Research Article

Assessing Dry Weight of Hemodialysis Patients via Sparse Laplacian Regularized RVFL Neural Network with L$_{2,1}$-Norm

Xiaoyi Guo,[1] Wei Zhou [ID],[1] Qun Lu,[2] Aiyan Du,[1] Yinghua Cai [ID],[2] and Yijie Ding [ID][3]

[1]*Hemodialysis Center, The Affiliated Wuxi People's Hospital of Nanjing Medical University, 214000 Wuxi, China*
[2]*Nursing Department, The Affiliated Wuxi People's Hospital of Nanjing Medical University, 214000 Wuxi, China*
[3]*School of Electronic and Information Engineering, Suzhou University of Science and Technology, 215009 Suzhou, China*

Correspondence should be addressed to Wei Zhou; 285403434@qq.com, Yinghua Cai; 179098331@qq.com,
and Yijie Ding; wuxi_dyj@163.com

Dry weight is the normal weight of hemodialysis patients after hemodialysis. If the amount of water in diabetes is too much (during hemodialysis), the patient will experience hypotension and shock symptoms. Therefore, the correct assessment of the patient's dry weight is clinically important. These methods all rely on professional instruments and technicians, which are time-consuming and labor-intensive. To avoid this limitation, we hope to use machine learning methods on patients. This study collected demographic and anthropometric data of 476 hemodialysis patients, including age, gender, blood pressure (BP), body mass index (BMI), years of dialysis (YD), and heart rate (HR). We propose a Sparse Laplacian regularized Random Vector Functional Link (SLapRVFL) neural network model on the basis of predecessors. When we evaluate the prediction performance of the model, we fully compare SLapRVFL with the Body Composition Monitor (BCM) instrument and other models. The Root Mean Square Error (RMSE) of SLapRVFL is 1.3136, which is better than other methods. The SLapRVFL neural network model could be a viable alternative of dry weight assessment.

1. Introduction

Fluid overload in patients with chronic renal failure is closely related to poor cardiovascular outcomes [1, 2]. Maintenance of hemodialysis (HD) is the main method for patients with renal failure [3]. However, the accurate assessment of body water volume is still a concern [4]. At present, dry weight has been used as an important indicator to assess the homeostasis of fluids in hemodialysis patients. Medical staff can use the patient's dry weight to estimate the amount of water needed for dialysis during hemodialysis. The conventional clinical-based dry weight assessment method is time-consuming and labor-intensive [1]. There are already some methods based on bioelectrical impedance analysis (BIA) [5] to determine dry weight, including body composition monitor (BCM) [6] and lung ultrasound (LUS). However, all the above methods require special instruments and pro-

fessional technicians to complete. Medical staff can use some clinical data to build predictive models [7] to accurately assess dry weight. Currently, machine learning (ML) or deep learning has solved many common clinical problems in medicine, such as brain diseases [8–10], cancer analysis, and diabetes.

Some scholars have used artificial neural networks (ANN) to predict the total water volume of hemodialysis patients and have obtained better results than conventional clinical calculation equations [11]. In addition, deep learning methods are also emerging in clinical diagnosis, including pixel-based convolutional neural networks to diagnose skin cancer [12]. In the biological field, microbiology analysis [13], CircRNAs [14], microRNAs, and cancer association prediction [15–17], lncRNA-miRNA association prediction, O-GlcNAcylation site prediction [18], DNA methylation site [19–21], protein remote homology [22], function prediction

of proteins [23–29], electron transport proteins [30], breast cancer [31], cell-specific replication [32], osteoporosis diagnoses [33], and drug complex network analysis [34–38].

In our previous research, a Multiple Kernel Support Vector Regression (MKSVR) [39] predictor was proposed to assess the dry weight and obtain good predictive performance. Inspired by the previous work and baseline Random Vector Functional Link (RVFL) network [40], we propose a new dry weight assessment model, called Sparse Laplacian regularized RVFL neural network with $L_{2,1}$-norm (SLapRVFL), which considers the topological relationship between samples and more sparse connections between the input layer and the hidden layer.

2. Materials and Methods

2.1. Materials. This work collects demographic and anthropometric data and bioimpedance spectroscopy (BIS) from historical data (2018-9 to 2019-9) from Wuxi people's hospital and the northern Jiangsu people's hospital. This study has been approved by the ethics committees of the hospitals (Nos. KYLLKS201813 and 2018KY-001). The collected patient data meet the following requirements: age greater than 18 years; ESRD for more than three months and maintenance hemodialysis [41]; no heart failure, no metal implants, no pregnancy, no disability, no infection, and no edema and other diseases; and hemodialysis treatment 3 times a week, 4 hours each time. Finally, we obtain a data set of 476 hemodialysis patients. DW is the normal body weight after clinical diabetes. DW is obtained by a clinician under strict clinical supervision using a clinical scoring system (using trial and error method) [42, 43].

We choose 7 features, including age, gender (binary feature), systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI), heart rate (HR), and years of dialysis (YD) to build our predictive model. Table 1 shows the information of the data set. BMI is measured before hemodialysis treatment.

2.2. Methods. The baseline RVFL was proposed for regression or classification. The schematic diagram of RVFL is shown in Figure 1. The basic information of the patient is put into the RVFL neural network model for processing, and the predicted dry weight is the output.

