Kidney Transplant Risk Predictor

500065071, 490096600, 500065071, 500050686, 500614994, 500448045, 490293759

GitHub code repository is here

This version was compiled on June 1, 2022

Executive Summary:

End-Stage Kidney Disease patients awaiting kidney transplantation and recent transplant recipients are exposed to the risk of death from surgical complications in the early post-transplantation stage [1-3]. To make the most of kidney resources and optimise the chance of survival, clinicians need reliable tools to foresee the possible outcome for the post-transplantation kidney. And they can benefit from the recent genomic technology advancement and the computation power of data science. Statistically, we discovered that the patient's top 10 relevant genes and serum creatinine levels are the most significant variables in predicting kidney graft conditions. And we managed to build a predictive model of 81.4% accuracy. Not all medical practitioners are familiar with gene data interpretation, so we designed an interactive application, Kiddo, on these findings. Kiddo provides clinicians with a predictive platform for the stability of transplanted kidneys. In the project breakdown, we selected datasets with descriptions of the transcriptional genes of kidney transplant biopsies and their clinical information. We began with an initial analysis of multiple datasets, repeated cross-validation methods, and algorithm optimisation. We used biopsy results in the model building because our result shows that the analysis of kidney biopsy is far more accurate than the one from a blood test. We selected the top 10 genes to optimise the regression model by ranking the most relevant genes with adjusted p-value and the clinical data, serum creatinine. Finally, we used the confusion matrix to tune the model performance and got 81.4% accuracy.

Kidney | Adjust p-value | Serum creatinine | Regression model | Confusion matrix

Background and aim

Background. End-Stage Kidney Disease (ESKD) is the most severe form of Chronic Kidney Disease where the kidney is almost entirely dysfunctional, requiring any patient with ESKD to survive either on regular dialysis or from kidney transplantation. Multiple studies have shown that the survival rate of patients is better with kidney transplantation than with maintenance dialysis, despite an increased risk of death in the early post-transplant period [1-3]. According to the study of 14,241 primary kidney-only transplants performed in Australia and New Zealand between April 1997 and June 2017, long-term results show a mortality rate of up to 20% over time, with a corresponding complication rate of 17%. The study finds patients with early Acute Rejection (AR) are more vulnerable to functional death and graft loss due to chronic allograft nephropathy (CAN) and recurring AR [4]. Upon recognising menacing challenges of complication risk, we presume that there is an emerging demand to maximise the survival of kidney transplantation recipients from graft loss and other surgical complications in their early post-transplant period.

Aim. The relatively recent introduction of genetic testing paves the way for the broader application of data analytics in clinical decision making. We acknowledge the feasibility of a gene data-driven product in response to the emerging medical demand discussed in the background. This research project aims to offer clinicians, whether with experience in clinical practice on genetic information or not, a simple and reliable tool that predicts the prognosis of kidney transplant patients with minimal input. The tool is presented on the interactive Shiny app platform with detailed user instructions, and it is driven by a comprehensive prediction model. We realised such a model sources knowledge from multiple disciplines. Rooting in the understanding of patients' genetic and clinical information from a biomedical science perspective, we implemented methods from data science to build the predictive model and validated its accuracy with various evaluations. We believe that the platform can support clinician decisions and prepare them in advance by mapping out the most likely outcome for kidney transplant patients.

