



# Efficient machine learning model to predict fineness, in a vertical raw meal of Morocco cement plant

Fahd Belmajdoub, Souad Abderafi\*

MOSEM2PI, Mohammadia Engineering School, Mohammed V University in Rabat, Morocco



## ARTICLE INFO

**Keywords:**

Vertical raw mill  
Soft sensor  
Bayesian regularization neural network  
Support vector regression  
Bayesian optimization

## ABSTRACT

Soft sensor enables computing parameters that can be physically impossible to measure. This work aims to develop a soft sensor for raw meal fineness in a vertical roller mill of a cement plant. In previous research, some key indicators of the process were ignored. It is well known that vertical raw meal is feeded with heterogeneous material that impacts raw meal fineness, and therefore, can deteriorate prediction accuracy of the soft sensor. To deal with this constraint, one-year history of industrial data was exploited and processed by a robust multivariate outlier identification method, followed by process expert outlier verification. Then a particular attention was given to process and material quality parameters using a statistical method to test parameters significance on fineness based on data variance calculation. Eight process and four quality parameters were retained to develop robust nonlinear model using artificial neural network: Bayesian regularization and Levenberg Marquart, then compared with Support vector regression with different acquisition function and kernels to tune and select optimum model hyperparameters. The best model performance was obtained with the Bayesian regularization algorithm.

## 1. Introduction

Nowadays, raw meal fineness is still monitored every hour with manual sampling by laboratory staff. Operators in the control room adjust the process and do the monitoring in real time relying on manual raw meal fineness laboratory measures. At the same time, many factors impact significantly fineness variation between two sampling period such as material grindability, mill circulating load, mill gas flow coming from preheater tower, etc... Therefore, a great optimization potential is lost because of lack of information between fineness analyses, and remains in most of time to operator's experience. Raw meal fineness prediction model can be a solution for the operator in the control room to achieve a stable quality and a better mill performance. It is to highlight that it is difficult to describe and optimize the process with an adequate mathematical model. This is even more complex when the mill is fed with different material types or when there is heterogeneous physicochemical characteristics in materials to grind.

In the cement grinding and mining industry, product fineness prediction has been subject to various industrial applications. It is used as a first step before an online mill monitoring implementation and optimization of the specific electricity consumption [1–3]. Industrial

applications range from horizontal ball mills to vertical roller mills, the two most widely used grinding technologies in the cement industry. There are therefore several studies for online fineness prediction, especially for horizontal ball mills. Many of these studies are available for online estimation of fineness by Blaine specific surface area or sieve residue, which allows a continuous and reliable fineness estimation to ensure a better monitoring and control of the mill. Blaine fineness prediction is done continuously for 1-min cycle in a horizontal ball mill for cement grinding. According to Sankar [1], this led to a considerable increase in output and a reduction in the specific electricity consumption when coupled with a model predictive control for process optimization. A similar study consists of testing the effect of nine process parameters on the cement fineness using a virtual neural sensor in a cement horizontal ball mill [2]. Model predictive control is used afterwards for adjustment control of fineness and return material to optimize the process. Fineness prediction in ore industry has shown a 10% increase potential in horizontal ball mills throughput when it is connected to an expert system [4,5]. There are also studies that deal exclusively with cement fineness prediction without automating the monitoring of the mill. This led to an improvement of the mill control and monitoring by the operator in the control room. Blaine's fineness prediction with kernel principal component regression showed a great prediction

\* Corresponding author.

E-mail address: [abderafi@emi.ac.ma](mailto:abderafi@emi.ac.ma) (S. Abderafi).

## Nomenclature

VRM	Vertical Raw Meal
BPNN	Back Propagation Neural Network
DL	Deep Learning
BR	Bayesian Regularization
LM	Levenberg-Marquardt
SVM	Support Vector Machine
SVR	Support vector regression
PCA	Principal Component Analysis
AF	Acquisition Function
PI	Probability of Improvement
EI	Expected Improvement
GP-LCB	Gaussian Process Lower Confidence Bound
MSE	Mean square error
MAE	Mean Absolute Error
R <sup>2</sup>	Coefficient of determination
R	Correlation coefficient

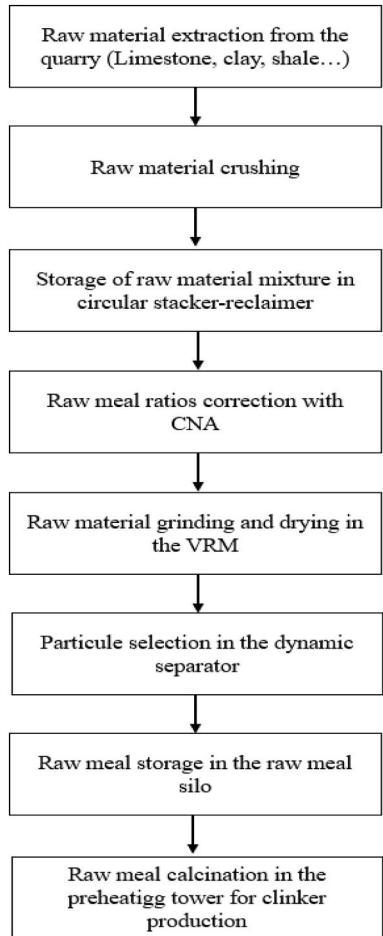


Fig. 1. Steps of raw meal process making.

accuracy when online monitoring in an industrial ball mill is carried out [6]. According to Stanišić et al. [7], two types of models have been developed using multilayer perceptron neural networks. It turned out that the neural model retrained using fineness laboratory data gives better long-term results than the model trained with just predicted fineness and does not deteriorate with time. Another study done by

Vogel et al. [8] reduced the computing time of two fineness control interval from 30 min to less than 30 s using a multi-layered perceptron network model. It is well known that Vertical Raw Meal (VRM) has a higher efficiency than horizontal ball mill [9]. However, fineness soft sensing studies in vertical roller mills are less common than horizontal ball mills in the literature. The authors Ajaya and Hare have successfully made a benchmark with different nonlinear models to predict Blaine's fineness in vertical cement mills [10,11]. Accurate model performance was obtained using Back Propagation Neural Network (BPNN) and Adaptive Neuro Fuzzy Inference System to predict cement Blaine in vertical mill [10,11]. A more recent study has shown that soft-sensing method combined with Distributed principal component analysis similarity-based Time series segmentation technique can accurately predict fineness in VRM [12]. An exhaustive feature selection was performed indeed using 26 input variables. Nevertheless, material quality characteristics in the mill input were not taken into consideration in those studies. Machine learning has been more common for sensor and signal applications, but some authors use Deep Learning (DL) models to predict cement and raw meal fineness [13–15]. The DL popularity remains to a better understanding of domain and feature engineering issues, not to forget the amount of data that has a huge impact on DL model performance. In this study, neither the above-mentioned reasons seem to be interesting while small data samples are used to train the soft sensor model, and an appropriate feature engineering approach will be addressed with an appropriate and robust machine learning model.