Suppose, there are N training samples with $\{x_i, y_i\}$, $i = 1, 2, \cdots, N$. The output value is $y_i \in R^{1\times c}$ and the input data is $x_i \in R^{1\times d}$. d denotes the dimension of x_i. As per Figure 1, RVFL randomly initializes all weights and deviations between the hidden layer and the input layer. These parameters are fixed during the training process and do not need to be tuned. There are connections between the output layer, input layer, and hidden layer. This part of the weight needs to be obtained by training RVFL. The output layer of RVFL is connected to both the input layer and the hidden layer, so as to ensure the nonlinear and linear relationships between the input and the output. The RVFL network with P hidden nodes are formulated as

$$H\beta = Y, \tag{1}$$

Table 1: The information of data set.

Feature	Value	r^*
Age (years)	54.17 ± 14.22	-0.2341
Gender (males/females)	312/164	-0.4489
BMI	22.96 ± 2.95	0.9558
Systolic blood pressure (mmHg)	150.64 ± 29.36	-0.1739
Diastolic blood pressure (mmHg)	88.32 ± 19.56	-0.1249
Heart rate (times/min)	73.41 ± 8.92	0.1862
Years of dialysis (years)	5.97 ± 3.22	-0.1069

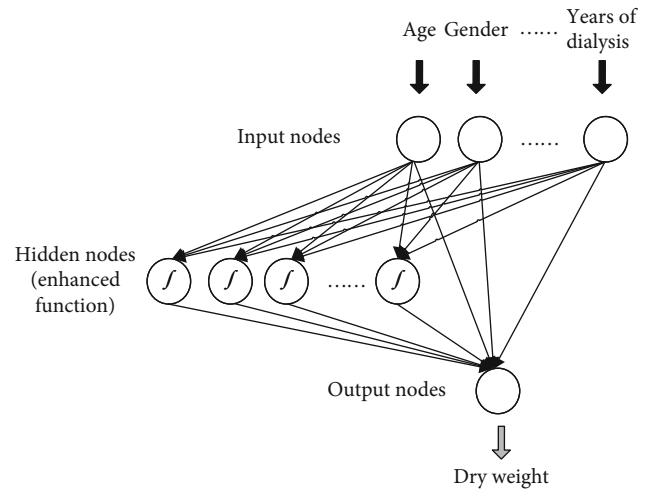*Denotes that each feature correlated with dry weight using Pearson correlation coefficient (r).

Figure 1: Schematic of our proposed method.

where β denotes the output weight matrix; H is the concatenated matrix, which combines the output of the hidden layer and the input layer; and Y denotes the label matrix. H and β can be represented as

$$H = [H_1\ H_2], \tag{2}$$

$$H_1 = \begin{bmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nd} \end{bmatrix}_{N\times d}, \tag{3}$$

$$H_2 = \begin{bmatrix} G(a_1 x_1 + b_1) & \cdots & G(a_P x_1 + b_P) \\ \vdots & \ddots & \vdots \\ G(a_1 x_N + b_1) & \cdots & G(a_P x_N + b_P) \end{bmatrix}_{N\times P}, \tag{4}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_{d+P}^T \end{bmatrix}_{(d+P)\times C}. \tag{5}$$

Require: Training set $\{x_i, y_i\}$, $i = 1, 2, \cdots, N$, test set $\{x_j^{te}\}$, $j = 1, 2, \cdots, M$, the numbers of hidden layer nodes (P), the maximum number of iterations tmax, coefficients of λ_1 and λ_2;

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

(1) Randomly initializing all weights and deviations between the hidden layer and the input layer. Calculating the hidden layer output matrix H (training set)and Laplacian matrix L by Equations (2), (12), and (13);

(2) Set $t = 0$, estimate the initial β^0 using Equation (7);

Repeat

(3) Update the diagonal matrix G with

$$G^{t+1} = \begin{bmatrix} 1/2\|\beta_1^t\|_2 & & \\ & \ddots & \\ & & 1/2\|\beta_{d+P}^t\|_2 \end{bmatrix}_{(d+P) \times (d+P)},$$

(4) Update β via Equation (11d);

Until $t > $ tmax;

(5) Calculate the hidden layer output matrix H^{te} (test set);

(6) Estimate $\{y_j^{te}\}$, $j = 1, 2, \cdots, M$ by $Y^{te} = H^{te}\beta$.

ALGORITHM 1. Algorithm of SLapRVFL

In Equation (4), a_j and b_j are the weights and bias of the hidden and input layers. C and P are numbers of output and hidden layer nodes. In general, the activation function is a Gaussian function: $g(x) = e^{-x^2}$. The activation function has a nonlinear approximation effect. To consider the potential linear relationship between the input data and the output value, RVFL adds a direct connection weight between the input layer and the output layer. Therefore, RVFL is a model that contains both linear and nonlinear approximations to improve prediction performance. For optimal β, the RVFL can be formulated as a regularized least-squares:

$$\beta^* = \arg\min \ \frac{1}{2}\|H\beta - Y\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2, \tag{6}$$

where λ is the parameter of regularization term. The solution of Equation (6) can be found by setting its gradient to 0:

$$\beta^* = \left(H^T H + \lambda I\right)^{-1} H^T Y, \tag{7}$$

where I denotes the identity matrix. However, the RVFL network did not consider the topological relationship between samples. For the output node, it must be connected to both the input and the hidden layer.

