Research Article

Modeling and prediction of wear rate of grinding media in mineral processing industry using multiple kernel support vector machine

Asghar Azizi[1] · Reza Rooki[2] · Nader Mollayi[3]

Received: 14 May 2020 / Accepted: 15 July 2020 / Published online: 3 August 2020

Abstract
In this study, we investigates the application of three powerful kernel-based supervised learning algorithms to develop a global model of the wear rate of grinding media based on the input factors such as pH, solid percentage, throughout, charge weight of balls, rotation speed of mill and grinding time. It is found that there is a trade-off between the training and testing error when a single kernel function is used and therefore these methods cannot provide the generalization capability. However, this problem is solved utilizing the multiple kernel learning frameworks for support vector machine in which the kernel function was expressed as a combination of basis kernel functions. It is distinguished that compared to the single kernel and ANN-based techniques, the use of multiple kernel support vector machines benefit from a higher degree of correctness and generalization ability for prediction of wear rate of grinding media. Meanwhile, the findings indicate that in this state, the values of R^2 are achieved 0.99417 and 0.993 for training and testing datasets, respectively.

Keywords Wear rate · Grinding media · Mineral processing · Multiple kernel · Support vector machine

1 Introduction

The grinding media wear plays an important role in the economics of grinding processes in mineral processing plants. Wear is defined as a progressive loss of material from a solid body owing to its contact and relative movement against a surface [1]. It has been accepted that wear is resulted in a lower the operational efficiency of machinery and its components, and also it is a major source of costs in the various industries [2]. Meanwhile, it is known that the mining and metallurgy industries significantly depend on the comminution operations to increase mineral liberation. Comminution is one of the most important operational units in the mineral processing industry and it is well known that the direct operating costs in comminution circuits (including crushing and milling) are mainly the energy consuming and the metal lost through wear in the mineral industry [3, 4]. Radziszewski reported that typical operational costs may be divided into extraction (30–70%), separation (5–20%), and comminution (30–50%) [5]. Moema et al. stated that consumption of grinding media forms an important part of the operational costs and grinding medium wear can constitute up to 40–45% of the total operation cost in comminution process [6]. In addition, King et al. expressed that wear rate is one of the most significant parameters for appraising the overall performance of grinding medium [7]. Thus, grinding media should be produced to provide the highest performance

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s42452-020-03212-0) contains supplementary material, which is available to authorized users.

✉ Asghar Azizi, azizi.asghar22@yahoo.com; aazizi@shahroodut.ac.ir; Reza Rooki, rooki@birjandut.ac.ir; Nader Mollayi, mollayi@birjandut.ac.ir | [1]Faculty of Mining, Petroleum and Geophysics, Shahrood University of Technology, Shahrood, Iran. [2]Department of Mining Engineering, Birjand University of Technology, Birjand, Iran. [3]Department of Computer Engineering, Birjand University of Technology, Birjand, Iran.

that is the lowest wear magnitude and largest grinding transfer to cost ratio [8].

Total media wear in grinding media is attributed to three main mechanisms, including abrasion, impact and corrosion. Abrasion and impact wear are described as metal loss due to mechanical force on the grinding media. Erosion wear results from the friction between grinding media and particles. Additionally, corrosive wear is depicted as metal loss owing to chemical and/or electrochemical reactions of grinding media with the solution [9, 10]. The grinding media wear is influenced by the different operational factors. Many factors are beyond the control of the engineers, while, some of them cannot be even measured quantitatively with the available instruments [11–13]. A vast number studies were carried out on the influence of these factors in wear of grinding media. Chen et al. [14] performed a laboratory study of high chromium alloy wear on grinding mill in the Florida phosphate industry and reported that corrosive wear was a very serious problem. Also, their findings indicated that the solution pH had the most influence on the wear rate and the minimum wear rate were obtained at 8.7 solution pH, 61 rpm rotation speed, 65% solid percentage and 58% crop load. Azizi et al. [15] utilized response surface methodology for modeling and optimization of influential factors on the corrosive wear of grinding balls and reported that the interactive effects between factors had a significant role on corrosive wear and the highest corrosion rate could be obtained 78.38 and 40.76 mils per year for low alloy and high carbon chromium steel balls, respectively. Additionally, further investigations demonstrate using the classic routs of investigating the impacts of factors on the media wear in grinding mills. These methods rely on the empirical models which are achieved from statistical correlations among dependent and independent factors. One of the limitations with empirical models is series of simplifications and plurality of the parameters which should be considered in grinding media wear. In addition, studies indicate that linear models are well established and reliable, but have limited application. Hence, some researchers have employed neural networks models as a very useful and powerful tool of the modeling the complex systems to estimate wear rate especially in mills [16, 17]. Meanwhile, more recent attentions have been made to build models for obtaining a better understanding from the mechanistic principles of grinding media wear. Kor and coworkers have developed a fuzzy logic model to estimate the wear rate of a high chromium alloy [18]. Ashrafizadeh and Ashrafizadeh applied a numerical simulation including discrete element to predict the wear caused by solid particle impact [19]. Furthermore, developing the prediction capability of grinding media wear allows a more precise estimation of wear costs. Therefore, investigation of newer approaches to achieve a higher accuracy and generalization capability can be beneficial in mineral processing and metallurgy industries.

