*International Journal of Hospital Research* 2012, **2**(1):49-54
www.ijhr.iums.ac.ir

RESEARCH ARTICLE

IJHR
Open Access

# Data Mining Performance in Identifying the Risk Factors of Early Arteriovenous Fistula Failure in Hemodialysis Patients

Morteza Khavanin Zadeh [1], Mohammad Rezapour [2], Mohammad Mehdi Sepehri [2,3*]

[1] Hasheminejad Clinical Research Development Center (HCRDC), Iran University of Medical Sciences (IUMS), Tehran, Iran [2] Department of Industrial Engineering, School of Engineering, Tarbiat Modares University, Tehran, Iran [3] Hospital Management Research Center (HMRC), Iran University of Medical Sciences (IUMS), Tehran, Iran

## Abstract

**Background and Objectives:** Arteriovenous fistula is a popular vascular access method for surgical treatment of hemodialysis patients. The method, however, is associated with a high rate of early failure varying in the range of 20-60%. Predicting early Arteriovenous fistula failure and its risk factors can help reduce its incidence, its hospitalization rate, and associated costs. In this study, we examined performance of data mining in the prediction of early AVF failure and identification of its risk factors.

**Methods:** The data of 193 patients who underwent homodialysis in Hasheminejad Kidney Center were explored. Eight common attributes of the patients including age, sex, hypertension level, Diabetes Mellitus state, hemoglobin level, smoking behavior, location of Arteriovenous fistula, and thrombosis state were used in the machine learning process. Two learning operators including W-Simple Cart and WJ48 tree were used in data mining process.

**Findings:** Smoking was identified as a factor influencing the relationship between the outcome of vascular access surgery and hemoglobin level. Prediction accuracy varied within the range of 69.15-85.11%.

**Conclusions:** According to our results smoking is a crucial risk factor for early Arteriovenous fistula failure, even at normal levels of hemoglobin. Our results provide further supports for the notion that data mining can help medical decision-making process by deciphering the complex interactions between various biological variables and translating the hidden patterns in data into detailed decision-making criteria.

**Keywords:** Data Mining, Prediction, Hemodialysis, Arteriovenous Fistula, Vascular Access, Knowledge Discovery

## Background and Objectives

Knowledge Discovery in Databases (KDD) is described as the non-trivial process of identifying valid patterns in crude data [1]. Data Mining is one of the crucial steps in KDD, which includes use of explorative algorithms to identify meaningful patterns in data with acceptable computational efficiency [1].

Data mining is becoming increasingly popular in healthcare-related studies. This approach has proved useful in medical sciences for its performance in in-depth analysis of large and complex datasets and generation of testable evidence-based medical hypotheses. Data mining is increasingly used in areas such as medical diagnosis, prediction of intervention outcomes, and clinical decision making process [2, 3]. In the field of urology, data mining is used to identify the factors contributing to Arteriovenous Fistula (AVF) failure [4]. Temporal data mining techniques are used predicting of dialysis failure and deriving information from the dialysis data [5]. K-means and expectation maximization (EM) algorithms are used for clustering the attributes of hemodialysis (HD) patients [6]. Mining HD data has resulted in a successful patients' risk prediction [7,8]. In our previous study, we were able to cluster attributes of early fistula surgery failure for 99 patients using descriptive data mining methods [9]. In this study, in order to gain further insight into the factors contributing to early AVF failure, we adopted a predictive data mining approach to identifying risk factors.

*Corresponding author: Mohammad Mehdi Sepehri, Department of Industrial Engineering, School of Engineering, Tarbiat Modares University (TMU), Tehran, Iran, P.O.Box: 1411713114. Phone/Fax: +98 21 8288 3379, E-mail: mehdi.sepehri@modares.ac.ir

Anemia is a common complication of Chronic Kidney Disease (CKD) and an important clinical characteristic of progressive kidney disorders [10]. The disorder usually worsens with the development of renal failure [10]. Administration of recombinant human erythropoietin is considered as a common treatment to this illness [10, 11]. However, the effectiveness of this treatment is limited due to unclear optimal target hemoglobin (Hgb) level in the patients at different stages of CKD [12]. Indeed, the effect of treatment can vary in those who need dialysis and those who do not; the latter group generally encounters less advanced cardiovascular problems, lower risk of dialysis-related Hgb increase, and lower risk of vascular access thrombosis as compared with patients who need dialysis [12]. Although AVF is associated with lower hospitalization rates as compared with other methods [13], vascular access failure has been identified as the most common reason for hospitalization of HD patients [14]. In this study, we examine the performance of two data mining algorithms in identifying the relationship between Hgb and early AVF failure and predicting the factors influencing such a relationship.

