

Predicting Arteriovenous Fistula Non-Maturation in Hemodialysis Patients: Analytics of Inflammatory Markers and Serum Metabolic Values

Akram Nakhaei

Tarbiat Modares University

Mohammad Mehdi Sepehri (✉ mehdi.sepehri@gmail.com)

Tarbiat Modares University <https://orcid.org/0000-0002-9920-7452>

Pejman Shadpour

Tehran University of Medical Sciences

Morteza Khavanin Zadeh

Tehran University of Medical Sciences

Research article

Keywords: Arterio-venous fistula, maturation process outcomes, inflammatory markers, serum metabolic values, predictive analysis.

Posted Date: October 17th, 2019

DOI: <https://doi.org/10.21203/rs.2.16183/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Population aging has brought a rise in the prevalence of diabetes and hypertension, leading to more cases of renal failure. Hemodialysis, as a method of renal replacement therapy, by far prevails over peritoneal dialysis (93.5% vs. 6.5%). Although arteriovenous fistula (AVF) is frequently chosen as the vascular access route for chronic hemodialysis; it has limitations including non-maturation. As maintenance of an AVF is much more costly than its creation, foreseeing maturation failure can lead to a wiser allocation of patients to AVF surgery or other alternatives, with potential for significant cost containment.

Methods We investigated the relationship of routinely available systemic inflammatory markers and baseline metabolic values in 107 end-stage renal disease patients (over 35 years of age undergoing their first brachio-cephalic AVF access surgery at wrist level for chronic hemodialysis). In this study, for the first time to our knowledge, we applied predictive analytic tools such as Random Forest for retrospective analysis of prospectively collected data between 2011 and 2018.

Results Our results showed that a combination of inflammatory markers and serum metabolic values can prognosticate AVF maturation outcomes with an accuracy of 0.723, by the 95% confidence interval of (0.715, 0.731) and AUC of 0.853. Also, a combination of inflammatory markers, including albumin, c-reactive protein, erythrocyte sedimentation rate, hemoglobin, lymphocytes, neutrophils, white blood cells, platelets, and red blood cell distribution width, can prognosticate AVF maturation outcomes with an accuracy of 0.674, by the 0.95 confidence interval of (0.665, 0.684) and AUC of 0.824.

Conclusion Risk stratification of patients for AVF non-maturation before attempting the first AVF surgery may help prevent multiple surgical failures and costly endovascular interventions by allowing vascular surgeons to make an individualized choice of vascular access method for new patients.

Background

Population aging has increased the median age, causing a worldwide increment in the elderly ratio. In less-developed regions, the median age will rise from 26 years in 2010 to 35 years in 2050. Aging has increased the incidence of diabetes and hypertension, two leading causes of renal failure [1], [2].

In Iran, about 93.5% of end-stage renal disease (ESRD) cases are managed by hemodialysis, and only 6.5% by peritoneal dialysis [1]. Hemodialysis requires a robust and long-term vascular access such as arteriovenous fistula (AVF), synthetic vascular graft, and cuffed central venous catheter [3], [4]. Among these methods, AVF is arguably preferred because it causes fewer complications and mortality rate [3], [5].

After creating an AVF, progressive increment in the diameter and blood flow within downstream superficial veins makes them ready for hemodialysis. This process is known as maturation. Unfortunately, many factors, including inflammatory markers, may disturb the maturation process, resulting in non-maturation of 51-60% of AVFs [6]. The cost of AVF maintenance is much higher than its creation [7], [8]. So, assessing the risk of AVF non-maturation before surgery would be cost containment for both patients and the healthcare industry.

Inflammation can disturb the compensatory mechanisms of the maturation process and lead to intimal hyperplasia and occlusion [7], [9]. The goal of this study is to explore the relationship between an assortment of routinely available systemic inflammatory and metabolic markers on the outcomes of the AVF maturation process.

A few gaps exist in the data setting and data interpretation in published studies. They did not allude to the fact that statistical analyses are parametric and as such put restrictions on the data. One of these conditions is preferring the normal distribution of variables. This research intends to provide an analytical method appropriate for actual data features to extract trustable information.

We aimed to assess the risk of AVF non-maturation non-invasively. As systemic inflammatory markers and serum metabolic values are routinely available, we focused on them. In addition to these variables, we also gathered medications (anti-coagulant, anti-platelet, anti-hypertensive, anti-inflammatory, and hematopoietic drugs), disease histories (diabetes mellitus, heart disease, and hypertension), and systolic and diastolic blood pressures.

Methods

Research participants are ESRD patients dialyzed by primary AVF. We used statistical, and machine learning approaches to explore and analyze data.

1) Participants and Variables

We retrospectively analyzed the archived data and the electronic charts of ESRD patients at Hasheminejad Kidney Center (HCK); the national tertiary referral center for urology and nephrology in Tehran, Iran. People from around the country refer to this hospital for AVF surgery, so electronic archives at this center are a reasonable representation of the population. We applied the following inclusion and exclusion criteria on 943 patients who had undergone AVF creation at HCK in 2011-2018 (Fig 1).

Patients with non-primary AVF were excluded to remove vascular consequences of the preceding failed access. Patients under 35 were excluded to consider the projected population. Records, with over 50% of missing values were excluded to limit data noise. We limited this study to AVFs created on the wrist to form a more uniform case-mix (assuming that metabolic factors promoting accelerated atherosclerosis are likely to be more severe among failing larger vessels such as brachio-cephalic AVFs). All procedures were sutured in 6-0 Prolene, and the anastomosis technique was end-to-side. All the patients had initially provided informed consent for access to their archived data for research purposes. The Ethics Committee of Hasheminejad Clinical Research Development Center approved the research.

From 934 patients of the sampling frame, 549 had primary AVFs. Four hundred five of them had a determined maturation outcome (matured or failed). Three hundred thirty-eight of 405 patients aged above 35 years; among them, 315 did not experience early failure. Two hundred ninety-four of 315 patients had fewer than 50% missing values, and 190 of them had radio-cephalic AVFs. Since surgeon

skill is a risk factor in AVF maturation, we just considered 114 patients surgically treated by a specific vascular surgeon (Fig 2). Measurements were collected by (version 21.0.0.0, 32-bit edition). We used R (version 3.6.1, 2019-07-05) and Cran packages for analyzing data. Table I represents the characteristics of 114 included patients.

Table I- Characteristics of matured and failed classes

| Class | Number of patients | Age | Body mass index | Gender (male/female) | Smoking (yes/no) | Hypertension | Diabetes mellitus | Heart disease |
|---------|--------------------|-----|-----------------|----------------------|------------------|-----------------|-------------------------|-------------------------|
| Matured | 86 | | | 60/26 | 17/69 | 65 yes 21 no | 44 yes 41 no 1 NA | 6 yes 45 no 35 NA |
| Failed | 28 | | | 19/9 | 4/24 | 25 yes 3 no | 13 yes 15 no | 2 yes 12 no 14 NA |

Table II represents the explicit and implicit inflammatory markers mentioned in the literature. In this study, we considered inflammatory markers, other serum metabolic values, medications, pre-operation blood pressure, and disease histories which were available on the electronic archives (Table III, Table IV, and Table V). From among the lab tests recorded during hospitalization interval, we collected the closest test to the operation time (pre-opt, at dialysis initiation). The target variable is the outcome of the maturation process that captures the values of "matured" or "failed." Our criteria for maturation were (1) easily palpable superficial vein, (2) vein relatively straight, (3) adequate diameter for easy two-needle cannulation (3-4 mm), (4) proper length (≥ 10 cm, for sufficient distance between the two needles), and (5) uniform thrill to palpation and auscultation [10].

