particular, in renal transplantation, several kidney graft outcome models have been proposed using ML methodology [Bro+12], [AIG08]. Despite all these publications so far, the comparative performance of the models is still unclear, resulting in a narrower domain of applications in clinical practice [Sen+19].

Several studies implemented more than one ML method to predict the same outcome. Decision trees were the most commonly used ML method for predicted graft outcome, followed by artificial neural networks(ANN) and Bayesian belief networks [Sen+19]. Most studies included many input variables (features), and some used feature selection methods to identify the most important input variables before modeling their problem [Lin+08], [Top+18].

The review for this project report shows that researchers used various ML methods (artificial neural networks, decision trees, support vector machine, Bayesian belief networks) to develop the predictive models. Yet, the best ML method to develop a predictive model is a widely debated topic among practitioners [You15]. The overall conclusion is that no one method fits all data sets since everyone has different levels of complexities [Lor+19]. Investigators use multiple machine learning methods on single data sets to get around this uncertainty, and the best is chosen based on validation parameters. In particular [Alj+18], it used four machine learning methods to predict chronic kidney disease, emphasizing the importance of employing multiple ML methods on a single data set.

3. DATA

A data set collected from a different study on renal transplantation is considered for this project. The data set used includes 8 performance indicators and 12 outcome indicators, showing the performance of kidneys after the transplantation over time intervals of 3 months, 6 months, and so on. The 12 outcome indicators include graft rejection at various time intervals or infection at post-graft intervals, which are the quality of the graft. This data set includes the data gathered from 110 patients with 46 features, including clinical and non-clinical features.

Given that this data set is collected from a hospital in Tehran that did not have an organized data collection process, some experiments and measures lack data. Over a discussion I had with experts in this field, I was recommended to replace the blank data points with average (if the feature is continuous) and mode (if the feature is binary). Besides, some features had no numeric values, which had to be replaced with a numeric value such as sex which is replaced eventually with 1 and 2 for opposite genders.

Features (input variables) in this data set can be separated into two different categories, clinical and non-clinical. Clinical features are the direct result of experiments done with the patients before and after the transplantation in different intervals. These features show the performance of the kidney is doing its duties in the body (most clinical features represent the result of kidney performance when a specific solution is injected into the body. Thus, the name of each feature represents the solution name). In contrast, non-clinical features are the ones that are not the result of clinical experiments but gathered from the patients' profiles such as Age, Sex, blood type, height, weight, BMI, dialysis type, and dialysis duration.

Output variables are similarly defined to be specific performance of the kidney over injection of particular solutions to the body. Therefore, we have 8 performance matrices with which we evaluate the success or failure of kidney transplantation.

4. MODELING AND ANALYSIS

In this course project, various common machine learning methods in the literature are examined for the same outcome. Many features and methodologies that have shown their significance in the literature (as specified before) have been considered to make a firm model applicable to real-world problems. A such, each method is analyzed in separate subsections to make it easier to follow.

Kidney performance metrics	MAE of estimation
CG_eGFRD	6.9732
MDRD_eGFRD	9.6982
CG_eGFRM1	7.2229
MDRD_eGFRM1	10.3445
CG_eGFRM3	14.2664
MDRD_eGFRM3	13.4968
CG_eGFRM6	11.7082
MDRD_eGFRM6	12.4620

TABLE 1. Performance of ANN models on estimating output variables

4.1. Artificial Neural networks

In this project, we have used a sequential neural network with three hidden layers, each having 35, 25, and 35 neurons and a final layer with 8 neurons. The main metric used to evaluate the performance of our model is MAE (Mean average error), which is considered because of its prevalence in the literature. Fixing the random seed, optimization algorithms are checked Nadam showed promising results in train and error. We have split our data set into training and testing partitions, each with 70 and 30 percent of randomly sampled data. This ANN model is implemented using the Keras library in Python. To reduce the incidence of over-fitting in our model, we have proposed a drop-out layer to deactivate 30 percent of the neurons randomly after each layer. To facilitate faster convergence, all the data is normalized before being fed to the model. Activation functions on each neuron are selected to be leaky-relu to allow the model to correct its past performance. The figure 2 illustrates the training process. The table 1represents the result of estimating the output variables with artificial neural networks.

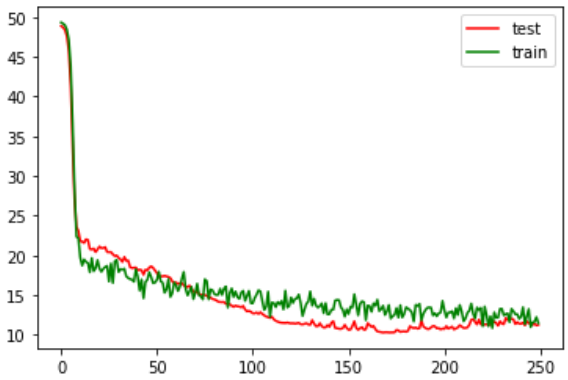

FIGURE 2. The test and training data; the horizontal axis represents the repetition number, and the vertical axis represents the Mean Absolute Error

.