In order to further improve the robustness of RVFL, we propose Sparse Laplacian regularized RVFL neural network with $L_{2,1}$-norm (SLapRVFL). The objective function is

$$\beta^* = \arg\min \ \frac{1}{2}\|H\beta - Y\|_2^2 + \frac{\lambda_1}{2} Tr\left((H\beta)^T LH\beta\right) + \frac{\lambda_2}{2}\|\beta\|_{2,1}^2, \tag{8}$$

where $L \in R^{N \times N}$ denotes the Laplacian matrix. λ_1 and λ_2 are the coefficients of Laplacian regularization the and L_{21}-norm term, respectively. Laplacian regularization is used to indicate the potential manifold between samples. It can better describe the topological association between samples to improve the
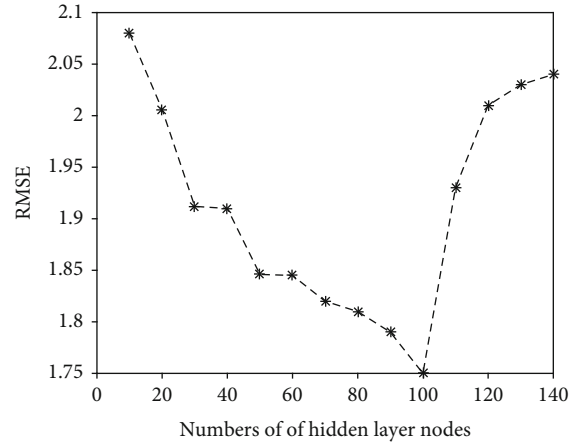

FIGURE 2: The RMSE under different numbers of hidden layer nodes (SLapRVFL network).

generalization ability of the model. Since the third term of $\|\beta\|_{2,1}^2$ is not diversified, we convert Equation (8) to

$$\beta^* = \arg\min \ \frac{1}{2}\|H\beta - Y\|_2^2 + \frac{\lambda_1}{2} Tr\left((H\beta)^T LH\beta\right) + \frac{\lambda_2}{2} Tr\left(\beta^T G\beta\right), \tag{9}$$

where $G \in R^{(d+P) \times (d+P)}$ denotes a diagonal matrix whose ith-diagonal element

$$G_{ii} = \frac{1}{2\|\beta_i\|_2}, \quad i = 1, 2, \cdots, (d+P). \tag{10}$$

We take the derivative of the formula Equation (10) as

$$H^T(H\beta - Y) + \lambda_1 H^T LH\beta + \lambda_2 G\beta = 0, \tag{11a}$$

$$H^T H\beta + \lambda_1 H^T LH\beta + \lambda_2 G\beta = H^T Y, \tag{11b}$$
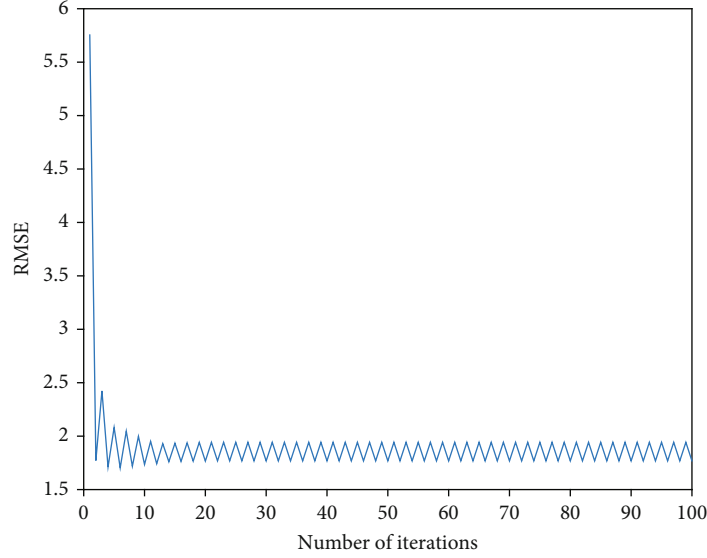
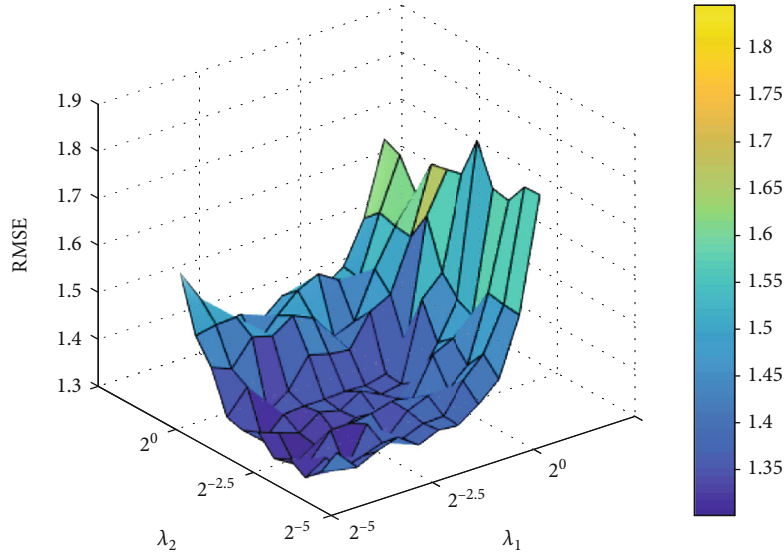FIGURE 3: The RMSE of iterations on the training set.