In recent years, data-driven soft sensors due to delay-free and low-cost properties have been widely developed and utilized to predict the behavior of the chemical processes, especially multi-grade processes [20–22]. It has also been accepted that among these methods, nonlinear soft sensors such as neural networks [23, 24], Gaussian process regression (GPR) [25] and support vector regression (SVR) [26–29] are more attractive mainly because of the nonlinear relation existing between the response variable of process and operating conditions. In fact, these techniques can relatively easily develop without deep understanding of the process mechanism [29]. Today, support vector machines (SVM) are a new achievement in the field of data driven modelling, and along with other learning based-kernel algorithms indicate better performance than artificial neural networks and other intelligent or statistical models on the most popular benchmark problems and has been successfully implemented in classification, regression and function estimation [30, 31]. Multiple kernel support vector machine (MK-SVM) is a newer formulation of SVM, which results in higher accuracy and generalization capability in many applications [32].

The wear rate of grinding media is mainly influenced by various parameters such as pH, solid percentage, throughout, charge weight of balls, rotation speed of mill and grinding time which the relationship between the wear rate and the effective parameters is a highly-nonlinear and coupled multivariable relationship and cannot be expressed by an explicit mathematical model. On the other hand, the consumption of grinding media in grinding circuits forms a remarkable part of the operating costs and also the required experimental and analytical tasks are costly, time consuming and complex. Therefore, it is very necessary to find a simple, reliable, capable and accurate approach or model for predicting the loss rate of girding media from these operating parameters. Hence, this study was aimed for the prediction of wear rate of grinding media using MK-SVM modeling as a state-of-the-art estimation approach. Also, it needs to be pointed out that application of this algorithm is studied for wear rate prediction of grinding media in mills for first time.

The present paper is organized in the following fashion. Section 1 expresses the literature and a theoretical background of research. Experimental data is presented in Sect. 2. Section 3 provides a detailed explanation of support vector machines (SVM) and multiple kernels learning

(MKL). Then, Sect. 4 describes results and discussion and in fact the comparison between the single kernel SVM, ANN and multiple linear regression methods for predicting the wear rate of girding media. Finally, the conclusions are drawn in Sect. 5.

2 Experimental data

A data set derived from 50 experiments on the wear rate of a low alloy steel ball was applied in this work. The experimental data were obtained from literature [17]. The experimental conditions, influential factors and the total wear rate for each experiment, which was determined by the grinding media weight loss, are implied in Table 1. To measure the grinding media mass losses through total wear, 15 steel balls were handpicked and marked and then before and after each grinding test were weighted to determine the ball losses. Ultimately, the total wear rate was calculated from following Equation [33].

$$\mathrm{CR} = \frac{534 \times \mathrm{W}}{\rho \times \mathrm{A} \times \mathrm{t}} \tag{1}$$

where *CR* is the wear rate in mils penetration per year (mpy), *W* denotes weight loss in milligrams, ρ represents density in grams per cubic centimeter, *A* depicts area in square inches, and *t* exhibits time in hours [33].

3 Support vector machines (SVM)

Rapid advances in information processing systems in recent decades, has led to a demanding need to systems that can learn from limited information and solve complex decision problems. The investigation and production of algorithms that can learn from a series of observed data and make predictions based on them are explored by a subfield of computer science known as machine learning. In supervised learning, given a set of *N* input vectors $\{x_n\}_{n=1}^{N}$ and the corresponding targets $\{t_n\}_{n=1}^{N}$, we want to learn a model of the dependency of the targets on the inputs in order to make predictions of in cases which have not been observed [34].

The SVM, in its present form, was developed in 1990s at AT&T Bell Laboratories [35] and has been proven to be efficient in many practical applications for classification and regression analysis. The major advantage of the support vector machines compared to the neural networks is minimization of the structural risk besides the empirical risk which leads to a better generalization capability.

A SVM-based classifier system is described by Eq. (2), in which *w* and *b* represent the weights and bias vector, respectively. The goal is to find the hyper-plane which results in an equivalent maximum margin between the samples of the two classes in the training dataset. Maximization of the margin between the two classes is performed via minimization of the risk function *R(w)* described by Eq. (3), subjected to the constraints of Eq. (4), for the *N* samples of the (x_i, y_i) in the training dataset.