Bayesian regularized artificial neural networks are more robust than standard back-propagation networks and can reduce or eliminate the need for lengthy cross-validation [16]. Bayesian Regularization (BR) has been subject of many successful studies. High accuracy and excellent generalization were obtained using BR neural network model for Permeability prediction of multi-stage tight gas sandstones [17]. Another comparison study for predicting Oil-gas Drilling Cost using BR algorithm with Levenberg-Marquardt (LM), Momentum Back Propagation, and Variable Learning Rate Back Propagation models, showed that BR has better comprehensive performances in terms of prediction precision, convergence rate and generalization ability [18]. The authors [19] presented BR neural networks to render shape optimization problems for fluid flow processes, with a considerable improvement over conventional neural network. BR has first been presented by David MacKay [20] and has been applied to LM network training by Foresee and Hagan [21]. According to Ref. [19], the power of this method has not yet been explored by applications in context of engineering implementations. In conventional neural network training, an optimal set of weights is chosen by minimizing an error function [22–24]. But the BR involves a probability distribution of neural network weights, and minimizes the over-fitting problem by taking into account the goodness-of-fit as well as the network architecture [24]. According to the authors, the power of BR has not yet been explored soft sensor applications or grinding circuit field. In this work, a comparative study of BR algorithm with LM and Support vector regression (SVR) algorithms will be performed to predict raw meal fineness in the VRM.

While there is few research work on raw meal fineness soft sensing in VRM established [12], the main contribution in this current work is to propose an original soft sensor study based on robust machine learning models, especially the BR algorithm that outperforms other back-propagation algorithms in many successful studies, but never adopted in cement soft sensor studies. On the other hand, no fineness soft sensing study in grinding circuits considers material analysis in the mill feed, therefore, a particular attention will be addressed to the right choice of parameters that impact raw meal fineness, including material quality characteristics, to build an accurate and reliable long-term model. This will be carried out using prior knowledge of the process, especially right choice of process variables that impact fineness, and checked out with an appropriate statistical method in order to test parameters significance on fineness based on data variance calculation.

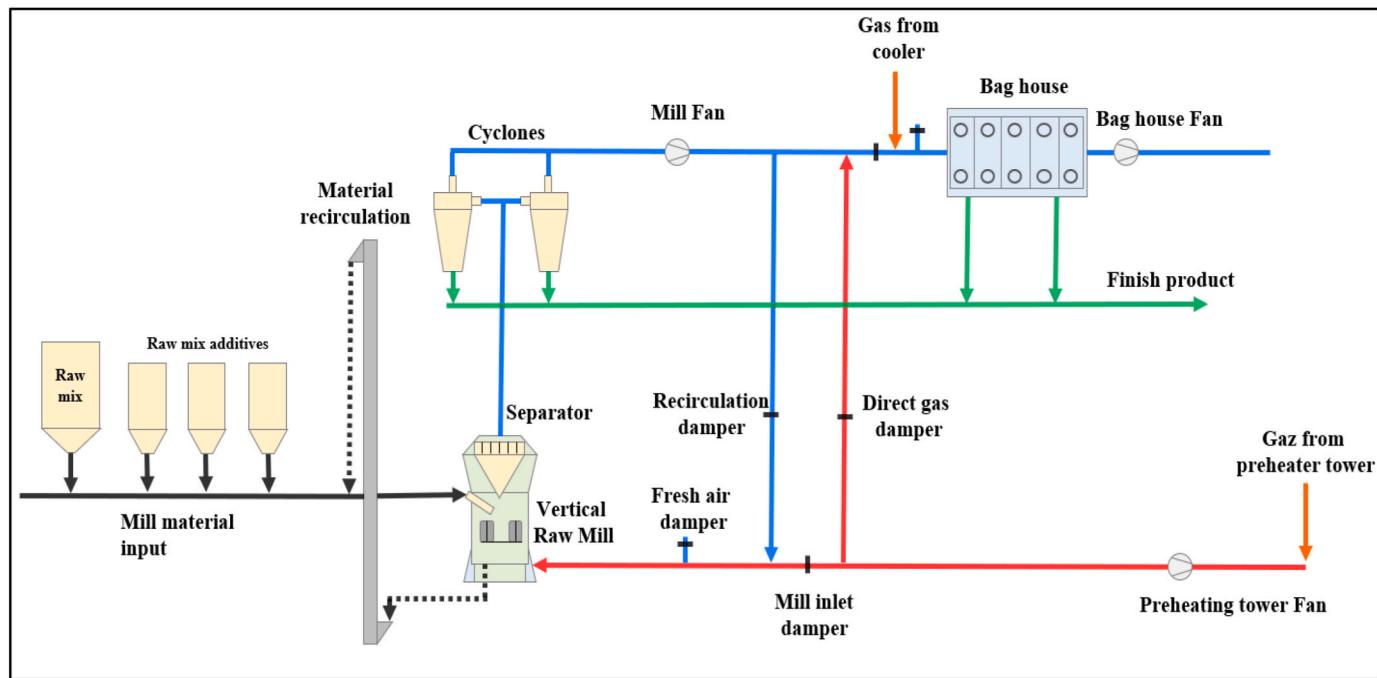


Fig. 2. Raw meal workshop schemes.

## 2. Materials and methods

### 2.1. Process description

Cement is produced from different materials added to enhance its properties, as it plays a vital and important role due to its cementitious and adhesive characteristics [25]. Raw material is extracted from quarries that are usually close to the cement plant. This raw material is composed mainly of limestone, as well as other assets as clay, shale, sand, iron ore, pozzolana and pyrrhotite. The crusher is fed with specific raw material proportions calculated from material's mineral composition by the laboratory. Raw meal ratios are introduced to simplify the control of the chemical composition during raw meal grinding and clinker production. Most important ratios are lime saturation factor, silica ratio and alumina ratio [26]. These ratios are controlled by an online neutron analyzer for online quality monitoring before being stored in a circular stacker-reclaimer and fed into the vertical mill. The different steps of the raw meal process manufacturing are described in Fig. 1. The process of grinding the raw mixture and the additives is achieved by a pressure exerted by the rollers on the grinding table. A fan draws gas flow from the preheating tower to feed the VRM as shown in Fig. 2. Then, a second fan located downstream of the mill draws grinded material to the dynamic separator located above the mill. Coarse particles are rejected to the center of the grinding table, and the fines are transported by the flow gas to a cyclone system and then to the process filter to be stored in the raw meal silo. Raw meal fineness is monitored by a defined mesh sieve residue instead of Blaine specific surface area fineness. The sieve residue at 90 µm mesh is commonly used in cements plants for monitoring the fineness. Blaine is highly correlated with limestone rate indeed, and consequently gives a fake Blaine [27]. Therefore, monitoring with Blaine to measure raw meal fineness will lead to biased results. For this reason, the output parameter to predict in our soft sensing study will be raw mill fineness at 90 µm mesh sieve residue.

