

Research Article

A Self-Representation-Based Fuzzy SVM Model for Predicting Vascular Calcification of Hemodialysis Patients

Xiaobin Liu,¹ Xiran Zhang,¹ Xiaoyi Guo,¹ Yijie Ding ,² Weiwei Shan,¹ Liang Wang ,¹ Wei Zhou ,¹ and Hua Shi ³

¹Department of Nephrology, The Affiliated Wuxi People's Hospital of Nanjing Medical University, 214023, Wuxi, China

²School of Electronic and Information Engineering, Suzhou University of Science and Technology, 215009, Suzhou, China

³School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, 365001, Xiamen, China

Correspondence should be addressed to Liang Wang; wangliang_wuxi@126.com, Wei Zhou; 285403434@qq.com, and Hua Shi; shihua@xmut.edu.cn

Received 4 May 2021; Revised 30 June 2021; Accepted 8 July 2021; Published 28 July 2021

Academic Editor: Chuan Lu

Copyright © 2021 Xiaobin Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In end-stage renal disease (ESRD), vascular calcification risk factors are essential for the survival of hemodialysis patients. To effectively assess the level of vascular calcification, the machine learning algorithm can be used to predict the vascular calcification risk in ESRD patients. As the amount of collected data is unbalanced under different risk levels, it has an influence on the classification task. So, an effective fuzzy support vector machine based on self-representation (FSVM-SR) is proposed to predict vascular calcification risk in this work. In addition, our method is also compared with other conventional machine learning methods, and the results show that our method can better complete the classification task of the vascular calcification risk.

1. Introduction

Chronic kidney disease-mineral bone disease (CKD-MBD) is one of the most serious complications in patients with end-stage renal failure, including an abnormal metabolism of calcium, phosphorus, parathyroid hormone, vitamin D, abnormal bone transformation, vascular calcification, and ultimately cardiovascular disease.

In recent years, fibroblast growth factor (FGF23) has been recognized as a protein that plays an important role in phosphate regulation. Klotho protein is the receptor protein of FGF23. It participates in regulating the body's bone metabolism, calcium and phosphorus metabolism, protecting the integrity of blood vessels, and inhibiting vascular calcification through the formation of FGF23-klotho complexes. Therefore, FGF23 and klotho are key participants in CKD-MBD, and they are closely related to the occurrence of vascular calcification and cardiovascular disease. Existing evidence shows that there is a clear correlation between FGF23 and the occurrence of vascular calcification and cardiovascular

disease (CVD). The increase of FGF23 can be used as a risk factor for CVD in patients with end-stage renal disease (ESRD) [1].

Fetuin-A is considered to be an inhibitor of the progression of vascular calcification and can delay the progression of abdominal aortic calcification [2]. Studies have shown that there is a close correlation between fetuin-A and the malnutrition-microinflammatory state of ESRD patients [2]. It is currently believed that low serum fetuin-A levels in ESRD patients is an independent risk factor for vascular calcification.

Malnutrition is a common complication in ESRD patients, and it is closely related to vascular calcification, cardiovascular events, and all-cause mortality. Factors affecting the nutrition of ESRD patients include protein-energy expenditure, digestion and absorption, inflammation, and endocrine hormone level disorders [3].

There are a variety of tools available to assess the nutritional status of dialysis patients. Among them, the geriatric nutrition risk index (GNRI) is considered to be an important

predictor of cardiovascular death [4]. The latest research also shows that there is a certain positive correlation between GNRI and the degree of aortic calcification in CKD patients [5].

Vascular calcification (VC) scores of the artery or aorta on plain radiographs are associated with CVD events and may be predictive of CVD in dialysis patients [6]. Many research results show that abdominal aortic calcification as assessed on a lateral lumbar X-ray is predictive for the presence of significant coronary artery disease in asymptomatic dialysis patients [7].

The previous research results of our work show that patients with end-stage renal failure have abnormal levels of FGF23 and klotho and microinflammatory states. Their interaction and mutual influence are involved in the occurrence and development of vascular calcification and CKD-MBD [8].

Therefore, in order to further explore the risk factors of vascular calcification in patients with ESRD, this article studies the scientific and accurate prediction of vascular calcification risk factors in ESRD patients with different forecasting model, so as to help clinicians to detect and intervene early, thereby delaying the occurrence and development of CKD-MBD, reducing the incidence of CVD, and improving the prognosis. Machine learning (ML) has been widely used in the dry weight (DW) [9] of hemodialysis patients and has achieved good results. Lots of ML-based models also have been well used in drug discovery [10–12], protein function [13–16], and disease analysis [17, 18].

In this study, we employ a support vector machine (SVM) to build a predictive model. SVM has the following advantages: (1) Nonlinear mapping is the theoretical basis of the SVM method. SVM uses the inner product kernel function to replace the nonlinear mapping to high-dimensional space. (2) The optimal hyperplane to divide the feature space is the goal of SVM, and the idea of maximizing the classification margin is the core of the SVM method. (3) A small number of support vectors determine the final result, which can not only help us capture key samples but also “remove” a large number of redundant samples. For imbalanced datasets, the standard SVM is not good at classifying a small number of categories. In this work, we propose a fuzzy support vector machine based on self-representation (FSVM-SR) to identify vascular calcification of hemodialysis patients under imbalanced data. FSVM can estimate a weight for each training sample. When constructing the hyperplane of classification, FSVM avoids some low-weight samples (noise samples) to alleviate the influences of imbalanced datasets.