$$f(x) = sign\left(w^T x + b\right) \tag{2}$$

$$R(w) = \frac{1}{2} w^T w = \frac{1}{2} w^2 \tag{3}$$

$$y_i\left(w^T x_i + b\right) \geq 1, i = 1, \ldots, N \tag{4}$$

The support vectors lay on a hyper-plane satisfying the condition of

$$y_{sp}\left(w^T x_{sp} + b\right) = 1 \tag{5}$$

In case of a training database with samples which are not linearly separable, another factor will be added to the risk function in Eq. (3) for the inevitable error in case of the samples which lay outside the permitted borders, resulting in the risk function expressed in Eq. (6) subjected to the restrictions of Eq. (7), in which ξ_i is the distance between the support vectors' hyper-plane and the samples which lay outside it, as depicted in Fig. 1. The variable *C* implies the regularization factor which trades off the relative significance of maximizing the margin and training error.

$$R(w) = \frac{1}{2}\left\| w^2 \right\| + C \sum_{i=1}^{N} \xi_i \tag{6}$$

$$s.t.\ y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, \ldots, N \tag{7}$$

This concept can be applied for nonlinear classification by mapping the original feature space to some higher-dimensional feature space where the training set is separable by a hyper-plane, through a nonlinear function known as the kernel function [36].

Thus, SVM-based classification is described based on the following equation:

$$f(x) = sign\left(w^T \phi(x) + b\right) \tag{8}$$

Table 1 The conducted experiments conditions and calculated values of wear rate [17]

Run	pH	Solid percentage (%)	Throughout (g)	Charge weight (kg)	Speed (rpm)	Grinding time (min)	Wear rate (mpy)
1	8	35	720	12	70	10	462.11
2	10	45	360	12	80	10	351.45
3	10	35	360	12	70	10	425.36
4	10	45	360	12	70	15	363.53
5	9	50	540	10	75	12.5	317.66
6	9	40	540	6	75	12.5	373.8
7	9	40	540	10	75	12.5	400.65
8	10	45	360	8	70	10	292.91
9	9	30	540	10	75	12.5	491.74
10	9	40	540	10	75	12.5	383.4
11	10	35	720	8	70	10	378.3
12	8	45	360	12	80	15	515.51
13	8	45	360	12	70	10	376.58
14	9	40	540	10	65	12.5	337.95
15	10	35	720	12	70	15	447.67
16	9	40	540	10	75	12.5	417.79
17	8	35	360	8	70	15	478.46
18	8	35	720	8	80	10	480.45
19	10	45	720	12	70	10	335.18
20	8	45	720	8	80	15	469.95
21	9	40	540	10	85	12.5	451.8
22	8	45	360	8	80	10	401.55
23	9	40	540	10	75	12.5	387.04
24	10	45	720	8	70	15	342.48
25	9	40	540	10	75	12.5	391.58
26	10	45	360	8	80	15	393.94
27	8	35	360	8	80	10	488.36
28	8	45	720	8	80	10	403.43
29	9	40	540	10	75	12.5	405.45
30	9	40	900	10	75	12.5	361.8
31	9	40	540	10	75	17.5	461.06
32	8	35	360	12	70	15	531.79
33	10	45	720	8	80	10	347.4
34	8	45	720	12	70	15	458.29
35	8	35	360	8	80	15	555.53
36	10	35	720	12	80	10	442.73
37	9	40	540	10	75	7.5	340.46
38	8	45	720	8	70	10	366.3
39	10	35	360	12	80	15	475.28
40	8	35	720	8	80	15	521.7
41	8	35	360	8	70	10	463.27
42	9	40	540	14	75	12.5	451.68
43	9	40	180	10	75	12.5	436.46
44	8	35	720	8	70	15	484.98
45	10	45	720	12	80	15	385.35
46	8	45	360	8	70	15	435.97
47	10	35	360	8	80	10	433.46
48	11	40	540	10	75	12.5	333.22
49	7	40	540	10	75	12.5	499.87

Table 1 (continued)

Run	pH	Solid percentage (%)	Throughout (g)	Charge weight (kg)	Speed (rpm)	Grinding time (min)	Wear rate (mpy)
50	10	35	720	8	80	15	408.26