FIGURE 4: The RMSE under different λ_1 and λ_2.

$$\left(H^T H + \lambda_1 H^T L H + \lambda_2 G\right)\beta = H^T Y, \qquad (11c)$$

$$\beta = \left(H^T H + \lambda_1 H^T L H + \lambda_2 G\right)^{-1} H^T Y. \qquad (11d)$$

We use the baseline RVFL solution with Equation (7) as the initial β^0. In addition, the Laplacian matrix can be calculate as

$$L = D^{-1/2} \Delta D^{-1/2}, \qquad (12a)$$

$$\Delta = D - S, \qquad (12b)$$

where D is diagonal matrix, $D_{ii} = \sum_{j=1}^N S_{ij}$. Similarity matrix S is built by Radial Basis Function (RBF):

$$S_{ij} = \exp\left(-\gamma \|x_i - x_j\|^2\right). \qquad (13)$$

The process of SLapRVFL is list in Algorithm 1.

3. Results

We test our model on the benchmark data set and obtain the optimal parameters of the predictor through cross-validation. The SLapRVFL network is compared to other machine learning-based models. In addition, the body composition monitor (BCM) device (Fresenius Medical Care, Baden Humboldt, Germany) is also compared with the SLapRVFL network.

3.1. Evaluation Measurements. The 10-fold cross-validation (10-CV) is employed to evaluate the robustness of methods. Root Mean Square Error (RMSE), R square, correlation coefficient (R), Bland–Altman analysis, and Empirical Cumulative Distribution Plot (ECDP) [44] are all used in our study. To evaluate the agreement of two different methods, the Bland–Altman analysis usually can obtain whether the two

TABLE 2: Comparison on existing methods via 10-fold cross-validation.

Method	R	R squared	RMSE	Empirical cumulative distribution plot		
				Highest value	Lowest value	Median value
BCM*	0.9473	0.9137	1.9694	3.2235	-6.2776	-0.9863
LR*	0.9403	0.9308	1.4335	4.2524	-4.4014	0.1418
ANN (BP)*	0.9398	0.9295	1.4794	7.3661	-4.7447	0.1324
MKRR*	0.9399	0.9289	1.5015	4.9227	-4.2604	0.1104
MKSVR*	0.9412	0.9321	1.3817	4.3962	-4.1273	0.0082
RVFL	0.9389	0.9300	1.3828	6.7004	-4.3557	0.0704
SLapRVFL (our method)	0.9632	0.9501	1.3136	3.1940	-3.5066	0.1014

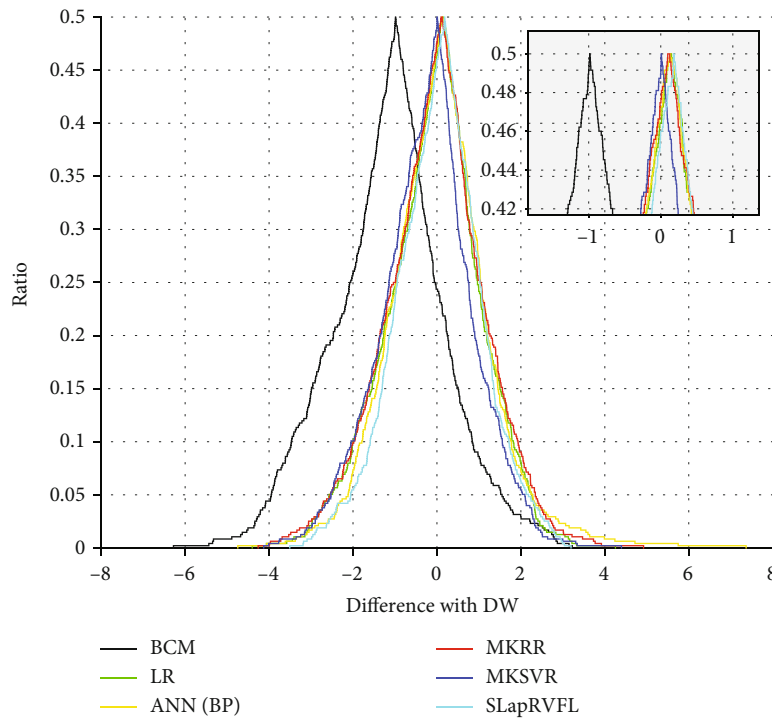*The results are from previous work on MKSVR [39].

FIGURE 5: Folded empirical cumulative distribution plot between different methods.

TABLE 3: Bland–Altman plot analysis for different models.

Model	Differences with DW (%)			Limits of agreement (%)		
	Mean	SD	95% confidence interval	Lower limit	Upper limit	Number (ratio) of outside agreement interval
BCM*	-1.8232	2.7466	-2.0706 to -1.5759	-7.2066	3.5601	30/476 (6.30%)
LR*	0.0002	2.4269	-0.2184 to 0.2187	-4.7566	4.7569	21/476 (4.41%)
ANN (BP)*	0.1152	2.5139	-0.1112 to 0.3416	-4.8119	5.0424	22/476 (4.62%)
MKRR*	-0.0801	2.5007	-0.3053 to 0.1451	-4.9814	4.8212	23/476 (4.83%)
MKSVR*	-0.2638	2.3372	-0.4743 to -0.05329	-4.8446	4.3171	22/476 (4.62%)
SLapRVFL (our method)	0.0867	2.2202	-0.1133 to 0.2866	-4.2650	4.4383	20/476 (4.20%)

*The results are from previous work on MKSVR [39].

