Chemical Papers (2023) 77:2509–2524
https://doi.org/10.1007/s11696-022-02642-x

ORIGINAL PAPER

Machine learning-guided prediction and optimization of precipitation efficiency in the Bayer process

Abbas Bakhtom[1] · Saeed Ghasemzade Bariki[1] · Salman Movahedirad[1] · Mohammad Amin Sobati[1]

Received: 26 September 2022 / Accepted: 21 December 2022 / Published online: 6 January 2023

Abstract

Machine learning approaches were used to predict and optimize the precipitation efficiency in the Bayer process. One thousand five hundred and sixty real operating data points of the precipitation efficiency from Iran Alumina Company were used for the model's development. Radial basis function (RBF) and support vector machine (SVM) networks were applied to develop a black-box model of the process. The input parameters of the models were the concentrations of sodium oxide (Na_2Oc) and aluminum oxide (Al_2O_3), tank temperature, ambient temperature, residence time, and solid content. To create an optimal model, a trial-and-error strategy based on analyzing all potential configurations was used. The network's prediction performance is further demonstrated through model generalization inside the training data domain. The outcomes of both RBF and SVM networks demonstrate a good agreement between the industrial data and the model predicted values when considering statistical measures such as correlation coefficients of more than 0.99999, mean square errors, the absolute average deviation, and the absolute average relative deviation of less than 0.01%. The outcome of the models was used to optimize the operating parameters in such a way as to maximize precipitation efficiency with a minimum concentration of sodium oxide. The results show that the average precipitation efficiency of 42% was increased to 47% at optimized conditions.

Keywords Machine learning · Precipitation efficiency · Bayer process · Radial basis function · Support vector machine

Introduction

The Bayer process was studied and patented by Karl Josef Bayer and is still the most economic procedure for producing 1 ton of alumina from 1.9–3.6 t of bauxite ores (Hind et al. 1999). The process consists of three steps (Bahrami et al. 2012a; Sidrak and research 2001):

- Bauxite is extracted from the mine and mixed with a hot stream of sodium hydroxide before entering the dissolution tank with digester slurry.
- Precipitation of cooled supersaturated sodium aluminate liquor
- Calcination of the aluminum hydroxide to Al_2O_3 and classification of the deposited crystals based on their size

The most important step for the production of alumina in the Bayer process is the precipitation stage. The precipitation step determines the final production efficiency of aluminate liquor. Super-saturation of aluminum oxide (Muhr et al. 1997), temperature (Liu et al. 2020), seed concentration (Paspaliaris et al. 1999a), impurities (Dorin and Frazer 1988; Ostap 1986; Rosenberg 2017; Vogrin et al. 2020; Wellington et al. 2007), and mixing rate affect (Bahrami et al. 2012b; Misra 2016) the precipitation process (Totten and MacKenzie 2003a).

The main step in the Bayer process for producing alumina trihydrate is precipitation. This technique involves rapidly forming a frugally soluble solid phase from a liquid solution phase (Seecharran 2010). The performance of the precipitation step strongly affects the produced aluminum's quality, productivity, and selectivity. The appropriate control of this stage is essential to achieving maximum production efficiency, minimum operating cost, optimal particle size distribution, and high product purity (Geoff Bearne 2017). Significant parameters in the precipitation step include super-saturation (Ilievski and Livk 2006; Veesler et al. 1994; Yu et al. 2020), temperature (Huang et al. 2019), caustic

✉ Salman Movahedirad
movahedirad@iust.ac.ir

[1] School of Chemical Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran

ratio (Satpathyc 2014), stirring rate (Ilievski and Livk 2006; Zhou et al. 2018), seed concentration (Zhang et al. 2006), residence time, solids content, the presence of any impurities (Hui-bin Yang 2020; Smeulders et al. 2001; Zhang et al. 2018, 2006, 2020), and additives (Zeng J 2007; Liu et al. 2018; Lü et al. 2010; Sahu and Meyrick 2015; Paulaime et al. 2003; Yin et al. 2006; Zhang et al. 2009; Zeng et al. 2008) that decelerate the hydrate growth/agglomeration (Sonthalia et al. 2013).

In the precipitating solution, super-saturation degree is the most important parameter, which is affected by temperature, concentration, and caustic ratio and affects the rates of various processes, such as nucleation and growth. The solubility of alumina in the precipitation step strongly depends on both the tank and ambient temperatures (Bahrami et al. 2012a). In this regard, as temperatures rise, super-saturation and precipitation efficiency decrease. It should be noted that reducing the solution's viscosity reduces the temperature's effect and increases efficiency. Super-saturation and precipitation rates decrease when the caustic ratio increases, resulting in a reduction in particle size and morphological irregularities. Another important and limiting parameter in the precipitation process is the residence time of aluminum hydroxide seed precipitation, which varies approximately between 48 and 72 h. As the residence time increases, the super-saturation degree decreases and the precipitation efficiency increases (Seecharran 2010).

Zhang et al. (2009) and Sahu and Meyrick (2015) studied the effect of the caustic ratio on precipitation efficiency. The results show that increasing the caustic ratio reduced the precipitation rate.

Yu et al. (2020) measured the influence of super-saturation on the precipitation rate. Due to the increase in driving force for the migration of aluminum from the soluble phase to the crystal surface, the higher super-saturation led to an increase in the precipitation rate. Furthermore, Veesler et al. (1994), Ilievski et al. (2006) Barata et al. (1996), and Zhang et al. (2009) investigated the super-saturation parameter and found comparable results.

Huang et al. (2019) studied the temperature effect on the efficiency of the precipitation rate. The results show that with increasing temperatures, the precipitation rate decreases. Also, Zhang et al. (2009) and Yu et al. (2020) reported similar results. However, Sahu et al. (2015) obtained results that contradicted those of other researchers. They claimed that increasing the temperature would increase the precipitation rate.

Many studies have been conducted to investigate the effect of the stirring rate on the precipitation rate. As the stirring rate increases, the precipitation efficiency increases to an optimal point and then decreases (Ilievski and Livk 2006; Veesler et al. 1994; Zhang et al. 2009; Zhou et al. 2018).

Zhang et al. (2006) studied the effect of seed concentration on the precipitation rate. The results showed that as the seed concentration increased, so did the available surface area of ions on the seed surface. Bahrami et al. (Seecharran 2010) detected a similar trend in the effect of seed concentration on precipitation efficiency.

The effect of different additives (i.e., methanol, monohydroxyl alcohol, ether crone, ethylene diamine-tetra-acetic acid (EDTA), urea, and tetracarbon) on precipitation rate was also investigated. The positive effect of these additives on the precipitation efficiency in the Bayer process is evident (Satpathyc 2014; Zhang et al. 2007; LÜ et al. 2010; Yin et al. 2006; Ying Zhang et al. 2009; Zeng et al. 2008). Table 1 illustrates the effect of different parameters on the precipitation efficiency in the Bayer process.

Various methods, such as theoretical, experimental, semi-experimental, numerical, artificial neural networks (ANN), and machine learning approaches, can be used to model the precipitation efficiency in the Bayer process. For this purpose, experimental data on the parameters affecting the precipitation

Table 1 Effects of different parameters on the precipitation efficiency

Parameters		Effect	Refs
Super-saturation		Increase	Huang et al. (2019), Hui-bin Yang (2020), Ilievski and Livk (2006), Ostap (1986), Sidrak and research (2001)
Temperature		Increase	Huang et al. (2019), JingTao and Tan (2001), Sahu and Meyrick (2015), Ostap (1986)
Caustic ratio		Decrease	Sahu and Meyrick (2015), Ostap (1986)
Stirring rate		Increase–decrease	Hui-bin Yang (2020), Ilievski and Livk (2006), Liu et al. (2018), Ostap (1986)
Seed concentration		Increase–decrease	Heidari et al. (2016), Liu et al. (2020)
Additives	Methanol	Increase	Ostap (1986)
	Ether Crone	Increase	Paspaliaris et al. (1999a)
	Tetracarbon	Increase	Paspaliaris et al. (1999b)
	Monohydroxy alcohol	Increase	Paulaime et al. (2003)
	EDTA	Increase	Rosenberg (2017)

step were obtained from the Iran alumina factory. Several configurations were evaluated in the development of the machine learning algorithms, and the network performance was checked by changing the number of hidden layers, the number of neurons, the training algorithm, and the variance to obtain the best network for predicting precipitation.

According to some authors, ANN has been used to estimate alumina recovery in the Bayer process. Multiple linear regression analysis (MLRA) was used by Đurić et al. (2012) to model alumina recovery during the leaching process. They investigated the effect of bauxite composition, sodium aluminate solution composition, and caustic module composition on the degree of alumina recovery before and after the leaching process. In another study, radial basis function (RBF) and multilayer perceptron (MLP) as neural networks and the multiple linear regression (MLR) method were used to predict the alumina recovery efficiency (A.R.E.), the amount of produced red mud (A.P.R.), the red mud settling rate (R.S.R.), and bound soda losses (B.S.L.) based on the lime-to-bauxite ratio and chemical analyses in the Bayer process (Mahmoudian et al. Spring 2016). MLP, RBF, and MLR networks were used to estimate bound soda losses in Iran Alumina Complex solid residue (red mud) using regression and neural networks (AAN) (Ghaemi et al. 2018). Chelgani et al. (2009) investigated the relationship between alumina leaching recovery and the chemical modules of bauxite using regression and ANN methods. Since the modeling of the precipitation stage in the Bayer process has not been done so far with machine learning approaches, an attempt was made to observe the performance of available operating parameters affecting the precipitation efficiency with two RBF and SVM networks.

