

Predicting the excretion of feces, urine and nitrogen using support vector regression: A case study with Holstein dry cows

Qiang Fu¹, Weizheng Shen^{1,2}, Xiaoli Wei^{1,2}, Yanling Yin^{1,2}, Ping Zheng¹,
Yonggen Zhang³, Zhongbin Su¹, Chunjiang Zhao^{1,4*}

(1. College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China;

2. Key Laboratory of Pig-breeding Facilities Engineering, Ministry of Agriculture, Harbin 150030, China;

3. College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China;

4. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China)

Abstract: Predicting the excretion of feces, urine and nitrogen (N) from dairy cows is an effective way to prevent and control the environmental pollution caused by scaled farming. The traditional prediction methods such as pollutant generation coefficient (PGC) and mathematical model based on linear regression (LR) may be limited by prediction range and regression function assumption, and sometimes may deviate from the actual condition. In order to solve these problems, the support vector regression (SVR) was applied for predicting the cows' feces, urine and N excretions, taking Holstein dry cows as a case study. SVR is a typical non-parametric machine learning model that does not require any specific assumptions about the regression function in advance and only by learning the training sample data, and also it can fit the function closest to the actual in most cases. To evaluate prediction accuracy effectively, the SVR technique was compared with the LR and radial basis function artificial neural network (RBF-ANN) methods, using the required sample data obtained from actual feeding experiments. The prediction results indicate that the proposed technique is superior to the other two conventional (especially LR) methods in predicting the main indicators of feces, urine, and N excretions of Holstein dry cows.

Keywords: cow farming pollution, feces/urine excretion prediction, nitrogen excretion prediction, non-parametric model, SVR technique

DOI: 10.25165/j.ijabe.20201302.4781

Citation: Fu Q, Shen W Z, Wei X L, Yin Y L, Zheng P, Zhang Y G, et al. Predicting the excretion of feces, urine and nitrogen using support vector regression: A case study with Holstein dry cows. *Int J Agric & Biol Eng*, 2020; 13(2): 48–56.

1 Introduction

Nowadays, agricultural environmental pollution and management have become an important problem in the world^[1]. Among them, the environmental pollution caused by the discharge of pollutants from livestock husbandry has become increasingly severe, so it is necessary to effectively evaluate and scientifically treat the discharge of pollutants from livestock and poultry^[2,3]. In particular, with the continuous increase in the scale and intensification of dairy cows breeding, more and more excreta such as feces and urine have been produced, and a large amount of fecal nitrogen (FN) and urinary nitrogen (UN) have been discharged into the environment^[4,5]. If these cannot be managed in time, the excreta will pollute the soil, air and water sources^[6-9]. In order to effectively prevent and control the environmental pollution caused by feces, urine and nitrogen (N) excrete from dairy cows, and to

further realize the harmless treatment and resource utilization of pollutants, it is very important to predict the excretion of feces, urine and N of dairy cows accurately^[10].

At present, there are two main methods to predict the pollutant discharge of livestock and poultry, namely, the pollutant generation coefficient (PGC) and the mathematical modeling^[11]. Among them, the PGC method is used to estimate the average content of main pollutants in livestock and poultry excreta, and it is mainly divided into the two categories: the country-wide or the provincial-city level. Gan and Hu^[12] estimated the annual pollutant productions of eight livestock and poultry species at the country-wide level in 2005 and 2013 using the PGC. Zhou et al.^[13] studied the optimization of the PGC aiming at the scale composition of livestock and poultry production at provincial and national levels from 2002 to 2010. Fu et al.^[14] estimated the annual discharge of livestock and poultry in Henan Province from 2000 to 2014 through an optimized PGC. In summary, the PGC method is generally applicable to statistics for large-scale areas. Therefore, its estimated results are relatively rough and are usually used as a policy guidance at the macro level, while it was not applied to accurately solve practical prediction applications at the micro level^[11]. Compared with the PGC method, the mathematical modeling is a more accurate quantitative prediction method by analyzing and modeling the data of animal breeding process, and is mainly applicable to the prediction of the production of pollution in a small area such as breeding farm or animal individual^[11]. Generally, the mathematical modeling used for livestock and poultry pollution prediction is mainly based on linear regression (LR) method^[15]. More specifically, taking dairy cows' pollution prediction as a case, Wilkerson VA et al.^[16]

Received date: 2018-11-27 **Accepted date:** 2019-12-10

Biographies: Qiang Fu, PhD, Lecturer, research interests: intelligent animal husbandry, Email: fuqiangphd@126.com; Weizheng Shen, PhD, Professor, research interests: animal husbandry informatization, Email: alienmars@163.com; Xiaoli Wei, PhD, Associate professor, research interests: dairy industry informatization, Email: 34749669@qq.com; Yanling Yin, PhD, Lecturer, research interests: intelligent animal husbandry, Email: 569278707@qq.com; Ping Zheng, PhD, Associate professor, research interests: intelligent agriculture, Email: zhengping@neau.edu.cn; Yonggen Zhang, PhD, Professor, research interests: dairy nutrition, Email: zhangyonggen@sina.com; Zhongbin Su, PhD, Professor, research interests: intelligent agriculture, Email: suzb001@163.com.

*Corresponding author: Chunjiang Zhao, PhD, Professor, CAE Academician, research interests: agricultural information technology. Beijing Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China. Tel: +86-13936253160, Email: zhaocjnercita@sina.com.

predicted the average feces and nitrogen excretion of Holstein dairy herds relatively early. Nennich et al.^[17,18] predicted the excretion of feces, urine, and N from heifers, dry cows and lactating cows, respectively. Yan et al.^[19] predicted the excretion of feces, urine, and N from Holstein-Friesian and Norwegian lactating cows. Knowlton KF et al.^[20] predicted the excretion of feces, urine, and N from Jersey and Holstein cows. Higgs et al.^[21] predicted the excretion of FN and UN for lactating cows. Jiao et al.^[22] predicted the FN and organic matter from Holstein steers and heifers. Basically, the aforementioned prediction models for the excretion of feces, urine and N of dairy cows mainly adopt the LR-based model. It is well known that the LR-based model is a typical parametric model and usually assumes the form of the objective function followed by the prediction data, and then estimates the parameters of the objective function during the training process to determine the previously proposed hypothesis model^[23,24]. However, since the animal body including dairy cows is itself a complex system, it is difficult to assume an appropriate form of the objective function in advance and thus lead to the unsatisfactory prediction results caused by improper function form^[25-27].

In recent years, the research and application of non-parametric models have been a hot topic in the field of machine learning research. Specifically, compared with the classical parametric models including LR model, non-parametric models usually do not make any specific assumptions about the objective function when modeling, and can fit the function closest to the actual by learning the training sample data. Up to now, machine learning algorithms based on non-parametric models have been widely used in solving regression problems of prediction applications. Among them, support vector regression (SVR), is a machine learning algorithm of typical non-parametric model. In fact, SVR is support vector machine (SVM) used to solve the regression problems, which has been widely used in many fields such as electric power^[28], transportation^[29], engineering^[30], and securities^[31]. In most of the above cases, SVR generalization performance either matches or is

significantly better than competing methods. However, to the best of our knowledge, its application in animal husbandry environment was seldom reported, especially in the prediction of dairy cows' feces, urine and N excretions. In addition, for the practical predicting the excretion of feces, urine and N of dairy cows, it is usually difficult to obtain a large number of measured samples due to the various constraints such as animal numbers, feeding environment, manpower and financial conditions^[11]. But fortunately, compared with other non-parametric prediction models, SVR-based prediction methods usually have the advantages of high accuracy and fewer samples required for modeling^[32], which is very accord with the actual needs of dairy cows' feces, urine and N excretions prediction.