Table II- Serum inflammatory markers

| Markers | | |
|--------------------------------|-------------------|-----------------------------------|
| Albumin | Interleukin 6 | Platelets |
| C-reactive protein | Interleukin 8 | Red blood cell distribution width |
| Erythrocyte sedimentation rate | Lymphocytes | Tumor necrosis factor-alpha |
| Fibrinogen | Neutrophils | |
| Hemoglobin | White blood cells | |

Table III- Blood test variables

| Variables | Units | Variables | Units |
|--------------------------------|--------|-----------------------------------|----------|
| C-reactive protein | mg/l | Phosphorous | mg/dl |
| Erythrocyte sedimentation rate | Mm | Sodium | MEq/l |
| Total protein | mg/dl | Serum Iron | µg/dl |
| Fasting blood sugar | mg/dl | Potassium | MEq/l |
| Blood urea nitrogen | mg/dl | White blood cells | *10^3/µl |
| Albumin | g/L | Red blood cells | *10^3/µl |
| Creatinine | mg/dl | Hemoglobin | gr/dl |
| Cholesterol | mg/dl | Platelet | % |
| Triglyceride | mg/dl | Red blood cell distribution width | % |
| High-density lipoprotein | mg/dl | Neutrophils | % |
| Low-density lipoprotein | Mmol/l | Lymphocytes | % |
| Lactic Acid Dehydrogenase | IU/L | Partial thromboplastin time | Sec |
| Calcium | mg/dl | Uric acid | mg/dl |
| Ferritin | ng/ml | Blood group | n |
| Parathyroid hormone | ng/l | Total Iron Binding Capacity | µg/dl |
| Creatine phosphokinase | IU/L | | |

Table IV- Medications

| Medicines | Units | Medicines | Units |
|---------------------|-------|-------------|-------|
| Erythropoietin | IU/ml | Losartan | mg |
| Ferfolic | mg | Amlodipine | mg |
| Atorvastatin | mg | Material | mg |
| Aspirin | mg | N.C. | mg |
| Plavix_Osvix | mg | Amiodarone | mg |
| Enalapril_Captopril | mg | Warfarin | mg |
| Venofer | mg | Digoxin | mg |
| Atenolol | mg | Carvedilol | mg |
| Prednisolone | mg | Simvastatin | mg |

Table V- Blood pressure and disease histories

| Variables | Units | Variables | Units |
|----------------------------|-------|-----------------------|-------|
| Systolic blood pressure | mmHg | Smoking | y/n |
| Diastolic blood pressure | mmHg | Heart disease | y/n |
| Diabetes mellitus | y/n | Hypertension | y/n |
| Diabetes mellitus duration | yrs | Hypertension duration | yrs |

2)Data Analysis

In the pre-processing phase, we removed two variables (Creatine phosphokinase and Lactic acid dehydrogenase) with more than 80% missing values, seven records with more than 50% missing values, and five medications with zero-variance (Plavix-Osvix, Warfarin, Digoxin, Enalapril-Captopril, and Simvastatin). In the end, missing values were imputed, and the data were balanced (Fig 3).

Missing values frequently occur in medical studies (Burton and Altman, 2004) and can bring about a significant amount of bias and inefficiency (Barnard and Meng, 1999). Removing records or variables with missing values often leads to bias, decreased coverage of confidence intervals, and a decreased discriminative ability of the multivariable model (Janssen et al., 2010). So, it is better to handle missing values carefully based on the mechanism behind missingness. There are three possible mechanisms: 1) missing completely at random, 2) missing at random (MAR), and 3) missing not at random. Literature shows that in medical science, usually the mechanism behind missingness is the MAR mechanism (Schafer, 1997; Schafer and Graham, 2002; Toutenburg, 1990). In MAR mechanism, missingness is often related to other observed variables. So, we can predict missing values using other available data (Janssen et al., 2010). In the present study, we used the random forest model for prediction of missing values. For each variable, a model was built using other variables as predictors; classification trees were grown for categorical and regression trees for continuous variables.

In medical studies, we often encounter imbalanced classes of data (Belarouci and Chikh, n.d.). Class imbalance heavily compromises the learning process of predictive models because they tend to focus on the prevalent samples and ignore the rare events (Menardi and Torelli, 2014). Hence, prediction models trained on imbalanced data are highly susceptible to produce inaccurate results (He and Garcia, 2009). Prior studies applied class rebalancing techniques to mitigate the risk of imbalanced data.

Tantithamthavorn and Matsumoto investigated the impact of rebalancing techniques on the performance and interpretation of prediction models and found that these techniques have little effect on popularly-used classification models like logistic regression and random forest (Tantithamthavorn et al., 2018). In this study, we used Synthetic Minority Oversampling Technique (SMOTE) to over-sample the minority and under-sample the majority class. After rebalancing, we had 34 failed and 34 matured samples.

After pre-processing, the data is ready for analysis. For choosing the appropriate analysis method, we did some exploratory analysis. Based on exploratory analysis results, we concluded that the relationships between variables are non-linear, and the variables are almost non-normal. In medical studies, logistic regression is a commonly used method for classification. However, when we have a non-linear classification problem with few samples and lots of features, it is not the right choice. In this situation, it ends up with too many features that may lead to over-fitting and is computationally expensive. The random forest model is a much better way to learn complex non-linear hypotheses, even when the feature space is ample, and the sample size is small. This model involves a group of decision trees to achieve better accuracy and stability. Predictive models usually suffer from bias and variance errors. As the complexity of the predictive model increases, bias will decrease, and the variance will increase. The random forest model tries to balance between bias and variance this way that every tree in the forest is randomly built and fully grown. These trees are at the highest complexity to reduce the bias. The

randomness of trees mounts the noise uniquely on each tree and prevents the noise from being aggregated. We used this model to identify informative factors in predicting maturation outcomes. Also, we extracted frequent and accurate rules from the model to use for decision making.

Results

1) Descriptive Analysis

In this section, statistical methods such as normality check, and principal component analysis (PCA) were employed to explore the data.

A. Central Tendency and Dispersion

The mean and standard deviation () calculated for independent variables are given in Tables VI, VII, and VIII.

Table VI- Mean and standard deviations of the blood test variables

| Variables | Failed () | Matured () | Variables | Failed () | Matured () |
|--------------------------------|--------------------|----------------|-----------------------------|--------------------|--------------------|
| C-reactive protein | | | Calcium | | |
| Erythrocyte sedimentation rate | | | Phosphorous | | |
| Total protein | 7.183 0.702 | 6.803 0.951 | Total Iron Binding Capacity | 235.722± 38.703 | 241.781± 48.371 |
| Fasting blood sugar | | | Sodium | | |
| Blood urea nitrogen | | | Potassium | | |
| Albumin | | | White blood cells | | |
| Creatatinine | | | Red blood cells | | |
| Uric acid | | | Hemoglobin | | |
| Cholesterol | | 30.757 | Platelets | | |
| Triglyceride | | | Red cell distribution width | | |
| High density lipoprotein | | | Neutrophils | | |
| Low density lipoprotein | | | Lymphocytes | | |
| Ferritin | 388.316 | | Partial thromboplastin time | | |
| Parathyroid hormone | 217.164 108.734 | | Serum iron | 56.818 9.927 | 56.813 16.740 |