2. Materials and Methods

2.1. Materials. This work employs 29 features to describe the patient’s information, which includes gender, age, body mass index (BMI), diabetes mellitus (DM), cerebral infarction (CI), and coronary heart disease (CHD). Table 1 shows the details of our dataset. The mean and standard deviation of samples is also list in it.

During the data collection process, we classified 59 patients into risk levels. We roughly classify the 7 risks according to two classification methods. Some adjacent levels will be grouped into one category. The classification results are shown in Table 2. In the first classification scheme (CS1), levels {0, 1, 2} and levels {3, 4, 5, 6} are classified into classes 1 and 2, respectively. In addition, the levels {0, 1}, {2, 3}, and {4, 5, 6} are classified into classes 1, 2, and 3 (in CS2), respectively.

2.2. Methods

2.2.1. Abdominal Aortic Calcification. All patients need to undergo lateral lumbar X-ray examination within 1 week of blood biochemical examination to assess the calcification of the abdominal aorta corresponding to levels 1-4 [7]. According to the length of the calcified plaques on the anterior and posterior walls of the abdominal aorta, for scores of 0 to 3: no calcification is 0 points, calcification range $<1/3$ arterial wall length is 1, calcification range $1/3-2/3$ arterial wall length is 2, calcification range $>2/3$ arterial wall length is 3, and total score is between 0 and 24. Two radiologists separately scored and averaged. The calculation of geriatric nutrition risk index (GNRI) [19] can be estimate by

$$\text{GNRI} = [14.89 \times \text{serum albumin}] + \left[41.7 + \left(\frac{\text{actual body weight}}{\text{ideal body weight}} \right) \right]. \quad (1)$$

Serum levels of intact FGF23, soluble Klotho, Fetuin-A, and interleukin-6 were received by using two-site enzyme-linked immune assays (reagents from Elabscience Biotech, Wuhan, China).

2.2.2. Fuzzy Support Vector Machine. SVM is a robust machine learning method based on statistical learning, which considers empirical risk and adds a regularization term to reduce structural risk. It is a sparse and robust classifier [20]. SVM also can perform nonlinear classification through the kernel method, which is one of the common kernel learning methods. In many practical classification tasks, the number of samples in different categories is often different. Under the imbalanced dataset, the SVM model will produce a large deviation. In order to avoid the above situation, Lin and Wang proposed fuzzy SVM (FSVM) [21]. Different from SVM, FSVM uses membership value to describe the weight of the training sample. In general, the membership value of outlier samples is lower, and it is easier for the algorithm to weaken the contribution to the decision hyperplane during the training process.

For FSVM, a training sample can be defined as $\{\mathbf{x}_i, y_i, s_i\}$, $i = 1, 2, \dots, N$, where N is the number of training samples, $\mathbf{x}_i \in \mathbb{R}^{d \times 1}$, y_i , and $s_i \in [0, 1]$ are feature vector, label, and membership value of sample i , respectively. The feature vector dimension of the model is d . The objective optimization function of FSVM is

TABLE 1: The information of dataset.

No.	Feature	Value	r^*
1	Gender (males/females)	32/27	-0.0455
2	Age (years)	55.83 ± 15.60	0.4010
3	Smoking (yes/no)	1/58	-0.0847
4	BMI (kg/m ²)	23.56 ± 3.12	0.1639
5	DM (yes/no)	24/35	0.4847
6	CI (yes/no)	3/56	0.0025
7	CHD (yes/no)	5/54	0.0433
8	Systolic blood pressure (mmHg)	155.69 ± 23.63	-0.1150
9	Diastolic blood pressure (mmHg)	88.11 ± 13.71	-0.2043
10	Phosphate binder (yes/no)	36/23	-0.2141
11	Hemoglobin (g/L)	85.38 ± 18.12	-0.2584
12	C-reactive protein (mg/L)	11.74 ± 35.61	0.3016
13	Serum creatinine (μmol/L)	785.09 ± 368.62	-0.4252
14	Serum glucose (mmol/L)	5.48 ± 2.28	0.2608
15	Serum calcium (mmol/L)	2.08 ± 0.24	0.0520
16	Serum phosphorus (mmol/L)	1.81 ± 0.38	-0.0862
17	Total glyceride (mmol/L)	1.69 ± 1.08	-0.0542
18	Total cholesterol (mmol/L)	4.53 ± 1.42	-0.0466
19	Low-density lipoprotein-C (mmol/L)	2.45 ± 0.96	-0.0252
20	High-density lipoprotein-C (mmol/L)	0.97 ± 0.55	0.0866
21	HbA1c (%)	5.82 ± 1.03	0.2151
22	Serum albumin (g/L)	34.31 ± 6.61	-0.1308
23	25-OH vitamin D3 (ng/mL)	7.86 ± 4.55	0.3850
24	iPTH (pg/mL)	274.50 ± 306.31	-0.0225
25	GNRI	96.06 ± 12.76	-0.0078
26	FGF-23 (pg/mL)	32.21 ± 53.02	-0.0966
27	Klotho (ng/mL)	2.38 ± 2.33	0.0443
28	Interleukin-6 (pg/mL)	25.37 ± 53.69	0.2634
29	Fetuin-A (pg/mL)	3.0320e + 05 ± 2.0606e + 05	-0.0234

*Denotes that each feature correlated with ascular calcification level using the Pearson correlation coefficient (r).

TABLE 2: The information of patient risk classification scheme.