In the present study, the modeling of precipitation efficiency in the Bayer process for alumina production in the Iran Alumina factory was investigated using data collected during the period 2020–2021 (521 days). The precipitation efficiency is predicted using regression and machine learning algorithms. To achieve this goal, radial basis function (RBF) and support vector machine (SVM) networks were used. The results of the two networks were compared in terms of the correlation coefficient (R^2), mean square error (MSE), absolute average deviation (AAD), and absolute average relative deviation (AARD). Sensitivity analysis was applied to investigate the effect of input parameters on precipitation efficiency. The prediction performance is further demonstrated through model generalization inside the training data domain.

Practical implications and data set

The Hall–Heroult process converted alumina into the primary raw material for aluminum production. Alumina is typically extracted from bauxite using the Bayer process. Depending on the ore's quality, one ton of alumina typically requires between 1.9 and 3.6 tons of bauxite. Digesting, clarifying, precipitating, and crystallizing are the main steps in the Bayer process (Paspaliaris et al. 1999b).

The precipitation step is the most crucial step in the Bayer process for producing alumina. The precipitation process is primarily influenced by the degree of aluminum oxide supersaturation in the input solution, variations in temperature during precipitation, the specific surface area of the bud, impurities in the liquor solution, and the mixing intensity in the precipitation tanks (Totten and MacKenzie 2003b). The production conditions influence the chemical quality and physical properties of alumina products, such as particle size, wear resistance, fracture, and shear properties. Understanding the mechanisms of the precipitation process, which include nucleation, growth, and agglomeration, is essential due to the conflict between the production of a high-quality product and a large quantity of the product.

The phenomena of nucleation, growth, and agglomeration necessitate that the solution be supersaturated with the dissolved substance. There are two distinct categories of nucleation: primary and secondary. Primary nucleation is accomplished in a supersaturated solution devoid of crystals of the soluble substance and is referred to as "nucleation." Obviously, primary nucleation is also classified as homogeneous or heterogeneous. In homogeneous type, primary nucleation occurs in a solution devoid of solid particles, whereas in heterogeneous type, primary nucleation is caused by exogenous solid particles. Secondary nucleation involves the formation of nuclei in the vicinity and the presence of soluble material particles. By adding the bud to the solution, secondary nucleation is induced, making this form of nucleation the primary cause of nucleation during the precipitation process (Ilievski and Livk 2006). After nucleation in the precipitation process, particle growth occurs. Shape, behavior, size, purity, and tensile and wear resistance are the particle qualities that influence the growth stage. There are three distinct stages of crystal growth. The first stage, called volumetric diffusion, involves the growth units diffusing the fluid mass to reach the particles' surfaces. The second stage, called surface diffusion, involves the growth units diffusing the particles' faces. The third stage, called junction diffusion, involves the growth units diffusing the kink sites, where the energy is the lowest. The diffusion phase controls the rate in systems when the solubility of the dissolved substance is very high in the solvent. The surface reaction phase controls the growth process in low-solubility solutions. Agglomeration is a crucial step in the Bayer process that raises the original particle size. During this stage, the fine crystals added to the solution and the fine nuclei formed are connected. The solution must be supersaturated, which is one of the requirements for agglomeration (Bahrami et al. 2012b). According to the stated concepts, it can

be concluded that the precipitation step is the most critical step in the aluminum hydroxide production process.

In order to model the precipitation efficiency with machine learning algorithms, the precipitation unit daily data set of the Iran Alumina Factory has been used from 2020 to 2021 (1560 data points from 521 days). As mentioned before, the precipitation unit efficiency is affected by various parameters. Due to the importance of the sodium oxide concentration (Na_2O_c), aluminum oxide concentration (Al_2O_3), solid content, residence time, ambient/tank temperatures, their significant effect on the precipitation efficiency, and limitation of collected data from Iran Alumina Factory were used as input parameters for the network. It is obvious that the precipitation efficiency considers the output parameter of the network.

Various definitions have been proposed for precipitation efficiency, but the simplest definition is based on precipitation rate per volume according to Eq. (1) (Bahrami et al. 2012a):

$$\eta = \frac{\alpha_{\text{final}} - \alpha_{\text{primary}}}{\alpha_{\text{final}}} \times 100 \tag{1}$$

α (%) is the alumina module and is generally used to express efficiency as follows (Bahrami et al. 2012a):

$$\alpha = 1.645 \times C/A \tag{2}$$

where C and A represent the Na_2O_c concentration and Al_2O_3 concentration in the solution phase (gr/lit), respectively.

The ranges of the data set are given in Table 2. To limit the range of a set of values, the data set must be normalized before being submitted to the network and regression models. As a result, the data were normalized using the following equations to adapt them to the transfer or activation function and place the input and output values in the same order between 0 and 1. (Ganguly 2003). The normalized values of the data are also given in Table 2.

$$x_{\text{normalized}} = s.x + o \tag{3}$$

$$s = \frac{d_{\max} - d_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

$$o = \frac{x_{\max} d_{\min} - x_{\min} d_{\max}}{x_{\max} - x_{\min}} \tag{5}$$

Here, $\boldsymbol{x_{max}}$ and $\boldsymbol{x_{min}}$ are the minimum and maximum of data sets, respectively. $\boldsymbol{d_{max}}$ and $\boldsymbol{d_{min}}$ are the maximum and minimum desired values of the output range, respectively ($\boldsymbol{d_{max}} = 1, \boldsymbol{d_{min}} = 0$). $\boldsymbol{s}$ and $\boldsymbol{o}$ are the scaling and offset parameters, respectively. Also, Eq. (3) can be used to return the network response from the normalized range to the original range (inverse parameter scaling). The normalized

Table 2 Data set's range and normalized value

Input parameters										Output parameter	
Concentration of Na_2O_c (gr/lit)		Concentration of Al_2O_3 (gr/lit)		Tank Temperature (°C)		Ambient Temperature (°C)		Residence Time (hr)		Solid Content (gr/lit)		Precipitation efficiency (%)	
Real value	Normalized	Real value	Normalized	Real value	Normalized	Real value	Normalized	Real value	Normalized	Real value	Normalized	Real value	Normalized
129.00	**0.8158**	**128.00**	**0.7052**	**56.00**	**0.7246**	**10.5**	**0.5185**	**4.76**	**0.4270**	**386**	**0.9763**	**45.77**	**0.6300**
133.50	+1	**135.70**	**0.9925**	**56.40**	**0.7826**	**9**	**0.4815**	**4.78**	**0.4382**	**328**	**0.6331**	**46.47**	**0.7102**
130.20	**0.8609**	**132.10**	**0.8582**	**51.00**	**0**	**13.5**	**0.5926**	**5.55**	**0.9962**	**327**	**0.6272**	**47.78**	**0.8603**
132.70	**0.9549**	**131.60**	**0.8396**	**55.00**	**0.5797**	**−1.5**	**0.2222**	**4.16**	**0**	**282**	**0.3609**	**45.36**	**0.5830**
129.00	**0.8158**	**128.00**	**0.7052**	**55.10**	**0.5942**	**27.5**	**0.9383**	**5.49**	**0.9533**	**390**	+1	**47.13**	**0.7858**
132.20	**0.9361**	**135.00**	**0.9664**	**56.00**	**0.7246**	**2.5**	**0.3210**	**5.09**	**0.6649**	**282**	**0.3609**	**48.74**	**0.9702**
$Max = 133.90$	$Max = +1$	$Max = 135.90$	$Max = +1$	$Max = 57.90$	$Max = +1$	$Max = 30$	$Max = +1$	$Max = 5.56$	$Max = +1$	$Max = 390$	$Max = +1$	$Max = 49.00$	$Max = +1$
$Min = 107.30$	$Min = 0$	$Min = 109.10$	$Min = 0$	$Min = 51$	$Min = 0$	$Min = -10.5$	$Min = 0$	$Min = 4.16$	$Min = 0$	$Min = 221$	$Min = 0$	$Min = 40.27$	$Min = 0$
$Avg = 128.87$	–	$Avg = 128.61$	–	$Avg = 56.07$	–	$Avg = 15.54$	–	$Avg = 5.18$	–	$Avg = 308.20$	–	$Avg = 42.02$	–

operating data were fed as inputs for training RBF and SVM, whereas normalized precipitation efficiency was kept as an output.

Network model implementation

The artificial neural network (ANN) is a modern computational method for machine learning, knowledge demonstration, and the use of learned knowledge to predict output responses from complex systems. The main idea of ANN is based on the biological neural system method for data processing and information learning (JingTao and Tan 2001; Yao et al. 1999). The key component of this idea is the creation of new structures for data processing systems. The structure is made up of a large number of interconnected processing elements called neurons that operate together to model the system. Radial basis function (RBF), support vector machine (SVM), and multilayer perceptron (MLP) networks are the most common and practical structures of machine learning approaches. Figure 1 shows the work cycle algorithms of RBF and SVM networks.