Kidney performance metrics	MAE of estimation
CG_eGFRD	8.5608
MDRD_eGFRD	4.0262
CG_eGFRM1	7.9610
MDRD_eGFRM1	8.2444
CG_eGFRM3	16.4429
MDRD_eGFRM3	12.8444
CG_eGFRM6	15.8503
MDRD_eGFRM6	13.8095

TABLE 2. Performance of random forest models on estimating output variables

Kidney performance metrics	MAE of estimation
CG_eGFRD	9.1023
MDRD_eGFRD	7.3053
CG_eGFRM1	10.2225
MDRD_eGFRM1	7.9559
CG_eGFRM3	14.1370
MDRD_eGFRM3	13.1454
CG_eGFRM6	11.5357
MDRD_eGFRM6	14.8491

TABLE 3. Performance of SVM model with rbf kernel function on estimating output variables

4.2. Random forests and decision trees

Given the prevalence of this branch of classical machine learning applications in clinical diagnosis extensively defined in the literature review, in this project, random forests and decision trees are considered for estimating the outcome variables and prioritizing and ranking input variables. The most important features and benefits of using this branch of ML are that they are interpretative, they do not require scaling the data, filling the missing data, and correcting them. To implement random forests, the sci-kit-learn library is used within python to facilitate the process, and most hyper-parameters are selected automatically. The table 2represents the estimation performance of random forests on our data.

4.3. Support vector machines

Following the objective of comparing various machine learning methodologies, support vector machines (SVM) are applied to our model is employed. Given that SVMs have kernel functions that transform the feature space of our input variables, the three most commonly used kernel functions are applied, and the results are shown in tables3, 4 and 5

5. FEATURE SELECTION AND PRIORITIZATION

Feature selection methods are intended to reduce the number of input variables to those believed to be most useful to a model to predict the target variable.

Kidney performance metrics	MAE of estimation
CG_eGFRD	9.7201
MDRD_eGFRD	8.0249
CG_eGFRM1	10.47088
MDRD_eGFRM1	6.9541
CG_eGFRM3	16.2440
MDRD_eGFRM3	15.9740
CG_eGFRM6	16.3300
MDRD_eGFRM6	16.3518

TABLE 4. Performance of SVM model with poly kernel function on estimating output variables

Kidney performance metrics	MAE of estimation
CG_eGFRD	9.2124
MDRD_eGFRD	7.7630
CG_eGFRM1	10.2536
MDRD_eGFRM1	7.6115
CG_eGFRM3	14.4201
MDRD_eGFRM3	13.4094
CG_eGFRM6	12.0352
MDRD_eGFRM6	15.6293

TABLE 5. Performance of SVM model with linear kernel function on estimating output variables

"Feature selection is primarily focused on removing non-informative or redundant predictors from the model." [K+13]. Some predictive modeling problems have many variables that can slow the development and training of models and require a large amount of system memory. Additionally, the performance of some models can degrade when including input variables that are not relevant to the target variable.

As specified in the literature review section, feature selection and prioritization could enhance the model's performance and, in the meantime, make the input variables as concise as possible. In this part, directly in line with the literature, we would use various methods to select the most important feature set and then compare the results of each method.

5.1. Feature selection based on random forest structure

As specified on Page 487 [K+13], some models are naturally resistant to non-informative predictors. Tree-based models, for example, intrinsically conduct feature selection. Random Forests is a kind of Bagging Algorithm that aggregates a specified number of decision trees. The tree-based strategies used by random forests naturally rank by how well they improve the purity of the node or a decrease in the impurity (Gini impurity) over all trees. Nodes with the greatest decrease in impurity happen at the start of the trees, while notes with the slightest decrease in impurity occur at the end of trees. Thus, we can create a subset of the most important features by pruning trees below a particular node. The figure 3 represents the importance of each feature during the random-forest process for predicting the most important variables. The table 6 provides a ranking for the most important features affecting the performance of our random forest model. Interestingly, a mix of clinical and non-clinical features is ranked among the most important factors, which means both groups play an important role in estimating the patients' quality of life after transplantation.

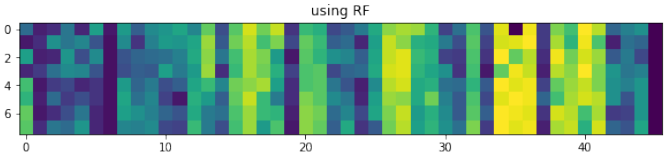

FIGURE 3. Importance of each feature for estimating the outcome variables, the most important features are the darkest ones.

5.2. Feature selection based on Linear SVM structure

Given the linear function attributed between the features, and the scale of each of the multipliers, we could propose another feature selection approach. The figure 4shows the heatmap of the most important features based on the Linear SVM structure. The table also illustrates the ranking of the most important features for estimating each of the outcome variables.