(a) ANN

(b) LR

(c) MKRR

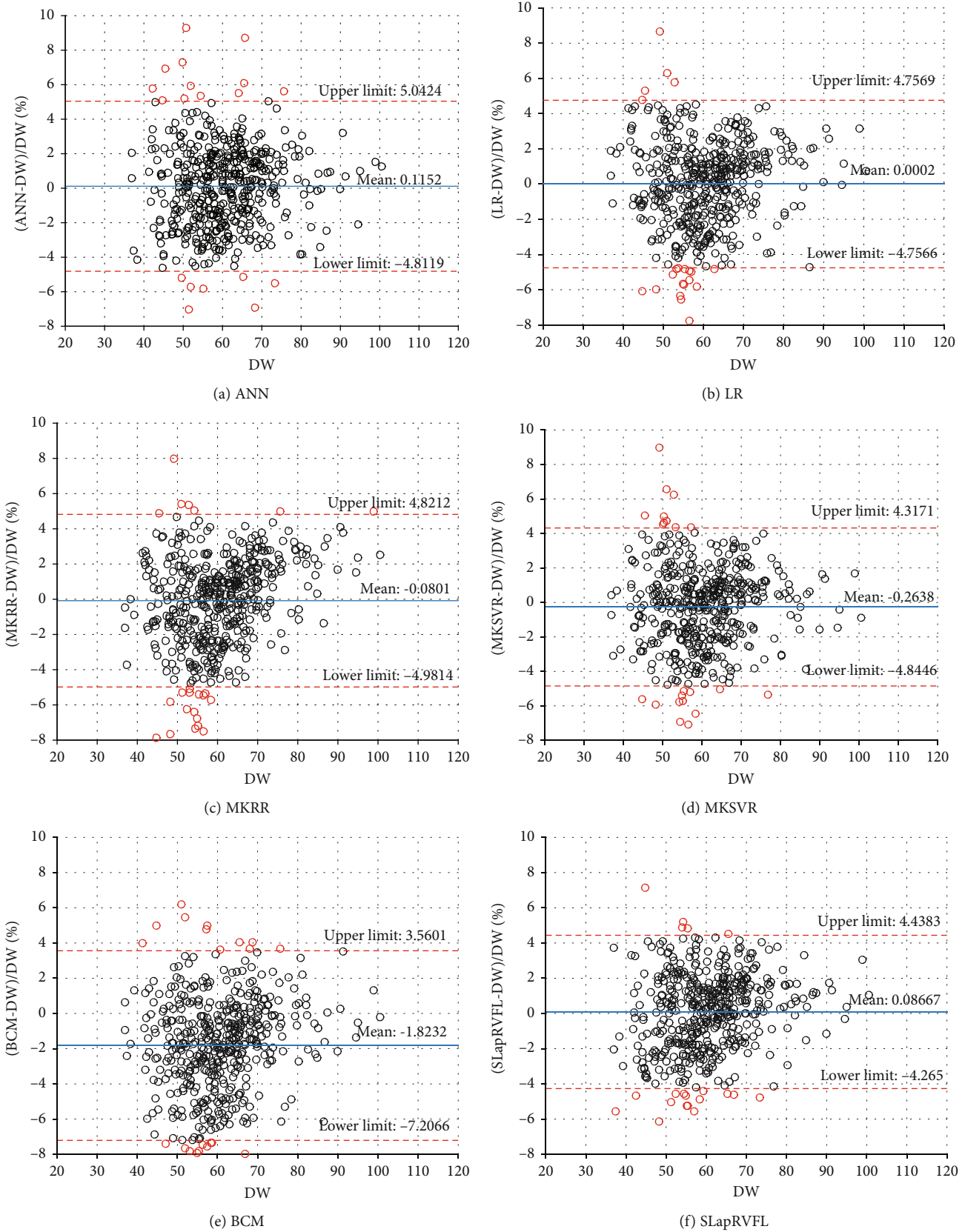(d) MKSVR

(e) BCM

(f) SLapRVFL

Figure 6: Bland–Altman plot analysis.

methods can be substituted for each other (equivalence). Evaluating the agreement of the two methods can answer the question, "Can these two methods replace each other?"

3.2. Selection of Optimal Parameters. To get the optimal parameters of the predictive method, we obtain them through a grid search method. The parameters that need to be determined include the numbers of hidden layer nodes P, maximum iterations, and coefficients of λ_1 and λ_2. For the numbers of hidden layer nodes P, we fix the iterations, λ_1 and λ_2. Setting the maximum number as 50, $\lambda_1 = 1$ and $\lambda_2 = 1$. The value of P is from 10 to 140 with step of 10. The results are shown in Figure 2. From 10 to 100, the more neurons in the hidden layer, the lower the RMSE. Since then, RMSE has gradually increased. So, we get the lower RMSE under $P = 100$.

Next, $P = 100$, $\lambda_1 = 1$, and $\lambda_2 = 1$. We gradually increase the number of iterations from 1 to 100 (shown in Figure 3). After the number of iterations reaches 10, the RMSE value drops to a minimum and slightly oscillates within a certain value. In our study, maximum number of iterations is 10.

Then, we use the better number of hidden layer nodes and iterations to search for the best λ_1 and λ_2. The search range of parameters is from 2^{-5} to 2^0 (with step of $2^{0.5}$). Figure 4 shows the results of different parameters. When λ_1 and λ_2 are 2^{-3} and $2^{-2.5}$, RMSE is the lowest.