Methods

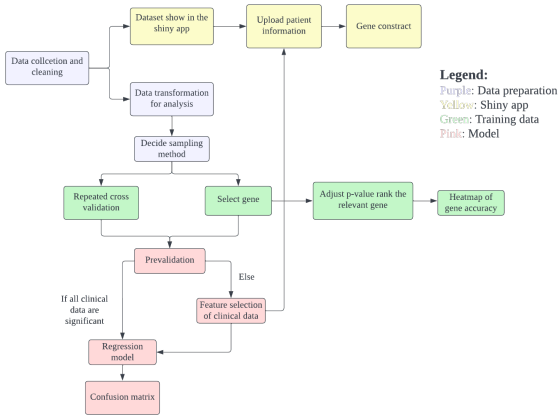

Fig. 1. Overview of project

Data collection and cleaning

We initiated the project on two datasets: GSE9493[5][6] and GSE14346[7][8], for their comprehensive recordings of patient information. GSE9493 collects data on patient gene expression, type of rejection, and clinical information such as sex and age by biopsy sampling. The dataset provided by Novartis Pharma AG, was used to analyse conservative rejection profiles and molecular pathways shown in kidney transplant biopsies. The GSE14346 dataset uses blood tests to collect patient genetic and peripheral blood samples. The dataset is provided by the Department of Paediatric Nephrology at Stanford University. Researchers analysed the dataset to identify biomarkers that can be used to diagnose and predict acute rejection. The two datasets are named and scaled differently. Therefore, we need to clean the data for consistency in future steps. Firstly, we changed gene IDs to gene names and removed all missing values in the datasets. Secondly, we renamed the column that the transplant results are either Reject or Stable. Moreover, we discovered that the measurement of gene values is scaled differently between two datasets and hence significantly impacts the quantitative metrics. We normalised the data by changing all the gene values to the same relative range of measurement

before building any model. (Normalisation graph shown in the appendix)

Research questions analysis

Sampling method. We supposed that different methods of sampling a patient's genes would affect the accuracy of the predictive model. Therefore, for the diversity and comparison, we selected two datasets, GSE14346, sampling from blood tests, and GSE9493, sampling from the biopsy. Due to thousands of gene recordings in both datasets and the limited computing power, we screened the top 2000 genes from each dataset by the large variance method and found all matching genes by gene names. We compared the results with KNN, SVM and random forest algorithms. With KNN, the prediction accuracy of kidney biopsy is 79%, and of blood test is 74%; with SVM, the accuracy of kidney biopsy results is 84%, and of blood test results is 80%; with random forest, the accuracy is 74% for kidney biopsy and 78% for a blood test. Furthermore, with more genes, the random forest algorithm also produces more accurate predictions of kidney biopsy results than blood test results. Therefore, we concluded that biopsy sampling is the more accurate method and should be thus featured in the model building. (SVM boxplot available in the appendix)

Repeated cross-validation. Once again, a large number of gene recordings in datasets and the various available models require evidence-based gene pooling and model selection. First, we used topTable to obtain various data for each gene (e.g.: *logFC*, adjusted p-value), and we ranked genes by adjusted p-values. According to the Volcano Plot in the appendix, we concluded that more than 9,000 genes were significant by logFC. However, due to the limitation of the computing power, we can only test the most powerful 200 genes in the following steps. We performed gene selection and model selection simultaneously, testing various genetic combinations with different models (KNN with different Ks, SVM, random forest) for cross-validation. We then presented all final results in a heatmap. The down-right corner highlighted in the heatmap indicates the highest accuracy rate with the top 10 significant genes and the SVM model, building a solid assumption for the prevalidation step.

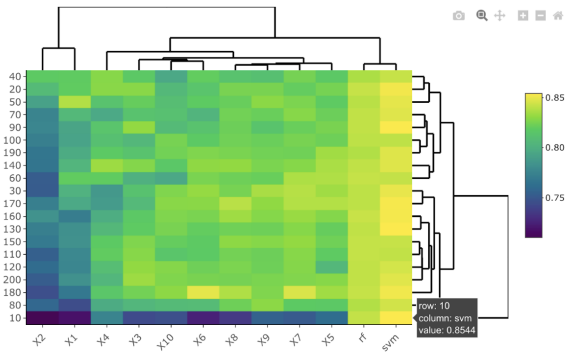

Fig. 2. Heatmap

Prevalidation method. The goal of prevalidation is to build a "fairer" prediction rule than in the cross-validation from the same dataset without biased results towards the new rule or the problem of overfitting [9]. When we constructed the new predictor, we first split the dataset of the top 10 significant genes from the cross-validation into five groups. Then we put one group away and ran the other four groups in the proven SVM model for predictive outcomes (*Rejection or Stable*) of the left-out group. We repeated this step until each group had a predictor fitting all other groups but themselves. In this way, we set up a new prediction rule while the clinical results from the left-out group would not impact its predictor. We set this predictive outcome as our *apv* (Adjusted Present Value, Rejection or Stable predictor from the SVM model). Then we choose a fitted linear or logistic regression model to compare the apv with the existing clinical outcomes. If the coefficient of the prevalidated predictor is significant, it proves that the new prediction rule significantly enhanced the predictive model.[9] We will use this hypothesis in combination with other clinical variables to build the final regression model. The graph of the prevalidation process is in the appendix.