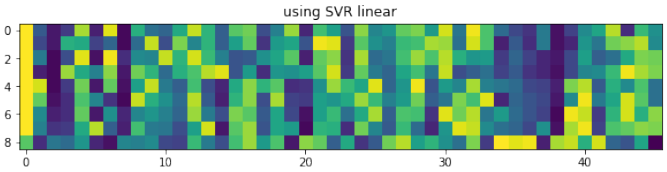

FIGURE 4. Importance of each feature using linear SVM structure for estimating the outcome variables, the most important features are the darkest ones

5.3. Recursive Feature Elimination (RFE)

Recursive Feature Elimination, or RFE for short, is a popular feature selection algorithm. RFE is popular because it is easy to configure and use and effective at selecting those features in a training data set that are more or most relevant in predicting the target(output) variable. There are two important configuration options when using RFE: the choice in the number of features to select and the choice of the algorithm used to help choose features. Both of these hyperparameters can be explored, although the performance of the method is not strongly dependent on these hyperparameters being configured well.

Rank	CG_eGFRD	MDRD_eGFRD	CG_eGFRM1	MDRD_eGFRM1	CG_eGFRM3	MDRD_eGFRM3	CG_eGFRM6	MDRD_eGFRM6
1	TACCDRatio	TACCDRatio	TACCDRatio	TACCDRatio	TACCDRatio	TACCDRatio	TACCDRatio	TACCDRatio
2	BMI	BMI	Sex	BMI	BMI	BMI	BMI	BMI
3	Sex	DGF	BMI	DGF	Donor_type	IBW	Donor_type	TACDkg
4	Weight	Sex	DGF	Sex	Sex	TACDkg	Sex	IBW
5	Donor_type	ATG	mTOR_level	mTOR_level	IBW	RatioD	IBW	RatioD
6	Age	Donor_type	Donor_type	TACDD	mTOR_level	Donor_type	TACDD	Sex_Similarity
7	PCR2group	D_Inotrpe	TACDD	TACDD	Height	Dialysis_type	PCR2group	Sex
8	mTOR_Level	D_Weight	Dialysis_Type	IBW	Weight	TACCD	Height	mTOR_Level

TABLE 6. Ranking most important and detrimental input features in estimating each of kidney performance measures using random forest structure

Rank	CG_eGFRD	MDRD_eGFRD	CG_eGFRM1	MDRD_eGFRM1	CG_eGFRM3	MDRD_eGFRM3	CG_eGFRM6	MDRD_eGFRM6
1	SCrD	SCrD	Age	SCrD	SCrD	SCrD	SCrD	SCrD
2	Age	Kidney_trans	SCrD	Age	IBW	D_Age	D_Age	D_Age
3	IBW	D_Age	IBW	Kidney_trans	Age	LOS	IBW	LOS
4	LOS	LOS	LOS	LOS	LOS	D_DM	Dialysis_type	BMI
5	D_Age	ATG_tota_dose	Kidney_trans	Sex	D_Age	Age	LOS	D_HTN
6	Kidney_trans	Age	D_Age	D_Age	Donor_type	TACCDRatio	Age	D_DM
7	D_HTN	Dialysis_type	Donor_type	D_HTN	D_DM	ATG_tota_dose	CYP3A4_22_Genotype	CYP3A4_22_Genotype
8	Height	D_Weight	Dialysis_Type	IBW	Weight	TACCD	Height	mTOR_Level

TABLE 7. Ranking most important and detrimental input features in estimating each of kidney performance measures using linear SVM structure

5.3.1. Using SVM estimator in RFE algorithm

Figures 5 to 12 illustrate the cross-validation score for a different number of features. As mentioned before, more features do not necessarily mean better outcomes in the process. The figures clearly show that there are a minimum number of features for each outcome variable for which our model performs best. This will allow practitioners to minimize their clinical experiments through time and do them when needed for them to be done. The table 8 shows the outcome variable estimation performance results after using the RFE algorithm with the SVM estimator.

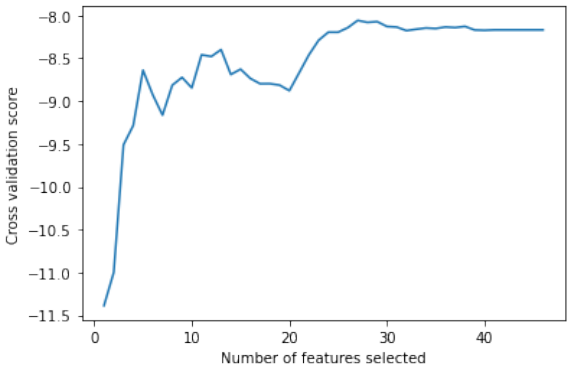

FIGURE 5. The score achieved with a different number of features when estimating 1st outcome variable using SVM estimator