Table VII- Mean and standard deviations of the medication variables

| Medicines | Failed () | Matured () | Medicines | Failed () | Matured () |
|----------------------------|-----------------|----------------|---------------------|---------------|----------------|
| Erythropoietin dosage | 666.667 960.769 | 625 | Losartan dosage | 11.111 | 11.563 |
| Ferfolic dosage | 0.926 | 1.9 | Amlodipine dosage | 2.222 | 4.313 |
| Atorvastatin dosage | 5.926 | 4.875 | Metoral dosage | 1.852 | 11.875 |
| Enalapril_Captopril dosage | 0.185 | 6.094 | Amiodarone dosage | 0 | 0.375 |
| Venofer dosage | 11.111 | 17.5 38.236 | Aspirin dosage | 2.963 | 8 |
| Atenolol dosage | 0 0 | 3.75 15.457 | Prednisolone dosage | 0 | 2.063 10.333 |
| N.C. dosage | 0.385 1.388 | 0.52 2.01 | Carvedilol dosage | | 0 |

Table VIII- Mean and standard deviations of the blood pressure, duration of hypertension and diabetes mellitus

| Variables | Failed () | Matured () |
|----------------------------|---------------|----------------|
| Systolic blood pressure | 139.746 | 138.504 |
| Diastolic blood pressure | 81.767 | 81.347 |
| Hypertension duration | 3.962 3.691 | 3.953 4.319 |
| Diabetes mellitus duration | 7.074 8.223 | 5.260 6.665 |

B. Normality Checks

We investigated the normality of variables' distribution by Shapiro-Wilk's test. The p-value > 0.05 implies that the distribution is close to normal. Here, almost all of the p-values are under 0.05, except for cholesterol, low-density lipoprotein, total protein, hemoglobin, lymphocytes, and serum iron (Tables IX, X, and XI). So, we can conclude that the data almost do not obey the normal distribution.

Table IX- Normality test results of blood test variables

| Variables | W | P-values | Variables | W | P-values |
|--------------------------------|--------------|----------------|-----------------------------|--------------|---------------|
| C-reactive protein | 0.683 | 3.332e-09 | Phosphorous | 0.898 | 1.544e-06 |
| Erythrocyte sedimentation rate | 0.896 | 3.58e-05 | Sodium | 0.939 | 0.0001007 |
| Fasting blood sugar | 0.869 | 8.431e-07 | Potassium | 0.905 | 1.21e-06 |
| Blood urea nitrogen | 0.926 | 2.035e-05 | White blood cells | 0.883 | 1.678e-07 |
| Albumin | 0.709 | 1.665e-09 | Red blood cells | 0.949 | 0.0007656 |
| Creatinine | 0.771 | 1.416e-11 | <u>Hemoglobin</u> | <u>0.983</u> | <u>0.2315</u> |
| Uric acid | 0.960 | 0.01748 | Platelets | 0.945 | 0.0003697 |
| <u>Cholesterol</u> | <u>0.974</u> | <u>0.09207</u> | Red cell distribution width | 0.948 | 0.002368 |
| Triglyceride | 0.890 | 2.871e-06 | Neutrophils | 0.962 | 0.03868 |
| High density lipoprotein | 0.907 | 1.727e-05 | <u>Lymphocytes</u> | <u>0.973</u> | <u>0.1657</u> |
| <u>Low density lipoprotein</u> | <u>0.972</u> | <u>0.07078</u> | Partial thromboplastin time | 0.340 | < 2.2e-16 |
| Calcium | 0.961 | 0.005231 | Ferritin | 0.854 | 1.057e-06 |
| <u>Total protein</u> | <u>0.951</u> | <u>0.1359</u> | Parathyroid hormone | 0.852 | 8.176e-05 |
| Total Iron Binding Capacity | 0.934 | 0.01191 | <u>Serum iron</u> | <u>0.955</u> | <u>0.1314</u> |

Table X- Normality test results of medical variables

| Medicines | W | P-values | Medicines | W | P-values |
|----------------------------|-------|-----------|---------------------|-------|-----------|
| Erythropoietin dosage | 0.606 | 1.618e-15 | Losartan dosage | 0.668 | 3.313e-14 |
| Ferfolic dosage | 0.479 | < 2.2e-16 | Amlodipine dosage | 0.496 | < 2.2e-16 |
| Atorvastatin dosage | 0.523 | < 2.2e-16 | Metoral dosage | 0.433 | < 2.2e-16 |
| Enalapril_Captopril dosage | 0.237 | < 2.2e-16 | Venofer dosage | 0.439 | < 2.2e-16 |
| Amiodarone dosage | 0.154 | < 2.2e-16 | Aspirin dosage | 0.310 | < 2.2e-16 |
| Atenolol dosage | 0.214 | < 2.2e-16 | Prednisolone dosage | 0.169 | < 2.2e-16 |
| N.C. dosage | 0.232 | < 2.2e-16 | Carvedilol dosage | 0.071 | < 2.2e-16 |

Table XI- Normality test results of blood pressure and disease history variables

| Variables | W | P-values |
|----------------------------|-------|-----------|
| Systolic blood pressure | 0.961 | 0.004015 |
| Diastolic blood pressure | 0.905 | 1.824e-06 |
| Hypertension duration | 0.784 | 1.587e-08 |
| Diabetes mellitus duration | 0.765 | 2.028e-11 |

C. Principal Component Analysis

Using PCA, we investigated the relationship between variables. Measures showed that there is no dominant direction of variation, so the variables almost vary independently, and there is no redundant dimension (Table XII).

Table XII- Percent of variance addressed by each principal component extracted from the independent variables of this study

| Principal Components | Variance percent | Cumulative variance percent | Principal Components | Variance percent | Cumulative variance percent |
|----------------------|------------------|-----------------------------|----------------------|------------------|-----------------------------|
| PC.1 | 0.0831 | 0.0831 | PC.27 | 0.0121 | 0.8691 |
| PC.2 | 0.0771 | 0.1602 | PC.28 | 0.0120 | 0.8810 |
| PC.3 | 0.0637 | 0.2239 | PC.29 | 0.0117 | 0.8928 |
| PC.4 | 0.0582 | 0.2821 | PC.30 | 0.0104 | 0.9032 |
| PC.5 | 0.0484 | 0.3305 | PC.31 | 0.0102 | 0.9134 |
| PC.6 | 0.0447 | 0.3752 | PC.32 | 0.0093 | 0.9228 |
| PC.7 | 0.0407 | 0.4159 | PC.33 | 0.0091 | 0.9319 |
| PC.8 | 0.0375 | 0.4533 | PC.34 | 0.0078 | 0.9397 |
| PC.9 | 0.0356 | 0.4889 | PC.35 | 0.0065 | 0.9462 |
| PC.10 | 0.0333 | 0.5222 | PC.36 | 0.0062 | 0.9524 |
| PC.11 | 0.0310 | 0.5533 | PC.37 | 0.0056 | 0.9580 |
| PC.12 | 0.0284 | 0.5816 | PC.38 | 0.0055 | 0.9636 |
| PC.13 | 0.0259 | 0.6075 | PC.39 | 0.0052 | 0.9688 |
| PC.14 | 0.0257 | 0.6333 | PC.40 | 0.0046 | 0.9734 |
| PC.15 | 0.0243 | 0.6576 | PC.41 | 0.0041 | 0.9775 |
| PC.16 | 0.0240 | 0.6816 | PC.42 | 0.0037 | 0.9812 |
| PC.17 | 0.0222 | 0.7038 | PC.43 | 0.0035 | 0.9847 |
| PC.18 | 0.0213 | 0.7250 | PC.44 | 0.0032 | 0.9879 |
| PC.19 | 0.0208 | 0.7458 | PC.45 | 0.0028 | 0.9907 |
| PC.20 | 0.0192 | 0.7650 | PC.46 | 0.0023 | 0.9930 |
| PC.21 | 0.0172 | 0.7822 | PC.47 | 0.0023 | 0.9953 |
| PC.22 | 0.0164 | 0.7986 | PC.48 | 0.0016 | 0.9968 |
| PC.23 | 0.0154 | 0.8140 | PC.49 | 0.0013 | 0.9981 |
| PC.24 | 0.0153 | 0.8293 | PC.50 | 0.0011 | 0.9992 |
| PC.25 | 0.0144 | 0.8437 | PC.51 | 0.0005 | 0.9997 |
| PC.26 | 0.0132 | 0.8570 | PC.52 | 0.0003 | 1.0000 |

PC, Principal component.