Regression model. Before the regression model building, we further cleaned the dataset by selecting the clinical data columns: age, sex, and serum creatinine, and converted them to the correct variable types, for instance, extracting the numeric part of age and serum creatinine. For the regression model, we ought to find the most significant variables among sex, age, serum creatinine, and apv. We used stepwise algorithms to make the feature selection and pick the best model by AIC. The *step()* function helped us with a stepwise regression, which selected some variables featured in the result section.

Model Evaluation. After the supposedly successful regression model building, the model evaluation will help us judge its performance retrospectively. We proposed two model evaluation strategies: the Confusion Matrix and the Receiver Operating Characteristic method (ROC). The Confusion Matrix is one of the most powerful and commonly used, which allows us to compute the accuracy. The Confusion Matrix in R is a table that categorises the predictions against the actual values. It visualises results in two dimensions, one for the predicted and the other for the actual values. The matrix outputs the model accuracy by calculating truly predicted values over total values. Besides, we proposed ROC as the other evaluation strategy in logistic regression, in which the values are predicted based on probability. In this project, we constructed the curve by plotting the true positive rate (TPR) against the false positive rate (FPR). As a baseline, a random classifier is expected to give points lying along the diagonal (FPR = TPR). The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test. And we can calculate AUC(area under the curve) to determine the score of our model.

Results

Regression model. As discussed in the method section, we chose the stepwise algorithm to screen the most significant variables among sex, age, serum creatinine and apv, and select the best-fitted model by AIC. In the backward and forward model illustration (see appendix), the serum creatinine and apv have the lowest p-value: 0.001 for serum creatinine and 0.048 for the apv(stable). The low p-values indicate their higher significance than other variables. Besides, AIC is a standard measure of the goodness of a statistical model's fit. The model with serum creatinine and apv has an AIC of 48.869, the lowest of all combinations. Since the lower the AIC measures, the better the model fits. We concluded that our final regression model is with serum creatinine level and apv results

while finding the sex and age variables relatively insignificant to predictive outcomes.

Our statistical findings of serum creatinine level as a significant clinical predictor of prognosis in kidney transplant patients resonate with the conclusion of a 2003 biomedical study, in which the author discovered that "6- and 12-month serum creatinine levels, as well as the change between 6 and 12 months, are strongly associated with long-term graft survival"[10]. This study also corroborates the validity of our model.

```
Call:
glm(formula = outcome ~ serum_creatinine + apv, family = binomial(link = "logit"),
    data = combined_df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9383  -0.5590  -0.1385   0.7242   2.0479

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       4.622844   1.471628   3.141 0.001682 **
serum_creatinine -0.034317   0.009897  -3.467 0.000526 ***
apvstable         2.820808   1.286167   2.193 0.028294 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 74.15  on 58  degrees of freedom
Residual deviance: 44.97  on 56  degrees of freedom
  (4 observations deleted due to missingness)
AIC: 50.97

Number of Fisher Scoring iterations: 7
```

Fig. 3. Regression model

The following illustration shows the summary of our logistic regression model. The graph shows that the low p-values of serum creatinine and apv significantly influence model outcomes. In conclusion, our regression model states:

$$\text{Outcome} = 4.93 - 0.03(\text{serum\_creatinine}) - 2.62(\text{apvstable})$$

Model Evaluation.

1. **Confusion Matrix**

To evaluate the performance of the final model, we calculated values with caret::ConfusionMatrix and presented the values in an intuitive graph with *plot()* function. It allocates model predictions with respect to their actual values of rejection or stable. As shown in the graph, the prediction model returns 34 true rejections and 15 true stables out of 59 cases, with an accuracy of 81.4%.