## Methods

### Settings

Data was collected from Hasheminejad Kidney Center (HKC), one of the largest urology hospitals in Iran. Patients' demographic and medical data were extracted from their clinical records. Medical variables and the status of AVF were extracted for two conditions: during surgery and after surgery.

### Dataset

Our initial dataset consisted of two parts. The first part included the records of 36 variables related to 99 patients who referred to HKC during 2005 to 2006, and had experienced early AVF failure. The second part of the dataset included the records of 25 variables related to 94 patients under HD in HKC during December and November 2010, of whom 87 had successful surgery and the rest were subject to early AVF failure. The two datasets were combined to produce an integrated database, including 193 samples accounting for 106 failure and 87 success cases.

Eight common attributes of the patients including age, sex, hypertension level (Htn), Diabetes Mellitus (DM) state, hemoglobin level (Hgb), smoking behavior, location of AVF, and thrombosis state were used in the data mining process.

### Data Mining Process

Our two datasets differed in inclusion of the *Location* field, which is the site of VA determined during surgery. Given that, the data mining process was carried out in a two-stage procedure; stage 1 where the *Location* field was excluded from the records, and stage 2 where all target variables, including Location were accounted for. A supervised learning approach was adopted for system training based on fortuitous sampling of 99 records. Two algorithms were used for machine learning; W-Simple Cart that enables a rule-based classification of the data, and WJ48 tree that provides a tree-based description of the extracted patterns. After training, the algorithms were executed on the remaining part of the dataset (94 records) to predict AVF failure cases and the associated risk factors.

## Results and Discussion

### Mining AVF-related data, without including the surgery location

*W-Simple Cart*

Table 1 presents the results of execution of W-Simple Cart on the entire dataset from which Location field was excluded. The algorithms identified a threshold of 8.05 g/dL, above which there is a high risk of AVF failure for smoking patients. The algorithm was then trained on the first part of the dataset and then used to predict the frequency of AVF failure in the second part. From the total 94 records, 66 cases of AVF success and 28 cases of AVF failure were predicted, yielding a prediction rate of 69.15%.

*WJ48*

We replaced the learner operator with WJ48. Results of execution of the algorithm on the entire dataset are represented in Figure 1. Consistent with the results from W-Simple Cart, WJ48 identified smoker patients to be at a high risk of AVF failure. In addition, it was indicated that among female patients with no hypertension, an Hgb level of lower than 6.8 g/dL increases the risk of AVF failure.

We further trained WJ48 on a random sample of 100 records, and then used the algorithm to predict AVF failure using the rest of the dataset (Figure 2). It was found that smoking modulate the effect of Hgb level on AVF status. In addition, in contrast with non-smoking patients, an Hgb level of greater than the previously identified threshold (8 g/dL) [16] is associated with a higher risk of AVF failure in smoking patients. Executing the algorithm on the second part of the dataset predicted 71 successes vs. 23 failures corresponding to a prediction rate of 76.6%.