D. Correlation Analysis

Highly correlated independent variables may interfere with each other when a model is being interpreted [11]. Jiarpakdee et al. claimed that removing correlated independent variables improves the consistency of the highest-ranked variables regardless of how a model is specified and negligibly impacts the performance and stability of models [12]. Fig 4 shows the hierarchically clustered correlation matrix of independent variables. Three clusters have been formed that are represented in Table XIII. Since variables are highly correlated, we selected one variable in each group and removed cholesterol, neutrophils, and red blood cells.

Table XIII- Clusters extracted from hierarchical clustering of the correlation matrix of independent variables

| Cluster | Variables | Correlation |
|---------|--------------------------------------|-------------|
| 1 | Cholesterol, low-density lipoprotein | 0.90 |
| 2 | Neutrophils, lymphocytes | -0.968 |
| 3 | Red blood cells, hemoglobin | 0.840 |

2) Predictive Analysis

Data distribution is a determinative factor in the choice between parametric and non-parametric methods. The assumption of normality needs to be checked for many statistical procedures, especially parametric tests, because their validity depends on it [13]. If we are not sure or we suspect that distributions do not behave normally, it is better to use non-parametric methods. Also, considering the results of PCA, relationships between variables are non-linear, so we should choose a non-parametric method for analyzing non-linear data relationships. In the data analysis section, we argued that the random forest model is the right choice for non-linear and non-normal data. So, in the following, we modeled the data using this method.

A. Random Forest

A random forest model containing a thousand decision trees was built. Splitting criteria for this model was Gini decrease, and the number of variables randomly selected at each split was two. Mean accuracy of the model was 0.723, with a 95% confidence interval of (0.715, 0.731) and AUC of 0.809. Variables importances calculated by the mean Gini decrease criteria are presented in Table XIV.

Table XIV- Importance of variables in predicting fistula maturation outcome

| | Variables | Mean Gini decrease | | Variables | Mean Gini decrease |
|----|--------------------------------|--------------------|----|----------------------------|--------------------|
| 1 | Creatinine | 1.504 | 26 | Age | 0.758 |
| 2 | Red cells distribution width | 1.376 | 27 | White blood cells | 0.744 |
| 3 | High-density lipoprotein | 1.331 | 28 | Body mass index | 0.736 |
| 4 | Sodium | 1.154 | 29 | Fasting blood sugar | 0.716 |
| 5 | Ferritin | 1.147 | 30 | Diastolic blood pressure | 0.678 |
| 6 | Serum iron | 1.082 | 31 | C-reactive protein | 0.662 |
| 7 | Total protein | 1.081 | 32 | Atorvastatin dosage | 0.615 |
| 8 | Albumin | 0.946 | 33 | Blood group | 0.605 |
| 9 | Partial thromboplastin time | 0.925 | 34 | Metoral dosage | 0.439 |
| 10 | Parathyroid hormone | 0.923 | 35 | Amlodipine dosage | 0.413 |
| 11 | Hypertension duration | 0.922 | 36 | Diabetes duration | 0.371 |
| 12 | Calcium | 0.917 | 37 | Venofer dosage | 0.265 |
| 13 | Erythrocyte sedimentation rate | 0.883 | 38 | Erythropoietin dosage | 0.254 |
| 14 | Triglyceride | 0.881 | 39 | N.C. dosage | 0.217 |
| 15 | Hemoglobin | 0.870 | 40 | Ferfolic dosage | 0.189 |
| 16 | Total iron-binding capacity | 0.868 | 41 | Gender | 0.155 |
| 17 | Blood urea nitrogen | 0.857 | 42 | Smoking | 0.142 |
| 18 | Potassium | 0.815 | 43 | Heart disease | 0.141 |
| 19 | Lymphocytes | 0.810 | 44 | Prednisolone dosage | 0.125 |
| 20 | Platelets | 0.810 | 45 | Enalapril-Captopril dosage | 0.083 |
| 21 | Uric acid | 0.806 | 46 | Carvedilol dosage | 0.054 |
| 22 | Systolic blood pressure | 0.804 | 47 | Aspirin dosage | 0.039 |
| 23 | Phosphorous | 0.788 | 48 | Amiodarone dosage | 0.0000000000 |
| 24 | Losartan dosage | 0.786 | 49 | Atenolol | 0.0000000000 |
| 25 | Low-density lipoprotein | 0.774 | | | |

We reduced the feature space to 33 most important variables and then built a new random forest model (1000 trees, Gini criteria, 2 randomly selected variables at each split). This time, the mean accuracy was 0.71, with a 95% confidence interval of (0.701 0.719) and AUC of 0.853. We extracted the decision rules from all the decision trees and selected a concise set of rules with low length, high frequency, and small error rate (Table XV). For this task, we used “inTree” package of Cran.

Table XV- Important rules extracted from the random forest model

| Length | Freqency | Error rate | Condition | | Prediction | Importance |
|--------|----------|------------|-----------|---|------------|------------|
| 1 | 3 | 0.51 | 0.08 | Red cells distribution width>13.641 & High-density lipoprotein<=45.5 & Sodium>134.5 | failed | 1 |
| 2 | 4 | 0.48 | 0.06 | Red cells distribution width >13.641 & Sodium>132 & Calcium>7.95 & Total iron-binding capacity<=270.885 | failed | 0.712 |
| 3 | 4 | 0.48 | 0.09 | High-density lipoprotein<=40.551 & Serum iron>35 & Hypertension duration>0.65 & Calcium>7.95 | failed | 0.321 |
| 4 | 5 | 0.45 | 0.06 | Red cells distribution width>13.441 & Total protein>6.476 & Hypertension duration>0.65 & Uric acid<=11.05 & Diastolic blood pressure>65 | failed | 0.275 |
| 5 | 2 | 0.26 | 0 | Total protein>7.020 & Platelets<=213 | failed | 0.124 |
| 6 | 4 | 0.57 | 0.17 | High-density lipoprotein<=45.5 & Serum iron>50.31 & Systolic blood pressure<=172.889 & White blood cells>4.1 | failed | 0.107 |
| 7 | 2 | 0.14 | 0 | Red cells distribution width>13.835 & Systolic blood pressure>152.5 | failed | 0.0917 |
| 8 | 4 | 0.30 | 0 | Sodium>134.5 & Triglyceride>99.050 & Triglyceride<=305 & Uric acid>7.539 | failed | 0.0826 |
| 9 | 2 | 0.04 | 0 | Sodium>137.700 & Partial thromboplastin time>31.730 | matured | 0.0816 |
| 10 | 4 | 0.19 | 0 | Total protein<=7.025 & Hemoglobin<=11.212 & Uric acid<=8.152 & Systolic blood pressure<=160 | matured | 0.0591 |
| 11 | 2 | 0.05 | 0 | Fasting blood sugar>94.227 & Fasting blood sugar<=96.455 | matured | 0.0290 |
| 12 | 1 | 0.13 | 0 | Ferritin<=144 | matured | 0.0240 |
| 13 | 2 | 0.07 | 0 | Partial thromboplastin time>27 & Hypertension duration>8.476 | matured | 0.0229 |
| 14 | 2 | 0.17 | 0 | Calcium>7.95 & Erythrocyte sedimentation rate >60 | failed | 0.0228 |
| 15 | 2 | 0.13 | 0 | Phosphorous<=6.075 & Phosphorous>5.55 | failed | 0.0219 |

| Length | Freqency | Error rate | Condition | Prediction | Importance |
|--------|----------|------------|---------------|------------|----------------|
| 16 | 1 2 | 0.16 0 | Sodium<=133.5 | matured | 0. 01 92 |