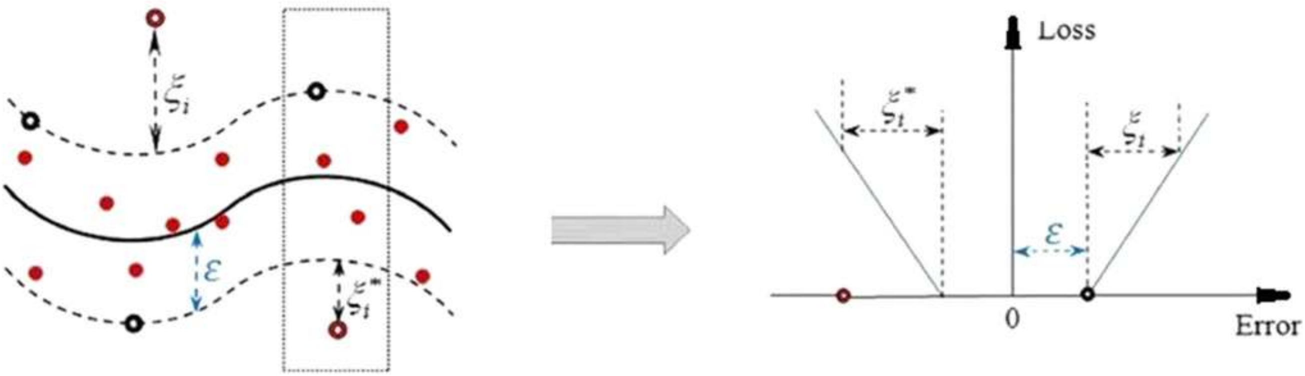

Fig. 1 SVM-based regression and the concept of ε—insensitivity

where w and b can be determined using minimizing the risk function *R(w)* in Eq. (6) with the limitation of Eq. (9):

$$y_i\left(w^T\phi(x_i)+b\right)\geq 1-\xi_i, \tag{9}$$

In general, the concept of this classification is generalized for the aims of regression with representing a margin of tolerance [27] and ultimately, the SVM-based regression (SVR) can be formulated as follow:

$$y=f(x)=\sum_{i=1}^{m} w_i\phi_i(x)+b_i=w^T\phi(x)+b \tag{10}$$

The error of SVM-based regression is measured based on Vapnik's ε-insensitive loss function, as shown in Fig. 1. Also, it can be expressed via below equation.

$$\xi=|y-f(x)|_{\varepsilon}=max\{0,|y-f(x)|-\varepsilon\} \tag{11}$$

The aim is to determine the magnitudes of w and b based on a set of available training data, so that the difference between the original function and the predicted function is minimized. For this purpose, in SVM-based regression, the parameters are calculated by minimizing the risk function *R(w)* formulated as:

$$\min\left\{R(w)=\frac{1}{2}\left\|w^2\right\|+C.\sum_{i=1}^{N}\xi_i+\xi_i^*\right\}$$
$$s.t.\begin{cases} y_i-\left(w^T\phi(x_i)+b\right)\leq\varepsilon+\xi_i \\ \left(w^T\phi(x_i)+b\right)-y_i\leq\varepsilon+\xi_i^* \\ \xi_i\xi_i^*\geq 0, i=1,\dots,N \end{cases} \tag{12}$$

By presenting the dual optimization problem [37], the convex optimization problem in (2011) is reformulated according to the following relation:

$$max\left\{L\left(\alpha_i,\alpha_i^*\right)=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\phi(x_i),\phi(x_j)\right\}-\varepsilon\sum_{i=1}^{N}\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{N}y_i\left(\alpha_i-\alpha_i^*\right)\right\} \tag{13}$$

$$s.t.\begin{cases} \sum_{i=1}^{N}\alpha_i-\alpha_i^*=0 \\ \alpha_i,\alpha_i^*\in[0,C], i=0,\dots,N \end{cases}$$

Based on the Mercer's theorem [35], the inner product $\langle\varphi(x),\varphi(x_i)\rangle$ can be defined through a kernel $K(x,x_i)$ as

$$K\left(x_i,x_j\right)=\phi\left(x_i\right),\phi\left(x_j\right) \tag{14}$$

Therefore, the dual optimization problem is expressed as:

$$max\left\{ L(\alpha_i,\alpha_i^*) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{ (\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j) \right\} - \varepsilon\sum_{i=1}^{N}(\alpha_i+\alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i-\alpha_i^*) \right\} \tag{15}$$

$$s.t.\begin{cases} \sum_{i=1}^{N}\alpha_i-\alpha_i^*=0 \\ \alpha_i,\alpha_i^*\in[0,C], i=0,\dots,N \end{cases}$$

Finally, the dual optimization problem is solved by quadratic programming optimization and based on the optimal obtained parameters. The estimated function is expressed as:

$$f(x)=\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)K(x,x_i)+b \tag{16}$$

The most popular relations for the kernel function are presented in Table 2.

3.1 Multiple kernel learning

In SVM and other kernel-based learning algorithms, the efficiency of the algorithm extremely depends on the data representation, selected through the kernel function. The nonlinear similarity between samples is measured by kernel function and therefore an efficient kernel must be able to represent data adaptively. In addition, the kernel function's parameters define a proper regularization term for the learning problem. In the most states, the parameters of a single kernel function are tuned for the whole data sets. Despite proper tuning of the kernel parameter can increase the generalization ability, learning with one kernel is not very data-adapted and does not lead to acceptable results. Recent developments in the literature on SVMs and other kernel approaches show the need to consider multiple kernels to improve flexibility based on the fact that typical learning problems often involve multiple, heterogeneous data sources [38]. Multiple kernel learning (MKL) aims at simultaneously learning a kernel and the associated predictor in a supervised learning problem. In multiple kernel learning framework, the kernel function is constructed by a linear convex combination of M functions [39], each one satisfying the Mercer's conditions, expressed as:

$$K(x,x_i)=\sum_{m=1}^{M}d_mK_m(x,x_i) \tag{17}$$

where d_m is the combining weight of the *m-th* basis kernel function, satisfying the constraints of:

$$\sum_{m=1}^{M}d_m=1,\ d_m\geq 0 \tag{18}$$

The vector of weights is defined as $d=[d_1,\dots,d_M]^T$.