5.3.2. Using random forests estimator in RFE algorithm

Figures 13 to 20 illustrate the cross-validation score for a different number of features. As mentioned before, more features do not necessarily mean better outcomes in the process. The figures clearly show that there is a minimum number of features for each outcome variable for which our model performs best. This will allow

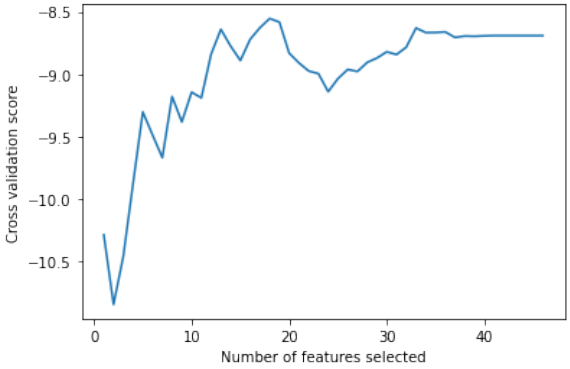

FIGURE 6. The score achieved with a different number of features when estimating 2nd outcome variable using the SVM estimator

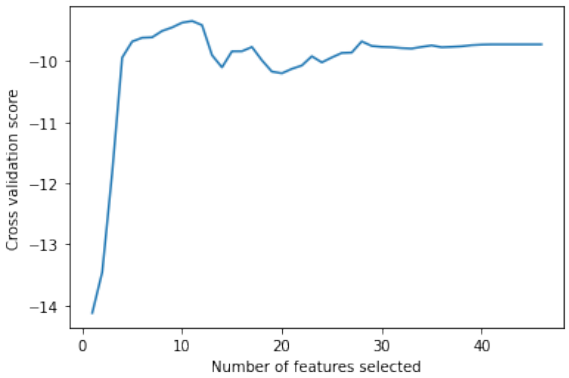

FIGURE 7. The score achieved with a different number of features when estimating 3rd outcome variable using the SVM estimator

practitioners to minimize their clinical experiments through time and do them when needed for them to be done. The tab 9shows the outcome variable estimation performance results after using the RFE algorithm with

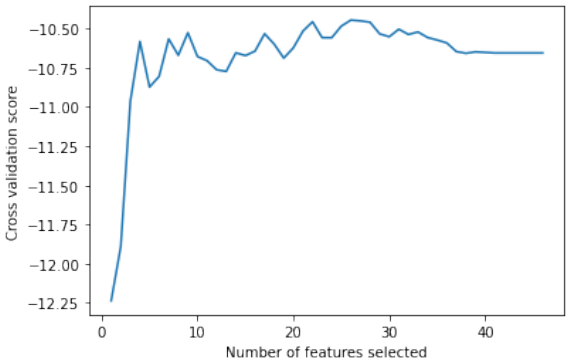

FIGURE 8. The score achieved with a different number of features when estimating the 4th outcome variable using the SVM estimator

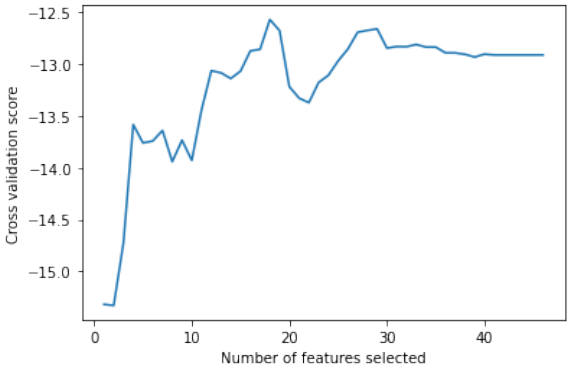

FIGURE 9. The score was achieved with a different number of features when estimating the 5th outcome variable using the SVM estimator

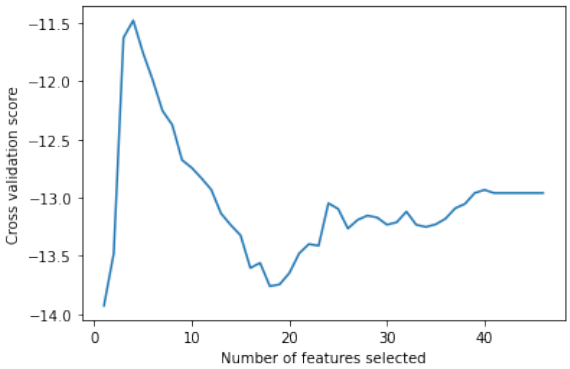

FIGURE 10. The score was achieved with a different number of features when estimating the 6th outcome variable using the SVM estimator

the SVM estimator.