In this study, a novel SVR-based prediction model was proposed for predicting dairy cows' feces, urine and N excretions, taking Holstein dry cows as a case. The experimental results demonstrate that the proposed model can effectively predict Holstein dry cows' feces, urine and N excretion indicators, and shows better prediction accuracy in comparison with the other two conventional (especially LR) methods.

2 Materials and methods

2.1 Experimental design

In order to accurately predict the feces, urine and N excretions of dairy cows, in this study, the sample data needed for the prediction algorithm are obtained through actual feeding experiments. Twelve healthy Holstein cows in the dry period with similar body weight were used in the experiment. They were randomly and equally divided into three groups, and they were fed by twelve total mixed rations (TMRs) in groups, these diets composition and nutritional level were shown in Table 1. The experimental cows were fed twice daily at 6:00 and 18:00 and ensuring that there was 5% remaining diets. The water was supplied for 24 h and they can drink freely. Each dietary feeding pre-trial period is ten days, formal trial period is five days, and the experimental data were collected during the formal trial period.

Table 1 Composition and nutrient levels of experimental diets (DM basis %)

Items	Experimental diets											
	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12
Ingredients												
Chinese wildrye	63.74	57.51	49.43	42.02	33.19	25.18	17.29	10.25	9.39	8.73	7.64	6.28
Alfalfa	6.21	6.26	6.73	7.22	8.98	9.36	10.24	10.79	9.33	8.62	7.37	6.16
Corn silage	13.52	14.16	15.52	16.64	17.82	19.36	20.46	20.47	17.18	12.29	8.38	5.21
Corn straw	8.58	9.76	10.87	11.62	12.24	13.25	14.12	15.51	16.14	17.45	18.65	19.49
Corn	2.25	4.69	6.72	9.21	12.15	14.25	16.15	18.55	20.13	22.12	24.08	25.83
Soybean meal	1.31	1.85	2.06	2.82	3.37	4.43	5.70	6.66	7.38	8.16	9.58	10.71
Wheat bran	1.29	1.33	1.80	2.34	2.59	3.60	4.32	4.52	5.05	6.22	7.16	8.36
DDGS	1.12	1.24	1.71	2.13	2.53	3.13	3.52	4.13	5.16	5.23	5.31	5.49
Cottonseed meal	0.00	1.18	1.78	2.06	2.32	2.47	2.70	3.15	3.49	3.89	4.02	4.52
Rice bran	0.00	0.00	1.25	1.68	2.39	2.46	2.87	3.21	3.87	4.33	4.73	4.84
Molasses	1.08	1.12	1.23	1.36	1.42	1.51	1.63	1.76	1.88	1.96	2.08	2.11
CaHPO ₄	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.10	0.05	0.05
Limestone	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.40	0.45	0.45
Premix	0.40	0.40	0.40	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Nutrient levels												
DM	82.80	82.22	81.13	80.22	79.19	77.99	77.07	76.86	78.85	81.89	84.28	86.20
OM	93.92	93.92	93.81	93.77	93.59	93.54	93.45	93.44	93.56	93.72	93.87	94.00
CP	10.11	10.65	11.14	11.67	12.21	12.84	13.57	14.22	14.79	15.28	15.85	16.46
NDF	62.11	59.76	57.00	54.16	50.84	48.10	45.22	42.52	41.02	39.44	37.66	36.01
ADF	35.34	34.04	32.45	30.81	28.99	27.37	25.74	24.17	22.93	21.63	20.27	18.99
Ca	0.57	0.56	0.54	0.53	0.53	0.52	0.51	0.50	0.48	0.51	0.50	0.48
P	0.27	0.28	0.31	0.32	0.34	0.36	0.38	0.40	0.42	0.41	0.43	0.44

Note: One kilogram of premix contained the following: Fe 1 650 mg, Cu 1 560 mg, Mn 3 590 mg, Zn 12 100 mg, I 170 mg, Co 60 mg, VA 800 000 IU, VD 700 000 IU, VE 10 000 IU.

2.2 Data acquisition and measurement

2.2.1 Collection and determination of diets nutrient intake

According to the requirements of the Cornell Net Carbohydrate and Protein System (CNCPS) standard^[33], in this experiment, the nutrient intake of dairy cows diets were used as independent variables to predict the excretion of feces, urine and N of dairy cows. During the formal trial period, the daily diets and residual samples were collected and determined, and the average intakes of dry matter and nutrients in the diets were calculated by the difference between the amount of inputs and residues of the daily diets. The contents of dry matter (DM), crude ash (Ash), ether extract (EE), crude protein (CP), and lignin (LIGNIN) were determined by the standard method of AOAC^[34], neutral detergent fiber (NDF), acid detergent fiber (ADF), acid detergent lignin (ADL), neutral detergent insoluble crude protein (NDICP), and acid detergent insoluble crude protein (ADICP) by Van Soest et al.^[35], non-protein nitrogen (NPN) by Licitra et al.^[36], crude protein (SCP) by Krishnamoorthy et al.^[37], and starch by Karkalas^[38]. In addition, According to the standards of CNCPS, the contents of various components of carbohydrates and proteins (PA, PB1, PB2, PB3, PC, CA, CB1, CB2, and CC) were calculated by the method of Sniffen et al.^[39].

2.2.2 Collection and determination of feces, urine and N

In order to obtain the data of the predicted variables, during the formal trial period, feces samples and urine samples were collected and weighed daily, and the dry matter of feces and urine excretions of dairy cows were measured and recorded every day according to whole feces and urine collection method^[40], the content of FN was determined according to general Kjeldahl nitrogen method, the content of UN was determined according to GB11891-89 method, and the FN and UN excretions of dairy cows were measured and recorded at the same time.

2.3 Support vector regression technique

Initially, SVM technique was developed to solve the classical binary classification problems and usually possess excellent generalization capabilities. Another advantage is that it can provide sparse solutions where only the most relevant samples of the training data are weighted and thus result in low computational cost and memory requirements^[41]. In addition, SVM also shows excellent capabilities in the field of prediction, and is usually expressed in the form of SVR when it is used to solve regression problems. In regression, the goal is to estimate an unknown continuous-valued function based on a finite number set of training samples. SVR uses the principle of structural risk minimization to simultaneously optimize empirical loss cost and generalization, and is often able to find non-linear and unique solutions^[42]. In general, SVR tries to locate a regression hyperplane with small risk in high dimensional feature space. Particularly, the standard SVR transforms the input data into a high-dimensional feature space using a non-linear function, solving the final model in the transformed feature space so that not only the training error but also the complexity of the model is minimized^[42]. It is worth mentioning that, the SVR trains the model in a non-parametric manner, and does not make any assumptions about the distribution of the training samples, so it can often fits well for both linear and non-linear data^[43]. In addition, SVR is especially suitable for small sample prediction problems^[32]. Therefore, SVR is considered as one of the most effective machine learning method in predicting applications.

2.3.1 SVR-based modeling

In general, a typical regression problem is learned from the

training samples and used to predict the target values of unknown input vectors. In order to intuitively demonstrate how SVR is used to solve the prediction problems, firstly it needs to introduce the concept of loss function. The loss function is a measure of the error generated by the SVR model during the learning process and is generally selected before the SVR model is learned. In this case, we choose the ε -insensitive loss function proposed by Vapnik^[44], as shown in Figure 1.