3.3. Comparison to Other Predictive Models and BCM. To evaluate our model, SLapRVFL is compared with our previous work of Multiple Kernel Support Vector Regression (MKSVR) [39], Multikernel Ridge Regression (MKRR), Linear Regression (LR), Artificial Neural Network based on Back Propagation algorithm (ANN with BP), and BCM measuring instrument. Clinical dry weight is our reference standard (also the regression target value of the prediction model). The comparisons are listed in Table 2, which shows that SLapRVFL achieves best performance of RMSE (1.3136). Although the ECDP median value (peak) of MKSVR (0.0082) is more close to zero, Figure 5 shows that SLapRVFL has the least bias and much less tails than MKSVR (smaller width). The RMSE of BCM is 1.9694, which is larger than SLapRVFL.

3.4. Bland–Altman Analysis. Bland–Altman plot is a useful tool to evaluate the agreement between predictive methods and clinical DW. In Table 3 and Figure 6, SLapRVFL, MKSVR, LR, ANN (BP), MKRR, and BCM are analyzed via Bland-Altman difference plot. SLapRVFL achieves the smallest range of 95% confidence interval (-0.1133 to 0.2866) and standard deviation (2.2202). In addition, the number (ratio) of outside agreement interval for predictive models is all less than 24 (5%) predictive samples. These results of models are clinically acceptable. SLapRVFL achieves least number (20) of the outside agreement interval in Table 3. As shown in Figure 6, two red horizontal dotted lines (upper and lower) denote the upper and lower limits of the 95% agreement limit, respectively. The middle blue solid line is the average value of the difference (between measurement methods and clinical DW). While one measurement method and clinical

method can be considered as a better agreement, they can be substituted for each other (equivalence). If 95% of the points of the data set are in the agreement range, the measurement method (predictive model) is clinically acceptable. The results of the evaluation show that SLapRVFL can help clinicians assess DW with low cost.

4. Discussion

Due to the limitations of clinical and BCM measurement (more time and cost), this study uses a machine learning method to assess the dry weight of hemodialysis patients. Based on the basic RVFL, we propose a sparse Laplace regularized RVFL network (SLapRVFL) model. SLapRVFL is compared not only with other machine learning methods (such as LR, MKRR, ANN with BP, and MKSVR) but also with BCM equipment (commonly used in hospitals). The RMSE and Bland–Altman analysis of the model are better than the BCM instrument. It is proven that the predictive model driven by data can provide reference for clinical dry weight assessment.

BCM requires the patient's information on weight (before hemodialysis) and height. It is a portable, inexpensive, and noninvasive technology that has been used to measure DW [45, 46]. For the Bland–Altman analysis, SLapRVFL achieves the least number (20) of outside agreement interval. However, BCM has 30/476 (6.30%) points (ratio) of the outside agreement interval. Obviously, our method has better agreement with the clinical method.

5. Conclusions

To further improve the robustness of RVFL, we introduce sparse Laplacian regular term with $L_{2,1}$-norm. In the training process, the graph topology information and the sparse weight matrix (output) are employed to improve the robustness of the RVFL. In fact, our work provides a new idea for assessing patients' dry weight. Not only that, in the fields of biology [47–57], pharmacy [58], and medicine [12, 59, 60], machine learning methods have helped solve many analysis tasks. In future research, we will consider collecting more samples, introducing more patient personal information, and building a predictor based on a deep learning model to more accurately assess the dry weight of hemodialysis patients.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical Approval

This study had been approved by the ethics committee of the hospital (ethical approval Nos. KYLLKS201813 and 2018KY-001). 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 and Northern Jiangsu 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

Xiaoyi Guo and Wei Zhou are joint first authors.

Acknowledgments

References

[1] A. Grassmann, I. Uhlenbusch-Körwer, E. Bonnie-Schorn, and J. Vienken, "Composition and management of hemodialysis fluids," *Good Dialysis Practice*, vol. 2, pp. 13–25, 2000.

[2] P. Wabel, P. Chamney, U. Moissl, and T. Jirka, "Importance of whole-body bioimpedance spectroscopy for the management of fluid balance," *Blood Purification*, vol. 27, no. 1, pp. 75–80, 2009.

[3] G. Alexiadis, S. Panagoutsos, S. Roumeliotis et al., "Comparison of multiple fluid status assessment methods in patients on chronic hemodialysis," *International Urology and Nephrology*, vol. 49, no. 3, pp. 1–8, 2016.

[4] Y. Ohashi, K. Sakai, H. Hase, and N. Joki, "Dry weight targeting: the art and science of conventional hemodialysis," *Seminars in Dialysis*, vol. 31, 2018.

[5] H. Asmat, R. Iqbal, F. Sharif, A. Mahmood, A. Abbas, and W. Kashif, "Validation of bioelectrical impedance analysis for assessing dry weight of dialysis patients in Pakistan," *Saudi Journal of Kidney Diseases & Transplantation*, vol. 28, no. 2, p. 285, 2017.

[6] C. Jiang, S. Patel, A. Moses, M. V. DeVita, and M. F. Michelis, "Use of lung ultrasonography to determine the accuracy of clinically estimated dry weight in chronic hemodialysis patients," *International Urology and Nephrology*, vol. 49, no. 12, pp. 2223–2230, 2017.

[7] P. Susantitaphong, S. Laowaloet, K. Tiranathanagul et al., "Reliability of blood pressure parameters for dry weight estimation in hemodialysis patients," *Therapeutic Apheresis and Dialysis*, vol. 17, no. 1, pp. 9–15, 2013.

[8] G. Liu, Y. Hu, Z. Han, S. Jin, and Q. Jiang, "Genetic variant rs17185536 regulates SIM1 gene expression in human brain hypothalamus," *Proceedings of the National Academy of Sciences*, vol. 116, no. 9, pp. 3347-3348, 2019.