The goal of this study is to use RBF and SVM to predict precipitation efficiency in the Bayer process. Four performance metrics were computed for each series to compare the results between the stated networks: the regression coefficient (R^2), mean square error (MSE), absolute average deviation (AAD), and absolute average relative deviation (AARD). The regression coefficient (R^2) is stated by (Heidari et al. 2016):

$$R^2 = \frac{\frac{\sum_i (x_i - \bar{x})\left(d_i - \bar{d}\right)}{N}}{\sqrt{\frac{\sum_i \left(d_i - \bar{d}\right)^2}{N}} \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}}} \tag{6}$$

The following equations define the mean square error (MSE), absolute average deviation (AAD), and absolute

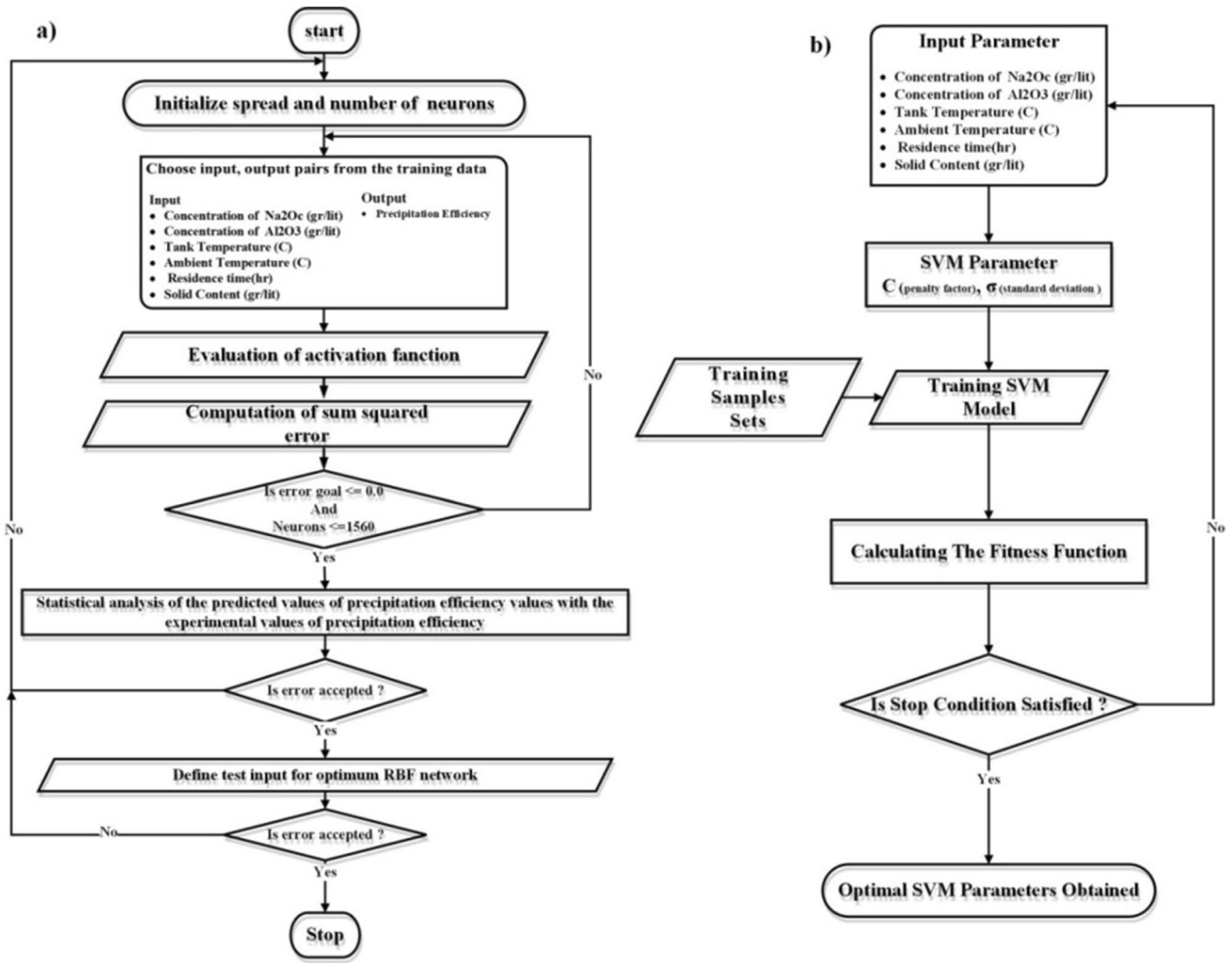


Fig. 1 Work cycle algorithms **a** RBF and **b** SVM networks

average relative deviation (AARD), respectively (Baş and Boyacı 2007; Heidari et al. 2016):

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - y_{di}\right)^2 \tag{7}$$

$$\mathrm{AAD} = \frac{1}{N}\sum_{i=1}^{N}\left(\left|y_i - y_{di}\right|\right) \tag{8}$$

$$\mathrm{AARD} = \left(\left(\sum_{i=1}^{N}\left|y_i - y_{di}\right|/y_i\right)/N\right)\times 100 \tag{9}$$

The experimental and computed responses are y_i and y_{di}, respectively, and N is the number of experimental data. ($N = 1560$ *data points*). Due to high-temperature changes during the year, the data were divided into two parts, and the modeling for the first half of 2020 and 2021 (1023 data points) and the second half of 2020 (537 data points) was done separately.

Radial basis function (RBF)

The radial basis function was used in 1988 due to its generalizability and simple structure, which prevented extended and unnecessary calculations. RBF is made up of three precursor layers that receive the input vector in the first layer and propagate the results to the middle layer to generate the network output in the last layer. In RBF, the general form of the approximation function is presented as follows (Vt and Shin, 1994):

$$G(x) = \sum_{i=1}^{M} w_i \varphi(\|x - c_i\|) \tag{10}$$

$$\varphi(s) = e^{-\frac{s^2}{2\sigma^2}} \tag{11}$$

Weights (w_i) can be obtained using the least squares method by having or estimating c_i and σ for each intermediate neuron.

Support vector machine (SVM)

Support vector machine (SVM) is a highly supervised machine learning technique that was originally developed for organization problems and then extended to regression tasks in 1995. In order to minimize error, SVM uses structural risk minimization (SRM) principles that work better than experimental risk minimization (ERM) principles (Andras 2002). By selecting a function, SVM estimates the target's true value as close to the reference value as possible. Then, it measures the overall pattern and the maximum permissible deviation of the targets from the estimated values for all training data. Finally, the input parameters are transferred to the multidimensional region using a nonlinear mapping function of the Gaussian nucleus, in which a linear model is estimated according to Eq. (13) (Byvatov et al. 2003):

$$f(x) = \sum_{i=1}^{l}\left(a_i^* - a_i\right).\varphi\left(x_i\right).\varphi\left(x_j\right) + b \tag{12}$$

where $\varphi\left(x_{i,j}\right)$ is the internal multiplications of the vectors in the multidimensional region, which can be represented by the kernel function according to Eq. (14) (Byvatov et al. 2003).

$$f(x) = \sum_{i=1}^{l}\left(a_i^* - a_i\right).K\left(x_i, x_j\right) + b \tag{13}$$

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

where l represents the support vectors number, a_i^*, a_i, and b are the Lagrange coefficients and bias, respectively, and $K\left(x_i, x_j\right)$ is the kernel function.

Results and discussion

Machine learning approaches are a set of interconnected mathematical neurons that create a model of complex functional relationships. With the weights and biases specified during training, this model should be able to compute output values from input values derived from internal calculations.

Application of RBF

Designing RBF networks depends on adjusting the radial function's parameters to provide the best possible approximation error for training data. Determining the number of radial, center, width, or variance functions and the weight coefficients for these functions is the key problem in designing such networks. The spread factor is used to measure the dispersion of data. The spread factor narrows the available options, whereas a higher number provides more flexibility. The centers of the basic functions need not be fully adapted to training samples, but they should be distributed uniformly through the feature vector space. To put it another way, RBF networks conceptually provide a local approximation of the main function in the neighborhood of each center, and by combining these local approximations, a general approximation of the considered function is supplied throughout the entire feature vector space.

Typically, the RBF network learning process consists of two steps: (i) Learning the centers and widths of radial functions in the hidden layer; (ii) Learning the weights connecting the middle layer to the output layer. Neurons in an RBF network are often chosen through trial and error. The learning method is initialized with a large number of neurons in the hidden layer, and then the number of neurons is gradually decreased. As the number of neurons in the network is reduced, the training method converges to the desired error, which may be determined by testing the network with its untrained inputs. Table 3 depicts the design of the RBF network and the results of the statistical tests (R^2, MSE, AAD, AARD, and run time) that compare the actual and predicted precipitation efficiency for the training and test data, regardless of the number of neurons and spread factor.

The spread factor, or variance, and the optimum number of neurons are the parameters gained by trial and error in RBF training. 85% and 15% of the total data were used for training and testing, respectively. Table 4 shows the optimal RBF network layout and the results of statistical tests (R^2, MSE, AAD, AARD, and run time) comparing actual and predicted precipitation efficiency for training and testing data. As a result, the effect of these two parameters on RBF network optimization can be observed.

Then, the RBF network with a spread factor of 0.1 and 100 neurons for the first half of 2020 and 2021 and a spread factor of 0.2 and 53 neurons for the second half of 2020 was trained for the inputs and target data of the precipitation process. The results of statistical tests between the experimental and predicted precipitation efficiency for training and testing data are shown in Fig. 2. It should be noted that the proposed ANN model's parameters, such as neuron values, weights, and biases, are accessible via an Excel file provided as supplementary material.