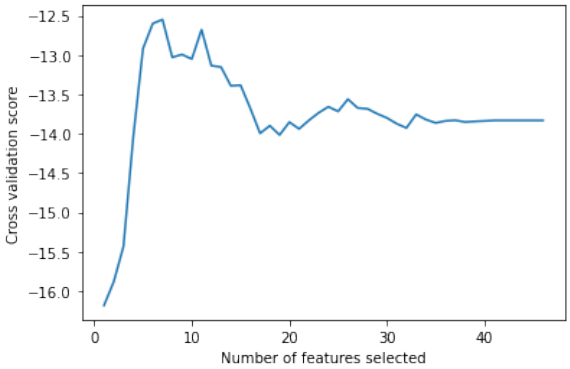

FIGURE 11. The score was achieved with a different number of features when estimating the 7th outcome variable using the SVM estimator

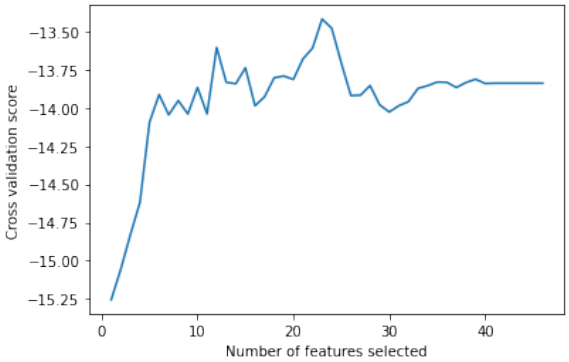

FIGURE 12. The score was achieved with a different number of features when estimating the 8th outcome variable using the SVM estimator

Kidney performance metrics	MAE of estimation
CG_eGFRD	7.5026
MDRD_eGFRD	7.7012
CG_eGFRM1	11.6803
MDRD_eGFRD1	7.5905
CG_eGFRM3	14.4201
MDRD_eGFRD3	13.4094
CG_eGFRM6	16.0838
MDRD_eGFRD6	16.7477

TABLE 8. Performance of model achieved after using RFE algorithms with SVM estimator

Kidney performance metrics	MAE of estimation
CG_eGFRD	5.6020
MDRD_eGFRD	2.0374
CG_eGFRM1	8.8831
MDRD_eGFRD1	8.7807
CG_eGFRM3	15.6330
MDRD_eGFRD3	12.9042
CG_eGFRM6	15.2464
MDRD_eGFRD6	14.7995

TABLE 9. Performance of model achieved after using RFE alforithm with random forests estimator

	ANN	RF	SVM-RBF	SVM-POLY	SVM-LINEAR	RFE-SVM	RFE-RF
CG_eGFRD	6.9732	8.5608	9.1023	9.7201	9.2124	7.5026	**5.6020**
MDRD_eGFRD	9.6982	4.0262	7.3053	8.0249	7.7630	7.7012	**2.0374**
CG_eGFRD1	7.2229	**7.9610**	10.2225	10.4708	10.2536	11.6803	8.8831
MDRD_eGFRD1	10.3445	8.2444	7.9559	**6.9541**	7.6115	7.5905	8.7807
CG_eGFRD3	14.2664	16.4429	**14.1370**	16.2440	14.4201	14.4201	15.6330
MDRD_eGFRD3	13.4968	**12.8444**	13.1454	15.9740	13.4094	13.4094	12.9042
CG_eGFRD6	11.7082	15.8503	11.5357	16.3300	12.0352	**11.0838**	15.2464
MDRD_eGFRD6	**12.4620**	13.8095	14.8491	16.3518	15.6293	16.7477	14.7995
Overall	12.4620	11.1495	11.0316	12.5087	11.2918	11.0460	**10.4858**

TABLE 10. Performance of different models in comparison to each other, best performances in bold

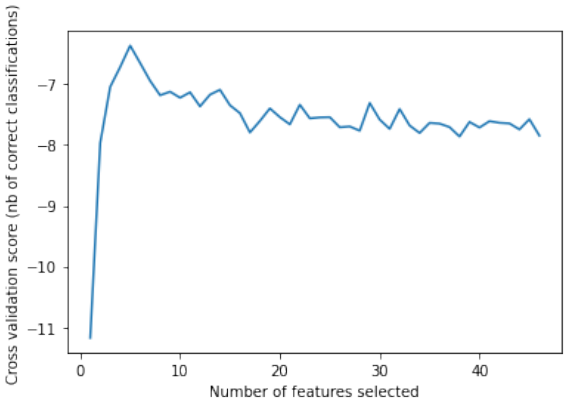

FIGURE 13. The score achieved with a different number of features when estimating 1st outcome variable using random forests estimator