Risk level	Number of patients	CS1	CS2
0	10		
1	15	Levels 0, 1, and 2 are class 1 (42 samples)	Levels 0 and 1 are class 1 (25 samples)
2	17		Levels 2 and 3 are class 2 (23 samples)
3	6	Levels 3, 4, 5, and 6 are class 2 (17 samples)	
4	6		
5	4		Levels 4, 5, and 6 are class 3 (11 samples)
6	1		

CS: classification scheme.

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N s_i \xi_i, \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, 2, \dots, N, \end{aligned} \quad (2)$$

where C denotes the regularization parameter, ξ_i is the error measure of \mathbf{x}_i . To build a robust model, different training samples should be given different regularization parameters. $s_i \xi_i$ is the error measure, which is weighted by the membership value. Outliers (noise) have a lower weight; on the contrary, important sample points will have a higher weight. In an imbalanced dataset, the type of data with a large number of samples often contains more outliers. In order to reduce the deviation, FSVM can well reduce its impact. Equation (2) also can be rewritten by the Lagrange dual problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \cdot y_i y_j \cdot K(\mathbf{x}_i, \mathbf{x}_j), \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq s_i C, \\ & \sum_{i=1}^N \alpha_i y_i = 0, i = 1, 2, \dots, N, \end{aligned} \quad (3)$$

where α_i is the Lagrange multiplier coefficient for sample \mathbf{x}_i . $K(\mathbf{x}_i, \mathbf{x}_j)$ is the value of samples i and j in the kernel matrix. And the kernel matrix can be calculated by the radial basis function (RBF):

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &= \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \\ i, j &= 1, 2, \dots, N, \end{aligned} \quad (4)$$

where γ is a Gaussian kernel bandwidth.

The final decision function of classification is

$$f(\mathbf{x}) = \text{sign} \left[\sum_{i=1}^N y_i \alpha_i \cdot K(\mathbf{x}, \mathbf{x}_i) + b \right]. \quad (5)$$

The basic SVM can only perform binary classification tasks. In this work, we use the one-against-one strategy to achieve multiple classifications.

2.2.3. Self-Representation-Based Membership Function. In this work, we propose a method based on a reconstruction error to construct the membership function. This method can measure the consistency between the overall data structure and a single data point. The reconstruction error can quantify the outlier degree of the noise sample, which helps to improve the robustness of the model.

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbf{R}^{d \times N}$, the self-representation function is defined as follows:

$$\mathbf{X} = \mathbf{XZ} + \mathbf{E}, \quad (6)$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N] \in \mathbf{R}^{N \times N}$ and $\mathbf{E} \in \mathbf{R}^{d \times N}$ are the coeffi-

cient and error matrix. \mathbf{z}_i is the new representation of sample i by other training samples. The self-representation formulation can be optimized by

$$\min J(\mathbf{Z}) = \|\mathbf{XZ} - \mathbf{X}\|_F^2 + \lambda \text{Tr}(\mathbf{ZLZ}^T), \quad (7)$$

where $\text{Tr}(\mathbf{ZLZ}^T)$ is the Laplacian regular term to smooth the coefficient \mathbf{Z} :

$$\text{Tr}(\mathbf{ZLZ}^T) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N W_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2, \quad (8)$$

where \mathbf{W} is the similarity matrix between samples. It also can be replaced by kernel matrix. $\mathbf{L} = \mathbf{D}^{-1/2} \Delta \mathbf{D}^{-1/2}$ is a normalized Laplacian matrix and $\Delta = \mathbf{D} - \mathbf{W}$. The $D_{ii} = \sum_{j=1}^N W_{ij}$ is an element of the diagonal matrix $\mathbf{D} \in \mathbf{R}^{N \times N}$. In this work, λ denotes the coefficient of the Laplacian regular term, which is set as 0.01. Setting $\partial J(\mathbf{Z}) / \partial \mathbf{Z} = 0$, the solution of Equation (7) can be obtained as follows:

$$\begin{aligned} \frac{\partial J(\mathbf{Z})}{\partial \mathbf{Z}} &= 0, \\ 2\mathbf{X}^T(\mathbf{XZ} - \mathbf{X}) + 2\lambda \mathbf{ZL} &= 0, \\ \mathbf{X}^T \mathbf{XZ} + \lambda \mathbf{ZL} &= \mathbf{X}^T \mathbf{X}, \end{aligned} \quad (9)$$

where $\mathbf{X}^T \mathbf{XZ} + \lambda \mathbf{ZL} = \mathbf{X}^T \mathbf{X}$ is a Sylvester equation. For each training sample, the reconstruction error of \mathbf{x}_i can be calculated as

$$r_i = \|\mathbf{Xz}_i - \mathbf{x}_i\|_2^2. \quad (10)$$

To map the value of reconstruction error in $0 \sim 1$. We define the following formula:

$$s_i = 1 - \frac{r_i - r_{\min}}{r_{\max} - r_{\min}}, \quad (11)$$

where r_{\min} and r_{\max} are the minimum and maximum reconstruction errors, respectively. The process of our method is list in Algorithm 1.

3. Results

3.1. Evaluation measurements. In our study, the accuracy (ACC) is employed to evaluate the predictive performance of our predictive model. In addition, a 10-fold crossvalidation method [22–25] was used in this work. The calculation method of ACC is as follows:

$$\begin{aligned} \text{Whole ACC} &= \frac{\sum_{i=1}^c \text{TP}^i}{M} \times 100\%, \\ \text{ACC}^i &= \frac{\text{TP}^i}{M^i} \times 100\%, \\ M &= \sum_{i=1}^c M^i, \end{aligned} \quad (12)$$

Require: training set $\{x_i, y_i\}, i = 1, 2, \dots, N$; ($\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$), test set $\mathbf{x}^{te} = \{\mathbf{x}_j^{te}\}, j = 1, 2, \dots, M$, the parameters of C and λ ; the Gaussian kernel bandwidth of γ .