Raw meal composition and raw mill fineness are key parameters for a stable kiln monitoring. Raw meal composition is monitored in real time by an online neutron analyzer, but raw meal fineness can't be available in real time because of manual measurements that should take 1 h in

average between two measurements. Therefore, fineness soft sensor development requires a relevant choice of industrial data.

### 2.2. Experimental data and their treatment

In order to develop a reliable model that takes into account different process configurations and material quality variations, one-year data processing history was collected from a Moroccan cement plant. Material's quality data and fineness are collected from the manual reporting of the cement plant. It is also to highlight that material analysis for the raw mill are monitored every first shift of every day. Therefore, it was mandatory to use daily data for process and material characteristics parameters to build the predictive model. A total of 360 daily average data was considered for model development. An adaptive screening method was used for the outlier identification and elimination. Dimension reduction technique will be used afterwards to select the appropriate parameters using Principal Component Analysis (PCA). The different methods followed are described below.

#### 2.2.1. Multivariate outlier detection method

Outliers can be defined in a simple way as observations that differ significantly from a normal data population. Most of time, they are measurement errors by sensors, defaults in transmission devices, sampling error or unusual disturbances. These data have an undesirable effect on model accuracy. Osborne and Overbay [28] summarize the impact of data contaminated by outliers on statistical analysis, especially by increasing the variance of the error and decreasing the statistical test capacity. There are several methods for detecting univariate outliers in a given dataset. The most common are 3sigma edit rule and Jolliffe parameters and analysis of linear regression [29]. Many other methods are described by Chiang et al. [30], Pearson [31] and Songwon Seo [32]. When multivariate outliers could not be detected, or normal values may be considered as outliers, known also as the masking and swamping effects [33], the outlier identification process won't be effective in practice. Many studies on soft sensing go briefly through this data preselection section and don't give the time and energy required. A recent study by Pani and Mohanta [34] highlights the importance of data screening using a multivariate method in a dynamic and non-linear

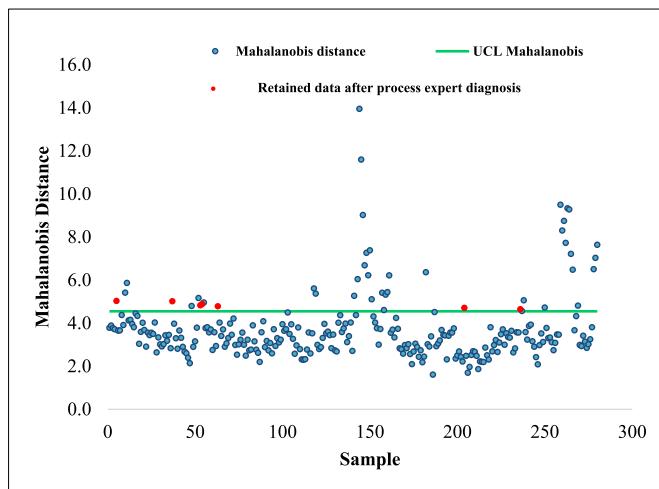


Fig. 3. Mahalanobis distance and upper control limit of row data.

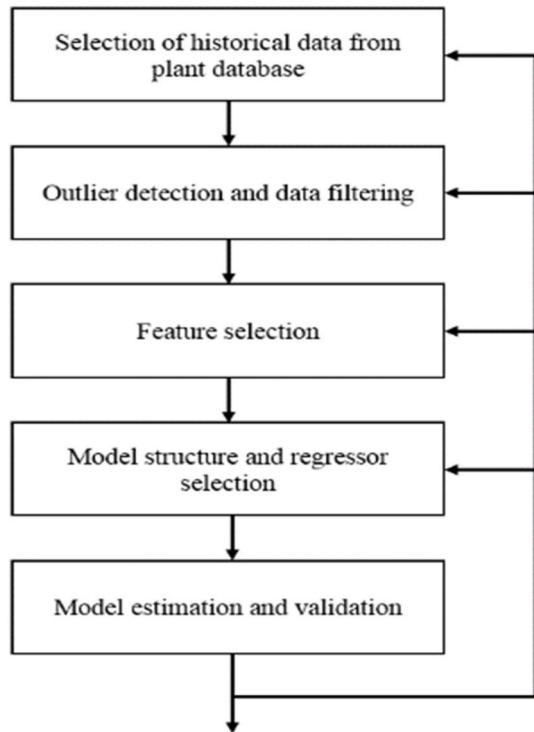


Fig. 4. Steps for soft sensor model design.

industrial process with several correlated variables. The use of multivariate techniques gives better model performance than univariate methods in this type of industrial application. Another study from Zeng and Gao [35] in blast furnace slag manufacturing process confirms that univariate methods are not adapted when industrial variables are highly correlated. There are several methods for multivariate data analysis such as Resampling by Half-Means [36] and Smallest Half Volume [37], robust distance and chi-square plot [37,38] and other methods [39]. In this study, three steps of data outliers will be processed:

- Visual data elimination corresponding to mill startup and shutdown periods.
- Robust multivariate method of Mahalanobis distance for data outlier detection [40].

- And compliance verification of software results by a process expert from the cement plant.

The first step is easy to operate while mill startup and shutdown periods are essentially characterized by a low mill throughput performance. The mill throughput minimum value can be obtained from mill process operators in the control room and proceed with outliers with a visual identification. In these transient regimes of mill, fineness samples are not representative and should be removed. Therefore, from the 360-initial data collected, only 280 data of mill steady-state monitoring were retained.

Subsequently, the second step is a robust multivariate outlier identification method based on Mahalanobis distance calculation from the remaining data using JMP software from SAS [41]. The Mahalanobis distance calculates the relative distance to the centroid that represents the multivariate mean. The distance of Mahalanobis between two points is defined according to Varmuza and Filzmoser [42] as follows:

$$M_i = \sqrt{(Y_i - \bar{Y}) \cdot S^{-1} \cdot (Y_i - \bar{Y})} \quad (1)$$

where,  $Y_i$  stands for the  $i$ th line of data;  $\bar{Y}$ , the average line, and  $S$  the estimate of the covariance matrix of data.