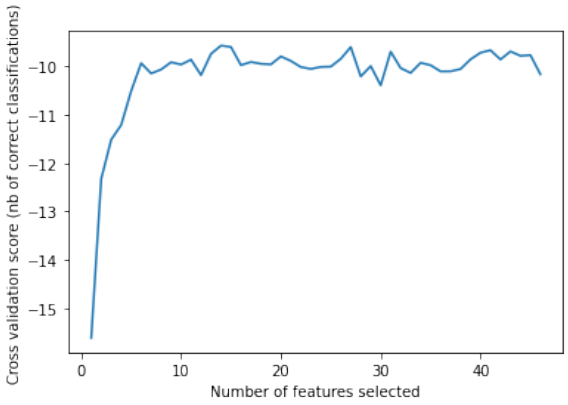

FIGURE 15. The score achieved with a different number of features when estimating 3rd outcome variable using random forests estimator

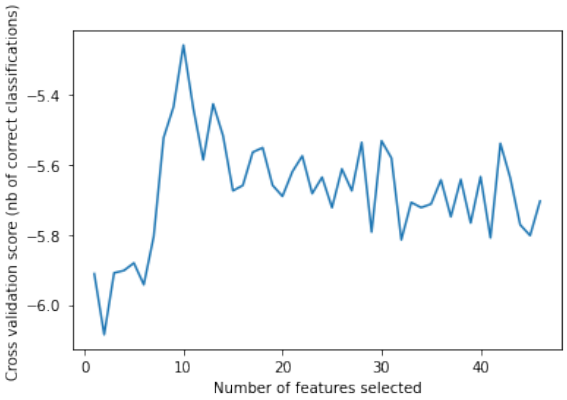

FIGURE 14. The score achieved with a different number of features when estimating 2nd outcome variable using random forests estimator

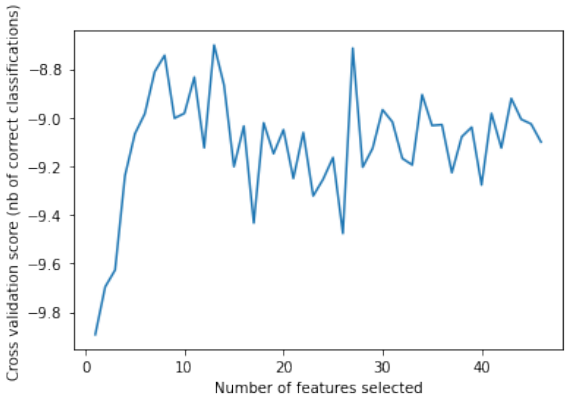

FIGURE 16. The score achieved with a different number of features when estimating the 4th outcome variable using random forests estimator

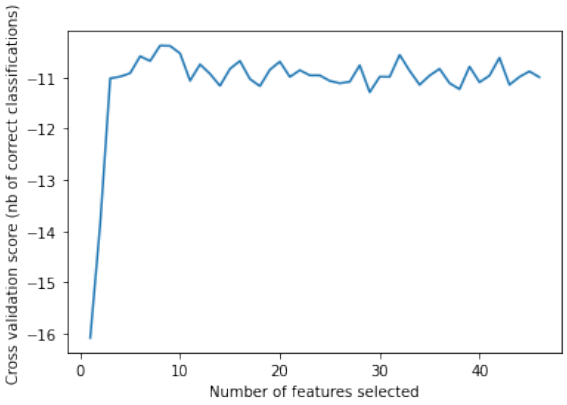

FIGURE 17. The score achieved with a different number of features when estimating the 5th outcome variable using random forests estimator

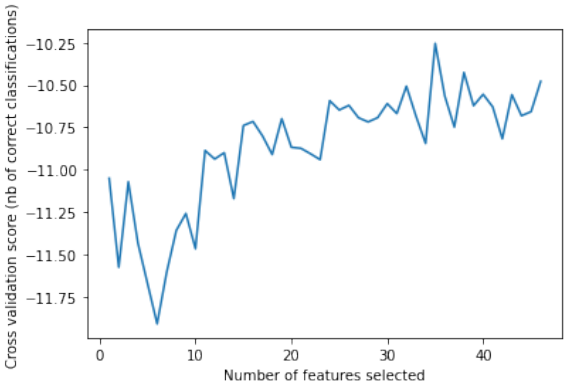

FIGURE 18. The score achieved with a different number of features when estimating the 6th outcome variable using random forests estimator

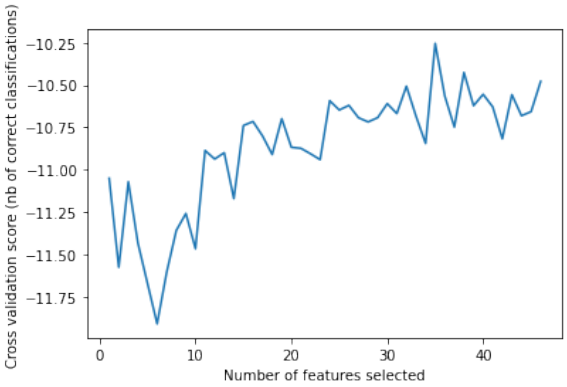

FIGURE 19. The score achieved with a different number of features when estimating the 7th outcome variable using random forests estimator