Ensure: The predictive values of $\{y_j^{te}\}, j = 1, 2, \dots, M$.

1. Calculate the training kernel ($K(\mathbf{X}, \mathbf{X})$), test kernel matrix ($K(\mathbf{x}^{te}, \mathbf{X})$) and Laplacian matrix \mathbf{L} by Equation (4) and $\mathbf{L} = \mathbf{D}^{-1/2} \Delta \mathbf{D}^{-1/2}$;
2. Estimate the self-representing coefficient matrix of $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N] \in \mathbf{R}^{N \times N}$ by Equation (9);
3. Calculate the reconstruction error $r_i, i = 1, 2, \dots, N$ for each training sample via $r_i = \|\mathbf{X}\mathbf{z}_i - \mathbf{x}_i\|_2^2$;
4. Obtain the final membership value of each training sample by $s_i = 1 - (r_i - r_{\min}) / (r_{\max} - r_{\min})$;
5. Train FSVM-SR (obtaining $\alpha_i, i = 1, 2, \dots, N$) via solving Eq. (3);
6. Predict $\{y_j^{te}\}, j = 1, 2, \dots, M$ by SVM decision function: $f(\mathbf{x}) = \text{sign} [\sum_{i=1}^N y_i \alpha_i \cdot K(\mathbf{x}, \mathbf{x}_i) + b]$.

ALGORITHM 1: Algorithm of FSVM-SR.

where TP^i is the number true positive (TP) in subclass i . c is the number of classes. M and M^i denote the number of whole test samples and subclass test samples. ACC^i is the accuracy of subclass i .

4. Selection of Optimal Parameters

In order to obtain the best prediction performance, we use the grid search method to obtain the optimal parameters C and γ . The search ranges are from 2^{-5} to 2^{10} (C), and from 2^{-10} to 2^5 (γ), with the step of 2^1 . Figures 1. and 2 show the average ACC with different C and γ (under CS1 and CS2), respectively.

As shown in the figures, the model reach ACC of 83.05% and 64.40%, when the optimal parameters $C = 27, \gamma = 2^{-6}$ and $C = 210, \gamma = 2^{-6}$, respectively.

4.1. Comparison to Other Classifiers. To further evaluate the performance of our model, we introduced other similar machine learning models [26, 27], including logistic regression, back propagation (BP) neural network, radial basis function (RBF) neural network, Takagi-Sugeno-Kang fuzzy system (TSK-FS) [28–30], and standard SVM. Under 2 classes (Table 3), logistic regression, BP network, RBF network, and TSK-FS achieve whole ACC of 71.18%, 66.10%, 77.96%, and 76.27%, respectively. The whole ACC of SVM (79.66%) and FSVM-SR (83.05%) are better than other models. In particular, FSVM-SR obtains the best prediction accuracy. In subclasses 1 and 2, FSVM-SR also achieves best accuracy of 95.23% and 52.94%, respectively. It can be seen from the results that for small sample learning, the SVM has more advantages than the neural network models. As the fuzzy model, FSVM has better performance than TSK-FS on this dataset. FSVM-SR can effectively suppress the influence of noise samples on the model. The receiver operating characteristic curves (ROC) of different models are shown in Figure 3. It can also be found that our method obtains the highest area under curve (AUC) value of 0.7955.

Under CS3, FSVM-SR also compares with these predictors, and the results of comparison are listed in Table 4. SVM and FSVM-SR achieve the best whole ACC of 64.40%. In subclass 1, the ACC of RBF neural network and SVM are 72.00% (best). FSVM-SR and TSK-FS obtain 65.21%, and FSVM-SR has smaller standard deviation (30.88%) in subclass 2. In addition, SVM and FSVM-SR have better per-

formance (54.54%) in subclass 3. It can be seen that FSVM-SR is also more stable and effective in the case of CS3.

Two-sample t -test is employed to evaluate the significance differences of average ACC value in CS1 and CS2, respectively. In our work, the significance level is 0.05. FSVM-SR is compared with other models via 10-fold cross-validation (20 times). The results of statistical significance are shown in Table 5. In CS1, the differences between FSVM-SR and other models are all significant (P value < 0.05). The max value is 0.0064 (for SVM) and min value is $3.4341e-11$ (for BP neural network). Except for SVM (P value 0.1965), the differences with other models are significant in CS2. It can be seen from the results that the proposed method (FSVM-SR) outperforms most methods in two patient risk classification schemes.

5. Discussion

Cardiovascular death is the main cause of ESRD patients. Studies have confirmed that the occurrence of abdominal aortic calcification (in ESRD patients) is extremely important for cardiovascular death [31]. In recent years, the influence of nontraditional risk factors, such as FGF23, klotho abnormality, microinflammatory state, and malnutrition on vascular calcification has attracted much attention from scholars.