Data is considered outlier if its Mahalanobis distance is greater than the upper limit control  $UCL_{Mahalanobis}$  defined according to Mason and Young [43] as follows:

$$UCL_{Mahalanobis} = \sqrt{\frac{(n-1)^2}{n} \times \beta_{\left[1-\alpha_{\frac{p(n-p-1)}{2}}\right]}} \quad (2)$$

where:  $n$ , the number of observations;  $p$ , the number of variables and  $\alpha$ , is the significance level.

$\beta_{\left[1-\alpha_{\frac{p(n-p-1)}{2}}\right]} = (1-\alpha)$  th quantile of  $\beta_{\left[\frac{p(n-p-1)}{2}\right]}$  distribution. For  $\alpha = 0.05$ , the number of outliers is 44 among 280 data corresponding to an upper limit control  $UCL_{Mahalanobis} = 4.55$  as illustrated in Fig. 3. The results of these outliers have been carefully analyzed by a process expert from the cement plant. Indeed, if an outlier carrying useful information is eliminated during this preliminary phase of developing the soft sensor, the iteration of the prediction procedure illustrated in Fig. 4 will not be reliable, resulting in a waste of time and resources [29]. Therefore, many authors insist on the importance of re-checking the software results by experts in the industrial sector. Fig. 3 shows the outliers retained after the plant expert inspection. The majority of outlier results obtained using the Mahalanobis distance seem relevant. The analysis shows that outliers correspond either to a VRM vibration which is a consequence of material's bed layer variation on the grinding table, or mill overfilling signs represented mainly by an increase in mill pressure loss. During these instability phases, there is a high chance to obtain a non-representative sample of raw meal, and therefore, false fineness values. Actual fineness values that correspond to outlier process and quality parameters within mill vibration or overfilling periods confirm this assumption with outlier fineness values, even if the software does not consider fineness data as a screening parameter. Among the 44 outlier vectors from Mahalanobis distance calculation, 7 were retained and considered as normal values after the expert's analysis as illustrated in Fig. 3. It is to highlight that Mahalanobis distance of these 7 data vectors are between 5.14 and 5.65 which are close to the upper limit bound  $UCL_{Mahalanobis}$ . The software considers these data as outliers even if they are not characterized neither by a vibration instability nor show signs of mill overfilling. However, process and quality analyze show that there are signs of material grindability variation as mill throughput decrease and raw meal fineness residue increase simultaneously. These data were deemed important and not to be eliminated so as not to lose useful information for our afterwards model development.

The next steps of modeling will therefore be with 243 data lines. In

**Table 1**

Parameters impacting product fineness in vertical roller mills according to literature.

Mill feed rate	Mill gas flow	Dynamic separator speed	Rollers hydraulic pressure	Mill outlet temperature	Recirculation damper	Other process variables	Reference
X	X	X					[10,11]
X		X		X	X		[12]
X		X	X	X	X	X	[15]

**Table 2**

Eigenvalues in descending order according to their contribution on data variance.

Number	Eigenvalue	Percentage	Cumulative Percentage
1	4.7	38.8	38.8
2	1.9	15.7	54.5
3	1.4	11.7	66.2
4	1.2	9.8	76
5	0.9	7.2	83.2
6	0.8	6.3	89.5
7	0.4	3.6	93.2
8	0.4	3	96.2
9	0.2	2	98.2
10	0.1	0.9	99.1
11	0.1	0.7	99.8
12	0	0.2	100

the next subsection, candidate variables impact on the soft sensor model will be checked.

### 2.2.2. Feature selection

Right choice of potential parameters candidates that correlate with raw meal fineness is an important step in soft sensor model development. Raw meal particles selection is achieved in the dynamic separator level above the mill. The product fineness is continuously regulated by the operator in the cement plant's control room by acting on the dynamic separator speed and mill gas flow. In other terms, it acts on the centrifugal and aerodynamic forces to reach fineness setpoint. The VRM is a nonlinear and dynamic system, there is actually more parameters that impact raw meal fineness in the industrial scale. Indeed, according to Boulvin et al. [44], experimental data from a grinding circuit show that mill throughput impacts the separating curve. A recent experimental study on parameters affecting the VRM performance shows a correlation between product fineness and dynamic separator speed, mill throughput, and the rollers hydraulic pressure applied to the grinding table of the VRM [45]. Another study shows that mill outlet temperature and recirculating damper can have less significant impact on product fineness [26].

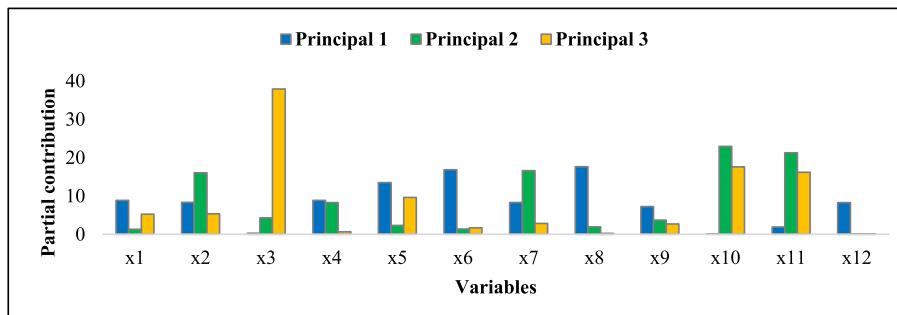
Vertical roller mills are particularly suitable for grinding medium-hard to soft mineral materials, but they can also be used for relatively hard materials. In addition to raw meal and cement clinker grinding, vertical roller mill is also used for grinding phosphate, chalk, etc. On the industrial scale, change in material grindability impacts significantly

the mill performance in terms of material flow and product fineness. Indeed, a decrease in material grindability results in an increase on mill circulating load at same fineness level, and, therefore, mill material throughput drops down [44]. In other words, to keep the feed rate in the same level, there should be an increase in fineness according to material's grindability in the mill feed. Table 1 shows parameters that affect product fineness in vertical roller mills according to the literature. It is noted that raw material characteristics were not considered in these models. Raw material is characterized by online chemical analysis from laboratory. In fact, material characteristics variations have a medium-term impact on raw meal performance and fineness. In this work, the effect of different parameters on raw meal fineness is tested, before analyzing their significance level.