[9] G. Liu, S. Jin, Y. Hu, and Q. Jiang, "Disease status affects the association between rs4813620 and the expression of Alzheimer's disease susceptibility gene TRIB3," *Proceedings of the National Academy of Sciences*, vol. 115, no. 45, pp. E10519–E10520, 2018.

[10] X. A. Bi, Y. Liu, Y. Xie, X. Hu, and Q. Jiang, "Morbigenous brain region and gene detection with a genetically evolved random neural network cluster approach in late mild cognitive impairment," *Bioinformatics*, vol. 36, no. 8, pp. 2561–2568, 2020.

[11] J. S. Chiu, C. F. Chong, Y. F. Lin, C. C. Wu, Y. F. Wang, and Y. C. Li, "Applying an artificial neural network to predict total body water in hemodialysis patients," *American Journal of Nephrology*, vol. 25, no. 5, pp. 507–513, 2005.

[12] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[13] K. Qu, F. Guo, X. Liu, Y. Lin, and Q. Zou, "Application of machine learning in microbiology," *Frontiers in Microbiology*, vol. 10, p. 827, 2019.

[14] Q. Zhao, Y. Yang, G. Ren, E. Ge, and C. Fan, "Integrating bipartite network projection and KATZ measure to identify novel CircRNA-disease associations," *IEEE Transactions on Nanobioscience*, vol. 18, no. 4, pp. 578–584, 2019.

[15] L. Jiang, Y. Xiao, Y. Ding, J. Tang, and F. Guo, "FKL-Spa-LapRLS: an accurate method for identifying human microRNA-disease association," *BMC Genomics*, vol. 19, no. S10, p. 911, 2018.

[16] X. Zeng, L. Liu, L. Lü, and Q. Zou, "Prediction of potential disease-associated microRNAs using structural perturbation method," *Bioinformatics*, vol. 34, no. 14, pp. 2425–2432, 2018.

[17] Y. Ding, L. Jiang, J. Tang, and F. Guo, "Identification of human microRNA-disease association via hypergraph embedded bipartite local model," *Computational Biology and Chemistry*, vol. 89, p. 107369, 2020.

[18] C. Jia, Y. Zuo, and Q. Zou, "O-GlcNAcPRED-II: an integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique," *Bioinformatics*, vol. 34, no. 12, pp. 2029–2036, 2018.

[19] L. Wei, S. Luan, L. A. Nagai, R. Su, and Q. Zou, "Exploring sequence-based features for the improved prediction of DNA N4-methylcytosine sites in multiple species," *Bioinformatics*, vol. 35, no. 8, pp. 1326–1333, 2018.

[20] Q. Zou, P. Xing, L. Wei, and B. Liu, "Gene2vec: gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA," *RNA*, vol. 25, no. 2, pp. 205–218, 2019.

[21] C. Dai, P. Feng, L. Cui, R. Su, W. Chen, and L. Wei, "Iterative feature representation algorithm to improve the predictive performance of N7-methylguanosine sites," *Briefings in Bioinformatics*, 2020.

[22] B. Liu, S. Jiang, and Q. Zou, "HITS-PR-HHblits: protein remote homology detection by combining PageRank and hyperlink-induced topic search," *Briefings in Bioinformatics*, vol. 21, no. 1, pp. 298–308, 2018.

[23] L. Wei, Y. Ding, R. Su, J. Tang, and Q. Zou, "Prediction of human protein subcellular localization using deep learning," *Journal of Parallel and Distributed Computing*, vol. 117, pp. 212–217, 2018.

[24] 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, p. 1, 2019.

[25] Y. Wang, Y. Ding, J. Tang, Y. Dai, and F. Guo, "CrystalM: a multi-view fusion approach for protein crystallization prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, p. 1, 2019.

[26] H. Wang, Y. Ding, J. Tang, and F. Guo, "Identification of membrane protein types via multivariate information fusion with Hilbert–Schmidt Independence criterion," *Neurocomputing*, vol. 383, pp. 257–269, 2019.

[27] Y. Shen, Y. Ding, J. Tang, Q. Zou, and F. Guo, "Critical evaluation of web-based prediction tools for human protein subcellular localization," *Briefings in Bioinformatics*, vol. 21, no. 5, pp. 1628–1640, 2019.

[28] Y. Ding, J. Tang, and F. Guo, "Human protein subcellular localization identification via fuzzy model on kernelized neighborhood representation," *Applied Soft Computing*, vol. 96, p. 106596, 2020.

[29] R. Su, L. He, T. Liu, X. Liu, and L. Wei, "Protein subcellular localization based on deep image features and criterion learning strategy," *Briefings in Bioinformatics*, 2020.

[30] X. Ru, L. Li, and Q. Zou, "Incorporating distance-based top-n-gram and random forest to identify electron transport proteins," *Journal of Proteome Research*, vol. 18, no. 7, pp. 2931–2939, 2019.

[31] J. Liu, R. Su, J. Zhang, and L. Wei, "Classification and gene selection of triple-negative breast cancer subtype embedding gene connectivity matrix in deep neural network," *Briefings in Bioinformatics*, 2021.

[32] L. Wei, W. He, A. Malik, R. Su, L. Cui, and B. Manavalan, "Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework," *Briefings in Bioinformatics*, 2020.