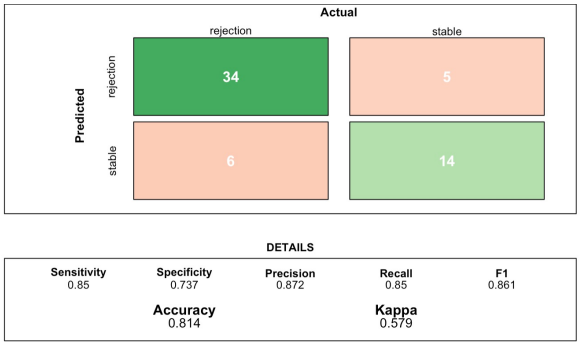

Fig. 4. Confusion matrix

2. **AUC & ROC Evaluation**

In the ROC model evaluation, we took the cut off of the probability as 0.5, i.e. any probability value greater than 0.5 will be accounted as 1 (stable), and any value less than 0.5 will be accounted as 0 (rejection). The plot shows that the x-axis is the false positive rate, and the y-axis is the true positive rate. The points represent the tradeoff between true positive and false positive. Looking at the graph, we can choose the optimal threshold depending on how many false positives (FP) we are willing to accept. After that, we can also calculate the area under the ROC curve (AUC); a single number summarises the model's accuracy. Typically we want to achieve a score higher than 0.75. And our final regression model scores 0.793, suggesting its high accuracy.

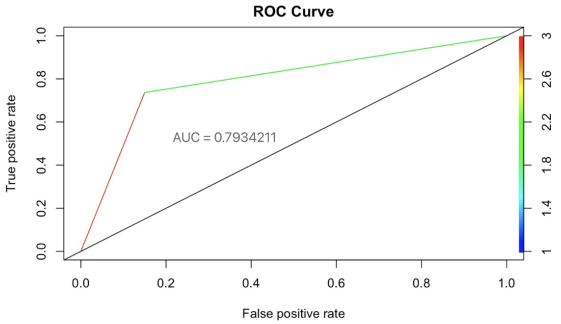

Fig. 5. Roc Curve

Interdisciplinary knowledge allows us to have a quantitative approach to a qualitative problem. We built the regression model to fulfil a clinical demand. To deploy our statistical findings, we designed three main functions for the shiny app: individual and group prediction of kidney graft conditions based on user inputs and a comparison of a single gene's impact on the predictive outcome. We designed the UI so that the front page can navigate clinicians with concise user instructions so that they can clearly find the functions they need. To drive *Kiddo*, we put the predictive model in an RMD file into the reserved position in shiny. We modified UI details to suit the practical need. For instance, the app could initially only output a text message about whether the kidney is rejected. We realised this plain and absolute result is not helpful to clinicians. We then added a pie chart to show the probability of rejection, which can offer clinicians a more intuitive and informed suggestion. In the shiny app construction, we constantly asked ourselves what we can offer from a data science perspective and what clinicians need from the biomedical science perspective. *Kiddo* is the final product of interdisciplinary thinking.

Discussion

Limitations.

- The final analysis results are generated from the gse9493 dataset. The data sample of gse9493 in 78 patients, so insufficient data may lead to biases in the accuracy of the prediction model. And only selecting a single clinical data is easy to form data selection bias.
- We tried to find the number of significant genes by *logFC*, but there were as many as 9000 such genes, and we could not test the optimal gene selection due to the limitation of computer performance. The results now are just the best of the 200 most important genes.
- Confirmation bias exists in the analysis of whether physicians perform biopsies or blood sampling for genes. We believe that

sampling with biopsy is more accurate. By using repeated cross-validation combined with SVM and KNN algorithms, biopsies have higher accuracy, but occasionally the random forest algorithm yields better results for blood test samples.

Future work.

- Try to combine more datasets, enrich the number of patient samples to prevent selection bias, and combine more clinical data to improve the accuracy of the model. It is hoped that in this way, the accuracy rate of our model can exceed 95%, and the model is more accurate and can help patients.
- Because our model can only predict the rejection and stability of kidney transplantation, but in fact, there are many reactions of patients after transplantation, including acute rejection and various complications, such as AR IA, AR IB, AR+CAN, CAN I, CAN II et al. In the future, we will add the ability of predictive models to be able to predict the probability of various conditions from patient data.

Conclusion

To conclude the project above, the prediction model of post-transplant kidney adaptability analysed multiple data sets, used repeated cross-validation methods and applied the SVM (Support vector machines) algorithm to generate a stability suggestion with 81.4% accuracy. The research outcome fulfils the basic need by giving clinicians a preliminary evaluation of the kidney adaptability during the preoperative preparation. However, every per cent of accuracy matters and 81.4% is far from a responsible clinical product since the clinical judgement involves taking the life of a human being. Therefore, a larger amount of patient samples is required to prevent selection bias, while combining more clinical data to train the prediction model and improve the accuracy. Nevertheless, the research gives us a glimpse of the correlation between the patient's genetic factors and the kidney transplant adaptability, while providing a preliminary preoperative suggestion to help clinicians allocate the target renal resource.