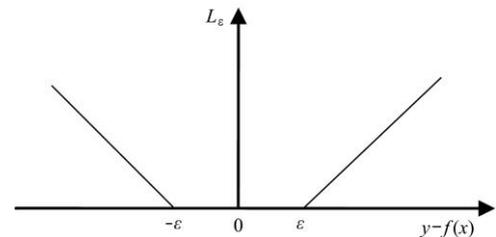


Figure 1 ε -insensitive loss function

The ε -insensitive loss function $L_\varepsilon(f(\mathbf{x}), y)$ is defined as:

$$L_\varepsilon(f(\mathbf{x}), y) = \max\{0, |y - f(\mathbf{x})| - \varepsilon\} \quad (1)$$

In Equation (1), $f(\mathbf{x})$ is a regression function constructed by learning the training sample set and is used to fit the training sample set (\mathbf{x}_i, y_i) , $i=1,2,\dots,n$, where $\mathbf{x}_i \in \mathbf{R}^d$ represents the i th sample value of the input vector, $y_i \in \mathbf{R}$ denotes the output value for a given value of the input variable, n is the number of training samples, $\varepsilon > 0$ is insensitivity loss coefficient and is used to control the fitting accuracy. When solving the regression problem, SVR needs to find an appropriate function $f(\mathbf{x})$ to minimize the error between the observed y and the predicted $f(\mathbf{x})$, and the fitting error can be expressed by the ε -insensitive loss function (also known as the ε -pipeline), as shown in Figure 2, when the training sample points are located in the pipeline represented by two dotted lines in the figure (the training sample points in the pipeline are indicated by hollow dots, and the outside of the pipeline is represented by solid dots), the fitting error of $f(\mathbf{x})$ is considered to be zero.

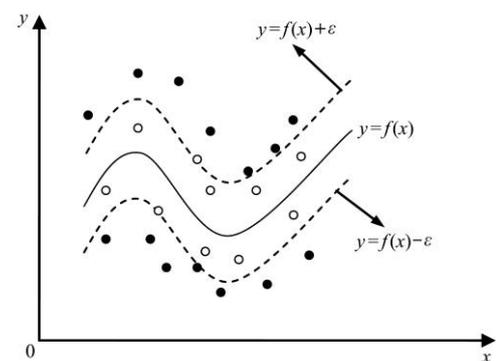


Figure 2 Prediction curve with ε -pipeline

For predicting of dairy cows' feces, urine and N excretions, under ideal conditions, we assume that all training samples are linearly distributed, and then SVR can use linear regression functions to fit the training sample data:

$$f(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} + b \quad (2)$$

where, $\boldsymbol{\omega}$ denotes the weight vector and b stands for the bias term, and the appropriate $\boldsymbol{\omega}$ and b can be determined by learning the training samples. In order to ensure the flatness of the Equation (2) and thus improve the generalization ability of the fitting function, it is necessary to find an optimal $\boldsymbol{\omega}$ by taking the minimization of the norm $\min \frac{1}{2} \|\boldsymbol{\omega}\|^2$ of the Euclidean space.

Next, according to the ε -insensitive loss function, the fitting error

accuracy of all training samples is assumed to be ε . In addition, considering the data that cannot be estimated under the error of ε , the relaxation factors $\zeta_i \geq 0$ and $\zeta_i^* \geq 0$ are introduced. Then, the minimization problem with respect to ω can be transformed into a convex optimization one^[42]:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*) \quad (3)$$

meanwhile, the corresponding constraints conditions are as follows:

$$\begin{cases} y_i - \omega^T x_i - b \leq \varepsilon + \zeta_i \\ \omega^T x_i + b - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \quad i=1,2,\dots,n \quad (4)$$

where, $C > 0$ is the penalty factor used to balance the complexity of the sample and algorithm beyond the error range. In general, the larger value of C usually indicate the greater the penalty for data points beyond the ε -pipeline. Equation (3) and Equation (4) belong to convex quadratic programming problem with linear inequality constraints, and usually solved by the Lagrange function method, that is, we need to establish the Lagrange equation:

$$\begin{aligned} L(\omega, b, \zeta, \zeta^*, \alpha, \alpha^*, \gamma, \gamma^*) &= \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*) - \\ &\sum_{i=1}^n \alpha_i [\zeta_i + \varepsilon - y_i + \omega^T x_i + b] - \sum_{i=1}^n \alpha_i^* [\zeta_i^* + \varepsilon - y_i + \\ &\omega^T x_i + b] - \sum_{i=1}^n [\zeta_i \gamma_i + \zeta_i^* \gamma_i^*] \end{aligned} \quad (5)$$

where, $\alpha_i, \alpha_i^* \geq 0, \gamma_i, \gamma_i^* \geq 0, i=1,2,\dots,n$, are the Lagrange multipliers and the partial derivatives of Equation (5) for parameters $\omega, b, \zeta, \alpha, \alpha^*, \gamma$ should be equal to zero. Then, the condition is brought into the Lagrange equation and the dual form of the convex quadratic programming problem is obtained:

$$\begin{aligned} W(\alpha, \alpha^*) &= \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i - \alpha_i^*)(\alpha_i - \alpha_j^*)(x_i \cdot x_j) + \\ &\sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varepsilon \end{aligned} \quad (6)$$

meanwhile, the corresponding constraints conditions are as follows:

$$\begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases} \quad i=1,2,\dots,n \quad (7)$$

In addition, the above equation is also a quadratic programming problem. According to the necessary and sufficient condition (KKT conditions) of the optimization, at the saddle point, the product of the Lagrange multiplier and the constraint is zero, namely:

$$\begin{aligned} \alpha_i [\varepsilon + \zeta_i - y_i + f(x_i)] &= 0 & \alpha_i^* [\varepsilon + \zeta_i^* - y_i - f(x_i)] &= 0 \\ \zeta_i \cdot \gamma_i &= 0 & \zeta_i^* \cdot \gamma_i^* &= 0 \end{aligned} \quad (8)$$

then, it can be obtained by Equation (8):

$$\begin{aligned} \alpha_i \cdot \alpha_i^* &= 0 \\ (C - \alpha_i) \zeta_i &= 0 \\ (C - \alpha_i^*) \zeta_i^* &= 0 \end{aligned} \quad (9)$$

In this case, as can be seen from Equation (9), if α_i is not 0, then α_i^* will be 0, and vice versa. When α_i and α_i^* are not both 0, the corresponding x_i sample is called Support Vector (SV), and only SVs can contribute to ω . In particular, if $\alpha_i = C$ or $\alpha_i^* = C$, then $|f(x_i) - y_i|$ may be greater than ε , and the corresponding x_i sample is called Boundary Support Vector (BSV), corresponding to the solid points outside the dotted line in Figure 2; In addition, if $\alpha_i = 0, \alpha_i^* \in (0, C)$ or $\alpha_i^* = 0, \alpha_i \in (0, C)$ then $|f(x_i) - y_i| = \varepsilon$, and the corresponding x_i called Normal Support Vector (NSV), corresponding to the points falling on the ε pipe in Figure 2; if $\alpha_i = 0, \alpha_i^* = 0$, the corresponding x_i is a non-support vector, corresponding to the

hollow points in the ε pipe Figure 2, and they have no contribution to ω . Therefore, the larger the ε , the smaller the number of SVs. For the NSV, the parameter b can be obtained from equation (8), namely:

$$b = y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) x_j \cdot x_i - \varepsilon \quad (10)$$