As a protein that plays a key role in phosphate regulation, FGF23 is involved in controlling the metabolism of phosphate, parathyroid hormone, and 1,25 dihydroxy vitamin D. FGF23 can not only regulate phosphate homeostasis but also further promote disease progression, left ventricular hypertrophy, and increase the occurrence and death of CVD [1]. Klotho, as an antiaging gene, has been confirmed by many studies that it participates in cardiovascular protection in ESRD patients by inhibiting phosphate-driven vascular calcification [32]. The results of previous studies of our center showed that serum FGF23 levels increased, and soluble klotho levels decreased. Moreover, after the combined abdominal aortic sclerosis, the abnormalities of serum FGF23 and klotho are more obvious [8]. This study is consistent with the results of previous studies. The levels of FGF23 and klotho are abnormal in ESRD patients. They are risk factors for vascular calcification. And the analysis results of different risk prediction models all support this conclusion.

Fetuin-A, as a protective factor for vascular calcification, can inhibit the process of vascular calcification [33]. The

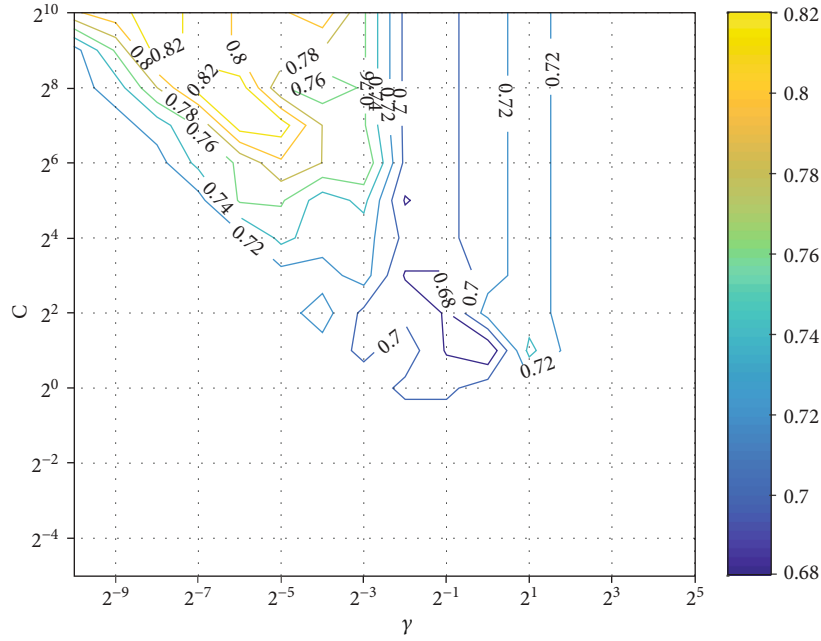


FIGURE 1: The average ACC with different C and γ (under CS1).

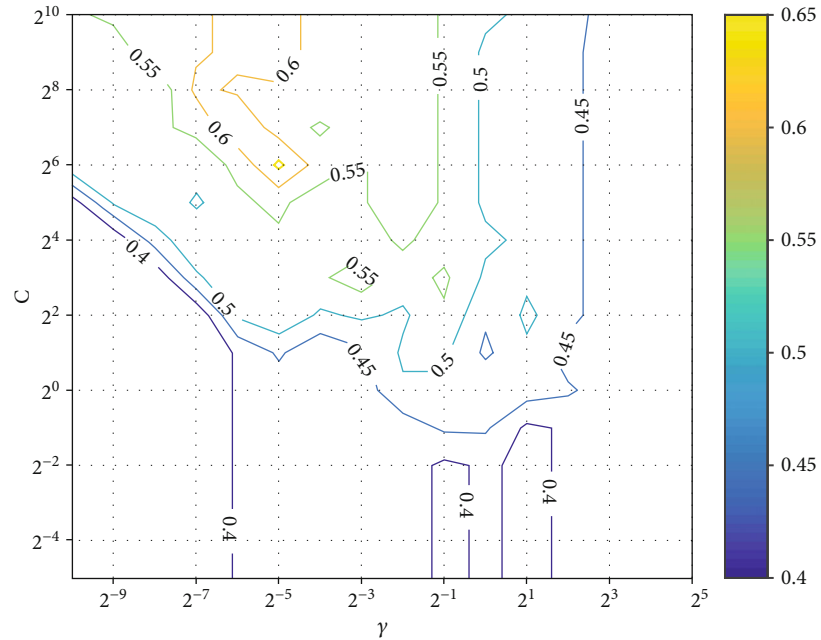


FIGURE 2: The average ACC with different C and γ (under CS2).

results of this study show that the serum fetuin-A of ESRD patients is significantly reduced. In addition, the serum fetuin-A decreases more significantly in patients with abdominal aortic calcification. The level of decrease has a certain early warning effect on vascular calcification. Both FSVM and traditional prediction models suggest that fetuin-A is an independent risk factor for abdominal aortic calcification. Therefore, clinicians should pay close attention to serum fetuin-A levels in the process of CKD. Once abnormalities occur, they should intervene as soon as possible.

Studies have shown that malnutrition is closely related to vascular calcification, cardiovascular death, and all-cause death in ESRD patients [4]. In this work, the results of different risk prediction models support that malnutrition is an independent risk factor for abdominal aortic calcification. Existing studies have shown that malnutrition and microinflammatory state, insulin resistance, FGF23/klotho axis abnormalities, and other factors are interconnected and ultimately jointly promote the occurrence of vascular calcification [19]. Therefore, many risk factors for vascular

TABLE 3: Comparison on existing models via 10-fold crossvalidation (under 2 classes).