Contribution

- **Dongsheng Han (490096600)**: Collect data and do filtering in the first stage, mainly including the data preprocessing part. And did the model evaluation part including transforming the output into the graph. At last, I take part in the report and slide production.
- **Haochen Zhang (490169971)**: First, I completed the data cleaning and processing, and analyzed and compared whether the doctor took a biopsy or a blood test. I also completed the comparison of various models and gene selections, and finally selected the best model. And I also provided a boxplot of the value range of each gene in stable and rejection. Finally, I also provided the fountain for the shiny app.
- **Shiyu Wu (500065071)**: Collect data to determine research direction, analyze project background and purpose, complete speech ppt, and speech. And analyzed the genetic sampling method, completed the preparation of the report, and the final layout and layout.
- **Qianze MAN (500050686)**: classification selection for prevalidation, do prevalidation by SVM, feature selection and build regression model, method code for the shiny app, report writing and illustrate plots for the report.
- **Chad Sun (500614994)**: Shiny app layout design and coding, embedded the prediction model and the corresponding graphical work into the interactive shiny dashboard, demonstration of the final shiny product during the presentation, writing a summary report.
- **Jialin Wang (500448045)**: First is projected background research, data collection and analysis by understanding the content. And to analyze the background of the datasets used in our topic, shiny app-text writing, PowerPoint producing, report writing.
- **Mike Lilie Xu (490293759)**: Background research and initial topic discussion. Data collection and interpretation. Writing presentation speech and presentation slides. Writing and proofreading the final report.

Bibliography

1. Wolfe, R. A., Ashby, V. B., Milford, E. L., Ojo, A. O., Ettenger, R. E., Agodoa, L. Y. C., Held, P. J., & Port, F. K. (1999). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a First cadaveric transplant. New England Journal of Medicine, 341(23), 1725–1730. https://doi.org/10.1056/nejm199912023412303
2. Rabbat CG, Thorpe KE, Russell JD, Churchill DN. (2000). Comparison of mortality risk for dialysis patients and cadaveric first renal transplant recipients in Ontario, Canada. J Am Soc Nephrol; 11: 917-22.
3. Oniscu GC, Brown H, Forsythe JL. (2005). Impact of cadaveric renal transplantation on survival in patients listed for transplantation. J Am Soc Nephrol. ; 16: 1859-65.
4. Clayton, P. A., McDonald, S. P., Russ, G. R., & Chadban, S. J. (2019). Long-term outcomes after acute rejection in kidney transplant recipients: An ANZDATA analysis. Journal of the American Society of Nephrology, 30(9), 1697–1707. https://doi.org/10.1681/asn.2018111101
5. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus(Edgar et al., 2002) and are accessible through GEO Series accession number GSE9493 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9493)
6. Saint-Mezard P, Berthier CC, Zhang H, Hertig A et al. (2009). Analysis of independent microarray datasets of renal biopsies identifies a robust transcript signature of acute allograft rejection. Transpl Int 2009 Mar. Available at: https://pubmed.ncbi.nlm.nih.gov/19017305/
7. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus(Edgar et al., 2002) and are accessible through GEO Series accession number GSE14346(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14346)
8. Li L, Khatri P, Sigdel TK, Tran T et al. (2012). A peripheral blood diagnostic test for acute rejection in renal transplantation. Am J Transplant 2012 Oct. Available at: https://pubmed.ncbi.nlm.nih.gov/23009139/
9. Hofling, H. and Tibshirani, R., (2007). A STUDY OF PRE-VALIDATION. [online] Tibshirani.su.domains. Available at: https://tibshirani.su.domains/ftp/PreValidationArticle.pdf .
10. First M. R. (2003). Renal function as a predictor of long-term graft survival in renal transplant patients. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association, 18 Suppl 1, i3–i6. https://doi.org/10.1093/ndt/gfg1027