The multiple kernel learning (MKL) problem is learning the combining weights d_m and the solutions of the original problem, for example, the solutions of α_i and α_i^* for SVR problem in Eq. (16) in a single optimization problem. The optimization problem of MKL-based SVR is obtained by substitution of Eq. (17) into Eq. (15) as:

$$max\left\{ L(\alpha_i,\alpha_i^*) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{ (\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\sum_{m=1}^{M}d_mK_m(x_i,x_j) \right\} - \varepsilon\sum_{i=1}^{N}(\alpha_i+\alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i-\alpha_i^*) \right\} \tag{19}$$

$$s.t.\begin{cases} \sum_{i=1}^{N}\alpha_i-\alpha_i^*=0 \\ \alpha_i,\alpha_i^*\in[0,C], i=0,\dots,N \\ \sum_{m=1}^{M}d_m=1, d_m\geq 0, m=1,\dots,M \end{cases}$$

Table 2 The most common formulations of the kernel function

Kernel type	Formulation
Linear	$K(x,x_i)=\langle x,x_i\rangle$
Gaussian radial basis (RBF)	$K(x,x_i)=exp(-\frac{\|x-x_i\|^2}{2\sigma^2})$
Polynomial of degree d	$K(x,x_i)=(\langle x,x_i\rangle+1)^d d\in N$
Multi-layer perceptron (MLP)	$K(x,x_i)=tanh(k.\langle x,x_i\rangle+\theta)k,\theta>0$
Gaussian radial basis function (RBF)	$K(x,x_i)=exp(-\frac{\|x-x_i\|^2}{2\sigma^2})$

Rakotomamonjy and co-workers [39] suggested an ordinary and efficient algorithm for MKL-based SVM and expressed that the objective function L in Eq. (19) is convex and differentiable and accordingly the gradient descent method can be utilized to solve the optimization problem. In this approach, the optimal vector of weights d, is achieved by updating it on the gradient descent direction of L. To calculate the gradient of the target function, the partial derivatives of L are measured from below equation.

$$\frac{\partial L}{\partial d_m} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)K_m\left(x_i, x_j\right) \tag{20}$$

Then the descent direction D of gradients is distinguished and d is updated via below relation.

$$d \leftarrow d + \gamma D \tag{21}$$

in which γ exhibits the step length. The gradient of the objective function is only updated when the target value reduces [40]. The described process is repeated until some stopping criterions are met, as discussed in ref [39].

4 Results and discussion

4.1 Predicting the wear rate of steel balls by single and multiple kernel SVM regression analysis

In this study, a database of 50 experimental measurements for the wear rate of grinding balls in mils penetration per year (mpy) and the corresponding input parameters of pH, solid percentage (%), throughout (weight of samples input ball mill) (g), charge weight of balls (kg), rotation speed of mill (rpm) and grinding time (minutes) was employed. The experimental conditions and measured values of input and output parameters are given in Table 1.

The models were trained using 40 randomly selected data (accounting for 80% of the total data) and the remaining 10 data (accounting for 20%) were applied for testing purposes.

To improve the accuracy, all the input and target values were normalized between − 1 and + 1 via below Equation.

$$pn = 2 * \frac{p - \left(\frac{max+min}{2}\right)}{(max - min)} \tag{22}$$

in which, *max* and *min* imply the maximum and minimum value of the input or the output from the whole datasets, respectively, p depicts the input or output and pn is corresponded to normalized value.

Table 3 The single kernel SVM parameters

SVM kernel parameter	0.2
SVM regularization factor (C)	10,000
SVM insensitivity parameter (ε)	10^{-5}

Table 4 Comparison of the statistical indices using the MK-SVM, SK-SVM and ANN methods

Database	Method	RMSE	NRMSE	R^2
Training	MK-SVM	4.64	0.0111	0.99417
	SK-SVM	4.6279	0.0110	0.9942
	ANN [17]	12.75	0.0304	0.9545
Testing	MK-SVM	16.003	0.0382	0.933
	SK-SVM	62.314	0.1486	0.0163
	ANN [17]	22.74	0.0542	0.912

According to the normalized dataset, the single kernel and multiple kernel SVM models were implemented utilizing the SVM-KM and the SimpleMKL toolboxes [41, 42], respectively. Training the models and calculating the predicted normalized outputs, they were scaled to their original range:

$$\hat{y} = y_n * \left(\frac{max - min}{2}\right) + \left(\frac{max + min}{2}\right) \tag{23}$$

where $\hat{y}$ is the predicted output in the original range and y_n is the normalized predicted output.