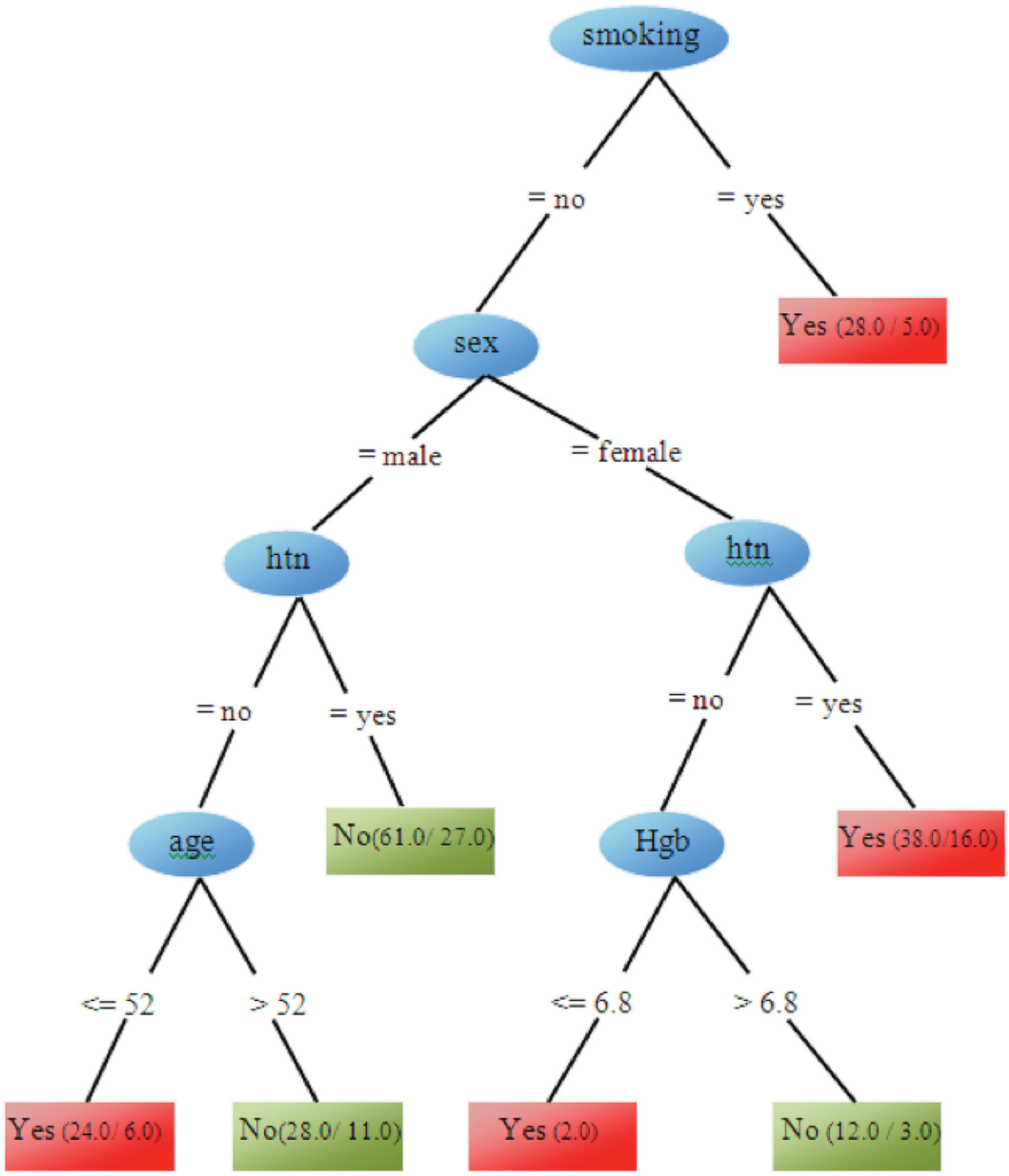


**Figure 1** The Impact of Low Hgb on Early AVF Failure (Yes: failure; No: success)

### Mining AVF related data, including surgery location

#### W-Simple Cart

We further conducted a mining study using W-Simple Cart algorithm by including the Location field (Table 2). When trained on the first part of the dataset and applied to the second part, the algorithm predicted 87 successes vs. 7 failures, yielding a prediction rate of 85.11%.

#### *WJ48*

We also trained WJ48 tree on the first part of the dataset and then examined the prediction capability of the learned operator using the the rest of the data (Figure 3). As seen, Location is identified as a determinant of AVF status. In the branch 'radial', smoking modulates the impact of Hgb level on AVF status. Hgb threshold