Appendix

Normalisation

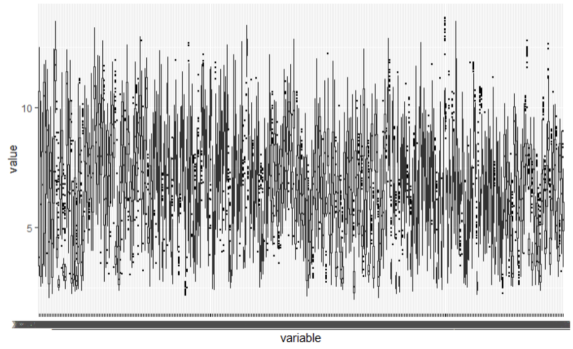

Fig. 6. Before normalisation

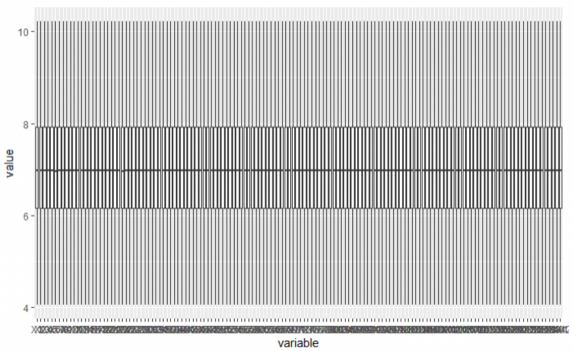

Fig. 7. After normalisation

Sampling method

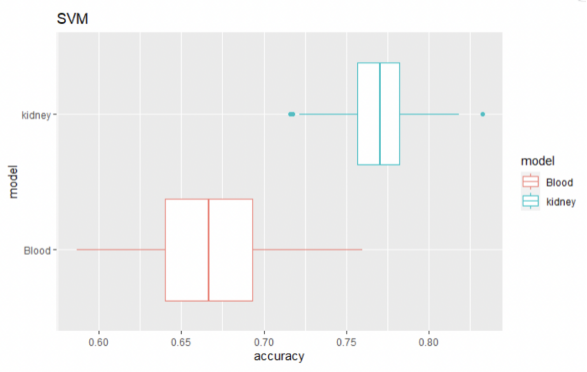

Fig. 8. Boxplot of SVM

Volcano Plot for significant genes

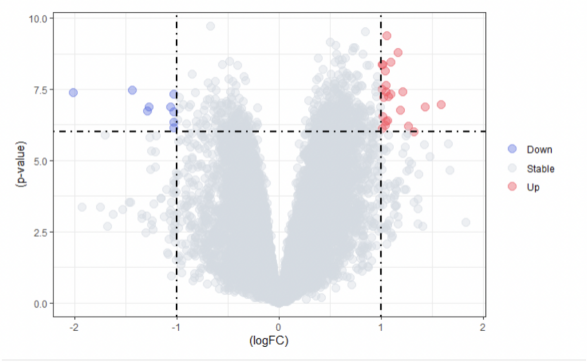

Fig. 9. Most significant genes

Feature selection

	Backward Model		Forward Model	
Predictors	*Odds Ratios*	*p*	*Odds Ratios*	*p*
(Intercept)	128.98	**0.002**	128.98	**0.002**
serum creatinine	0.97	**0.001**	0.97	**0.001**
apv [stable]	11.35	**0.048**	11.35	**0.048**
Observations	56		56	
R^2 Tjur	0.404		0.404	
AIC	48.869		48.869	

Fig. 10. Feature selection by AIC in a stepwise algorithm

Prevalidation

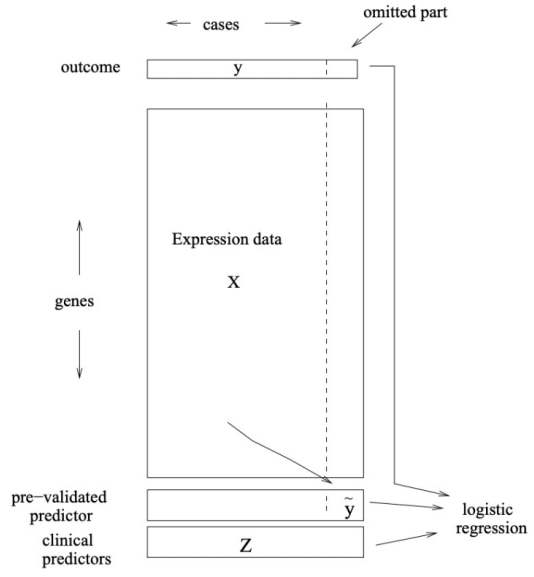

Fig. 11. The process of the prevalidation