The validity of the developed models was assessed using the root means square error (RMSE), normalized root means square error (NRMSE) and the coefficient of determination (R^2) statistical indices, defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{N}} \tag{24}$$

$$NRMSE = \frac{RMSE}{\overline{Y}} \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2} \tag{26}$$

In above equations, y_i and $\hat{y}_i$ represent the observed values and the estimated outputs, respectively, N denotes the number of training or testing samples, $\overline{Y}$ implies the average value of the total outputs and $\overline{y}$ is the mean value of the corresponding training or testing measured outputs.

In case of the single kernel (SK-SVM) method, using the Gaussian kernel function, the kernel and model variables were obtained by minimizing the training root mean square error, as listed in Table 3. The multiple kernel algorithm was trained with the same value of insensitivity parameter (ε) and the regularization factor set to $C = 100$. It needs to be pointed out that the kernel function was chosen as a mixture of ten Gaussians and three polynomial functions with parameters of 1, 2, 3.

4.2 Performance comparison of SVM

Using the MK-SVM method described above, all necessary computations were implemented by supplying extra codes in MATLAB software. The calculated values of the statistical indices for the training and testing databases based on these methods and the ANN-based method reported by Azizi and co-workers [17] are presented in Table 4. The results highlighted the superior performance of MK-SVM method for prediction of the wear rate towards other methods investigated. In addition, the performance of SVM techniques were compared with the results estimated and modeled by response surface methodology (RSM) based on the central composite design (CCD) [43]. The findings proved that the MK-SVM model suggested had comparable and relatively similar results with RSM Model (with predicted R^2 value of 0.93). Therefore, it can be concluded that the multiple kernel SVM methodology can be successfully applied for predicting and simulating the wear rates of grinding media.

In addition, further investigations exhibit that the SVM testing normalized root means square error can be decreased to 16.06 by changing the kernel parameter to 40, but with this kernel parameter, the training root mean square error became 21.053. Meanwhile, the SK-SVM model is overlearned in training process and cannot be trained to make the best predictions for the test data besides the training data. Thus, the MK-SVM method benefits from better generalization capability as well as higher overall precision. The measured outputs associated with the outputs estimated by the MK-SVM approach are illustrated in Figs. 2 and 3. Ultimately, the findings show an excellent agreement between measured and predicted values with R^2 magnitudes of 0.99417 and 0.933 for train and test datasets, respectively.