Model	Whole ACC (%)	ACC ¹ (%)	ACC ² (%)
Logistic regression	71.18 ± 15.14	83.33 ± 19.01	41.17 ± 33.75
BP neural network	66.10 ± 13.70	73.80 ± 26.78	47.05 ± 43.78
RBF neural network	77.96 ± 8.67	88.09 ± 15.47	52.94 ± 36.89
TSK-FS	76.27 ± 15.31	85.71 ± 15.42	52.94 ± 43.78
SVM	79.66 ± 10.40	90.47 ± 14.15	52.94 ± 36.89
FSVM-SR (our method)	83.05 ± 9.71	95.23 ± 10.54	52.94 ± 36.89

ACC¹: the accuracy of class 1; ACC²: the accuracy of class 2.

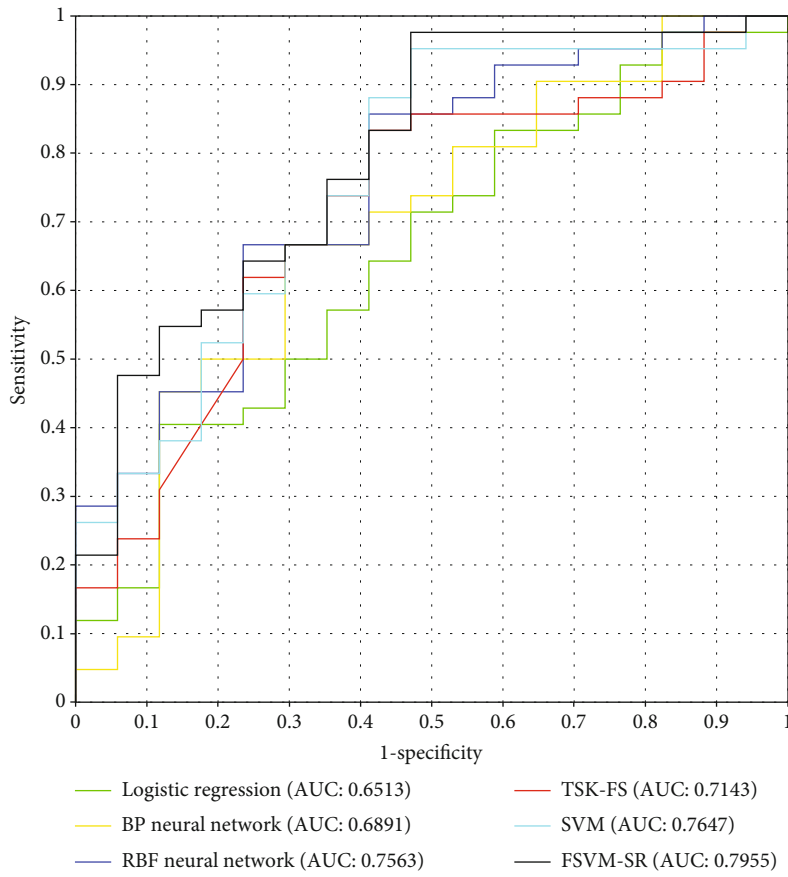


FIGURE 3: The ROCs for different models (under CS2).

TABLE 4: Comparison on existing models via 10-fold crossvalidation (under 3 classes).

Model	Whole ACC (%)	ACC ¹ (%)	ACC ² (%)	ACC ³ (%)
Logistic regression	57.62 ± 15.25	64.00 ± 32.20	60.86 ± 32.58	36.36 ± 48.30
BP neural network	55.93 ± 19.62	60.00 ± 32.44	60.86 ± 28.60	36.36 ± 47.43
RBF neural network	59.32 ± 21.99	72.00 ± 29.61	60.86 ± 41.91	27.27 ± 48.30
TSK-FS	62.71 ± 24.21	68.00 ± 28.81	65.21 ± 39.13	45.45 ± 49.72
SVM	64.40 ± 17.39	72.00 ± 21.94	60.86 ± 32.44	54.54 ± 49.72
FSVM-SR (our method)	64.40 ± 17.39	68.00 ± 24.85	65.21 ± 30.88	54.54 ± 49.72

ACC¹: the accuracy of class 1; ACC²: the accuracy of class 2; ACC³: the accuracy of class 3.

TABLE 5: Analysis of statistical significance for different methods via 10-fold crossvalidation (20 times).

Model	CS1 (P value)	CS2 (P value)
Logistic regression	$8.4386e - 10$	$1.1262e - 05$
BP neural network	$3.4341e - 11$	$6.9318e - 07$
RBF neural network	$7.9770e - 06$	0.0013
TSK-FS	$3.3813e - 08$	0.0024
SVM	0.0064	0.1965

calcification are often mixed, and it is difficult for clinicians to accurately determine the main risk factors and provide precise treatment intervention. By comparing the accuracy of different prediction models for predicting the risk of abdominal aortic calcification, it is found that the FSVM and SVM models are more accurate in identifying the main risk factors than the traditional logistic regression model. Under CS2, FSVM-SR achieves best accuracy of 95.23% and 52.94%, respectively. Our method (FSVM-SR) is significantly better than other methods (P value, logistic regression: $8.4386e - 10$, BP network: $3.4341e - 11$, RBF network: $7.9770e - 06$, TSK-FS: $3.3813e - 08$, and SVM: 0.0064). What is more, FSVM-SR and SVM achieve the best whole ACC of 64.40% in CS3. Self-representation-based membership function can estimate weight for training sample. The reconstruction error of outliers is relatively large, and the corresponding membership value is low. When constructing the hyperplane of classification, FSVM avoids some low-weight samples (outliers) to alleviate the influences of imbalanced datasets. The fuzzy methods [34] improve the interpretability and robustness of the model. There are related applications in the medical fields [35, 36].