Next, calculate the value of b for all the NSVs, and then calculate the average value, namely:

$$\begin{aligned} b &= \frac{1}{N_{NSV}} \left\{ \sum_{0 < \alpha_j < C} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) x_j \cdot x_i - \varepsilon \right] + \right. \\ &\left. \sum_{0 < \alpha_j^* < C} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) x_j \cdot x_i - \varepsilon \right] \right\} \end{aligned} \quad (11)$$

where, N_{NSV} is the number of the NSVs. Therefore, the SVR linear fitting function obtained from the sample point (x_i, y_i) is:

$$f(x) = \sum_{x_i \in SV} (\alpha_j - \alpha_j^*) x_i \cdot x + b \quad (12)$$

However, in reality, the samples are not necessarily linearly distributed. In order to solve this problem, SVR usually maps the input vectors to a high-dimensional feature space (Hilbert space) by a pre-determined non-linear mapping, and then perform linear regression in this high-dimensional space to obtain the effect of non-linear regression in the original space. To do this, the input quantity x is first mapped into the high-dimensional feature space H through the mapping $\Phi: R^n \rightarrow H$, next, using $f(x) = \omega^T \Phi(x) + b$ instead of Equation (1) to fit the training sample set $(x_i, y_i), i=1,2,\dots,n$. Then the convex quadratic programming Equation (6) becomes:

$$\begin{aligned} W(\alpha, \alpha^*) &= \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i - \alpha_i^*)(\alpha_i - \alpha_j^*)(\Phi(x_i) \cdot \Phi(x_j)) + \\ &\sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varepsilon \end{aligned} \quad (13)$$

Equation (13) involves dot product operation $\Phi(x_i) \cdot \Phi(x_j)$ in high-dimensional feature space, while function Φ is unknown. Fortunately, the SVR theory will replace the point product operation in the high-dimensional feature space with the kernel matrix $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ instead of using the function Φ directly. There are several types of kernel functions such as polynomial, sigmoid, and Gaussian kernel function. In particular, Gaussian kernel function is one of the most commonly used modeling and was selected in this case. The Gaussian kernel function is defined as:

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \quad (14)$$

where, σ indicates the Gaussian kernel width. Therefore, Equation (13) becomes:

$$\begin{aligned} W(\alpha, \alpha^*) &= \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i - \alpha_i^*)(\alpha_i - \alpha_j^*) \cdot K(x_i, x_j) + \\ &\sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varepsilon \end{aligned} \quad (15)$$

Finally, the expression of the SVR non-linear fitting function is:

$$f(x) = \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (16)$$

where, b is calculated as follows:

$$\begin{aligned} b &= \frac{1}{N_{NSV}} \left\{ \sum_{0 < \alpha_j < C} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j \cdot x_i) - \varepsilon \right] + \right. \\ &\left. \sum_{0 < \alpha_j^* < C} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j \cdot x_i) - \varepsilon \right] \right\} \end{aligned} \quad (17)$$

2.3.2 SVR model parameter selection

When applying SVR technique to predict the cows' feces, urine and N excretions, it is also necessary to determine some parameters of the SVR model. These parameters that need to be determined usually include penalty factor C , Gaussian kernel width σ , and

insensitivity loss coefficient ε . Specifically, the penalty factor C can control the tradeoff between the generalization ability of the model and the samples fitting degree; the Gaussian kernel width σ affects the number of SVs, and the insensitivity loss coefficient ε is used to control the fitting accuracy. In this case, for four predictive indicators of feces, urine, FN, and UN, the parameter sets $\{C, \sigma\}$ are determined by 5-fold cross-validation grid search in $C=\{2^{-2}, 2^{-1}, 1, 2^1, 2^2, 2^3, 2^6, 2^8, 2^{10}\}$ and $\sigma=\{0.1, 0.5, 0.7, 1.0, 1.2, 1.5, 2, 2.5, 3\}$. The final determined parameter sets $\{C, \sigma\}$ are $\{2^6, 1.5\}$, $\{2^3, 1.2\}$, $\{2^2, 1.0\}$ and $\{2^3, 0.7\}$ for feces, urine, FN, and UN, respectively. The insensitivity loss coefficient of the four predictive indicators are empirically set to be $\varepsilon=0.01$.

2.3.3 SVR model training and testing process

The whole process for predicting Holstein dry cows' feces, urine FN, and UN excretions using SVR technology is shown in Figure 3.

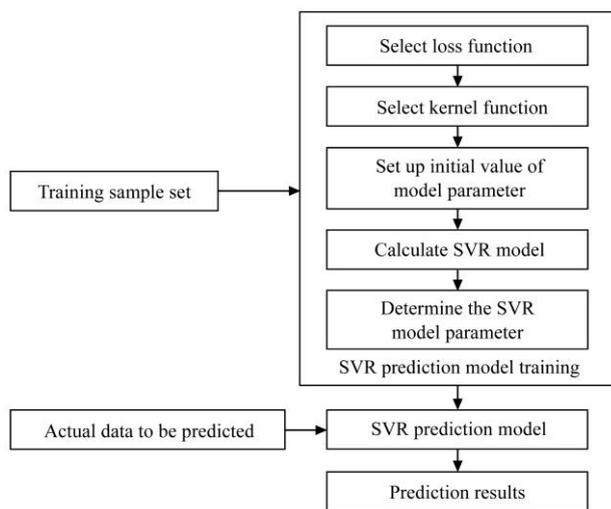


Figure 3 SVR modeling and testing process

In this study, the whole process for predicting four excretions indicators of Holstein dry cows based on the SVR technology can be categorized into training phase and testing phase, and both phases require a certain amount of experimental data for model training and testing. In order to make the model more suitable for the practical needs, the sample data needed for training and testing SVR model are obtained by elaborate experimental design, actual digestion experiments and standard measurement methods. Next, in the training phase of the SVR prediction model, we first select the loss function and kernel function as ε -insensitive loss function and Gaussian kernel function, respectively. Moreover, we set up initial value of corresponding parameters such as C , σ , and ε . When actually starting to calculate the SVR model, it is necessary to input the training set sample data into the established SVR algorithm. In this case, according to the requirements of the CNCPS standard^[33], the intakes of CNCPS components (PA, PB1, PB2, PB3, PC, CA, CB1, CB2, and CC) in the diet of dairy cows were taken as the input x_i , meanwhile, the excretions of dairy cows (feces, urine, FN, and UN) were taken as the predicted $f(x)$, respectively. Next, through the calculation of the training set samples, the parameters such as ε , C and σ of SVR prediction model are determined, in addition, the value of the parameter b can be calculated according to Equation (17), thereby, the trained SVR model is obtained and can be used for actual prediction. Finally, in the testing phase, the testing set sample data is input into the SVR prediction model and calculated according to Equation (16), and the corresponding predicted result can be obtained ultimately.

3 Results and discussion

In the experiment, sixty samples were obtained during the whole formal trial period, as shown in Table 2. For clarity of presentation, the samples data listed in the table were the average of the daily experimental results for each group of dairy cows during the formal trial period.