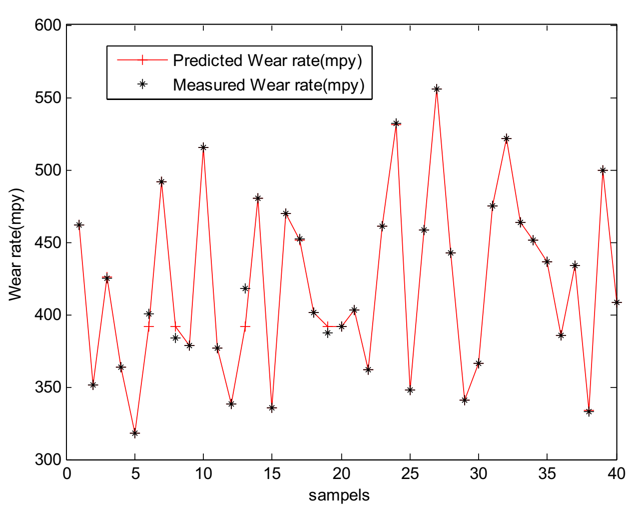

Fig. 2 MK-SVM predicted and measured wear rate (mpy) values for the train data

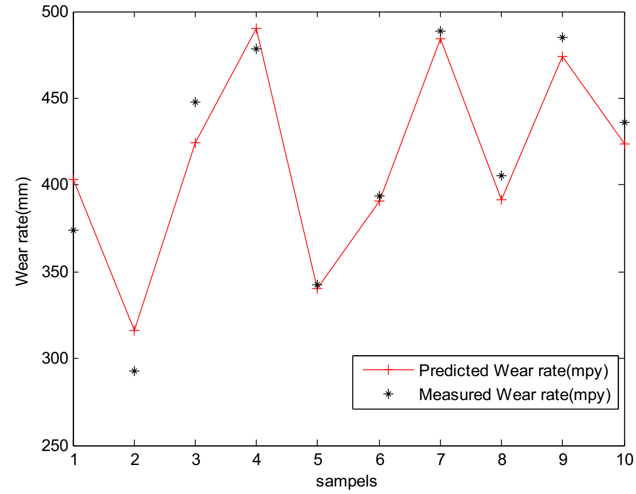

Fig. 3 MK-SVM predicted and measured wear rate (mpy) values for the test data

5 Conclusion

In present research, the use of multiple kernel SVM regression analysis was evaluated for modeling and predicting the wear rate of grinding media in mineral processing industry based on the input factors of pH, solid percentage, throughout, charge weight of balls, rotation speed of mill and grinding time. It is found that although a single kernel function's parameter cannot be tuned to provide a good accuracy for both the training and test data in kernel based approaches, using an optimal linear combination of basis kernel functions as the kernel function in support vector regression results in a good accuracy for both the test and training data. For this purpose, the kernel function was constructed based on a linear mixture of polynomial and RBF kernel functions and consequently applying a simple MKL algorithm, the optimized combination of the kernel function and the solutions of the SVR problem were achieved. Prediction results demonstrated a high accuracy of the multiple kernel SVM and also its progressive generalization capability towards the single kernel SVM, relevance vector machine (RVM), artificial neural network (ANN) and multiple linear regression methods. In this condition, the correlation coefficients (R^2) were determined to be 0.99417 and 0.933 for training and testing stages, respectively. Thus, it can be concluded that the MK-SVM technique can be efficiently utilized for predicting and modeling the wear rates of grinding balls in mineral processing industry. Additionally, Application of SVM method to modeling the wear and corrosion rates of liners in grinding mills and its performance comparison with ANFIS model can be considered for future researches.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

References

1. Zum Gahr KH (1987) Microstructure and wear of materials, vol 10. Elsevier, Amsterdam
2. Tylczak JH, Hawk JA, Wilson RD (1999) A comparison of laboratory abrasion and field wear results. Wear 225:1059–1069. https://doi.org/10.1016/S0043-1648(99)00043-5
3. Massola CP, Chaves AP, Albertin E (2016) A discussion on the measurement of grinding media wear. J Mater Res Technol 5(3):282–288. https://doi.org/10.1016/j.jmrt.2015.12.003
4. Bürger R, Bustamante O, Fulla MR, Rivera IE (2018) A population balance model of ball wear in grinding mills: an experimental case study. Miner Eng 128:288–293. https://doi.org/10.1016/j.mineng.2018.09.004
5. Radziszewski P (2002) Exploring total media wear. Miner Eng 15(12):1073–1087. https://doi.org/10.1016/S0892-6875(02)00228-5
6. Moema JS, Papa MJ, Slabbert GA, Zimba J (2009) Grinding media quality assurance for the comminution of gold ores. In: World Gold Conference, The Southern African Institute of Mining and Metallurgy
7. King J, Li Q, Wang A, He C, Zho J, Deng H, Xu R (2015) Evaluation of grinding media wear-rate by a combined grinding method. Miner Eng 73:39–43. https://doi.org/10.1016/j.mineng.2014.12.002
8. Jankovic A, Wills T, Dikmen S (2016) A comparison of wear rates of ball mill grinding media. J Min Metall 52A(1):1–10. https://doi.org/10.5937/JMMA1601001J
9. Chenje T, Simbi D, Navara E (2003) The role of corrosive wear during laboratory milling. Miner Eng 16(7):619–624. https://doi.org/10.1016/S0892-6875(03)00132-8
10. Aldrich C (2013) Consumption of steel grinding media in mills—a review. Miner Eng 49:77–91. https://doi.org/10.1016/j.mineng.2013.04.023
11. Moore J, Perez R, Gangopadhyay A, Eggert JF (1988) Factors affecting wear in tumbling mills: influence of composition and microstructure. Int J Miner Process 22(1–4):313–343. https://doi.org/10.1016/0301-7516(88)90071-3
12. Yelloji Rao M, Natarajan K (1991) Factors influencing ball wear and flotation with respect to ore grinding. Min Proc Ext Met Rev 7(3–4):137–173. https://doi.org/10.1080/08827509108952670
13. Azizi A, Shafaei SZ, Noaparast M, Karamoozian M, Greet CJ, Yarahmadi M, Jabbari M (2014) A study on the corrosive and abrasive wear of grinding media in grinding of minerals using fuzzy analytical hierarchy Delphi method. Arab J Sci Eng 39(5):3373–3382. https://doi.org/10.1007/s13369-014-0976-z
14. Chen G, Tao D, Parekh B (2006) A laboratory study of high chromium alloy wear in phosphate grinding mill. Int J Miner Process 80(1):35–42. https://doi.org/10.1016/j.minpro.2006.01.007
15. Azizi A, Shafaei SZ, Noaparast M, Karamoozian M (2013) Modeling and optimizing the corrosive wear of steel balls in ball grinding mill. IJMSE 10(4):58–71
16. Nakai ME, Aguiar PR, Guillardi H Jr, Bianchi EC, Spatti DH, D'Addonab DM (2015) Evaluation of neural models applied to the estimation of tool wear in the grinding of advanced ceramics. Expert Syst Appl 42(20):7026–7035. https://doi.org/10.1016/j.eswa.2015.05.008
17. Azizi A, Shafaei S, Rooki R (2016) Wear rate prediction of grinding media using BPNN and MLR models in grinding of sulphide ores. IJMSE 13(2):73–84
18. Kor M, Abkhoshk E, Tao D, Chen GL, Modarres HR (2010) Modeling and optimization of high chromium alloy wear in phosphate laboratory grinding mill with fuzzy logic and particle swarm optimization technique. Miner Eng 23(9):713–719. https://doi.org/10.1016/j.mineng.2010.04.009
19. Ashrafizadeh H, Ashrafizadeh F (2012) A numerical 3D simulation for prediction of wear caused by solid particle impact. Wear 276:75–84. https://doi.org/10.1016/j.wear.2011.12.003
20. Liu Y, Gao Z, Li P, Wang H (2012) Just-in-Time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes. Ind Eng Chem Res 51(11):4313–4327. https://doi.org/10.1021/ie201650u
21. Liu Y, Chen J (2013) Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes. J Process Control 23(6):793–804. https://doi.org/10.1016/j.jprocont.2013.03.008
22. Liu Y, Yang C, Liu K, Chen B, Yao Y (2019) Domain adaptation transfer learning soft sensor for product quality prediction. Chemometr Intell Lab Syst 192:103813. https://doi.org/10.1016/j.chemolab.2019.103813
23. Gonzaga JCB, Meleiro LAC, Kiang C, Filho RM (2009) ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. Comput Chem Eng 33(1):43–49. https://doi.org/10.1016/j.compchemeng.2008.05.019
24. Wang WC, Zhang M, Liu XG (2015) Improved fruit fly optimization algorithm optimized wavelet neural network for statistical data modeling for industrial polypropylene melt index prediction. J Chemom 29(9):506–513. https://doi.org/10.1002/cem.2729
25. Liu Y, Chen T, Chen J (2015) Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions. Ind Eng Chem Res 54(18):5037–5047. https://doi.org/10.1021/ie504185j
26. Chitralekha SB, Sha SL (2010) Application of support vector regression for developing soft sensors for nonlinear processes. Can J Chem Eng 88(5):696–709. https://doi.org/10.1002/cjce.20363
27. Kaneko H, Funatsu K (2011) Development of soft sensor models based on time difference of process variables with accounting for nonlinear relationship. Ind Eng Chem Res 50(18):10643–10651. https://doi.org/10.1021/ie200692m
28. Zhang M, Liu XG (2016) A real-time model based on optimized least squares support vector machine for industrial polypropylene melt index prediction. J Chemom 30(6):324–331. https://doi.org/10.1002/cem.2795
29. Liu Y, Yang C, Gao Z, Yao Y (2018) Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes. Chemometr Intell Lab Syst 174:15–21. https://doi.org/10.1016/j.chemolab.2018.01.008