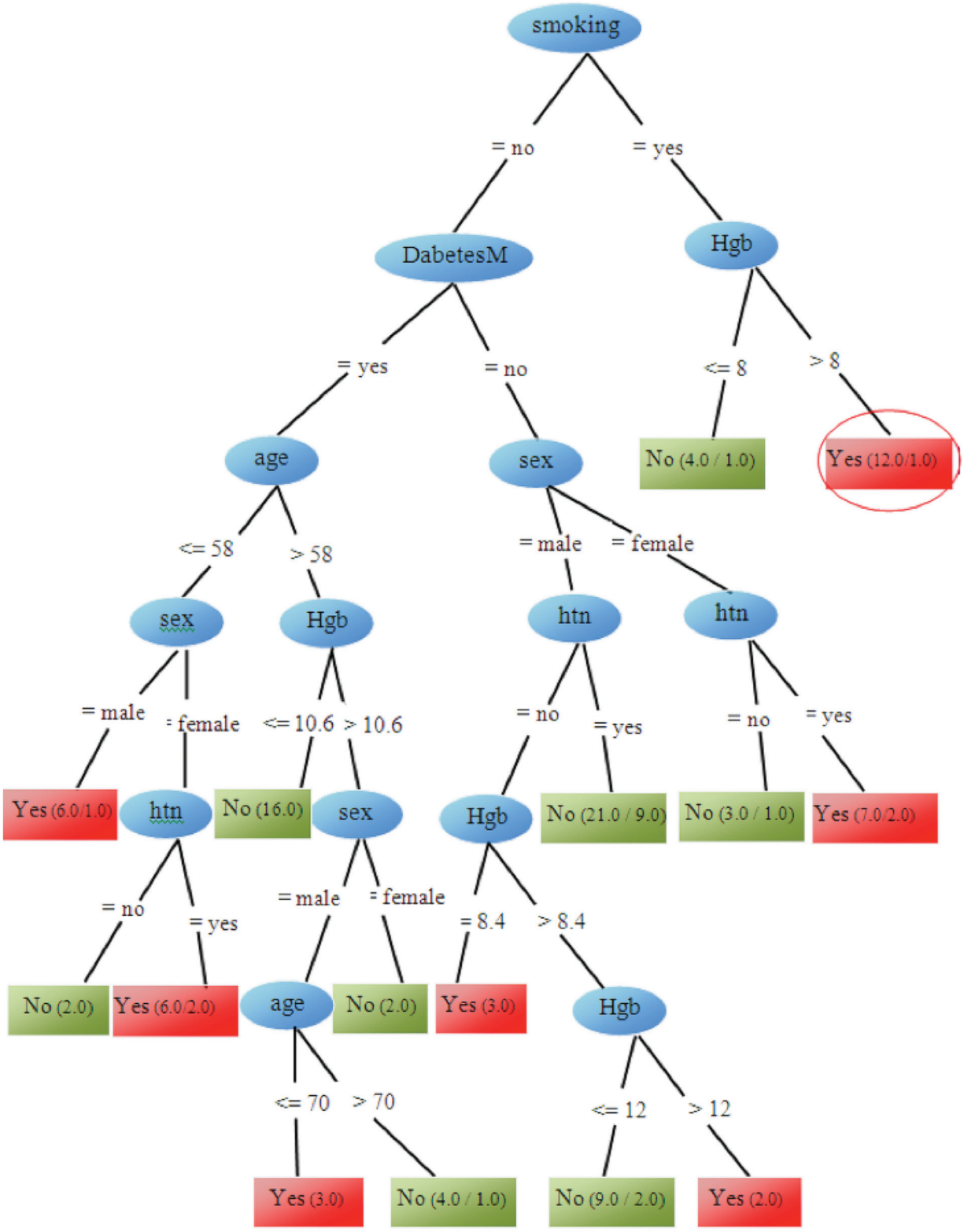


**Figure 2** Adverse effects of smoking where AVF Location was excluded

for classifying patients remained identical to the situation where Location was excluded from the records. In addition, at the lower levels of the branch, among males with no hypertension, Hgb level appeared again as a determinant of AVF status, though with a higher threshold. Executing the algorithm on the second part of the dataset predicted 81 successes vs. 13 failures corresponding to a prediction rate of 80.85%.