Discussion

We retrospectively investigated the archived data of 107 ESRD patients (27 failed samples, 80 matured samples, mean age 55.31 years, 68% male, 50% diabetic, 79% hypertensive). Several groups of variables were considered in this study, including blood lab tests, medications, disease histories, demographic data, and blood pressure. Blood lab tests included inflammatory markers and serum metabolic values. Among the lab tests recorded during the hospitalization interval, we considered the closest test to the operation time (pre-opt, at dialysis initiation). Prescribed medications included anti-coagulant, anti-platelet, anti-hypertensive, anti-inflammatory, and hematopoietic drugs. Histories of hypertension and diabetes mellitus and their duration were regarded. Presence of heart disease was considered to be a binary variable. Being smoker or not was considered as a binary variable, too. Demographic data included age and gender.

Regarding the characteristics of the data (non-linearity and non-normality), the random forest model was the right choice for modeling. Importances of variables in predicting the maturation outcomes showed that medications do not have considerable ability to discriminate two classes of “failed” and “matured” (Fig 4). In the following, we investigated the effect of highly important predicting variables such as creatinine, red cell distribution width, high-density lipoprotein, sodium, ferritin, serum iron, and total protein on the output of the AVF maturation process. Finally, we investigated the ability of inflammatory markers in predicting the outcomes of AVF maturation process.

- Creatinine

Fig 5 illustrates the boxplot of creatinine in two classes of “failed” and “matured.” This figure shows that creatinine has the same interval in two classes, but, the mean of creatinine is considerably lower in the “matured” class. There is no rule containing creatinine in table XV. Considering this figure, we can conclude that creatine may have a negative effect on the maturation process. However, Duque et al. did not find any association between creatinine level and AVF primary failure [14]. Conz et al. found that uremic patient (higher creatinine and blood urea nitrogen) have a slower maturation process [15] that is consistent with our results.

- Red blood cell distribution width

Fig 6 shows that red blood cell distribution width has a considerably lower mean in the “matured” class. Extracted rules containing red blood cell distribution width (Table XVI) confirm that high values of red blood cell distribution width result in AVF maturation failure. This finding is consistent with the study of Usman et al. that indicates high values of preoperative red blood cell distribution width will result in AVF failure [16].

Table XVI- Important rules containing red cells distribution width extracted from the random forest model

| Condition | Predict ion | Importa nce |
|---|-------------|-------------|
| Red cells distribution width>13.641 & High-density lipoprotein<=45.5 & Sodium>134.5 | failed | 1 |
| Red cells distribution width >13.641 & Sodium>132 & Calcium>7.95 & Total iron-binding capacity<=270.8 85 | failed | 0.712 |
| Red cells distribution width>13.441 & Total protein>6.476 & Hypertension duration>0.65 & Uric acid<=1 1.05 & Diastolic blood pressure>65 | failed | 0.275 |
| Red cells distribution width>13.835 & Systolic blood pressure>152.5 | failed | 0.0917 |

- High-density lipoprotein

Fig 7 shows that high-density lipoprotein has a higher mean in the “matured” class. Rules containing high-density lipoprotein (Table XVIII) indicate that low values of this variable will harm the maturation process. Kirkpantur et al. retrospectively investigated 99 patients to examine the effect of serum lipid profile on the AVF thrombosis. They found that patients with thrombosed AVF have a considerably lower level of high-density lipoprotein, and a higher level of low-density lipoprotein [17].

Table XVII- Rules containing high-density lipoprotein variable

| Condition | Predicti on | Importanc e |
|--|-------------|-------------|
| Red cells distribution width>13.641 & High-density lipoprotein<=45.5 & Sodium>134.5 | failed | 1 |
| High-density lipoprotein<=40.551 & Serum iron>35 & Hypertension duration>0.65 & Calcium>7.95 | failed | 0.321 |
| High-density lipoprotein<=45.5 & Serum iron>50.31 & Systolic blood pressure<=172.889 & White blo od cells>4.1 | failed | 0.107 |

- Sodium

Fig 8 shows the higher mean of sodium in the “failed” class. The dominant pattern in the rules containing sodium (Table XVIII) is that higher values of sodium will result in non-maturation of AVFs. The literature indicates that excessive sodium intake is thought to have direct toxic effects on blood vessels through mediating factors such as oxidative stress, inflammation, endothelial cell dysfunction, and vascular stiffness [18, p.], [19], [20].

Table XVIII- Important rules containing sodium extracted from the random forest model

| Condition | Predictio n | Importanc e |
|--|----------------|----------------|
| Red cells distribution width>13.641 & High-density lipoprotein<=45.5 & Sodium>134.5 | failed | 1 |
| Red cells distribution width >13.641 & Sodium>132 & Calcium>7.95 & Total iron-binding capacity<= 270.885 | failed | 0.712 |
| Sodium>134.5 & Triglyceride>99.050 & Triglyceride<=305 & Uric acid>7.539 | failed | 0.0826 |
| Sodium>137.700 & Partial thromboplastin time>31.730 | matured | 0.0816 |
| Sodium<=133.5 | matured | 0.0192 |

- Ferritin

Fig 9 shows that ferritin has almost lower values in the “matured” class. There is one rule containing ferritin (Table XIX) that shows more moderate amounts of ferritin will lead to successful AVF maturation. Morton et al. found no association between ferritin and AVF failure [21]. However, Ishii et al. found that ferritin may be associated with vascular events [22].

Table XIX- The important rule containing ferritin extracted from the random forest model

| Condition | Prediction | Importance |
|---------------|------------|------------|
| Ferritin<=144 | matured | 0.0240 |

- Serum iron

Fig 10 shows that the “matured” category has a slightly higher mean of serum iron, but, extracted rules containing serum iron (Table XX) say that higher values of serum iron will result in the AVF maturation failure. There is no study which investigated the effect of iron on the AVF maturation process, but the literature mentions a possible association between high iron marker levels and poor cardiovascular outcomes [23].

Table XX- Rules containing serum iron extracted from the random forest model

| Condition | Predicti on | Importanc e |
|--|----------------|----------------|
| High-density lipoprotein<=40.551 & Serum iron>35 & Hypertension duration>0.65 & Calcium>7.95 | failed | 0.321 |
| High-density lipoprotein<=45.5 & Serum iron>50.31 & Systolic blood pressure<=172.889 & White blo od cells>4.1 | failed | 0.107 |

- Total protein

Fig 11 shows the higher values of total protein in the “failed” class. Extracted rules containing total protein (Table XXI) indicate that more elevated amounts of total protein will lead to the AVF maturation failure. Kaygin et al. found no relationship between total protein and AVF maturation outcomes [24]. Also, Sari et al. found no correlation between total protein level and AVF thrombosis [25].