### 2.2.3. Principal component analysis for dimension reduction

PCA is a technique that transforms a n-dimensional input space to a mutually independent n-dimensional vector called the eigenvectors. It works by finding the linear combination  $a_{1x}, a_{2x} \dots a_{px}$  called the eigenvalues and having maximum variance of the data, in general,  $p$  is less than  $n$ , the initial dimension of our workspace. The parameters retained for our model development are mill feed rate ( $x_1$ ), mill drive power ( $x_2$ ), recirculation elevator amperage ( $x_3$ ), mill fan speed ( $x_4$ ), mill separator speed ( $x_5$ ), mill outlet temperature ( $x_6$ ), mill pressure difference ( $x_7$ ), rollers hydraulic pressure ( $x_8$ ) and raw materials quality in the mill feed:  $SiO_2$  ( $x_9$ ),  $Al_2O_3$  ( $x_{10}$ ),  $Fe2O3(x_{11})$  and  $CaO$  ( $x_{12}$ ). PCA method was adapted to analyze these 12 candidate variables correlation using JMP software from SAS. Eigenvalues according to their contribution to data variance are presented in Table 2.

Contribution to the most influential principal components was checked to quantify the impact of the candidate variables on model variance. Fig. 5 represents the partial contribution of the candidate variables to the first three principal components which represent 66% of model variation. It is noted that the first main component alone represents almost 39% of the overall model variance. All candidate variables are present in the first main component except the recirculation current and alumina in material  $Al_2O_3$ . Material's iron  $Fe_2O_3$  also has a lower and relatively less significant impact than the other variables in the first main component. However, these three variables are strongly present in the second and third main components, which represent almost 25% of the total model variance. Therefore, it is recommended to retain all candidate variables selected. Fig. 6 represents correlation map between inputs and output to double check PCA results. The last line in the table represents the impact of each factor on the output. It can be seen that

**Fig. 5.** Variables partial contribution on eigenvectors.

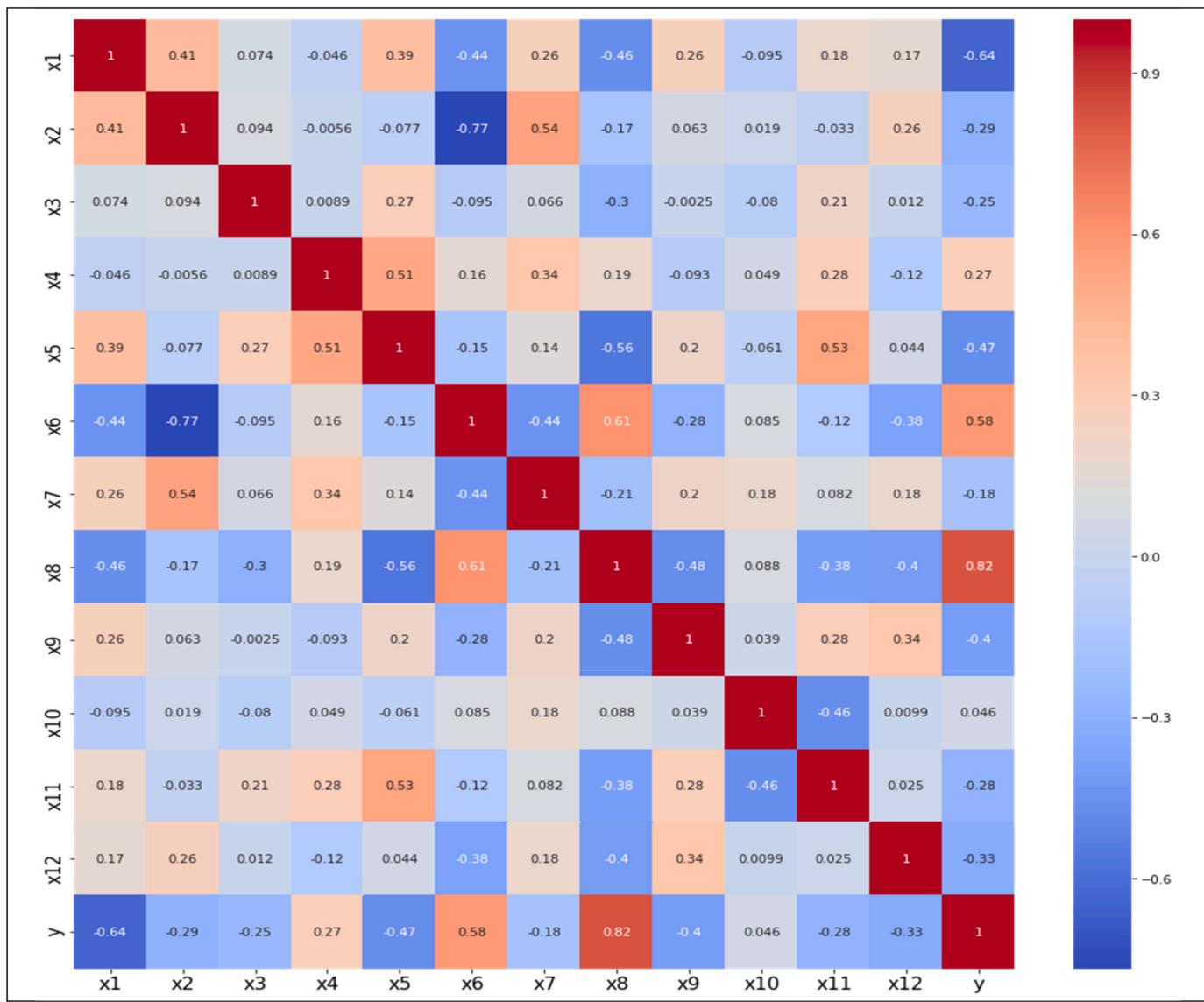


Fig. 6. Correlation map between inputs and output.

Table 3

The range of data set and their descriptive statistics, before and after treatment.