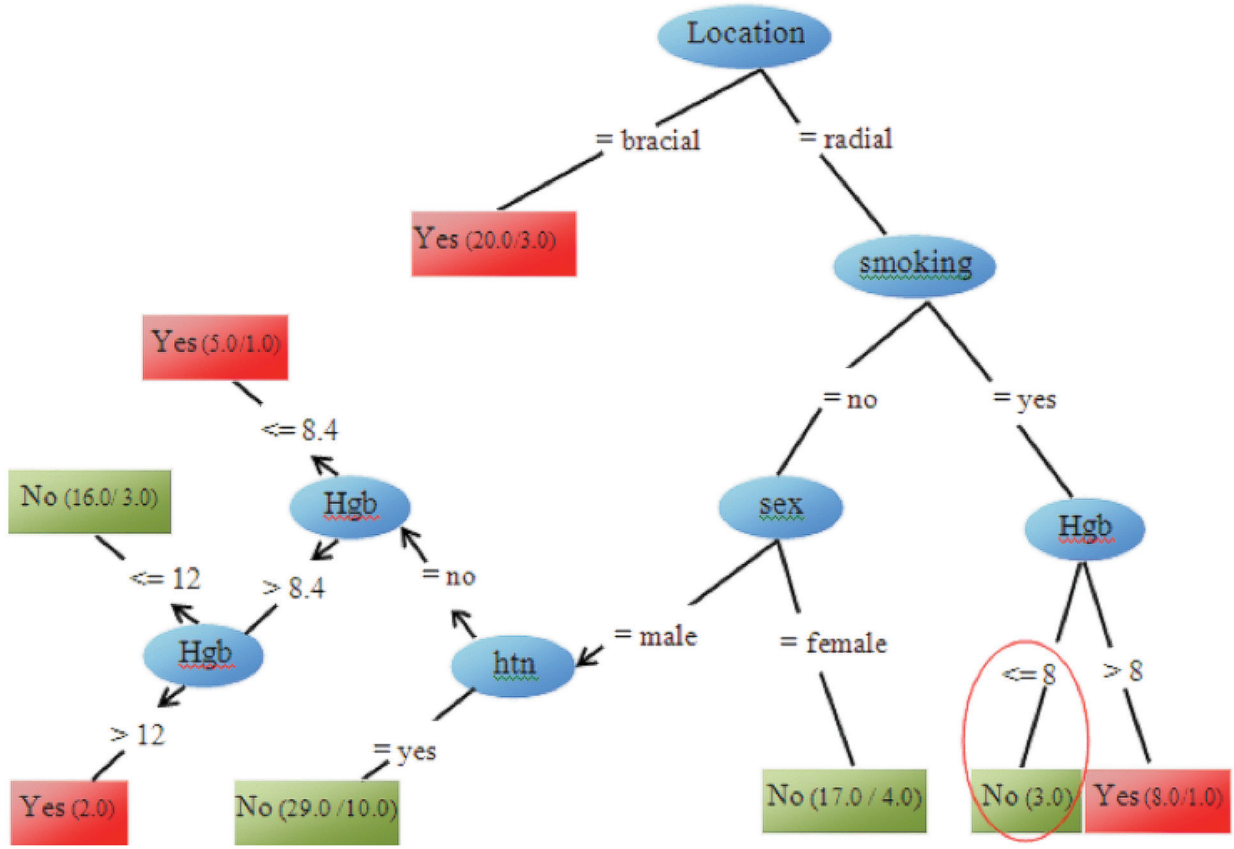


**Figure 3** Adverse effects of smoking where AVF Location was included

**Table1** Effect of Smoking on Hgb Impact

***W-SimpleCart***

CART Decision Tree

smoking=(no)

| age < 45.0: yes(15.0/9.0)

| age >= 45.0

| | Hgb< 7.45: yes(3.0/1.0)

| | Hgb>= 7.45

| | | Hgb< 9.75: no(25.0/6.0)

| | | Hgb>= 9.75

| | | | Hgb< 9.95: yes(5.0/0.0)

| | | | Hgb>= 9.95: no(14.0/6.0)

smoking!=(no)

| Hgb< 8.05: no(3.0/1.0)

| Hgb>= 8.05: yes(11.0/1.0)

Number of Leaf Nodes: 7

Size of the Tree: 13

**Table 2 The Extracted Rules where AVF Location Filed Was Included**

***W-SimpleCart***

CART Decision Tree

Location=(radial)

| Hgb< 12.6: no(50.0/26.0)

| Hgb>= 12.6: yes(4.0/0.0)

Location!=(radial): yes(17.0/3.0)

Number of Leaf Nodes: 3

Size of the Tree: 5

## Summary

Our data mining study identified smoking as the risk factors of early AVF failure. The study findings showed that smoking patients under HD are exposed to a higher risk of early AVF failure. In addition, it was found that high Hgb levels (greater than 8 g/dL) do not decrease the risk of

AVF failure in smoking patients. Hence, our study suggests that control of smoking in patients would be an effective strategy to reduce the risk of early AVF failure.

## Conclusions

This study indicated that data mining approach can predict the risk factors of AVF failure with a relatively high accuracy. Our findings provide further supports for the notion that data mining can help medical decision-making process by deciphering the complex interactions between various biological variables and translating the hidden patterns in data into detailed decision-making criteria.