30. Meyer D, Leisch F, Hornik K (2003) The support vector machine under test. Neurocomputing 55(1–2):169–186. https://doi.org/10.1016/S0925-2312(03)00431-4
31. Shi D, Gindy NN (2007) Tool wear predictive model based on least squares support vector machines. Mech Syst Signal Process 21(4):1799–1814. https://doi.org/10.1016/j.ymssp.2006.07.016
32. Bach FR, Lanckriet GR, Jordan MI (2004) Multiple kernel learning, conic duality, and the SMO algorithm. In: Proceedings of the twenty-first international conference on Machine learning, ACM. https://doi.org/10.1145/1015330.1015424
33. Jones DA (1996) Principles and prevention of corrosion, 2nd. Prentice Hall, Upper Saddle River, NY, pp 168–198
34. Tou JT, Gonzalez RC (1974) Pattern recognition principles. Addison-Wesley, Boston
35. Vapnik VN (1998) Statistical learning theory. Wiley, New York
36. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge
37. Parrella F (2007) Online support vector regression, Master's Thesis, Department of Information Science, University of Genoa, Italy
38. Sonnenburg S, Rätsch G, Schäfer C, Schölkopf B (2006) Large scale multiple kernel learning. J Mach Learn Res 7:1531–1565
39. Rakotomamonjy A, Bach F, Canu S, Grandvalet Y (2008) SimpleMKL. J Mach Learn Res 9:2491–2521
40. Xiang-rong Z, Long-ying H, Zhi-sheng W (2010) Multiple kernel support vector regression for economic forecasting. In: Management science and engineering (ICMSE), international conference on management science & engineering, 17th Annual Conference Proceedings, Melbourne, VIC, Australia, IEEE. https://doi.org/10.1109/ICMSE.2010.5719795
41. Rakotomamonjy A, Canu S (2008) SVM and kernel methods Matlab toolbox, UFR Sciences, Dept Physics, University of Rouen, French, https://asi.insarouen.fr/enseignants/~arakotom/toolbox/index.html. Accessed 1 Aug 2009
42. Toolbox, S., Interactive, *2010.11. 30*.https://asi.insarouen.fr/enseignants/~arakotom/code/mklindex.html
43. Azizi A (2015) Investigating the controllable factors influencing the weight loss of grinding ball using SEM/EDX analysis and RSM model. Eng Sci Technol Int J 18(2):278–285. https://doi.org/10.1016/j.jestch.2014.12.007