	Parameter	Before treatment				After treatment			
		Min	Max	Mean	Standard deviation	Min	Max	Mean	Standard deviation
Process input variables	$x_1$ (t/h)	242.3	318	282.5	9.5	250.1	318	283.7	9.1
	$x_2$ (KW)	2103.1	3065.7	2558.9	144.7	2213.8	2930.2	2551.3	140
	$x_3$ (A)	63.5	83.4	73	2.3	68.6	83.4	73.4	2.1
	$x_4$ (RPM)	910	998	931.8	7.4	910	950.4	930.3	6.8
	$x_5$ (RPM)	47.8	58	53.8	1.6	49.3	58	54.1	1.4
	$x_6$ (C)	96.4	126.7	112.6	6.4	96.4	126.7	112.7	6.7
	$x_7$ (mbar)	58.7	72.2	64.6	1.6	60.2	69.3	64.4	1.6
	$x_8$ (bar)	84.5	138	118	13.3	84.5	136.7	117.1	13.7
Quality input variables	$x_9$ (%)	12.7	13.4	13	0.1	12.8	13.3	13	0.1
	$x_{10}$ (%)	3.3	3.7	3.5	0.1	3.3	3.7	3.5	0.1
	$x_{11}$ (%)	1.8	2.1	2	0.1	1.8	2.1	2	0
	$x_{12}$ (%)	42.7	43.4	43	0.1	42.7	43.4	43	0.1
Output variable	$Y$ (%)	14.5	18.5	15.7	0.9	14.5	17.8	15.6	0.8

alumina in material  $\text{Al}_2\text{O}_3$  ( $x_{10}$ ) has the least impact on fineness variation, while it also spotted in the second and third principal component which represent almost only 25% of the total model variance.

The results obtained reflect the reality in the industrial scale. CaO

and  $\text{SiO}_2$  represent respectively the impact of limestone and sand used as additives in the VRM. Both are the most difficult materials to grind in the raw material mixture. It is to highlight that these two parameters are as important as the mill process variables: mill throughput, mill motor

power, mill outlet temperature, mill pressure loss and rollers hydraulic pressure, and have as much impact on model variance.

Experimental data corresponding to a matrix of 243 data rows and 12 variable columns will be exploited to predict raw meal fineness residue at sieve 90 µm ( $Y$ ), using neural network and SVR methods. Table 3 gives the range of data set and their statistic before and after treatment.

### 2.3. Support vector regression method

Support vector regression (SVR) method was developed as an extension of the theory of Support Vector Machine (SVM) to regression problems. SVM can give cleaner or more powerful way of learning complex non-linear functions. It was successfully applied to various classification and regression problems including time series models identification [46], applying SVR for multivariate nonlinear processes [47,48], and in numerous other engineering applications [49–52].

#### 2.3.1. Description of nonlinear SVR

In a simple regression model, the error rate between the prediction and the data is to be minimized, while in SVR, the goal is to make sure that error does not exceed a certain threshold. The support vector method for regression is formulated in solving a convex optimization problem, more specifically a quadratic programming problem. The regression is considered in the following set of function:

$$f(x) = w^T \varphi(x) + b \quad (3)$$

where,  $w \in \mathbb{R}^N$ ,  $\varphi(x)$  is a transformation that maps  $x$  to a higher dimensional space, and  $b \in \mathbb{R}$  is a scalar. For a set of training data where  $x_i$  is multivariate set of  $N$  observations with observed values  $y_i$ , the objective function, also known as the primal formula can be expressed as follows [53]:

$$\begin{aligned} \min J(w, b, \xi, \xi^*) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s.t. } y_i - w^T \varphi(x_i) - b &\leq \epsilon + \xi_i^* \quad i = 1, \dots, N \\ w^T \varphi(x_i) + b - y_i &\leq \epsilon + \xi_i \quad i = 1, \dots, N \end{aligned} \quad (4)$$

where, the constant  $C$  is a user regulated parameter tuning the trade-off between the training error and margin,  $\epsilon$  is a user defined small positive constant, the constraints involve two slack variables  $\xi_b$ ,  $\xi_i^* \geq 0$ . The introduced slack variables approach is similar to the « soft margin » concept in SVM classification, because slack variables allow regression errors to exist up to the value of  $\xi_i$  and  $\xi_i^*$  yet still satisfy the required conditions.

The optimization problem in (4) can be solved by constructing the Lagrangian for the objective function and the constraints. By solving the Lagrangian, the weight vector,  $w$ , can be derived as follows:

$$w = \sum_{i=1}^N (\alpha_i^* - \alpha_i) \varphi(x_i) \quad (5)$$

Where  $\{\alpha_i^*, \alpha_i\}$  are the Lagrangian nonnegative multipliers associated with the training point  $x_i$ . This leads to the dual formula, where the nonlinear SVR finds the coefficients that minimize:

$$\begin{aligned} L(\alpha) &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_j, x_i) + \epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) \\ &\quad - \sum_{i=1}^N y_i (\alpha_i^* - \alpha_i) \end{aligned}$$

$$\text{s.t. } \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0$$

$$0 \leq \alpha_i^* \leq C \quad i = 1, \dots, N$$

$$0 \leq \alpha_i \leq C \quad i = 1, \dots, N$$

Where  $k(x_j, x_i)$  is the kernel function represented as follows:

$$k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle \quad (7)$$

Using equation (7), the function used to predict new values is equal to:

$$f(x) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) k(x, x_i) + b \quad (8)$$

The most used kernel functions are the Gaussian Radial Basis Function-kernel (RBF-kernel) and the polynomial kernel with an order  $d$ , that can be expressed respectively as follows:

$$k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} = e^{-\gamma \|x_i - x_j\|^2} \quad (9)$$

$$k(x_i, x_j) = (\gamma x_i^T x_j + Constant)^d \quad (10)$$

Where,  $\gamma > 0$ , and  $\sigma$  defines the width of the kernel. Small value of  $\sigma^2$  parameter tends to over-fitting the SVR model with lower bias and higher variance. Whereas in contrast, if  $\sigma^2$  is large, higher bias and lower variance is obtained with possibility to cause under-fitting. Another parameter that needs to be selected when using SVR is the parameter  $C$  which was in the optimization objective. Assigning a large value to  $C$  means that not much regularization is being used. Therefore, model tend to have a hypothesis with lower bias and higher variance. In addition to that, it should be noted that increasing the  $C$  value can decrease the number of support vectors, but also can increase training time. Whereas hypothesis with small  $C$  has higher bias and is thus more prone to underfitting. As a result, SVR generalization performance and prediction accuracy depend on a good setting of the hyperparameters  $C$ ,  $\gamma$ , the kernel function and its parameters.

#### 2.3.2. Acquisition function for hyperparameters optimization

It is typically difficult and time-consuming to optimize manually SVR hyperparameters while they can strongly affect the performance of the regressor. Thus, intuitive hyperparameter optimization was introduced to adjust hyperparameters of Machine Learning algorithms and get us our desirable level of performance. The most common hyperparameter optimization techniques are Grid Search [54], Random Search [55] and Bayesian Search [56]. According to Ref. [57], Bayesian Search or Bayesian optimization has been shown to provide an excellent tool for finding good machine learning hyperparameters. For instance, Bayesian optimization was recently used for estimation of shear capacity of FRP reinforced concrete members [58]. The technique as the name suggests works on Bayes principle will be explained in subsection 3.3.1. One innovation in Bayesian optimization is the use of an Acquisition Function (AF), which the algorithm uses to determine the next point to evaluate, while constructing a probabilistic model for the objective function and integrating out uncertainty. There are several popular choices of AF for Bayesian optimization. This work will be subject of tree main AF: Probability of Improvement (PI) [59], Expected Improvement (EI) [60], and Gaussian Process Lower Confidence Bound (GP-LCB) [61].