Application of SVM

The support vector machine uses a method called the kernel trick. This method contains functions that convert the input space of low dimension to the space of higher dimension. This transformation turns a non-separable problem into a separable problem. These are referred to as kernel functions. Kernel functions are more useful in nonlinear separation problems. These functions transform some extremely complex data and then find a process that can be used to separate the data based on some user-defined labels.

Typical methods for RBF network design are unnecessary for the SVM network. The numbers of radial functions utilized, as well as the centers of those functions, determine the feature space dimensions in this network automatically. The complexity of the problem is independent of the dimensions

Table 3 Design of the RBF network and the results of the statistical tests (R^2, MSE, AAD, AARD, and run time), regardless of the number of neurons and spread factor

MSE		R^2		AAD	AARD (%)	Run time (s)
Training	Testing	Training	Testing			
First half of 2020 and 2021						
1.28E-02	9.00E-01	0.81602	0.82168	0.71964639	1.58	78.307343
9.99E-03	7.48E-01	0.8598	0.85361	0.604370464	1.33	91.599782
4.76E-03	5.01E-01	0.93586	0.90534	0.3618397	0.80	177.363896
1.32E-02	9.28E-01	0.80923	0.8149	0.734610257	1.61	60.721893
1.13E-02	8.12E-01	0.83921	0.8401	0.6515561	1.43	237.56233
1.13E-02	8.50E-01	0.84038	0.83285	0.677785304	1.49	55.297084
1.40E-03	1.23E-01	0.9815	0.97749	0.174182402	0.38	227.775486
1.45E-03	1.12E-01	0.98085	0.97949	0.164721367	0.36	218.911562
6.96E-05	1.04E-03	0.99909	0.99983	0.009095159	0.02	267.17947
4.42E-04	4.03E-02	0.99422	0.99273	0.093847186	0.21	182.922093
Second half of 2020						
6.00E-03	2.58E-01	0.90661	0.94351	0.45218235	0.98	29.593959
1.79E-03	9.44E-02	0.97302	0.97911	0.223141514	0.49	25.084861
5.26E-04	1.36E-02	0.99217	0.99703	0.105062931	0.23	30.216103
5.45E-03	1.96E-01	0.9155	0.95674	0.430863619	0.94	16.281988
1.50E-03	5.95E-02	0.97749	0.98699	0.195918125	0.43	24.571349
3.06E-04	3.05E-02	0.99544	0.99328	0.08467281	0.18	31.612298
1.79E-03	8.03E-02	0.97307	0.98252	0.20491677	0.45	28.809183
3.92E-04	3.14E-03	0.99416	0.99947	0.033523579	0.08	25.587263
1.45E-03	8.91E-02	0.97823	0.98025	0.104208139	0.24	21.614068
2.27E-03	1.61E-01	0.96575	0.96412	0.265305461	0.58	18.131575

Table 4 Comparison of RBF network performance in different spread factor with 100 and 53 neurons and different neurons with spread 0.1 and 0.2 for first half and second half of the years, respectively

Spread factor (-)	MSE		R^2		AAD	AARD (%)	Run time (s)
	Training	Testing	Training	Testing			
	First half of 2020 and 2021						
0.01	6.96E-05	1.04E-03	0.99909	0.99983	0.009095159	0.02	322.200698
0.05	5.41E-05	1.12E-02	0.99929	0.99797	0.010002	0.02	347.56328
0.1	**5.90E-07**	**1.72E-05**	**0.99999**	**1**	**0.001854046**	**0.00**	**351.12005**
0.2	2.98E-06	2.60E-04	0.99996	0.99995	0.008140031	0.02	364.19825
0.3	2.28E-05	1.91E-03	0.9997	0.99966	0.0204921	0.04	362.259941
0.5	5.33E-05	3.02E-03	0.9993	0.99946	0.02828117	0.06	381.253021
1	3.47E-04	2.31E-02	0.99546	0.99584	0.062840608	0.14	405.556302
S.F	Second half of 2020						
0.01	5.54E-05	5.79E-27	0.99918	1	0.004698056	0.01	42.399453
0.05	7.48E-05	7.99E-03	0.99889	0.99834	0.009274511	0.02	42.782471
0.1	4.41E-06	4.99E-04	0.99993	0.99989	0.004198698	0.01	40.406151
0.2	**7.96E-08**	**1.31E-06**	**1**	**1**	**0.000908316**	**0.00**	**43.258999**
0.3	3.06E-06	2.10E-04	0.99995	0.99996	0.007459987	0.02	44.629417
0.5	1.20E-06	2.70E-05	0.99998	0.99999	0.005058456	0.01	44.124597
1	3.67E-06	6.69E-05	0.99995	0.99999	0.00840624	0.02	37.929936
Number of neurons	First half of 2020 and 2021						
60	5.61E-03	4.17E-01	0.92382	0.92253	0.401553246	0.89	112.095022
80	9.04E-04	8.00E-02	0.98813	0.98542	0.142837788	0.32	213.496612
90	1.52E-04	1.19E-02	0.99801	0.99784	0.048434869	0.11	244.445826
100	**5.90E-07**	**1.72E-05**	**0.99999**	**1**	**0.001854046**	**0.00**	**351.12005**
120	1.42E-09	1.51E-27	1	1	1.77323E-14	0.00	380.317205
N.N	Second half of 2020						
35	5.56E-03	3.68E-01	0.9138	0.9161	0.369411428	0.81	22.544942
45	1.16E-03	2.70E-02	0.98265	0.99424	0.138743242	0.31	32.722913
50	1.60E-04	4.42E-03	0.99763	0.99906	0.03660029	0.08	38.734814
52	2.04E-05	1.16E-03	0.9997	0.99976	0.011648592	0.03	40.911074
53	**7.96E-08**	**1.31E-06**	**1**	**1**	**0.000908316**	**0.00**	**43.258999**

Bold indicates the best performance of the network

of the data, and the number of hidden layers and weights are chosen automatically.

In order to model the system with the support vector machine network and change the network settings, such as selecting the kernel function, initial guessing of alpha values, data standardization, and kernel function scale, appropriate expressions can be applied to the network in the form of the input argument of the SVM function. As mentioned earlier, a support vector machine is used for an optimization problem. In addition, modeling software has the ability to find the optimal parameters for this network. First, the network is run with the following command to determine the optimal values. Table 5 shows the optimal values of the SVM network for two data sets.

So, the selected values can be used for optimal network design according to Table 5. The SVM network has different solvers. To find the correct solver with minimum MSE and maximum R^2, the SVM network was performed with different types of solvers.

As shown in Table 6, the ISDA solver has the best performance among the other solvers. Therefore, using this solver and the optimal values in Table 5, the network regression diagram for precipitation efficiency is shown in Fig. 3.

Therefore, the SVM network was first obtained with optimal values and trained with a Gaussian function. Then, regression was performed between network output data and target values (experimental), and the regression coefficient was 0.99914 and 1 for the two data sets, respectively.

Sensitivity analysis

As can be seen from the previous sections, the RBF network has the best performance compared to the SVM network. In addition, the RBF convergence was guaranteed due to the same results in each run as opposed to the MLP network and the lack of need to examine the network convergence, unlike the SVM network. The longer run time of the RBF network