Table 2 Samples of experiment results

No.	Dietary nutrient intake/g d ⁻¹									Feces, urine and N excretions			
	PA	PB1	PB2	PB3	PC	CA	CB1	CB2	CC	Feces/kg d ⁻¹	Urine/kg d ⁻¹	FN/g d ⁻¹	UN/g d ⁻¹
1	250.28	182.97	224.78	238.23	99.75	1430.69	925.25	4480.22	1414.13	4.68	7.95	73.75	74.65
2	255.14	189.01	237.43	243.05	103.27	1449.54	940.80	4506.92	1436.02	4.38	8.83	76.68	82.24
3	253.46	186.14	232.74	240.32	101.35	1440.30	934.09	4495.83	1426.23	4.51	8.56	74.82	80.13
4	263.18	191.18	240.23	246.42	105.21	1459.72	954.54	4517.30	1438.34	4.24	9.59	78.36	85.33
5	254.87	187.22	233.21	242.62	102.80	1445.20	938.98	4499.46	1428.24	4.46	8.67	75.28	81.39
6	283.50	207.79	292.30	260.24	114.91	1650.52	1118.45	4599.26	1477.05	4.30	9.27	83.37	91.51
7	275.24	202.25	285.09	259.56	111.37	1635.18	1100.32	4590.60	1466.58	4.38	9.13	82.79	88.08
8	284.28	209.42	297.56	271.07	117.01	1656.07	1121.52	4606.97	1481.73	4.13	9.57	84.45	93.10
9	279.19	204.72	287.31	262.86	113.43	1645.67	1105.90	4592.47	1476.29	4.32	9.21	83.10	89.30
10	274.16	200.42	280.78	257.72	109.92	1628.23	1091.37	4583.64	1459.97	4.46	8.87	82.24	87.33
11	306.99	218.09	336.82	275.40	121.06	1845.24	1296.09	4590.73	1489.50	4.37	9.64	89.92	96.59
12	311.10	225.02	358.44	283.59	126.19	1871.73	1317.49	4612.60	1503.50	4.45	10.51	92.02	101.03
13	307.13	220.03	340.67	276.08	122.90	1851.52	1308.56	4597.05	1495.10	4.41	9.61	90.20	99.76
14	302.40	215.16	328.72	268.55	118.80	1832.29	1284.75	4587.61	1484.56	4.27	9.35	88.26	94.90
15	310.33	221.25	347.01	278.09	124.53	1864.30	1311.87	4604.50	1499.95	4.40	10.14	90.64	100.12
16	329.89	232.67	402.57	287.98	131.21	2052.31	1504.45	4518.92	1480.39	4.24	10.61	97.16	108.41
17	319.03	228.60	380.85	283.07	127.58	2031.93	1474.20	4528.54	1486.21	4.38	10.20	94.62	101.93
18	330.96	235.77	414.29	288.31	132.52	2059.89	1509.90	4512.47	1477.75	4.21	10.89	97.68	109.47
19	325.64	231.65	394.40	286.55	130.36	2044.46	1493.44	4525.59	1483.25	4.30	10.41	96.26	102.90
20	333.77	239.84	416.93	290.72	133.31	2071.80	1523.24	4488.17	1472.33	4.18	11.04	98.15	110.04
21	345.31	238.53	455.35	290.48	134.85	2268.46	1699.14	4356.55	1465.97	4.23	11.28	102.47	111.79

No.	Dietary nutrient intake/g d ⁻¹									Feces, urine and N excretions			
	PA	PB1	PB2	PB3	PC	CA	CB1	CB2	CC	Feces/kg d ⁻¹	Urine/kg d ⁻¹	FN/g d ⁻¹	UN/g d ⁻¹
22	348.79	240.10	467.96	292.07	136.96	2279.41	1712.38	4343.41	1461.16	4.13	11.39	103.30	112.29
23	349.31	243.16	469.40	293.58	138.73	2283.90	1728.10	4337.05	1454.58	4.10	11.48	103.12	114.29
24	355.66	247.15	488.36	295.92	143.83	2308.23	1742.02	4324.25	1445.54	3.92	12.09	104.82	119.31
25	351.80	246.09	471.15	294.99	139.04	2293.64	1730.66	4335.47	1453.09	4.02	12.06	103.38	115.83
26	370.32	259.83	556.21	300.22	149.10	2452.01	1944.74	4149.61	1386.38	3.76	13.20	112.21	125.73
27	365.41	255.13	542.57	298.17	146.62	2436.63	1922.61	4167.05	1407.05	3.82	12.23	109.63	123.87
28	366.22	256.04	549.72	299.41	147.81	2442.10	1930.21	4156.72	1399.78	3.79	12.51	110.13	124.57
29	358.67	249.84	529.00	294.29	142.81	2414.82	1896.27	4186.32	1415.22	3.94	11.34	108.09	117.22
30	362.45	251.32	535.93	297.50	144.09	2421.11	1914.50	4182.55	1413.68	3.93	11.96	109.12	121.47
31	384.79	269.57	637.77	303.17	155.62	2601.31	2124.97	3947.12	1350.30	3.59	11.72	118.27	133.42
32	379.31	263.74	611.06	298.62	150.98	2577.97	2083.39	3972.32	1371.52	3.73	12.47	115.12	127.26
33	385.29	270.49	641.76	305.11	157.85	2626.29	2135.87	3943.99	1345.02	3.49	11.50	119.09	134.39
34	382.90	265.51	617.34	299.56	151.90	2586.28	2103.54	3965.32	1363.83	3.67	12.14	116.10	129.80
35	383.46	266.07	620.27	302.22	153.04	2594.25	2116.14	3956.15	1357.70	3.65	11.77	117.22	132.95
36	388.40	269.92	697.83	302.25	158.68	2703.91	2278.12	3711.21	1296.30	3.65	11.40	121.58	136.90
37	389.98	271.48	704.37	301.36	159.44	2710.30	2288.28	3704.81	1287.51	3.62	11.21	122.76	137.09
38	385.50	267.15	683.36	303.62	157.25	2695.85	2260.08	3719.35	1304.03	3.67	11.70	121.10	135.96
39	391.21	272.62	710.12	298.53	161.91	2713.29	2297.37	3692.24	1280.83	3.59	11.01	123.25	139.40
40	392.73	274.65	712.12	297.86	162.80	2723.60	2313.62	3685.33	1272.41	3.52	10.94	123.77	141.21
41	375.98	283.81	762.97	311.54	168.60	2748.99	2397.12	3498.93	1211.91	3.44	10.54	127.86	145.19
42	379.36	280.30	741.20	308.92	164.57	2728.88	2364.29	3528.37	1231.43	3.56	11.18	124.56	141.35
43	378.05	281.22	751.35	309.17	166.50	2731.53	2370.56	3525.16	1226.18	3.50	10.98	126.12	142.38
44	381.94	276.32	729.28	307.60	161.88	2724.16	2351.53	3531.58	1242.10	3.59	11.23	123.95	138.36
45	377.13	282.41	757.82	310.55	167.10	2733.16	2378.54	3516.03	1222.42	3.46	10.82	127.46	144.78
46	369.64	293.31	818.93	322.40	170.08	2813.54	2543.19	3426.54	1207.76	3.51	11.25	131.64	148.97
47	368.62	295.43	828.04	328.16	175.18	2829.08	2578.85	3410.77	1200.93	3.48	11.96	132.01	151.88
48	370.17	292.89	795.78	316.57	169.46	2801.08	2532.98	3435.20	1217.58	3.53	10.61	129.81	142.99
49	366.13	297.48	848.43	329.98	177.88	2841.47	2594.67	3400.14	1185.12	3.43	12.04	134.13	154.74
50	368.64	294.14	822.88	324.58	174.93	2822.08	2554.34	3416.84	1205.39	3.49	11.49	131.72	149.52
51	366.00	316.33	922.57	344.14	184.22	2968.60	2807.42	3349.32	1189.36	3.70	12.27	140.63	159.60
52	363.02	319.21	943.14	349.37	188.07	2976.59	2847.01	3331.04	1174.65	3.83	12.91	142.98	164.09
53	366.34	315.50	908.23	343.70	183.06	2964.82	2790.30	3361.29	1190.67	3.66	12.12	139.22	158.06
54	364.97	316.74	932.19	347.10	185.06	2974.58	2814.45	3337.03	1181.39	3.76	12.56	141.69	162.88
55	367.47	309.71	899.33	339.86	182.71	2958.32	2769.96	3370.26	1196.16	3.51	11.69	137.80	156.28
56	351.07	316.94	972.07	344.70	183.61	2957.09	2920.71	3166.91	1139.17	3.63	12.48	148.57	162.49
57	346.17	323.78	990.71	351.09	189.01	2989.49	2945.25	3150.29	1125.92	3.86	13.00	143.48	164.04
58	344.64	327.24	1001.89	353.23	190.55	2991.50	2957.17	3144.39	1121.72	3.87	13.18	143.03	166.87
59	350.10	321.23	982.80	350.70	186.85	2965.86	2928.10	3163.73	1131.55	3.80	12.78	145.98	163.46
60	343.01	329.34	1005.33	355.10	191.60	3002.14	2889.10	3134.56	1100.84	3.91	13.61	141.32	168.70