Probability of Improvement is among the first AFs designed for Bayesian optimization. The PI AF is as a function of  $x$  can be computed analytically under gaussian model as follows:

$$PI(x) = \Phi(z(x)) \quad (11)$$

Where,  $z(x) = \frac{f(x_{Best}) - \mu(x)}{\sigma(x)}$ ,  $x_{Best} = \operatorname{argmin}_x f(x)$  is the best current value defined as the location of the lowest posterior mean,  $\mu(x)$  is the predictive mean at  $x$ ,  $\sigma^2(x)$  is the predictive variance at  $x$ , and  $\Phi$  denotes the

**Table 4**

SVR performance results with corresponding best point and minimum error hyperparameters results.

		EI			PI			LCB		
		Gaussian	Quadratic	Cube	Gaussian	Quadratic	Cube	Gaussian	Quadratic	Cube
Best point Hyper parameters	Box constraint	1.0286	1.0044	0.057	0.9262	0.0073	0.0326	2.4008	0.1669	0.0836
	Epsilon	0.0022	0.0033	0.0144	0.0017	0.0012	0.0038	0.0013	0.0012	0.2243
	Observed minimum MSE	0.1144	0.1235	0.1188	0.1143	0.1244	0.1281	0.1	0.1247	0.1232
Minimum error Hyper parameters	Box constraint	1.0819	0.3824	0.0906	1.2564	0.8766	0.0172	0.1	0.0842	0.0607
	Epsilon	0.0004	0.0024	0.0378	0.003	0.0012	0.0046	0.1	0.0012	0.2266
	Observed minimum MSE	0.1144	0.1235	0.1188	0.1143	0.1244	0.1278	0.1	0.1247	0.1232
Model performance	R <sup>2</sup>	0.84	0.82	0.83	0.84	0.76	0.82	0.83	0.82	0.83
	MSE	<b>0.1144</b>	0.1279	0.1203	0.1163	0.1735	0.1286	0.1216	0.1274	0.124
	MAE	0.2537	0.2583	0.2493	0.2549	0.3028	0.255	0.2636	0.2553	0.2623

**Table 5**

Simulation results of BR and LM algorithms with Logsig and Tansig activation functions.

Number of neurons	MSE of validation data			
	Logsig LM	Logsig BR	Tansig LM	Tansig BR
1	1.1311	1.0357	0.07716	0.12653
2	0.985	0.938	0.062	0.087
3	0.80739	0.9204	0.07087	0.069
4	0.941	0.9262	0.073578	0.09257
5	0.95641	1.0038	0.052858	0.08336
6	0.98402	0.97177	0.097119	0.1176
7	0.99048	1.011	0.0646	0.09783
8	1.017	1.1027	0.0866	0.1116
9	1.0194	1.128	0.0476	0.068
10	0.856	0.8636	0.076714	<b>0.0106</b>
11	0.6932	0.9054	0.0952	0.1412
12	1.0343	0.9478	0.0633	0.1569
13	1.0179	0.9613	0.02966	0.06863
14	0.935	0.9095	0.098107	0.1145
15	1.237	0.6593	0.09568	0.0579
16	0.9606	1.1	0.036658	0.0993
17	0.9131	0.907	0.042	0.0733
18	1.2551	0.78	0.085432	0.08061
19	1.0202	0.82	0.08979	0.03595
20	1.044	0.786	0.096295	0.13815

cumulative distribution function of standard normal. The point with the highest probability of improvement is selected.

Expected Improvement evaluates the expected amount of improvement in the objective function, ignoring values that cause an increase in the objective. EI has a closed-form expression [62]:

$$EI(x) = \sigma(x) (\zeta(x)\Phi(\zeta(x)) + \varphi(\zeta(x))) \quad (12)$$

Where,  $\varphi$  denotes the probability density function of standard normal. The first component of the function can be increased by reducing the mean function  $\mu(x)$ . The second one can be increased by increasing the variance function  $\sigma^2(x)$ . The EI criterion automatically captures the trade-off between exploitation (evaluating at points with low mean  $\mu$ ) and exploration (evaluating at points with high variance  $\sigma^2$ ) [63]. The point with the highest expected improvement is selected.Gaussian Process Lower Confidence Bound AF minimizes  $f$  according to the following formula:

$$LCB(x) = \mu(x) - k z(x) \quad (13)$$

where,  $k > 0$  is a trade-off parameter to balance exploitation against exploration. One drawback of the GP-LCB method compared to the others mentioned above is that it requires its own tuning parameter.

In the next section, a proper choice of different types of acquisition and kernel functions is intended to make comparison and further select the best model for this research. The model with minimum cross-validation loss of regression and the best performance indices is to be retained.

## 2.4. Artificial neural network method

The neural network technique is adapted to problems that involve the manipulation of multiple parameters and non-linear interpolation [64]. BPNN algorithm is the most popular and the oldest supervised learning multilayer feed forward neural network algorithm [65]. It is widely used in solving many practical problems.

Many algorithms can be used to train the artificial neural network model such as LM algorithm [66,67] and BR algorithm [21,68]. BR algorithm has given in many studies either moderate or best performance when compared with other training algorithms [69–73]. Meanwhile, according to the authors, BR algorithm was not exploited in soft sensor development and its performance's potential was not explored in this field. Therefore, BR algorithm was used and compared to LM, the most popular algorithm, which is widely used in many research studies.