Table XXI- Rules containing total protein extracted from the random forest model

| Condition | Prediction | Importance |
|---|------------|------------|
| Red cells distribution width>13.441 & Total protein>6.476 & Hypertension duration>0.65 & Uric acid<=1 1.05 & Diastolic blood pressure>65 | failed | 0.275 |
| Total protein>7.020 & Platelets<=213 | failed | 0.124 |
| Total protein<=7.025 & Hemoglobin<=11.212 & Uric acid<=8.152 & Systolic blood pressure<=160 | matured | 0.0591 |

- Inflammatory markers

Two inflammatory markers including red blood cell distribution width and albumin have considerable importance in predicting AVF maturation outcome. Other inflammatory markers are also among predicting variables but with a lower level of significance. We built a random forest model by a combination of inflammatory markers, including albumin, c-reactive protein, erythrocyte sedimentation rate, hemoglobin, lymphocytes, neutrophils, white blood cells, platelets, and red blood cell distribution width. The mean accuracy of the model was 0.674, with 0.95 confidence interval of (0.665, 0.684) and AUC of 0.824. This result shows that a combination of inflammatory markers can predict AVF maturation outcomes with acceptable accuracy.

Conclusion

The present study had several limitations. Our study was retrospective and restricted to one referral center. We had a small sample size and a considerable number of missing values and imbalanced data. Imputation and sampling added extra noise to the data. We did not control all the risk factors that may have biased the output.

Big data help machine learning algorithms to find better patterns in the data and improve their accuracy. So, it is better to repeat this study with much more sample size. Finally, it is concluded that serum values, including inflammatory markers, have an excellent power to predict AVF maturation outcomes. This prediction helps surgeons to avoid allocating patients in whom the AVF is destined to fail to this vascular access in the first place. This can potentially make high cost conserving decisions possible.

List Of Abbreviations

Table XXII- List of abbreviations used in this study

| Abbreviation | Phrase |
|--------------|------------------------------|
| AVF | Arteriovenous fistula |
| AUC | Area Under Curve |
| ESRD | End-stage renal disease |
| HKC | Hasheminejad Kidney Center |
| Pre-opt | Pre-operation |
| MAR | Missing at random |
| SD | Standard deviation |
| PCA | Principal component analysis |
| PC | Principal component |

Declarations

- Ethics approval and consent to participate

All procedures performed in this study were approved by the Ethics Committee of Hasheminejad Clinical Research Development Center. All the patients had initially provided written informed consent for access to their chart data for research purposes.

- Consent to publish

All the patients had initially provided informed consent for access to their chart data for research purposes.

- Availability of data and material

The data that support the findings of this study are available from Hasheminejad Kidney Center (HKC), but restrictions apply to the availability of these data, which were used under license for the current research, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of HKC.

- Competing interests

The authors declare that they have no competing interests.

- Funding

Not applicable.

- Authors' contributions

AN wrote the manuscript under the supervision and advice of MMS, PS, and doctor MK. All authors of this manuscript have read and approved the final version submitted.

- Acknowledgements

None.

References

- [1] M. Aghighi et al., "Dialysis in iran," *Iran J Kidney Dis*, vol. 2, no. 1, pp. 11–5, 2008.
- [2] D. of E. United Nations and P. D. Social Affairs, "World population ageing 2013," N. Y. U. N., 2013.
- [3] A. Irish et al., "Preventing AVF thrombosis: the rationale and design of the Omega-3 fatty acids (Fish Oils) and Aspirin in Vascular access OUTcomes in REnal Disease (FAVoured) study," *BMC Nephrol.*, vol. 10, no. 1, p. 1, 2009. DOI: <https://doi.org/10.1186/1471-2369-10-1>
- [4] J. VachharajaniTushar, "Dialysis vascular access selection in elderly patients," *US Nephrol.*, vol. 6, no. 2, pp. 128–130, 2011. DOI: <https://dx.doi.org/10.1111%2Fsdi.12037>
- [5] C. David, I. Peride, A. Niculae, A. M. Constantin, and I. A. Checherita, "Very low protein diets supplemented with keto-analogues in ESRD predialysis patients and its effect on vascular stiffness and AVF Maturation," *BMC Nephrol.*, vol. 17, no. 1, p. 131, 2016. DOI: <https://doi.org/10.1186/s12882-016-0347-y>
- [6] C.-C. Yen et al., "Factors affecting fistula failure in patients on chronic hemodialysis: a population-based case-control study," *BMC Nephrol.*, vol. 19, no. 1, p. 213, 2018. DOI: <https://doi.org/10.1186/s12882-018-1010-6>
- [7] A. J. Jackson, P. Coats, and D. B. Kingsmore, "Pharmacotherapy to improve outcomes in vascular access surgery: a review of current treatment strategies," *Nephrol. Dial. Transplant.*, vol. 27, no. 5, pp. 2005–2016, 2012. DOI: <https://doi.org/10.1093/ndt/gfr552>
- [8] C. Wong et al., "Liposomal prednisolone inhibits vascular inflammation and enhances venous outward remodeling in a murine arteriovenous fistula model," *Sci. Rep.*, vol. 6, 2016. DOI: <https://doi.org/10.1038/srep30439>
- [9] T. Lee, "Novel paradigms for dialysis vascular access: downstream vascular biology—is there a final common pathway?," *Clin. J. Am. Soc. Nephrol.*, p. CJN-03490413, 2013. DOI: <https://doi.org/10.2215/CJN.03490413>
- [10] M. Khavanin Zadeh, F. Gholipour, Z. Naderpour, and M. Porfakharan, "Relationship between vessel diameter and time to maturation of arteriovenous fistula for hemodialysis access," *Int. J. Nephrol.*, vol. 2012, 2012. DOI: <http://dx.doi.org/10.1155/2012/942950>
- [11] C. Tantithamthavorn, A. E. Hassan, and K. Matsumoto, "The impact of class rebalancing techniques on the performance and interpretation of defect prediction models," *IEEE Trans. Softw. Eng.*, 2018. DOI: <https://doi.org/10.1109/TSE.2018.2876537>

- [12] J. Jiarpakdee, C. Tantithamthavorn, and A. E. Hassan, "The impact of correlated metrics on defect models," ArXiv Prepr. ArXiv180110271, 2018.
- [13] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: a guide for non-statisticians," Int. J. Endocrinol. Metab., vol. 10, no. 2, p. 486, 2012. DOI: <https://dx.doi.org/10.5812%2Fijem.3505>
- [14] J. C. Duque et al., "Arteriovenous fistula maturation in patients with permanent access created prior to or after hemodialysis initiation," J. Vasc. Access, vol. 18, no. 3, pp. 185–191, 2017. DOI: <https://doi.org/10.5301/jva.5000662>
- [15] P. A. Conz, C. Crepaldi, and G. La Greca, "Slow maturation of arterio-venous fistula in seven uremic patients: use of Ash Split Cath® as temporary, prolonged vascular access," J. Vasc. Access, vol. 1, no. 2, pp. 51–53, 2000. DOI: <https://doi.org/10.5301/jva.5000662>
- [16] R. Usman, M. Jamil, and M. Naveed, "High preoperative neutrophil-lymphocyte ratio (NLR) and red blood cell distribution width (RDW) as independent predictors of native arteriovenous fistula failure," Ann. Vasc. Dis., p. oa-17, 2017. DOI: <https://doi.org/10.3400/avd.oa.17-00016>
- [17] A. Kirkpantur et al., "Association of serum lipid profile and arteriovenous fistula thrombosis in maintenance hemodialysis patients," Blood Purif., vol. 26, no. 4, pp. 322–332, 2008. DOI: <https://doi.org/10.1159/000132388>
- [18] Y. Al-Solaiman, A. Jesri, Y. Zhao, J. D. Morrow, and B. M. Egan, "Low-sodium DASH reduces oxidative stress and improves vascular function in salt-sensitive humans," J. Hum. Hypertens., vol. 23, no. 12, p. 826, 2009. DOI: <https://doi.org/10.1038/jhh.2009.32>
- [19] C. Rugale, S. Delbosc, J.-P. Cristol, A. Mimran, and B. Jover, "Sodium restriction prevents cardiac hypertrophy and oxidative stress in angiotensin II hypertension," Am. J. Physiol.-Heart Circ. Physiol., vol. 284, no. 5, pp. H1744–H1750, 2003. DOI: <https://doi.org/10.1152/ajpheart.00864.2002>
- [20] A. S. Todd et al., "Dietary salt loading impairs arterial vascular reactivity," Am. J. Clin. Nutr., vol. 91, no. 3, pp. 557–564, 2010. DOI: <https://doi.org/10.3945/ajcn.2009.28645>
- [21] S. K. Morton, A. J. Rodriguez, D. R. Morris, A. P. Bhandari, J. V. Moxon, and J. Golledge, "A systematic review and meta-analysis of circulating biomarkers associated with failure of arteriovenous fistulae for haemodialysis," PloS One, vol. 11, no. 7, p. e0159963, 2016. DOI: <https://doi.org/10.1371/journal.pone.0159963>
- [22] T. Ishii et al., "Duplex ultrasound for the prediction of vascular events associated with arteriovenous fistulas in hemodialysis patients," J. Vasc. Access, vol. 17, no. 6, pp. 499–505, 2016. DOI: <https://doi.org/10.5301/jva.5000595>