### Abbreviations

(AVF): Arteriovenous Fistula; (AVG): Arteriovenous Graft; (CKD): Chronic Kidney Disease; (CVC): Central Venous Catheter; (DM): Diabetes Mellitus; (EM): Expectation-Maximization; (Hct): Hematocrit; (HD): Hemodialysis; (Hgb): Hemoglobin; (HKC): Hasheminejad Kidney Center; (IUMS): Iran University of Medical Sciences; (KDD): Knowledge Discovery in Databases; (KDOQI): Kidney Disease Outcomes Quality Initiative; (TMU): Tarbiat Modares University; (VA): Vascular Access.

### Competing Interests

The authors declare no competing interests.

### Authors' Contributions

MMS and MKZ jointly designed the study. MKZ and MR contributed in collecting and
integrating the data and interpretation of the results. MR was involved in literature review, running data mining algorithms and preparation of the draft manuscript. All authors read and approved the final manuscript.

Received: 19 September 2012 Revised: 8 January 2013 Accepted: 21 March 2013

## References

1. Fayyad U, Piatetsky-Shapiro G, and Smyth P. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence 1996; 0738-4602.
2. Koh H.C., Tan G. Data Mining Applications in Healthcare, Journal of Healthcare Information Management, 2005; Vol. 19, No. 2.
3. Biafore, S. Predictive solutions bring more power to decision makers. Health Management Technology 1999; 20(10): 12-14.
4. Kusiak A., Dixon B. and Shah S. Predicting survival time for kidney dialysis patients: a data mining approach, Computers in Biology and Medicine 2005; 35: 311–327.
5. Bellazzi, R., Larizza, C., Magni, P., et al. Temporal data mining for the quality assessment of hemodialysis services, Artificial Intelligence in Medicine 2005; 34: 25-39.
6. BriesM. F. Modeling of hemodialysis patient hemoglobin: a data mining exploration, Master's thesis, University of Iowa, 2007.
7. Knorr T., Schmidt Thieme L., Johner C. Identifying Patients at Risk: Mining Dialysis Treatment Data. Cooperation in Classification and Data Analysis, Springer Berlin Heidelberg, 2009: 131-140.
8. Mohammad Rezapour, Morteza Khavanin Zadeh, and Mohammad Mehdi Sepehri, "Implementation of Predictive Data Mining Techniques for Identifying Risk Factors of Early AVF Failure in Hemodialysis Patients," Computational and Mathematical Methods in Medicine, vol. 2013, Article ID 830745, 8 pages, 2013.
9. Sepehri M. M., Khavanin Zadeh M., Rezapour M., et al. A Data Mining Approach to Fistula Surgery Failure Analysis in Hemodialysis Patients. ICBME 2011 proceeding, 2011; pp 21-26.
10. Garrancho JM, Kirchgessner J, Arranz M, Klinkner G, Rentero R, Ayala JA and Marcelli D. Haemoglobin level and vascular access survival in haemodialysis patients. Nephrol Dial Transplant. 2005; 20: 2453-7.
11. Gheith O. A. and Kamal M. M. Risk Factors of Vascular Access Failure in Patients on Hemodialysis, Iranian Journal of Kidney Diseases, October 2008; 2(4): 201-207.
12. Tilman B. Drueke TB, Locatelli F, Clyne N, Eckardt KU, Macdougall IC, Tsakiris D, Burger HU and Scherhag A. Normalization of hemoglobin level in patients with chronic kidney disease and anemia. New Engl Journal of Medicine. 2006; 355 :2071-84.
13. Dhingra RK, Young EW, Hulbert-Shearon TE, et al. Type of vascular access and mortality in U.S. hemodialysis patients. Kidney Int 2001; 60:1443.
14. Feldman HI, Kobrin S, Wasserstein A. Hemodialysis vascular access morbidity. J Am Soc Nephrol. 1996; 7:523–35.
15. Furuland H, Linde T, Ahlmen J, et al., A randomized controlled trial of haemoglobin normalization with epoetin alfa in predialysis and dialysis patients. Nephrol Dial Transplant. 2003;18:353-61.
16. Khavanin Zadeh M, Gholipour F, Hadipour R.,The effect of hemoglobin level on arteriovenous fistula survival in Iranian hemodialysis patients. Journal of Vascular Access, 2008 Apr-Jun;9(2):133-6.

**Please cite this article as:**
Morteza Khavanin Zadeh, Mohammad Rezapour, Mohammad Mehdi Sepehri. Data Mining Performance in Identifying Risk Factors of Early Arteriovenous Fistula Failure in Hemodialysis Patients. *International Journal of Hospital Research* 2013, **2**(1):49-54.