In order to evaluate the overall prediction performance of the SVR-based technique and facilitate comparison with other prediction algorithms, the two metrics, root mean square error (E_{RMSE}) and normalized root mean square error (E_{NRMSE}) were used to evaluate the prediction accuracy, the E_{RMSE} and E_{NRMSE} were defined as follows^[43]:

$$E_{RMSE} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\hat{y}(j) - y(j))^2} \quad (18)$$

$$E_{NRMSE} = \sqrt{\sum_{i=1}^n (\hat{y}(j) - y(j))^2} / \sqrt{\sum_{i=1}^n (y(j) - \overline{y(j)})^2} \quad (19)$$

In Equations (18) and (19), $\hat{y}(j)$ is the predicted value of $y(j)$, and E_{RMSE} indicates the average relative deviation of the predicted value $\hat{y}(j)$ relative to the real value $y(j)$, which reflects the accuracy of the prediction. E_{NRMSE} is a standardized representation of E_{RMSE} , which can eliminate the dimensional

impact between indicators and is suitable for comprehensive comparative evaluation. In general, the smaller values of these two metrics usually indicate the better predictive performance.

In order to effectively validate SVR-based prediction technique, two conventional prediction methods, LR and artificial neural networks (ANN) were utilized for comparison. Among them, the LR method is relatively simple and can be implemented according to [18]. For the ANN method, the radial basis function artificial neural network (RBF-ANN) model is selected here, and the RBF-ANN model can be established according to [45]. Especially, it is necessary to pre-specify the number of hidden nodes in each hidden layer. In this case, the number of hidden nodes is determined as the integer number closest to $\log(n)$, where n is the number of training samples.

In this study, All prediction methods (SVR, LR, and RBF-ANN) for comparison are implemented using MATLAB

R2010b software platform. Multiple training and testing were conducted for four predictive indicators of feces, urine FN, and UN, respectively. Ten testing samples were randomly selected for prediction each time, and the rest were used as training samples for learning. The prediction results of different indicators are shown in Figures 4-7, respectively.

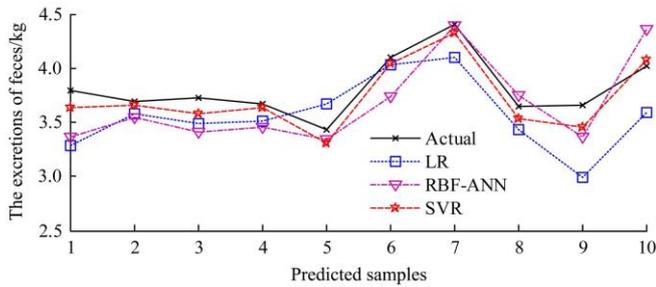


Figure 4 Results obtained by different methods for feces excretions prediction

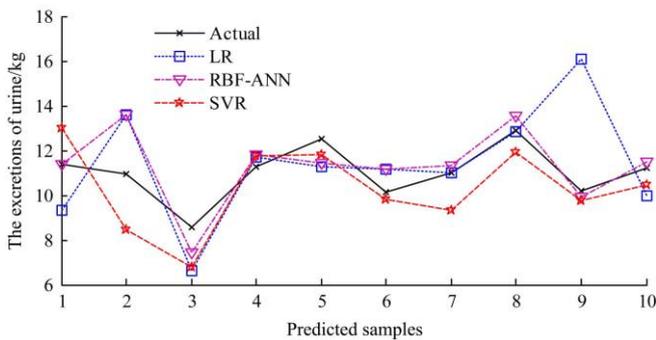


Figure 5 Results obtained by different methods for urine excretions prediction

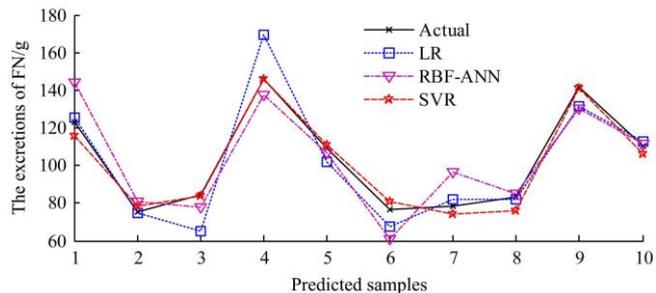


Figure 6 Results obtained by different methods for FN excretions prediction

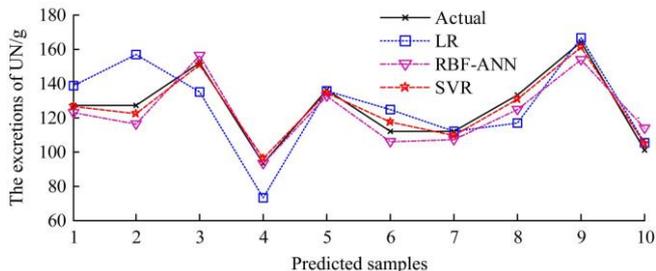


Figure 7 Results obtained by different methods for UN excretions prediction

The E_{RMSE} and E_{NRMSE} of the SVR model prediction results were compared with those of the LR and RBF-ANN methods, as shown in Tables 3-6. In order to eliminate the influence of randomness, each algorithm was run 20 times independently and the mean value of the prediction error was taken.

Figure 4 and Table 3 present the prediction results of the feces excretions indicator. It can be observed that for training samples,

the E_{RMSE} and E_{NRMSE} of the SVR method are slightly higher than the values of the RBF-ANN method, but lower than those of the LR method. However, for the testing samples, the E_{RMSE} and E_{NRMSE} of the SVR method are lower than those of other two methods, thus showing higher prediction accuracy. Especially, compared with the LR method, the prediction accuracy of the feces excretions indicator is improved by 66.01%.