6. Conclusions

In this work, we propose a FSVM based on a self-representation method to filter noise samples, improve the generalization ability of the model, and obtain good results. Although our method has achieved a better accuracy, it still has the following disadvantage: (1) The sample size needs to be further increased to minimize the prediction bias. (2) There is no detailed analysis of the various factors of the patient [37, 38]. (3) The interpretability of the model is not as good as that of the linear model. Based on the above, we will propose a sparse linear model in the next work to solve the problem of poor interpretability and factor analysis.

Data Availability

The data used to support the research can be obtained from the corresponding authors according to the requirements of the institution.

Ethical Approval

This study had been approved by the ethics committee of the hospital (ethical approval no. KS2019041). The experimental

protocol was established, according to the ethical guidelines of the Helsinki Declaration, and was approved by the Human Ethics Committee (Wuxi People's Hospital Ethics Committee).

Consent

Written informed consent for publication was obtained from all participants.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

Xiaobin Liu, Xiran Zhang, and Xiaoyi Guo are joint first authors.

Acknowledgments

We thank the Department of Nephrology of Wuxi People's Hospital for collecting data in this study. This work is supported by a grant from the Top Talent Support Program for young and middle-aged people of Wuxi Health Committee (HB2020008), the Scientific research project of Wuxi health committee (MS201927), the Scientific research project of Wuxi health committee (Z201914), the Scientific Research Projects of Jiangsu Provincial Health Commission (LGY201801), the Jiangsu Province "333" project (BRA2020142), National Natural Science Foundation of China (NSFC 61902271), and the Natural Science Research of Jiangsu Higher Education Institutions of China (19KJB520014).

References

- [1] D. Moldovan, I. Moldovan, C. Rusu, I. Kacso, I. M. Patiu, and M. Gherman-Caprioara, "FGF-23, vascular calcification, and cardiovascular diseases in chronic hemodialysis patients," *International Urology and Nephrology*, vol. 46, pp. 121–128, 2014.
- [2] R. A. Muzasti and R. Loesnihari, "High fetuin-A level as a protective factor to abdominal aortic calcification in Indonesian regular hemodialysis patients," *Open Access Macedonian Journal of Medical Sciences*, vol. 7, pp. 721–725, 2019.
- [3] H. Ishii, H. Takahashi, Y. Ito et al., "The Association of ankle brachial index, protein-energy wasting, and inflammation status with cardiovascular mortality in patients on chronic hemodialysis," *Nutrients*, vol. 9, p. 416, 2017.
- [4] T. Okamoto, S. Hatakeyama, H. Kodama et al., "The relationship between poor nutritional status and progression of aortic calcification in patients on maintenance hemodialysis," *BMC Nephrology*, vol. 19, p. 71, 2018.
- [5] K. Harada, S. Suzuki, H. Ishii et al., "Nutrition status predicts severity of vascular calcification in non-dialyzed chronic kidney disease," *Circulation Journal*, vol. 81, pp. 316–321, 2017.
- [6] H. S. Nam, S. M. Lee, E. G. Jeong et al., "Vascular calcification on plain radiographs is related with the severity of lesions detected by coronary angiography in dialysis patients," *The*