[33] R. Su, T. Liu, C. Sun, Q. Jin, R. Jennane, and L. Wei, "Fusing convolutional neural network features with hand-crafted features for osteoporosis diagnoses," *Neurocomputing*, vol. 385, pp. 300–309, 2020.

[34] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via multiple information integration with centered kernel alignment," *Neurocomputing*, vol. 325, pp. 211–224, 2019.

[35] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via semisupervised model and multiple kernel learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2619–2632, 2019.

[36] 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, p. 106254, 2020.

[37] Y. J. Ding, T. Jijun, and F. Guo, "Identification of drug-target interactions via fuzzy bipartite local model," *Neural Computing and Applications*, vol. 32, no. 14, pp. 10303–10319, 2020.

[38] X. Guo, W. Zhou, Y. Yu, Y. Ding, J. Tang, and F. Guo, "A novel triple matrix factorization method for detecting drug-side effect association based on kernel target alignment," *BioMed Research International*, vol. 2020, 11 pages, 2020.

[39] X. 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. 15, 2020.

[40] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163–180, 1994.

[41] J. Passauer, H. Petrov, A. Schleser, J. Leicht, and K. Pucalka, "Evaluation of clinical dry weight assessment in haemodialysis patients using bioimpedance spectroscopy: a cross-sectional study," *Nephrology Dialysis Transplantation*, vol. 25, no. 2, pp. 545–551, 2009.

[42] M. Kraemer, C. Rode, and V. Wizemann, "Detection limit of methods to assess fluid status changes in dialysis patients," *Kidney International*, vol. 69, no. 9, pp. 1609–1620, 2006.

[43] Y. Jian, X. Li, X. Cheng et al., "Comparison of bioimpedance and clinical methods for dry weight prediction in maintenance hemodialysis patients," *Blood Purification*, vol. 37, no. 3, pp. 214–220, 2014.

[44] J. S. Krouwer and K. L. Monti, "A simple, graphical method to evaluate laboratory assays," *European Journal of Clinical Chemistry and Clinical Biochemistry*, vol. 33, no. 6, pp. 525–527, 1995.

[45] K. Cha, G. M. Chertow, J. Gonzalez, J. M. Lazarus, and D. W. Wilmore, "Multifrequency bioelectrical impedance estimates the distribution of body water," *Journal of Applied Physiology*, vol. 79, no. 4, pp. 1316–1319, 1995.

[46] L. T. Ho, R. F. Kushner, D. A. Schoeller, R. Gudivaka, and D. M. Spiegel, "Bioimpedance analysis of total body water in hemodialysis patients," *Kidney international*, vol. 46, no. 5, pp. 1438–1442, 1994.

[47] Y. Wang, F. Shi, L. Cao et al., "Morphological segmentation analysis and texture-based support vector machines classification on mice liver fibrosis microscopic images," *Current Bioinformatics*, vol. 14, no. 4, pp. 282–294, 2019.

[48] M. N. F. Fajila, "Gene subset selection for leukemia classification using microarray data," *Current Bioinformatics*, vol. 14, no. 4, pp. 353–358, 2019.

[49] L. Wei, M. Liao, Y. Gao, R. Ji, Z. He, and Q. Zou, "Improved and promising identification of human microRNAs by incorporating a high-quality negative set," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 11, no. 1, pp. 192–201, 2014.

[50] L. Wei, R. Su, B. Wang, X. Li, Q. Zou, and X. Gao, "Integration of deep feature representations and handcrafted features to improve the prediction of N^6-methyladenosine sites," *Neurocomputing*, vol. 324, pp. 3–9, 2019.

[51] L. Wei, S. Wan, J. Guo, and K. K. L. Wong, "A novel hierarchical selective ensemble classifier with bioinformatics application," *Artificial Intelligence in Medicine*, vol. 83, pp. 82–90, 2017.

[52] L. Wei, P. Xing, J. Zeng, J. X. Chen, R. Su, and F. Guo, "Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble classifier," *Artificial Intelligence in Medicine*, vol. 83, pp. 67–74, 2017.

[53] L. Wei, C. Zhou, H. Chen, J. Song, and R. Su, "ACPred-FL: a sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides," *Bioinformatics*, vol. 34, no. 23, pp. 4007–4016, 2018.

[54] C. Yang, Y. Ding, Q. Meng, J. Tang, and F. Guo, "Granular multiple kernel learning for identifying RNA-binding protein residues via integrating sequence and structure information," *Neural Computing and Applications*, pp. 1–13, 2021.

[55] 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.

[56] H. Wang, Y. Ding, J. Tang, Q. Zou, and F. Guo, "Identify RNA-associated subcellular localizations based on multi-label

learning using Chou's 5-steps rule," *BMC Genomics*, vol. 22, no. 1, p. 56, 2021.

[57] Y. Zou, H. Wu, X. Guo et al., "MK-FSVM-SVDD: a multiple kernel-based fuzzy SVM model for predicting DNA-binding proteins via support vector data description," *Current Bioinformatics*, vol. 15, p. 1, 2020.

[58] J. Wang, H. Wang, X. Wang, and H. Chang, "Predicting drug-target interactions via FM-DNN learning," *Current Bioinformatics*, vol. 15, no. 1, pp. 68–76, 2020.

[59] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 1–9, 2017.

[60] Y. Huang, K. Yuan, M. Tang et al., "Melatonin inhibiting the survival of human gastric cancer cells under ER stress involving autophagy and Ras-Raf-MAPK signalling," *Journal of Cellular and Molecular Medicine*, pp. 1–13, 2020.