Table 3 Comparison of feces excretions prediction results in terms of metrics

Prediction method	Training E_{RMSE}	Training E_{NRMSE}	Testing E_{RMSE}	Testing E_{NRMSE}
LR	0.288	0.807	0.343	0.962
RBF-ANN	0.096	0.268	0.264	0.741
SVR	0.104	0.291	0.117	0.327

Table 4 Comparison of urine excretions prediction results in terms of metrics

Prediction method	Training E_{RMSE}	Training E_{NRMSE}	Testing E_{RMSE}	Testing E_{NRMSE}
LR	1.556	1.179	2.331	1.338
RBF-ANN	1.168	0.885	1.082	0.621
SVR	1.113	0.843	1.322	0.759

Table 5 Comparison of FN excretions prediction results in terms of metrics

Prediction method	Training E_{RMSE}	Training E_{NRMSE}	Testing E_{RMSE}	Testing E_{NRMSE}
LR	11.985	0.553	12.931	0.597
RBF-ANN	7.872	0.363	9.501	0.438
SVR	2.538	0.117	4.127	0.190

Table 6 Comparison of UN excretions prediction results in terms of metrics

Prediction method	Training E_{RMSE}	Training E_{NRMSE}	Testing E_{RMSE}	Testing E_{NRMSE}
LR	12.580	0.471	14.511	0.543
RBF-ANN	6.213	0.233	7.654	0.287
SVR	2.863	0.107	3.211	0.120

Figure 5 and Table 4 present the prediction results of the urine excretions indicator. Compared with the other two methods, the E_{RMSE} and E_{NRMSE} of the SVR method are slightly inferior to those of the RBF-ANN method. Nevertheless, the SVR method also shows excellent performance on both training samples and testing samples, and its two prediction accuracy metrics are significantly better than the LR method. In particular, its prediction accuracy is improved by 57.90% compared with the LR method.

It can be seen from Figures 6-7 and Tables 5-6 that the prediction accuracy of the SVR method is superior to other two methods for both training samples and testing samples in predicting FN and UN excretions indicators. Among them, the best predictive accuracy is the UN excretions indicator, reaching the testing samples E_{RMSE} value of 3.211 and E_{NRMSE} value of 0.120, which is 77.89% higher than that of the LR method.

Through comparison with LR, and RBF-ANN methods, it can be seen that all the methods can capture the tendency of the real data, but the developed SVR technique can fit slightly better to the real data in most cases. In view of the fact that most of the current predicted excretions indicators for dairy cows are still based on the LR method, the prediction accuracy of the developed SVR technique in terms of E_{NRMSE} are 2.94, 1.76, 3.14 and 4.53 times

that of the LR method in prediction of feces, urine, FN and UN excretions indicators, respectively. It is worth mentioning that, due to the limitations of actual conditions, the experiment in this case is carried out under the condition that the total number of training samples and testing samples does not exceed 60, which is a typical small sample prediction but is often the case in the prediction of dairy cows' feces, urine, and N excretions. The results demonstrated that the proposed SVR-based prediction technique can effectively predict Holstein dry cows' feces, urine and N excretions, and shows better prediction accuracy especially with small samples.

In addition, it should be noted that the depth optimization of the SVR algorithm is not discussed in this paper. Better performance would be expected if some factors such as hyper parameters are further optimized and multiple techniques are effectively integrated, which is the focus of our future research.

4 Conclusions

In this study, a prediction technique of feces, urine and N excretion from Holstein dry cows based on SVR was proposed. Unlike the traditional parametric prediction models such as LR, which requires the assumption of parametric model form, our proposed technique based on non-parametric machine learning model does not require any special assumptions about the predicted model but merely by learning the training samples to predict unknown data, and thus is more suitable for some complex system prediction cases such as cows' feces, urine and N excretion prediction. For the evaluation of the proposed SVR-based prediction technique, we obtain the required sample data through actual feeding experiments, where, a small part of the recorder data were used to train the prediction model and the rest were used for testing. By comparison with the conventional LR and RBF-ANN, the SVR-based prediction technique shows more excellent accuracy in most cases of predicting the main indicators of feces, urine, and N excretions of Holstein dry cows. In particular, under the conditions of sixty sample points, the prediction accuracies of SVR are significantly higher than that of LR in terms of E_{RMSE} and E_{NRMSE} for both training samples and testing samples. Therefore, it can be concluded that the proposed SVR-based prediction technique is an effective way to improve the prediction accuracy of feces, urine, and N excretions of Holstein dry cows. In addition, it is worth mentioning that the SVR-based prediction technique is more suitable for the prediction problems with small samples, which is often the case in the prediction of feces, urine, and N excretions of dairy cows, and then more coincide with actual needs. Finally, we hope that the present study has provided motivation for further study of machine learning technique applied to dairy cows' excretion indicators prediction.

Acknowledgments

The authors would like to acknowledge the financial support from the National Key R&D Program of China (2016YFD0700204-02), the China Agriculture Research System (CARS-36), the China Postdoctoral Science Foundation (2017M611346), and the Natural Science Foundation of Heilongjiang Province of China (C2018018).

[References]