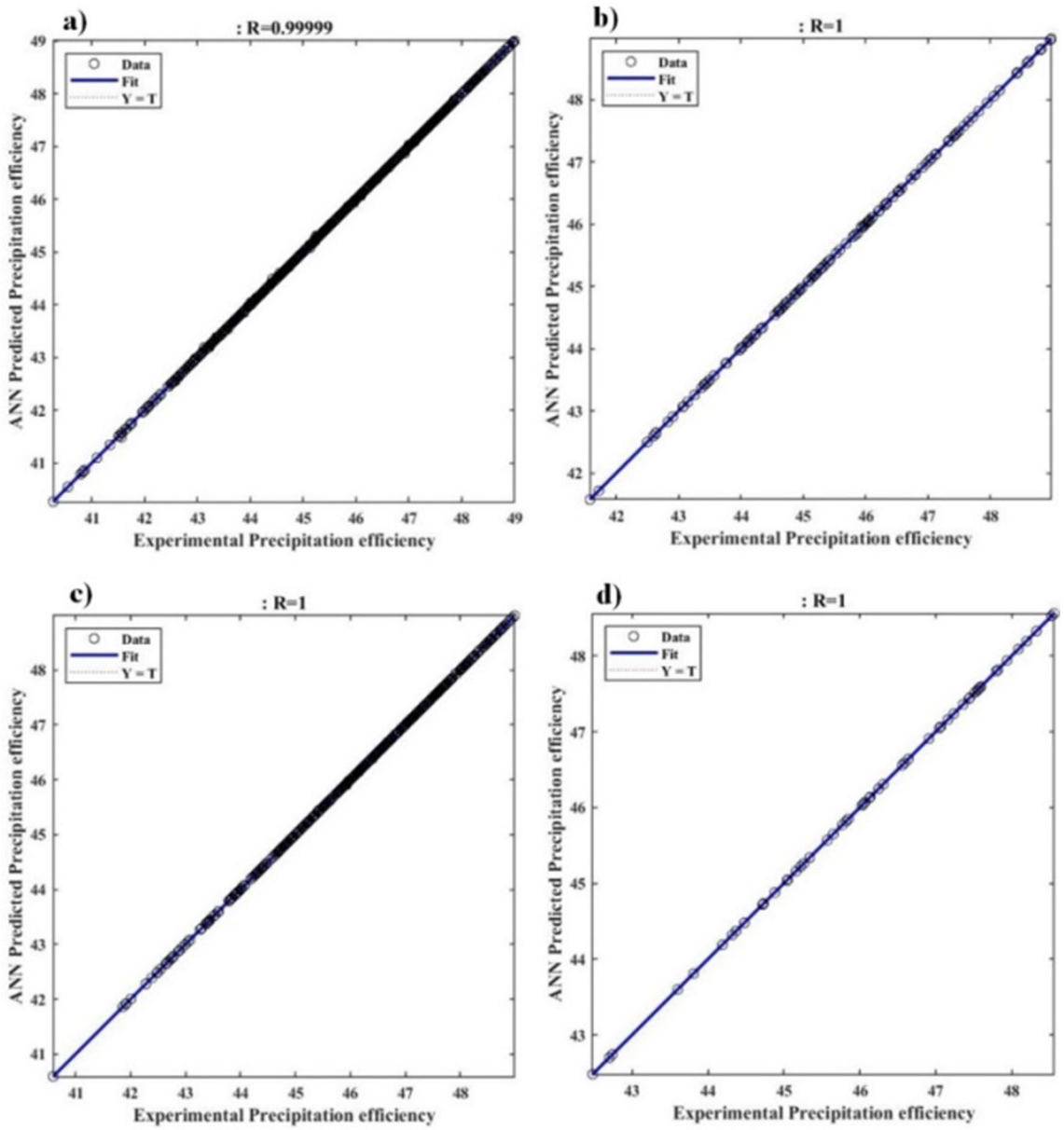


Fig. 2 RBF network regression performance for precipitation efficiency: **a** training and **b** testing data for the first half of 2020 and 2021; **c** training and **d** testing data for the second half of 2020

Table 5 Optimal values of SVM network parameters for two data sets

Fitrsvm function parameters	First half of 2020 and 2021		Second half of 2020	
	Optimal values (simulation)	Optimal values (selected)	Optimal values (simulation)	Optimal values (selected)
Box Constraint	2.2074	30	0.54479	5
Kernel scale	NaN	NaN	NaN	NaN
Epsilon	0.00050973	0.0084766	0.00032242	0.00032242
Kernel Function	Polynomial	Gaussian	Polynomial	Gaussian
Polynomial Order	3	NaN	3	NaN
Standardize	True	True	True	True

Table 6 Comparison of SVM network performance with different solvers for two data sets

Solver	Convergence	MSE		R^2		AAD	AARD (%)	Run time (s)
		Training	Testing	Training	Testing			
		First half of 2020 and 2021						
ISDA	**Yes**	**8.37E-05**	**5.35E-03**	**0.99914**	**0.99933**	**0.074752488**	**0.16**	**6.758433**
L1QP	Yes	8.37E-05	5.41E-03	0.99914	0.99932	0.074740053	0.16	6.437202
SMO	Yes	8.37E-05	5.41E-03	0.99914	0.99932	0.07475961	0.16	3.153944
Solver	Convergence	Second half of 2020						
ISDA	**Yes**	**1.20E-07**	**7.65E-06**	**1**	**1**	**0.002797855**	**0.01**	**2.04148**
L1QP	Yes	1.91E-07	7.34E-06	1	1	0.002806763	0.01	2.508422
SMO	Yes	1.93E-07	7.54E-06	1	1	0.002840346	0.01	1.811466

Bold indicates the best performance of the network

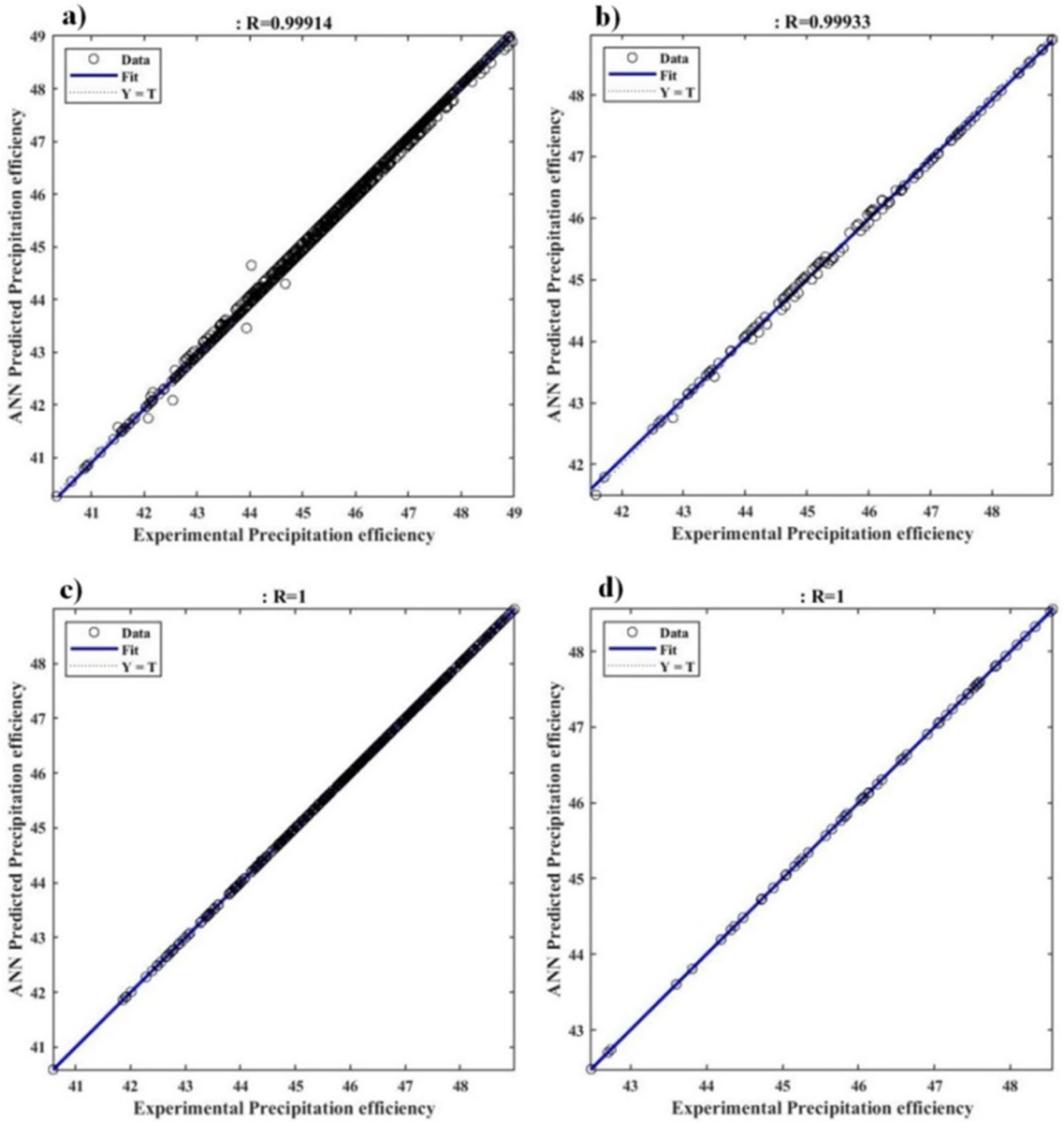


Fig. 3 SVM network regression performance for precipitation efficiency **a** training and **b** testing data for the first half of 2020 and 2021, **c** training and **d** testing data for the second half of 2020 with the ISDA solver

than the SVM network was ignored, and it was selected as the optimal network. Finally, this network was used for sensitivity analysis.

Because ANN is a black-box model, it cannot provide evidence about what is going on during the process. In most circumstances, determining which parameter has the highest impact on process performance is required. A well-trained network can analyze the relevancy importance or factor (r) of each input parameter. As the relevancy factor increases, the effect of input parameters increases. The relevance factor is calculated from Eq. (15) (Heidari et al. 2016; Mhurchú and Foley 2006):

$$\mathrm{r} = \frac{\sum_{i=1}^{N}(a_{z,i} - a_{z,\mathrm{average}})(Y_i - Y_{\mathrm{average}})}{\sqrt{\sum_{i=1}^{N}(a_{z,i} - a_{z,\mathrm{average}})^2 \times \sum_{i=1}^{N}(Y_i - Y_{\mathrm{average}})^2}} \quad (15)$$

where a and Y are the input and output parameters, respectively.

This ratio provides information about a process that is typically obtained through white-box modeling and is useful for optimization and control. In order to maintain high requirements in modeling, it is important for the modeler to provide a confidence assessment. To do so, it is necessary to first conduct an analysis of the level of uncertainty associated with the outcomes of the model (uncertainty analysis) and then determines the relative importance of the various inputs in generating that uncertainty. The second of these problems is tackled by this technique called sensitivity analysis (uncertainty analysis is typically a prerequisite), which ranks the inputs' intensities and relevance in determining the output's variability.