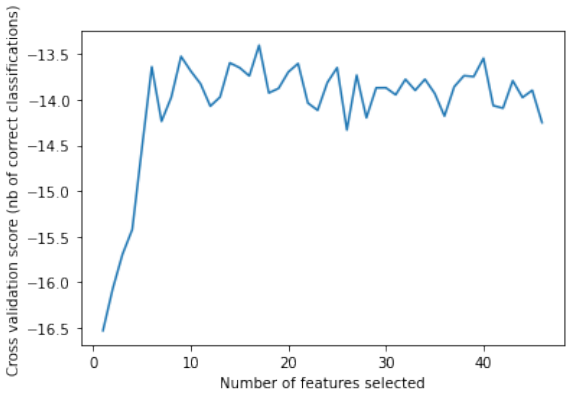

FIGURE 20. The score achieved with a different number of features when estimating the 8th outcome variable using random forests estimator

6. CLUSTERING THE PATIENTS

Having a clear understanding of various clusters of patients gives decision-makers tools to make policies for each of them. For instance, having a set of similar kidney matching would let us know if we should provide similar services to that particular group or perhaps whether matching is a good one or should be further studied to be improved.

In order to find a fairly just number of clusters, elbow method is used and the result is shown in figure 21. As can be seen the optimal number of clusters is 11. Since the clusters are multi-dimensional features are selected two by two for the sake of representation and as a result figures 22, 23, and 24 are achieved.

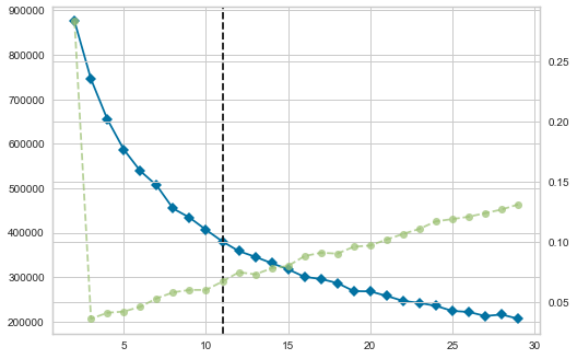

FIGURE 21. Elbow method and the optimal number of clusters

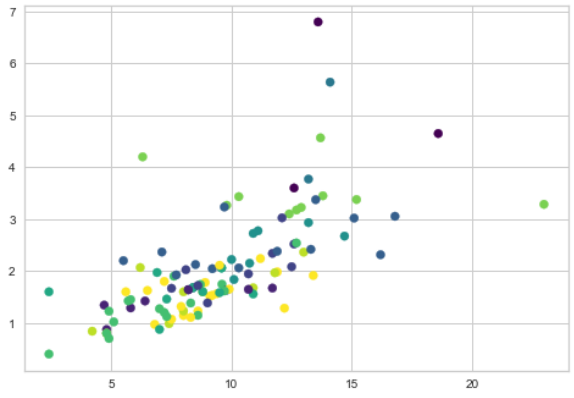

FIGURE 22. Two features, TACCD and TACCDkg are selected from various clusters and this plot is illustrated

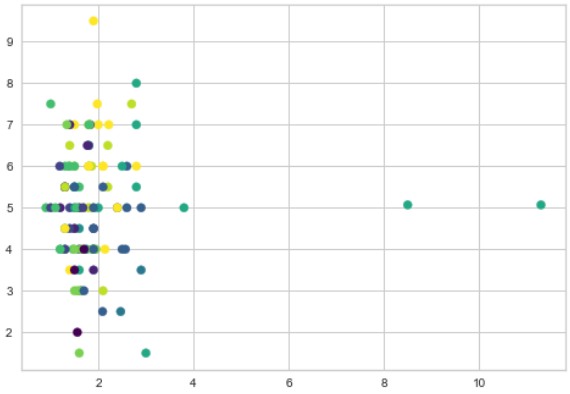

FIGURE 23. Two features, BMI and SCrD are selected from various clusters and this plot is illustrated

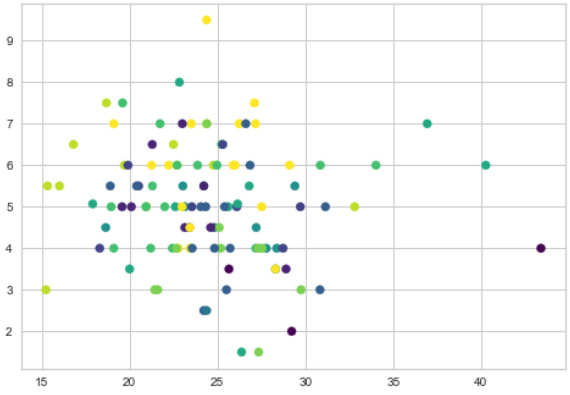

FIGURE 24. Two features, IBW and SCrD are selected from various clusters and this plot is illustrated

7. CONCLUSIONS

A literature review shows extensive applications of machine learning methods over statistical methods in medical sciences recently. Most of the authors focused on data sets archived from hospitals with enough laboratory equipment. This equipment assures medical decision-makers that they are making correct decisions about matching two patients. This will boost the quality of the life of the patients and results in fewer system costs. However, given the limitations in hospitals in Iran, one of the most important aspects of renal transplantation, the kidney tissue similarity index (HLA), is not available. As such, practitioners employ a mix of features including results from clinical experiments, and non-clinical features. Given these limitations, the importance of having an intelligent system is revealed more than before.