- [23] J. T. Salonen, K. Nyssönen, H. Korpela, J. Tuomilehto, R. Seppänen, and R. Salonen, "High stored iron levels are associated with excess risk of myocardial infarction in eastern Finnish men," *Circulation*, vol. 86, no. 3, pp. 803–811, 1992. DOI: <https://doi.org/10.1161/01.cir.86.3.803>
- [24] M. A. Kaygin et al., "The relationship between arteriovenous fistula success and inflammation," *Ren. Fail.*, vol. 35, no. 8, pp. 1085–1088, 2013. DOI: <https://doi.org/10.3109/0886022X.2013.815100>
- [25] F. Sarı, H. Taşkapan, A. Sığırıcı, and B. Akpinar, "Evaluation of Risk Factors for Arteriovenous Fistula Failure in Patients Undergoing Hemodialysis," *Erciyes Med. JournalErciyes Tip Derg.*, vol. 38, no. 1, 2016. DOI: <https://doi.org/10.5152/etd.2016.005>

Figures

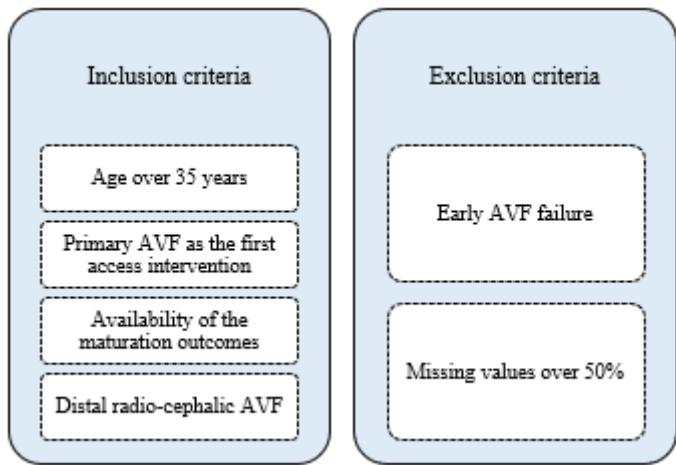


Figure 1

Inclusion and exclusion criteria for this study.

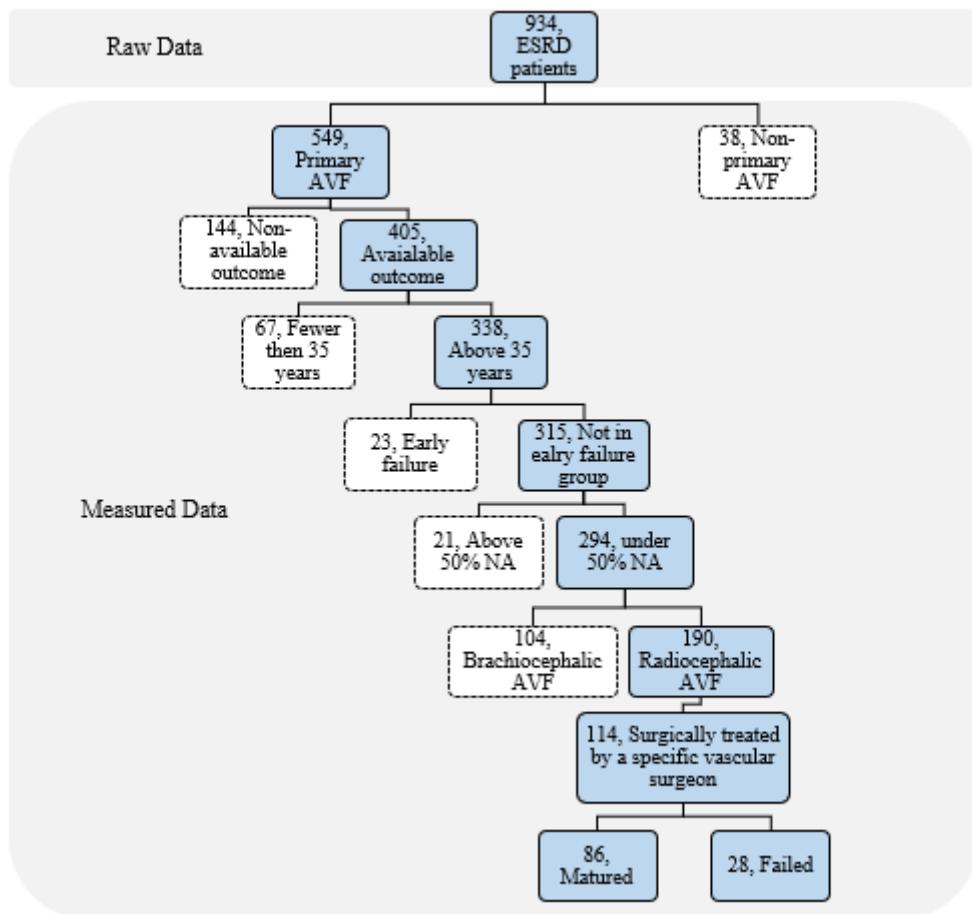


Figure 2

The inclusion/exclusion process for ESRD patients in this study.

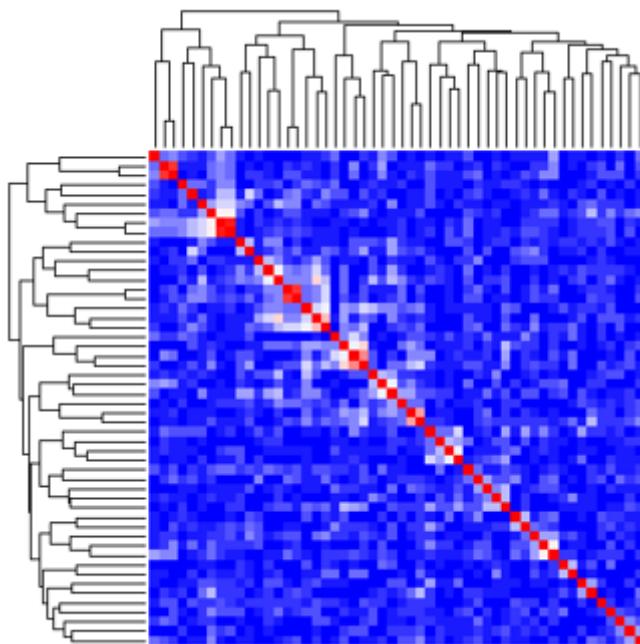


Figure 3

Hierarchically clustered correlation matrix of independent variables.