Figure 4 demonstrates that the solid content, concentration of Na_2O_c, and concentration of Al_2O_3 are all in the same range (0.130–0.145). This implies that the influence of these parameters on precipitation efficiency is very similar. Ambient temperature was predicted to have a greater effect on precipitation efficiency than other inputs because of the wide range of variation in ambient temperature data available from industry. As shown in Fig. 4, the relative factor of ambient temperature has the highest value (0.18). It follows that the relative factor is profoundly affected by the magnitude of the changes that can be made to a quantity. This is also why factors like tank temperature and residence time have a relatively minimal impact.

Model generalization

After choosing the best network with the optimal number of neurons and spread factor, it can be used to predict precipitation efficiency for a variety of inputs within the training data domain. In 3D plots, the precipitation efficiency is plotted against operating parameters. The generalization performances of the RBF network, as shown in Fig. 5, illustrate no oscillations, confirming the network's appropriate prediction capability. Figures 5a–h, represent the effects of various input operating parameters on precipitation efficiency.

The data gathered for this research originates from Jajarm's alumina hydroxide precipitation unit and analyzes the parameters that influence the efficiency of precipitation-based production. Production efficiency is directly affected by the parameters of temperature, solid content, residence time, aluminum oxide concentration, and sodium oxide concentration; however, some of these variables have an overlapping influence on each other, as is seen in the presented results.

According to the 3D plots, the ideal temperature range for the precipitation process is between 55 °C and 65 °C, where the precipitation effectiveness is suitable.

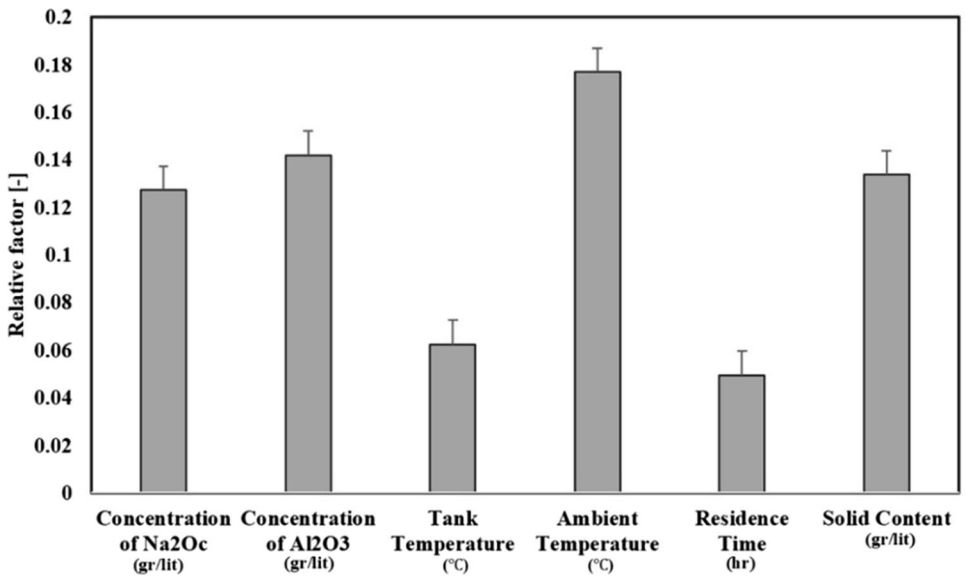
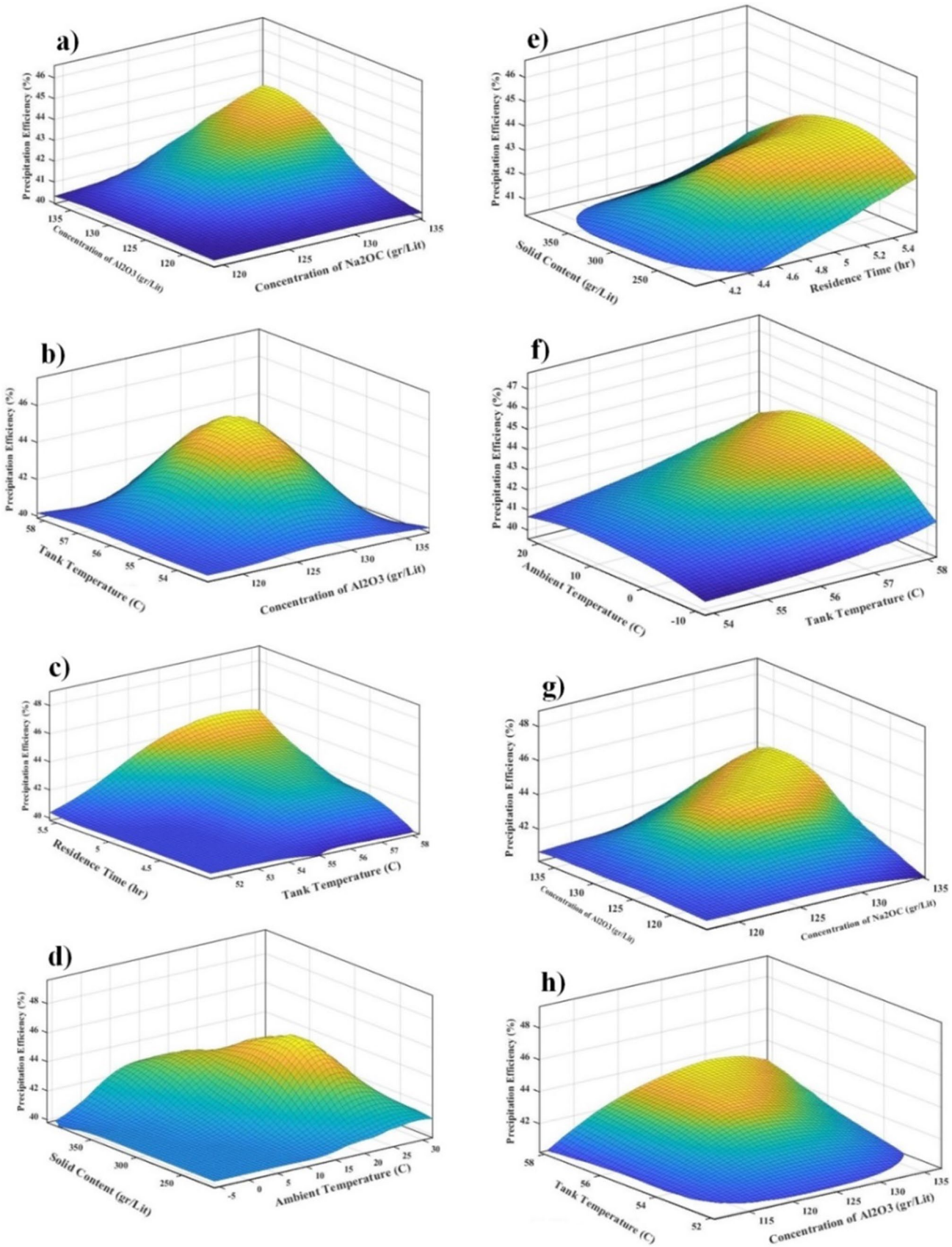


Fig. 4 Relative factor of operating parameters on precipitation efficiency. The error bars represent the standard deviation for 100 runs


a)
Precipitation Efficiency (%)
Concentration of Al2O3 (gr/Lit)
Concentration of Na2OC (gr/Lit)
b)
Precipitation Efficiency (%)
Tank Temperature (C)
Concentration of Al2O3 (gr/Lit)
c)
Precipitation Efficiency (%)
Residence Time (hr)
Tank Temperature (C)
d)
Precipitation Efficiency (%)
Solid Content (gr/Lit)
Ambient Temperature (C)
e)
Precipitation Efficiency (%)
Solid Content (gr/Lit)
Residence Time (hr)
f)
Precipitation Efficiency (%)
Ambient Temperature (C)
Tank Temperature (C)
g)
Precipitation Efficiency (%)
Concentration of Al2O3 (gr/Lit)
Concentration of Na2OC (gr/Lit)
h)
Precipitation Efficiency (%)
Tank Temperature (C)
Concentration of Al2O3 (gr/Lit)

◂**Fig. 5** Model generalization of optimal network, effect of **a** concentration of Al2O3 and Na2Oc; **b** tank temperature and concentration of Al2O3; **c** residence time and tank temperature; **d** solid content and ambient temperature on precipitation efficiency for the first half of 2020 and 2021, while other factor are kept constant at 129.6 gr/lit Na2OC concentration, 130.10 gr/lit Al2O3 concentration, 57 °C tank temperature, 13 °C ambient temperature, 5.50 h residence time and 379 gr/lit solid content. effect of **e** solid content and residence time; **f** ambient and tank temperatures; **g** concentration of Al2O3 and Na2Oc; **h** tank temperature and concentration of Al2O3 on precipitation efficiency for the second half of 2020, while other factors are kept constant at 124 gr/lit Na2OC concentration, 121.4 gr/lit Al2O3 concentration, 56.3 °C tank temperature, 17.5 °C ambient temperature, 5.42 h residence time and 321 gr/lit solid content

In accordance with Eq. (16), an increase in temperature increases the growth rate, and in accordance with Eq. (17)), it decreases super-saturation. A temperature range has been proposed to improve efficiency based on the interaction between these two variables (Ying Zhang 2009).

$$-\left(\frac{1}{S}\right) * \left(\frac{d\sigma}{dt}\right) = k\sigma^{n} \tag{16}$$

$$\sigma = \frac{A - A_{\mathrm{eq}}}{A_{\mathrm{eq}}} \tag{17a}$$

$$A_{\mathrm{eq}} = C * \exp\left(6.2106 - \frac{2486.7}{T + 273.15} + \frac{1.0875}{T + 273.15 \times C}\right) \tag{17b}$$

where σ is the degree of super-saturation, k is the growth rate constant, n is the degree of reaction, S is the total surface area of the particles, T is the temperature (°C), C is the caustic concentration (gr/lit), and A_{eq} is the equilibrium solubility of aluminum (gr/lit).