- Tohoku Journal of Experimental Medicine*, vol. 235, pp. 135–144, 2015.
- [7] M. K. de Bie, M. S. Buiten, J. I. Rotmans et al., “Abdominal aortic calcification on a plain X-ray and the relation with significant coronary artery disease in asymptomatic chronic dialysis patients,” *BMC Nephrology*, vol. 18, p. 82, 2017.
- [8] X. Zhang, X. Liu, L. Wang, M. Wu, W. Shan, B. Liu et al., “The relationship between abdominal aortic calcification and insulin resistance, micro-inflammatory state, and FGF23/klotho axis in end-stage renal failure patients undergoing dialysis,” *Shandong Medicine*, vol. 60, pp. 54–57, 2020.
- [9] X. Y. Guo, W. Zhou, B. Shi et al., “An efficient multiple kernel support vector regression model for assessing dry weight of hemodialysis patients,” *Current Bioinformatics*, vol. 16, pp. 284–293, 2021.
- [10] A. Munir, S. I. Malik, and K. A. Malik, “Proteome mining for the identification of putative drug targets for human pathogen *Clostridium tetani*,” *Current Bioinformatics*, vol. 14, pp. 532–540, 2019.
- [11] Y. Ding, J. Tang, and F. Guo, “Identification of drug-target interactions via dual Laplacian regularized least squares with multiple kernel fusion,” *Knowledge-Based Systems*, vol. 204, 2020.
- [12] J. Zhuang, S. Dai, L. Zhang et al., “Identifying breast cancer-induced gene perturbations and its application in guiding drug repurposing,” *Current Bioinformatics*, vol. 15, pp. 1075–1089, 2021.
- [13] Y. Ding, J. Tang, and F. Guo, “Protein crystallization identification via fuzzy model on linear neighborhood representation,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–1, 2019.
- [14] Y. Ding, J. Tang, and F. Guo, “Human protein subcellular localization identification via fuzzy model on Kernelized Neighborhood Representation,” *Applied Soft Computing*, vol. 96, 2020.
- [15] N. Ikram, M. A. Qadir, and M. T. Afzal, “SimExact - an efficient method to compute function similarity between proteins using gene ontology,” *Current Bioinformatics*, vol. 15, pp. 318–327, 2020.
- [16] M. Naveed, M. Z. Mehboob, A. Hussain, K. Ikram, A. Talat, and N. Zeeshan, “Structural and functional annotation of conserved virulent hypothetical proteins in *Chlamydia trachomatis*: an in-silico approach,” *Current Bioinformatics*, vol. 14, pp. 344–352, 2019.
- [17] H. Wang, J. Tang, Y. Ding, and F. Guo, “Exploring associations of non-coding RNAs in human diseases via three-matrix factorization with hypergraph-regular terms on center kernel alignment,” *Briefings in Bioinformatics*, 2021.
- [18] A. Khan, A. Zahra, S. Mumtaz, M. Q. Fatmi, and M. J. Khan, “Integrated in-silico analysis to study the role of microRNAs in the detection of chronic kidney diseases,” *Current Bioinformatics*, vol. 15, pp. 144–154, 2020.
- [19] S. Yamada, S. Yamamoto, S. Fukuma, T. Nakano, K. Tsuruya, and M. Inaba, “Geriatric nutritional risk index (GNRI) and creatinine index equally predict the risk of mortality in hemodialysis patients: J-DOPPS,” *Scientific Reports*, vol. 10, 2020.
- [20] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [21] Chun-Fu Lin and Sheng-de Wang, “Fuzzy support vector machines,” *IEEE Transactions on Neural Networks*, vol. 13, pp. 464–471, 2002.
- [22] H. Zhu, X. Du, and Y. Yao, “ConvsPPIS: identifying protein-protein interaction sites by an ensemble convolutional neural network with feature graph,” *Current Bioinformatics*, vol. 15, pp. 368–378, 2020.
- [23] J. Gao, L. Zhang, G. Yu, G. Qu, Y. Li, and X. Yang, “Model with the GBDT for colorectal adenoma risk diagnosis,” *Current Bioinformatics*, vol. 15, pp. 971–979, 2021.
- [24] M. Tahir, M. Sardaraz, Z. Mehmood, and M. S. Khan, “ESREEM: efficient short reads error estimation computational model for next-generation genome sequencing,” *Current Bioinformatics*, vol. 16, no. 2, pp. 339–349, 2021.
- [25] W. Liu, J. P. Haran, A. S. Ash et al., “High-dimensional causal mediation analysis with a large number of mediators clumping at zero to assess the contribution of the microbiome to the risk of bacterial pathogen colonization in older adults,” *Current Bioinformatics*, vol. 15, no. 7, pp. 671–696, 2020.
- [26] L. Chen, J. Li, and M. Chang, “Cancer diagnosis and disease gene identification via statistical machine learning,” *Current Bioinformatics*, vol. 15, pp. 956–962, 2021.
- [27] K. K. L. Wong, “Optimization in the design of natural structures, biomaterials, bioinformatics and biometric techniques for solving physiological needs and ultimate performance of bio-devices,” *Current Bioinformatics*, vol. 14, no. 5, pp. 374–375, 2019.
- [28] Z. Deng, Y. Jiang, K.-S. Choi, F.-L. Chung, and S. Wang, “Knowledge-leverage-based TSK fuzzy system modeling,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, pp. 1200–1212, 2013.
- [29] B. Rezaee and M. H. F. Zarandi, “Data-driven fuzzy modeling for Takagi-Sugeno-Kang fuzzy system,” *Information Sciences*, vol. 180, no. 2, pp. 241–255, 2010.
- [30] Y. Jiang, F. L. Chung, H. Ishibuchi, Z. Deng, and S. Wang, “Multitask TSK fuzzy system modeling by mining intertask common hidden structure,” *IEEE Transactions on Cybernetics*, vol. 45, pp. 548–561, 2015.
- [31] E. Honkanen, L. Kauppila, B. Wikstrom et al., “Abdominal aortic calcification in dialysis patients: results of the CORD study,” *Nephrology Dialysis Transplantation*, vol. 23, no. 12, pp. 4009–4015, 2008.
- [32] X. Lu and M. C. Hu, “Klotho/FGF23 axis in chronic kidney disease and cardiovascular disease,” *Kidney Diseases*, vol. 3, pp. 15–23, 2017.
- [33] H. Y. Chen, Y. L. Chiu, S. P. Hsu, M. F. Pai, J. Y. Yang, and Y. S. Peng, “Relationship between fetuin A, vascular calcification and fracture risk in dialysis patients,” *PLoS One*, vol. 11, article e0158789, 2016.
- [34] Y. Jiang, D. Wu, Z. Deng et al., “Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 2270–2284, 2017.
- [35] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C. T. Lin, “EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, 2021.
- [36] Y. Jiang, J. Zhu, X. Gu et al., “Recognition of large-scale ncRNA data using a novel multitask cross-learning 0-order TSK fuzzy classifier,” *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 2, pp. 502–507, 2020.
- [37] G. Zhang, P. Yu, J. Wang, and C. Yan, “Feature selection algorithm for high-dimensional biomedical data using information

gain and improved chemical reaction optimization,” *Current Bioinformatics*, vol. 15, no. 8, pp. 912–926, 2021.

- [38] F. Berrhail and H. Belhadef, “Genetic algorithm-based feature selection approach for enhancing the effectiveness of similarity searching in ligand-based virtual screening,” *Current Bioinformatics*, vol. 15, no. 5, pp. 431–444, 2020.