- [1] Li Y, Pan L G, Li A, Wang B H. Suitability evaluation of remediation technology for polluted farmland. *Int J Agric & Biol Eng*, 2015; 8(2): 39–45.
- [2] Franzluebbers A J, Lemaire G, de Faccio Carvalho P C, Sulc R M. Toward agricultural sustainability through integrated crop-livestock systems: Environmental outcomes. *Agric, Ecosyst & Environ*, 2014; 190: 1–3.
- [3] Fan M, Zhu H G, Ma J Q. Measurement and analysis of biogas fertilizer use efficiency, nutrient distribution and influencing factors of biogas residues and slurry on pig farms. *Int J Agric & Biol Eng*, 2014; 7(1): 60–69.
- [4] Abbasi I H R, Abbasi F, El-Hack M A, Abdel-Latif M A, Soomro R N, Hayat K, et al. Critical analysis of excessive utilization of crude protein in ruminants ration: impact on environmental ecosystem and opportunities of supplementation of limiting amino acids—a review. *Environ Sci & Pollut Res Int.*, 2018; 25(1): 181–190.
- [5] Anwar Z, Irshad M, Ping A, Hafeez F, Yang S. Water extractable plant nutrients in soils amended with cow manure co-composted with maple tree residues. *Int J Agric & Biol Eng*, 2018; 11(5): 167–173.
- [6] Mallin M A, Cahoon L B. Industrialized animal production—A major source of nutrient and microbial pollution to aquatic ecosystems. *Popul & Environ*, 2003; 24(5): 369–385.
- [7] Cambra-López M, Aarnink A J A, Zhao Y, Calvet S, Torres A G. Airborne particulate matter from livestock production systems: a review of an air pollution problem. *Environ Pollut*, 2010; 158(1): 1–17.
- [8] Mallin M A, Mciver M R, Robuck A R, Dickens A K. Industrial swine and poultry production causes chronic nutrient and fecal microbial stream pollution. *Water, Air, & Soil Pollut*, 2015; 226: 407.
- [9] Nansubuga I, Banadda N, Babu M, De Vrieze J, Verstraete W, Rabaey K. Enhancement of biogas potential of primary sludge by co-digestion with cow manure and brewery sludge. *Int J Agric & Biol Eng*, 2015; 8(4): 86–94.
- [10] Smith A P, Western A W. Predicting nitrogen dynamics in a dairy farming catchment using systems synthesis modelling. *Agric Syst*, 2013; 115(115): 144–154.
- [11] Qu Q B, Yang P, Zhai Z W, Zhang K Q. Prediction methods of major pollutants production in manure from large-scale livestock and poultry farms: A review. *J Agric Resour & Environ*, 2016; 33(5): 397–406. (in Chinese)
- [12] Gan L, Hu X. The pollutants from livestock and poultry farming in China-geographic distribution and drivers. *Environ Sci Pollut Res Int*, 2016; 23(9): 8470–8483.
- [13] Zhou T M, Fu Q, Zhu Y Q, Hu Z W, Yang F. Optimizing pollutant generation coefficients of livestock industry and mapping patterns of the pollutant constitution in China. *Geographical Research*, 2014; 33(4): 762–776. (in Chinese)
- [14] Fu Q, Wu G Y, Pan P, Wang W T. Analysis of livestock and poultry waste generation from 2000-2014 in Henan. *J Agro-Environ Sci*, 2017; 36(7): 1323–1329. (in Chinese)
- [15] Qu Q B, Yang P, Zhao R, Zhi S L, Zhai Z W, Ding F F, et al. Prediction of fecal nitrogen and phosphorus excretion for Chinese Holstein lactating dairy cows. *J Anim Sci*, 2017; 95(8): 3487–3496.
- [16] Wilkerson V A, Mertens D R, Casper D P. Prediction of excretion of manure and nitrogen by Holstein dairy cattle. *J Dairy Sci*, 1997; 80(12): 3193–3204.
- [17] Nennich T D, Harrison J H, Van Wieringen L M, Meyer D, Heinrichs A J, Weiss W P, et al. Prediction of manure and nutrient excretion from dairy cattle. *J Dairy Sci*, 2005; 88(10): 3721–3733.
- [18] Nennich T D, Harrison J H, Van Wieringen L M, St-Pierre N R, Kincaid R L, Wattiaux M A, et al. Prediction and evaluation of urine and urinary nitrogen and mineral excretion from dairy cattle. *J Dairy Sci*, 2006; 89(1): 353–364.
- [19] Yan T, Frost J P, Agnew R E, Binnie R C, Mayne C S. Relationships among manure nitrogen output and dietary and animal factors in lactating dairy cows. *J Dairy Sci*, 2006; 89: 3981–3991.
- [20] Knowlton K F, Wilkerson V A, Casper D P, Mertens D R. Manure nutrient excretion by Jersey and Holstein cows. *J Dairy Sci*, 2010; 93(1): 407–412.
- [21] Higgs R J, Chase L E, Van Amburgh M E. Development and evaluation of equations in the Cornell Net Carbohydrate and Protein System to predict nitrogen excretion in lactating dairy cows. *J Dairy Sci*, 2012; 95(4): 2004–2014.
- [22] Jiao H P, Yan T, McDowell D A. Prediction of manure nitrogen and organic matter excretion for young Holstein cattle fed on grass silage-based diets. *J Anim Sci*, 2014; 92(7): 3042–3052.
- [23] Kebreab E, France J, Mills J A N, Allison R, Dijkstra J. A dynamic

- model of N metabolism in the lactating dairy cow and an assessment of impact of N excretion on the environment. *J Anim Sci*, 2002; 80(1): 248–259.
- [24] Dong R L, Zhao G Y, Chai L L, Beauchemin K A. Prediction of urinary and fecal nitrogen excretion by beef cattle. *J Anim Sci*, 2014; 92(10): 4669–4681.
- [25] Dijkstra J, France J, Davies D R. Different mathematical approaches to estimating microbial protein supply in ruminants. *J Dairy Sci*, 1998; 81(12): 3370–3384.
- [26] Spek J W, Dijkstra J, Van Duinkerken G, Hendriks W H, Bannink A. Prediction of urinary nitrogen and urinary urea nitrogen excretion by lactating dairy cattle in northwestern Europe and North America: A meta-analysis. *J Dairy Sci*, 2013; 96(7): 4310–4322.
- [27] Schuba J, K H. Südekum, Pfeffer E, Jayanegara A. Excretion of faecal, urinary urea and urinary non-urea nitrogen by four ruminant species as influenced by dietary nitrogen intake: A meta-analysis. *Livest Sci*, 2017; 198: 82–88.
- [28] Yaslan Y, Bican B. Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting. *Meas*, 2017; 103: 52–61.
- [29] Moghadam M P A, Pahlavani P, Bigdeli B. A new car-following model based on the epsilon-support vector regression method using the parameters tuning and data scaling techniques. *Int J Civ Eng*, 2017; 15(1): 1–14.
- [30] Yoo K H, Back J H, Na M G, Kim J H, Hur S, Kim C H. Prediction of golden time using SVR for recovering SIS under severe accidents. *Ann Nucl Energy*, 2016; 94: 102–108.
- [31] Das S P, Padhy S. A novel hybrid model using teaching-learning-based optimization and a support vector machine for commodity futures index forecasting. *Int. J. Mach. Learn. Cybern.*, 2018; 9(1): 97–111.
- [32] Al-Anazi A F, Gates I D. Support vector regression to predict porosity and permeability: Effect of sample size. *Comput. Geosci.*, 2012; 39: 64–76.
- [33] Tylutki T P, Fox D G, Durbal V M, Tedeschi L, Russell J B, Van Amburgh M, et al. Cornell net carbohydrate and protein system: A model for precision feeding of dairy cattle. *Anim. Feed Sci. Technol.*, 2008; 143(1-4): 174–202.
- [34] Thiex N J. *Journal of AOAC International*. *J. AOAC Int.*, 2009; 97(3): 643. doi: urn:issn:1060–3271.
- [35] Van Soest P J, Robertson J B, Lewis B A. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *J Dairy Sci*, 1991; 74(10): 3583–3597.
- [36] Licitra G, Hernandez T M, Van Soest P J. Standardization of procedures for nitrogen fraction on ruminant feeds. *Anim Feed Sci Technol*, 1996; 57(4): 347–358.
- [37] Krishnamoorthy U, Sniffen C J, Stern M D, Van Soest P J. Evaluation of a mathematical model of rumen digestion and an in vitro simulation of rumen proteolysis to estimate the rumen-undegraded nitrogen content of feedstuffs. *Br J Nutr*, 1983; 50(3): 555–568.
- [38] Karkalas J. An improved enzymic method for the determination of native and modified starch. *J. Sci. Food Agric.*, 2010; 36(10): 1019–1027.
- [39] Sniffen C J, O'Connor J D, Van Soest P J, Fox D G, Russell J B. A net carbohydrate and protein system for evaluating cattle diets: II. Carbohydrate and protein availability. *J Anim Sci*. 1992; 70(11): 3562–3577.
- [40] Xia K, Wang Z B, Xi W B, Yao Q, Li F G, Wang Y, et al. Effects of forage combinations on nutrient digestibility, utilization of energy and nitrogen of diets for dairy cows. *Chin J Anim Nutr*, 2012; 24(4): 681–688. (in Chinese)
- [41] Cortes C, Vapnik V. Support-vector networks. *Mach. Learn.*, 1995; 20(3): 273–297.
- [42] Smola A, Schoelkopf B. A tutorial on support vector regression. *Stat. & Comput.*, 2004; 14(3): 199–222.
- [43] Ma J, Theiler J, Perkins S. Accurate On-line Support Vector Regression. *Neural Comput.*, 2003; 15(11): 2683–2703.
- [44] Vapnik V N. *The nature of statistical learning theory*. Springer, New York, 1995. doi: 10.1007/978-1-4757-2440-0.
- [45] Ghorbani M A, Zadeh H A, Isazadeh M, Terzi O. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environ. Earth Sci.*, 2016; 75(6): 476.