Due to the inclusion of Al_2O_3 in the calculation of alumina module, α (Eq. (2)), the concentration of Al_2O_3 has a direct effect on precipitation efficiency, as shown by the results of the Al_2O_3 concentration level. Increasing the amount of Al_2O_3 causes α to decrease at the inlet and increase at the outlet of the tank (super-saturation to increase), which in turn improves precipitation efficiency. The direct effect of Na_2O_c on α follows the same pattern as the prior experiment (concentration of Al_2O_3). When Na_2O_c is reduced, α is decreased at the tank's entrance (super-saturation is raised), and precipitation efficiency improves.

The precipitation process is based on mass transfer from the solution phase to the solid phase, and as stated earlier, this process consists of the time-consuming stages of nucleation, growth, and agglomeration. It takes between 4 and 5.5 h, depending on the temperature, the degree of super-saturation, and the concentration. The most efficient precipitation rates are achieved during this timeframe.

According to the liquor input to the precipitation tanks and the mass transfer of Al_2O_3 from the solution phase to the solid phase during the precipitation process, the solid content in the solution is low, the degree of super-saturation of the liquor solution is high, and as a result, precipitation efficiency is improved.

The super-saturation rate and precipitation efficiency increase as the Na_2O_c concentration rises, and the Al_2O_3 concentration falls at the reactor's inlet. Similarly, as the temperature difference increases, less Al_2O_3 remains in the liquor solution, leading to faster mass transfer. As a result, as tank and ambient temperatures rise, so does precipitation efficiency. Mass transfer and precipitation efficiency improve with a higher residence time, which gives more opportunity to the nucleation, growth, and agglomeration steps. As the solid content of the liquor solution decreases, more Al_2O_3 dissolves in the solution and consequently increases its separation rate. As a result, lowering the solid content improves precipitation efficiency.

Optimization of the precipitation efficiency

Precipitation efficiency must be maximized in the Bayer process. These input parameters, however, do not respond the same within the operating condition domain, as indicated by 3D plots in Fig. 5. Raising the tank temperature, for example, increases the precipitation efficiency, whereas increasing the ambient temperature decreases the precipitation efficiency. In the Bayer process, there are two methods for determining the optimal operating parameters for precipitation efficiency: (i) minimizing the concentration of Na_2O_c, and (ii) limiting the total consumed energy in the reactor.

The optimization technique for the aforementioned process in an industrial precipitation reactor is based on minimizing the concentration of Na_2O_c. According to the sensitivity analysis diagram, Na_2O_c has a significant effect on precipitation efficiency, and we can control this parameter in the Bayer process. The trained network, which was used to estimate precipitation efficiency in an industrial reactor, may be securely employed to predict the optimal values of input parameters in the Bayer process to get the best outcomes. When the network's input is in the range of the minimum to maximum operating conditions, we can trust the anticipated output data by employing the trained network.

For this purpose, a new data set containing various network input parameters was produced as the main and available parameters that affect reactor operating conditions (concentration of Na_2O_c, concentration of Al_2O_3, tank and ambient temperatures, residence time, and solid content). All of the data were within the range of the reactor's actual working conditions. Five of the six parameters in the generated data set were held constant, while the sixth parameter was adjusted throughout a variety of actual

operating conditions. This was done for each of the six parameters, and a total of 1800 data points were collected. The network was then employed to estimate precipitation efficiency using this data set as input.

The 1800 data points used in the optimization of precipitation efficiency are proposed points. These 1800 data points fall between the minimum and maximum of industrial data. Since industrial data are reliable, selecting 1800 points at random for each input is also trustable. Finally, it is possible to assert that the 1800 data points between the minimum and maximum ranges of reliable industrial data were generated at random. In addition, the machine learning-based model has made accurate predictions for this range of industrial data. Therefore, 1800 data points or any number of data points selected from industrial data within the interval of minimum and maximum inputs are reliable.

Precipitation efficiency is defined as the concentration of Na_2O_C and Al_2O_3 in the entered feed, as previously stated. This quantity must not exceed a set limit and must be adjusted in accordance with the required specification. Therefore, based on the available data of the Jajarm Refinery Company, the minimum and maximum calculated precipitation efficiency should be 40% and 49%, respectively. The best parameters in the Bayer process were explored within the data set with the lowest permissible value (40%) of precipitation efficiency to minimize total production cost. As given in Table 7, the process's optimum conditions were chosen as follows: concentration of $Na_2O_C = 107.3$ g/lit; concentration of $Al_2O_3 = 128.34$ g/lit; tank temperature = 55.93 °C; ambient temperature = −1.5 °C; residence time = 5.22 h; and solid content = 254.22 g/lit for the first half of 2020 and 2021; and concentration of $Na_2O_C = 107.9$ g/lit; concentration of $Al_2O_3 = 128.81$ g/lit; tank temperature = 56.31 °C; ambient temperature = −3 °C; residence time = 5.13 h; and solid content = 275.43 g/lit for the second half of 2020.

Since the sensitivity analysis demonstrated that the ambient temperature had a significant impact on the precipitation efficiency, it was considered one of the network inputs, despite the fact that the controller had no control over it. Two cases with constant ambient temperature are investigated to provide a set of best operating conditions for various fixed ambient temperature values: 1. the case of the first half of 2020 and 2021 with the average ambient temperature, 2. the case of the second half of 2020 with the average ambient temperature. Then, the RBF network is used to determine the optimal available values for the remaining parameters based on these two cases. Table 8 displays the precipitation efficiency under optimum operating conditions and a constant ambient temperature for the first half of 2020 and 2021 and the second half of 2020.

Conclusion

In this paper, the efficiency of precipitation as a function of operating parameters was explored and modeled using machine learning algorithms. The model input parameters were concentration of Na_2O_c, concentration of Al_2O_3, tank temperature, ambient temperature, residence time, and solid content. The results show that the machine learning approaches may be employed safely in the Bayer process for modeling precipitation efficiency. The findings of the modeling revealed that the estimated outcomes and industrial data for the training and testing data sets are in good agreement. The general results of the study are summarized below:

- In order to find an appropriate algorithm and predict the precipitation efficiency, the structures of RBF and SVM networks were examined.
- Based on the trial-and-error algorithm, the RBF network with a spread factor of 0.1 and 0.2 and a number of neu-

Table 7 Precipitation efficiency under various operational conditions with a precipitation efficiency of 40% and the minimum Na_2O_c concentration (the optimum point shown in bold)

Concentration of Na_2O_c (gr/lit)	Concentration of Al_2O_3 (gr/lit)	Tank Temperature (°C)	Ambient Temperature (°C)	Residence Time (hr)	Solid Content (gr/lit)	Precipitation efficiency (%)
First half of the 2020 and 2021						
107.3	110.06	51.25	−4.5	4.19	227.21	41.25
107.3	114.61	52.44	−1.5	4.43	254.22	43.75
107.3	119.98	54.03	3.3	4.81	294.35	45.11
107.3	**128.34**	**55.93**	**−1.5**	**5.22**	**254.22**	**46.99**
107.3	135.53	57.83	5	5.49	389.11	48.15
Second half of the 2020						
107.9	112.85	51.83	−10.5	4.23	234.03	40.26
107.9	120.06	54.15	−6	4.65	275.43	42.72
107.9	126.61	55.74	−3	5.00	323.92	43.61
107.9	**128.81**	**56.31**	**−3**	**5.13**	**275.43**	**46.62**
107.9	135.43	57.90	0	5.55	385.90	47.20

Table 8 Precipitation efficiency under optimum operating conditions and a constant ambient temperature for the first half of 2020 and 2021 and the second half of 2020 (the optimum point shown in bold)

Concentration of Na_2O_c (gr/lit)	Concentration of Al_2O_3 (gr/lit)	Tank temperature (°C)	Ambient temperature (°C)	Residence Time (hr)	Solid content (gr/lit)	Precipitation efficiency (%)
First half of the 2020 and 2021						
119.57	124.05	52.61	20.24	4.25	334.72	42.56
114.26	129.84	53.69	20.24	4.94	296.44	45.98
111.04	**129.93**	**56.22**	**20.24**	**5.38**	**279.80**	**47.12**
108.69	132.68	53.69	20.24	4.25	367.38	48.17
Second half of the 2020						
115.51	128.75	52.30	6.57	4.98	348	41.59
111.71	125.47	54.98	6.57	5.29	299.28	44.92
110.84	**128.75**	**57.90**	**6.57**	**5.55**	**258.96**	**46.46**
108.98	134.95	56.35	6.57	5.13	323.96	47.33

rons of 100 and 53 for the first half of 2020 and 2021 and the second half of 2020, respectively, was chosen as the best RBF structure to predict precipitation efficiency within the training data domain.