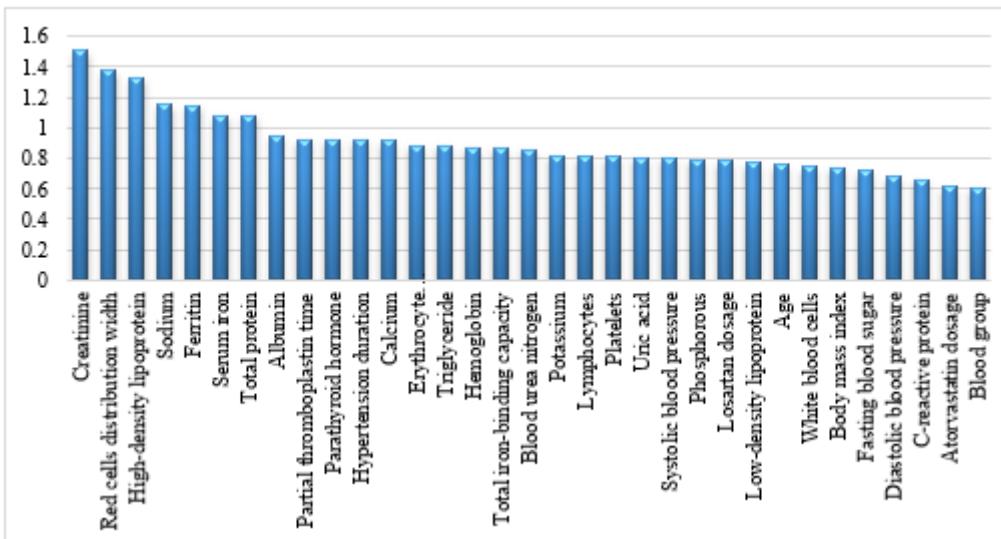


Figure 4

Variables importance based on mean decrease Gini extracted from the random forest model.

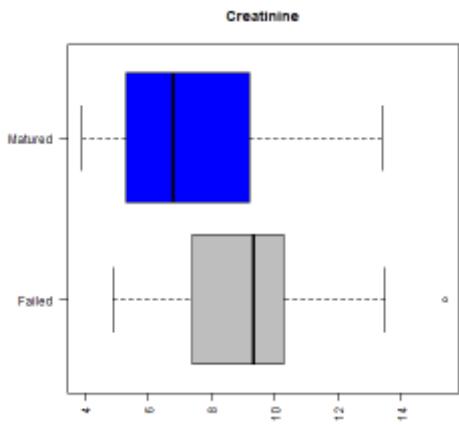


Figure 5

The boxplot of creatinine in two classes of "failed" and "matured."

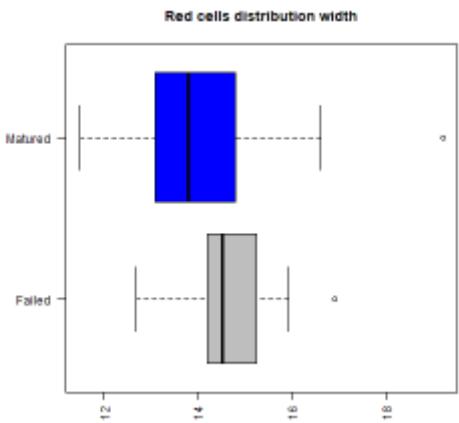


Figure 6

The boxplot of red blood cell distribution width separated by classes of "failed" and "matured."

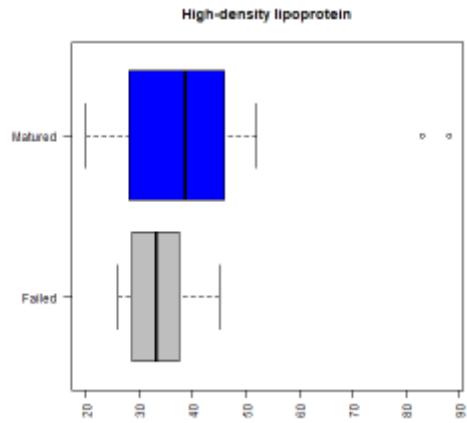


Figure 7

The boxplot of high-density lipoprotein in two classes of "failed" and "matured."

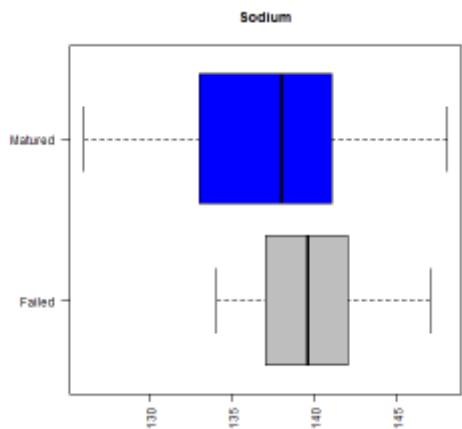


Figure 8

The boxplot of Sodium in two classes of "failed" and "matured."

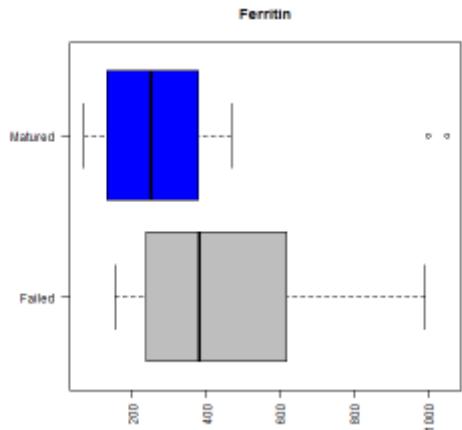


Figure 9

The boxplot of ferritin in two classes of "matured" and "failed."

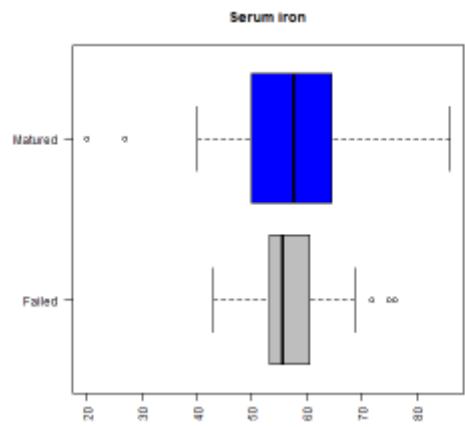


Figure 10

The boxplot of serum iron in two classes of "failed" and "matured."

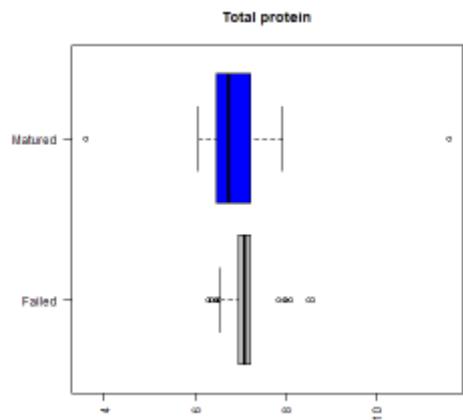


Figure 11

The boxplot of total protein in two classes of "failed" and "matured."