Effective application of these features, which might vary extensively from one hospital to another, is crucial. This study has focused on different methodologies widely used in the literature to find the most accurate method to forecast kidney performance metrics after the surgery. Besides, feature selection and ranking have been made with various methods to show minimal features with the best performance.

The table 10 compares the performance of each of the models in comparison to each other. Some models perform better in comparison to other models in certain time intervals. For example, take the CG_eGFRD experiment done in month three (CG_eGFRD3); the performance of the SVM-RBF model is superior compared to the rest in this case. Overall, however, the RFE_RF method had the best performance, and interestingly the average of best performances in each time interval is approximately the same as our overall best performance. This reveals that if only one method is going to be applied during this process RFE-RF model is the best option. Else, different methods could be applied on the data set for different time intervals.

8. FUTURE WORK

Extensive feature analysis in different time intervals could yield diverse managerial insights. As such, future studies could focus on the relevance of each feature set to each time interval and extracting insights on how is that in different time intervals, different features are more important. Perhaps predicting the residual errors with a set of different regression models could also result in better performance, which could be considered a future work.

ACKNOWLEDGEMENTS

I would like to thank my brother for his support at every stage, without whom hard times could have been intolerable for me. I also thank my friend Behzad who assisted me with the original data source directly from a hospital in Tehran, and for his kind support about medical aspects of my project, which I was not informed about.

REFERENCES

[Kar+03] Vincent H Karam et al. “Quality of life in adult survivors beyond 10 years after liver, kidney, and heart transplantation”. In: *Transplantation* 76.12 (2003), pp. 1699–1704.

[AIG08] Ahmed Akl, Amani M Ismail, and Mohamed Ghoneim. “Prediction of graft survival of living-donor kidney transplantation: nomograms or artificial neural networks?” In: *Transplantation* 86.10 (2008), pp. 1401–1406.

[Lin+08] Ray S Lin et al. “Single and multiple time-point prediction models in kidney transplant outcomes”. In: *Journal of biomedical informatics* 41.6 (2008), pp. 944–952.

[Pat+09] Vimla L Patel et al. “The coming of age of artificial intelligence in medicine”. In: *Artificial intelligence in medicine* 46.1 (2009), pp. 5–17.

[Fou+10] Yohann Foucher et al. “A clinical scoring system highly predictive of long-term kidney graft survival”. In: *Kidney international* 78.12 (2010), pp. 1288–1294.

[Moo+11] Jason Moore et al. “Development and evaluation of a composite risk score to predict kidney transplant failure”. In: *American journal of kidney diseases* 57.5 (2011), pp. 744–751.

[Bro+12] Trevor S Brown et al. “Bayesian modeling of pretransplant variables accurately predicts kidney graft survival”. In: *American journal of nephrology* 36.6 (2012), pp. 561–569.

[K+13] Max Kuhn, Kjell Johnson, et al. *Applied predictive modeling*. Vol. 26. Springer, 2013.

[You15] Ahmed H Yousef. “Extracting software static defect models using data mining”. In: *Ain Shams Engineering Journal* 6.1 (2015), pp. 133–144.

[Alj+18] Ahmed J Aljaaf et al. “Early prediction of chronic kidney disease using machine learning supported by predictive analytics”. In: *2018 IEEE congress on evolutionary computation (CEC)*. IEEE. 2018, pp. 1–9.

[Top+18] Kazim Topuz et al. “Predicting graft survival among kidney transplant recipients: A Bayesian decision support model”. In: *Decision Support Systems* 106 (2018), pp. 97–109.

[Lor+19] Ana C Lorena et al. "How Complex is your classification problem? A survey on measuring classification complexity". In: *ACM Computing Surveys (CSUR)* 52.5 (2019), pp. 1–34.

[Sen+19] Sameera Senanayake et al. "Machine learning in predicting graft failure following kidney transplantation: A systematic review of published predictive models". In: *International Journal of Medical Informatics* 130 (2019), p. 103957. ISSN: 1386-5056. DOI: https://doi.org/10.1016/j.ijmedinf.2019.103957. URL: https://www.sciencedirect.com/science/article/pii/S1386505619302977.

[Wan+19] Tongtong Wang et al. "Chronic kidney disease among US adults with type 2 diabetes and cardiovascular diseases: A national estimate of prevalence by KDIGO 2012 classification". In: *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 13.1 (2019), pp. 612–615.