- The average relative error in predicting the precipitation efficiency of training points was less than 0.001%.
- ANN's model generalization within the domain of training data also exhibits superior prediction ability.
- The evaluated model can be used to determine the optimal operating conditions for the Bayer process based on the minimum concentration of Na_2O_c by generating new datasets covering a broad range of possible operating conditions.
- The comparison findings (R^2, MSE, AAD, AARD, and run time) demonstrated the network's capacity to predict precipitation efficiency accurately over a wide range of operating parameters.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11696-022-02642-x.

Acknowledgements The authors would like to express their gratitude to the Iran Alumina Company (IAC) for providing the operational data.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Andras PJNPL (2002) The equivalence of support vector machine and regularization neural networks. Neural Process Lett 15(2):97–104

Bahrami M, Nattaghi E, Movahedirad S, Ranjbarian S, Farhadi F (2012a) The agglomeration kinetics of aluminum hydroxide in Bayer process. Powder Technol 224:351–355

Bahrami M, Nattaghi E et al (2012) The agglomeration kinetics of aluminum hydroxide in Bayer process. Powder Technol 224:351–355

Barata PA, Serrano ML (1996) Salting-out precipitation of potassium dihydrogen phosphate (KDP) II. Influence of agitation intensity. J Cryst Growth 163:426–433

Baş D, Boyacı İH (2007) Modeling and optimization II: Comparison of estimation capabilities of response surface methodology with artificial neural networks in a biochemical reaction. J Food Eng 78(3):846–854

Bearne G, Dupuis M, Tarcy G (eds) (2016) Essential Readings in Light Metals: Volume 2 Aluminum Reduction Technology. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-48156-2

Bl LÜ et al (2010) Effects of Na4EDTA and EDTA on seeded precipitation of sodium aluminate solution. Transact Nonferrous Met Soc China 20:s37–s41

Byvatov E, Fechner U, Sadowski J, Schneider G (2003) Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. Inf Comput Sci 43(6):1882–1889

Chelgani SC, Jorjani EJH (2009) Artificial neural network prediction of Al2O3 leaching recovery in the Bayer process—Jajarm alumina plant (Iran). Hydrometallurgy 97(1–2):105–110

Dorin R et al (1988) The electrodeposition of gallium from synthetic Bayer-process liquors. J Appl Electrochem 18(1):134–141

Đurić I et al (2012) Artificial neural network prediction of the aluminum extraction from bauxite in the Bayer process. J Serb Chem Soc 77(9):1259–1271

Ganguly S (2003) Prediction of VLE data using radial basis function network. Comput Chem Eng 27(10):1445–1454

Ghaemi A et al (2018) Processing, Comparing the capability of various models for predicting of the Bayer process parameters. J Adv Mater Process 6(1):71–86

Heidari E, Sobati MA, Movahedirad S (2016) Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN). Chemom Intell Lab Syst 155:73–85

Hind AR, Bhargava SK, Grocott SCJC (1999) Physicochemical, s.A., aspects, e., The surface chemistry of Bayer process solids: a review. 146(1–3): 359–374

Huang Wq et al (2019) Effect of lithium ion on seed precipitation from sodium aluminate solution. Transact Nonferrous Met Soc China 29(6):1323–1331

Hui-bin Yang et al. (2020) Characteristics of Sodium Oxalate Precipitates from the Bayer Precipitation Process. In: TRAVAUX 49, Proceedings of the 38th International ICSOBA Conference, China

Ilievski D, Livk IJCES (2006) An agglomeration efficiency model for gibbsite precipitation in a turbulently stirred vessel. Chem Eng Sci 61(6):2010–2022

Zeng J, Yin Z, Chen Q (2007) Intensification of precipitation of gibbsite from seeded caustic sodium aluminate liquor by seed activation and addition of crown ether. Hydrometallurgy 89(1–2):107–116

JingTao Y, Tan CL (2001) Guidelines for financial forecasting with neural networks, In: International Conference of Neural Information Processing, Shanghai, China

Liu G, Wu G, Chen W, Li X et al (2018) Increasing precipitation rate from sodium aluminate solution by adding active seed and ammonia. Hydrometallurgy 176:253–259

Liu Z et al (2020) Digestion behavior and removal of sulfur in high-sulfur bauxite during bayer process. Minerals Engineering 149:106237

Mahmoudian M, Ghaemi A, Hashemabadi H (2016) Prediction of red mud bound-soda losses in bayer process using neural networks. Iran J Chem Eng Spring 13:46–56

Mhurchú JN, Foley G (2006) Dead-end filtration of yeast suspensions: Correlating specific resistance and flux data using artificial neural networks. J Membr Sci 281(1–2):325–333

Misra C (2016) Agitation effects in precipitation. In: Donaldson D, Raahauge BE (eds) Essential readings in light metals. Springer International Publishing, Cham, pp 541–549. https://doi.org/10.1007/978-3-319-48176-0_75

Muhr H et al (1997) A rapid method for the determination of growth rate kinetic constants: application to the precipitation of aluminum trihydroxide. Ind Eng Chem Res 36(3):675–681

Ostap SJCMQ (1986) Control of silica in the Bayer process used for alumina production. Can Metall Q 25(2):101–106

Paspaliaris I, Panias D, Amanatidis A, Mordini J, Werner D, Panou G, Ballas D (1999a) Precipitation and calcination of monohydrate alumina from the Bayer process liquors. Eurothen 99:532–547

Paspaliaris I, Panias D, Amanatidis A, Mordini J, Werner D, Panou G, Ballas DJE (1999b) Precipitation and calcination of monohydrate alumina from the Bayer process liquors. 99:532–547

Paulaime AM, Seyssiecq I, Veesler SJPT (2003) The influence of organic additives on the crystallization and agglomeration of gibbsite. Powder Technol 130(1–3) 345–351

Rosenberg S (2017) Impurity removal in the bayer process. In: Travaux 46 proceedings of the 35th international ICSOBA conference, Hamburg, Germany, pp 175–196

Sahu NK, Sarangi CK, Tripathy BC, Bhattacharya IN, Satpathy BK (2014) Effect of urea on decomposition of sodium aluminate solution. J Taiwan Instit Chem Eng 45(3):815–822. https://doi.org/10.1016/j.jtice.2013.09.001

Sahu NK, Sarangi CK, Dash B, Tripathy BC, Satpathy BK, Meyrick D, Bhattacharya IN (2015) Role of hydrazine and hydrogen peroxide in aluminium hydroxide precipitation from sodium aluminate solution. Transact Nonferrous Met Soc China 25(2):615–621

Seecharran KR (2010) Bayer process chemistry. Alumina Plant, Guymine, Linden

Sidrak YLJI (2001) Dynamic simulation and control of the Bayer process. A review 40(4):1146–1156

Smeulders DE, Wilson MA et al. (2001) Insoluble organic compounds in the Bayer process. Ind Eng Chem Res 40(10): 2243–2251

Sonthalia R, Behara P, Kumaresan T et al. (2013) Review on alumina trihydrate precipitation mechanisms and effect of Bayer impurities on hydrate particle growth rate. Int J Miner 125: 137–148

Totten GE, Scott MacKenzie D (eds) (2003) Handbook of aluminum: alloy production and materials manufacturing. CRC Press. https://doi.org/10.1201/9780429223259

Totten GE, Scott MacKenzie D (eds) (2003) Handbook of aluminum: alloy production and materials manufacturing. CRC Press. https://doi.org/10.1201/9780429223259

Veesler S et al. (1994) General concepts of hydrargillite Al (OH) 3, agglomeration. J Cryst Growth 135(3–4): 505–512

Vogrin J et al. (2020) The anion effect on sodium aluminosilicates formed under Bayer process digestion conditions. Hydrometallurgy 192: 105236

Vt SE, Shin YC (1994) Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems 5(4): 594–603

Wellington M, Valcin F (2007) Impact of Bayer process liquor impurities on causticization. J Cryst Growth 46(15): 5094–5099

Yao J et al. (1999) Neural networks for technical analysis: a study on KLCI. IJTAF 2(02): 221–241

Yin J et al. (2006) Effects of monohydroxy-alcohol additives on the seeded agglomeration of sodium aluminate liquors. Light Met 3:153–157

Yu Hy et al. (2020) Effect of oxalate on seed precipitation of gibbsite from sodium aluminate solution. J Cent South Univ 27(3):772–779

Zeng, Js et al. (2008) Effect of tetracarbon additives on gibbsite precipitation from seeded sodium aluminate liquor. J Cent South Univ Technol 15(5): 622–626

Zhang Y, Zheng S, Du H, Xu H, Wang S, Zhang Y (2009) Improved precipitation of gibbsite from sodium aluminate solution by adding methanol. Hydrometallurgy 98(1–2):38–44

Zhang B, Pan X, Haiyan Y, Ganfeng T, Bi S (2018) Effect of organic impurity on seed precipitation in sodium aluminate solution. In: Martin O (ed) Light Metals 2018. Springer International Publishing, Cham, pp 41–47. https://doi.org/10.1007/978-3-319-72284-9_7

Zhang B et al. (2006) Influences of seed size and number on agglomeration in synthetic bayer liquors. J Cent South Univ Technol 13(5): 511–514

Zhang Y, Xu R, Tang H, Wang L et al. (2020) A review on approaches for hazardous organics removal from Bayer liquors 397: 122772

Zhou X, Yin J, Chen Y, Xia W, Xiang X, Yuan XJH (2018) Simultaneous removal of sulfur and iron by the seed precipitation of digestion solution for high-sulfur bauxite. Hydrometallurgy 181:7–15