### 2.4.1. Description of BR and LM algorithms

In contrast to conventional network training where an optimal set of weights is chosen by minimizing an error function of the mean sum of squares of the network error, the Bayesian approach involves adding a term to our loss function that penalizes for large weights. The cost function to be minimized is [16]:

$$S(w) = \beta \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{N_w} w_j^2 \quad (14)$$

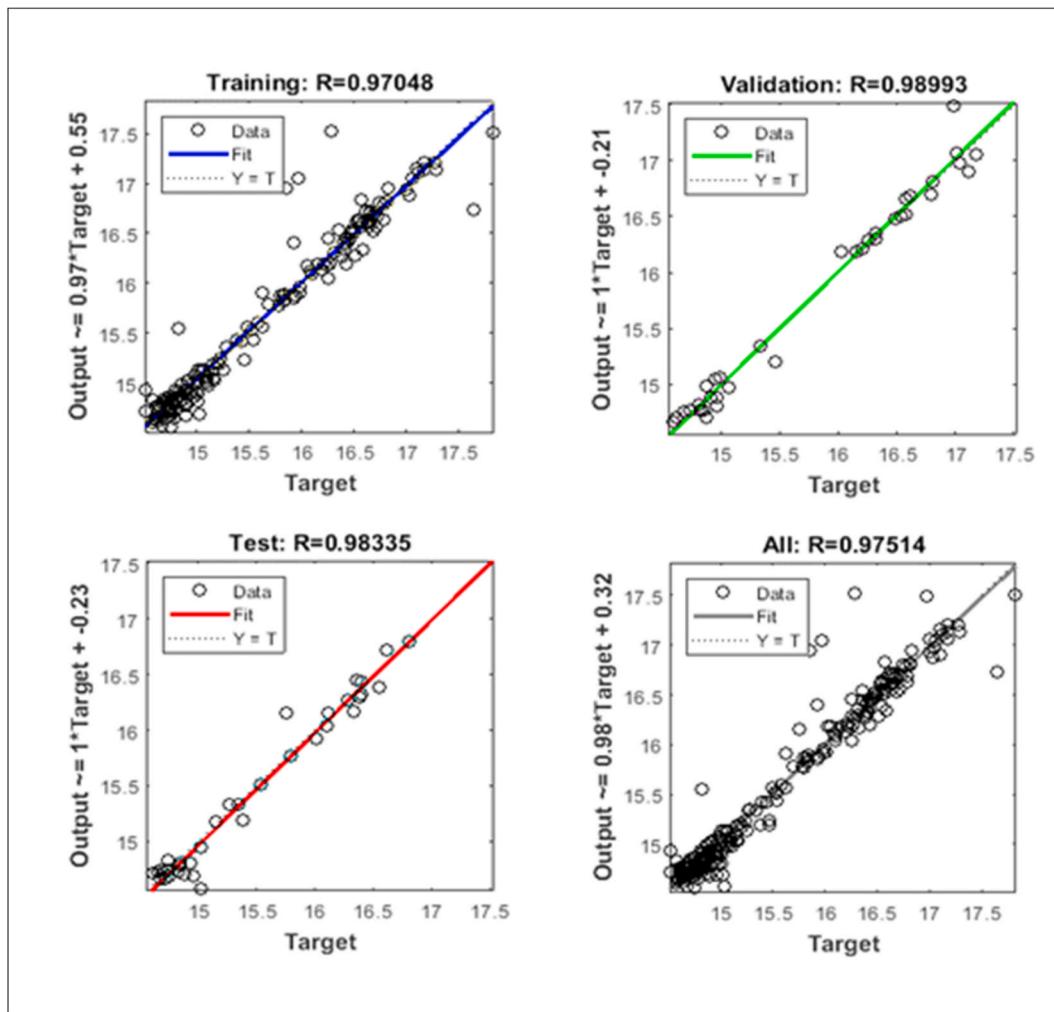
Where,  $y$  and  $\hat{y}$  are the  $i$ th experimental and predicted values, respectively;  $w_j$  is the  $j$ th value of the vector of networks weights;  $N_w$  is the number of weights, and  $n$  is the number of rows;  $\alpha$  and  $\beta$  are hyperparameters that need to be estimated function parameters; The  $\alpha$  and  $\beta$  factors are defined by the Bayes' theorem. Assuming the weight and data probability distributions are Gaussian, the posterior distribution is evaluated according to the Bayes' theorem [74]:

$$P(\alpha, \beta | D, M) = \frac{P(D | \alpha, \beta, M) P(\alpha, \beta | M)}{P(D|M)} \quad (15)$$

where:  $D$ , is the weight distribution;  $M$ , is the particular neural network architecture;  $(D|M)$  is the normalization factor;  $P(\alpha, \beta | M)$  is the uniform prior density for the regularization parameters, and  $P(D | \alpha, \beta, M)$  is the likelihood function. According to Ref. [75], minimizing  $S(w)$  in equation (3) is equivalent of maximizing the posterior probability function (15), and therefore, maximizing the likelihood function  $P(D|\alpha, \beta, M)$ . Thus, optimum values for  $\alpha$  and  $\beta$  for a given weight space are found. From this step, this Bayesian regularization takes place within the LM algorithm by calculating the Hessian or Jacobian Matrix. According to Ref. [74], the likelihood function can be expressed as follows:

$$P(D|\alpha, \beta, M) = \frac{P(D|w, \beta, M) P(w|\alpha, M)}{P(w|D, \alpha, \beta, M)} = \frac{Z_F(\alpha, \beta)}{(\pi/\beta)^{n/2} (\pi/\alpha)^{m/2}} \quad (16)$$

Where:  $n$  and  $m$  are the number of observations and total number of network parameters, respectively.  $Z_F(\alpha, \beta)$  can be developed according



**Fig. 7.** Correlation coefficient performance results for learning, testing and validation of the neural model with the BR algorithm.

to Laplace approximation, in terms of the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique [76] that is much less complex than computing the Hessian matrix. If the convergence is not met, algorithm estimates new values for  $\alpha$  and  $\beta$  and the whole procedure repeats itself until convergence is reached [77].

In summary, the main difference between BR and LM algorithms is that BR introduces network weights into the training objective function by adding a term to our loss function that penalizes for large weights, thereby removing the unnecessary parameters. If  $\alpha$  is very large than  $\beta$ , then the training algorithm will make the errors smaller. Otherwise, when  $\beta$  is much larger than  $\alpha$ , training will emphasize weight size reduction at the expense of network errors, thus producing a smoother network response [75]. In practice, Bayesian methods can solve the overfitting problem effectively and complex models are penalized in the Bayesian approach. A detailed analysis of BR in combination with training LM may be useful for interested researchers [78].

#### 2.4.2. Model topology

Proper choice of neural network model topology is an important step to build an accurate predictive model. This can be achieved by a proper choice of number of neurons, number of hidden layers and activation function. Indeed, black box model is often attributed to neural network method [79]. Moreover, there is no clear method or rule concerning the right choice of number of hidden layers or neurons to use in an

application [80]. Since the number of hidden layers increases the learning time without significantly improving the model error, only one single hidden layer is adopted with different number of neurons. Two types of activation functions in the hidden and output layers were used: logistic sigmoid and hyperbolic tangent sigmoid transfer functions expressed as follows:

$$\text{Logsig}(n) = \frac{1}{1 + e^{-n}} \quad (17)$$

$$\text{Tansig}(n) = \frac{2}{1 + e^{-2n}} - 1 \quad (18)$$

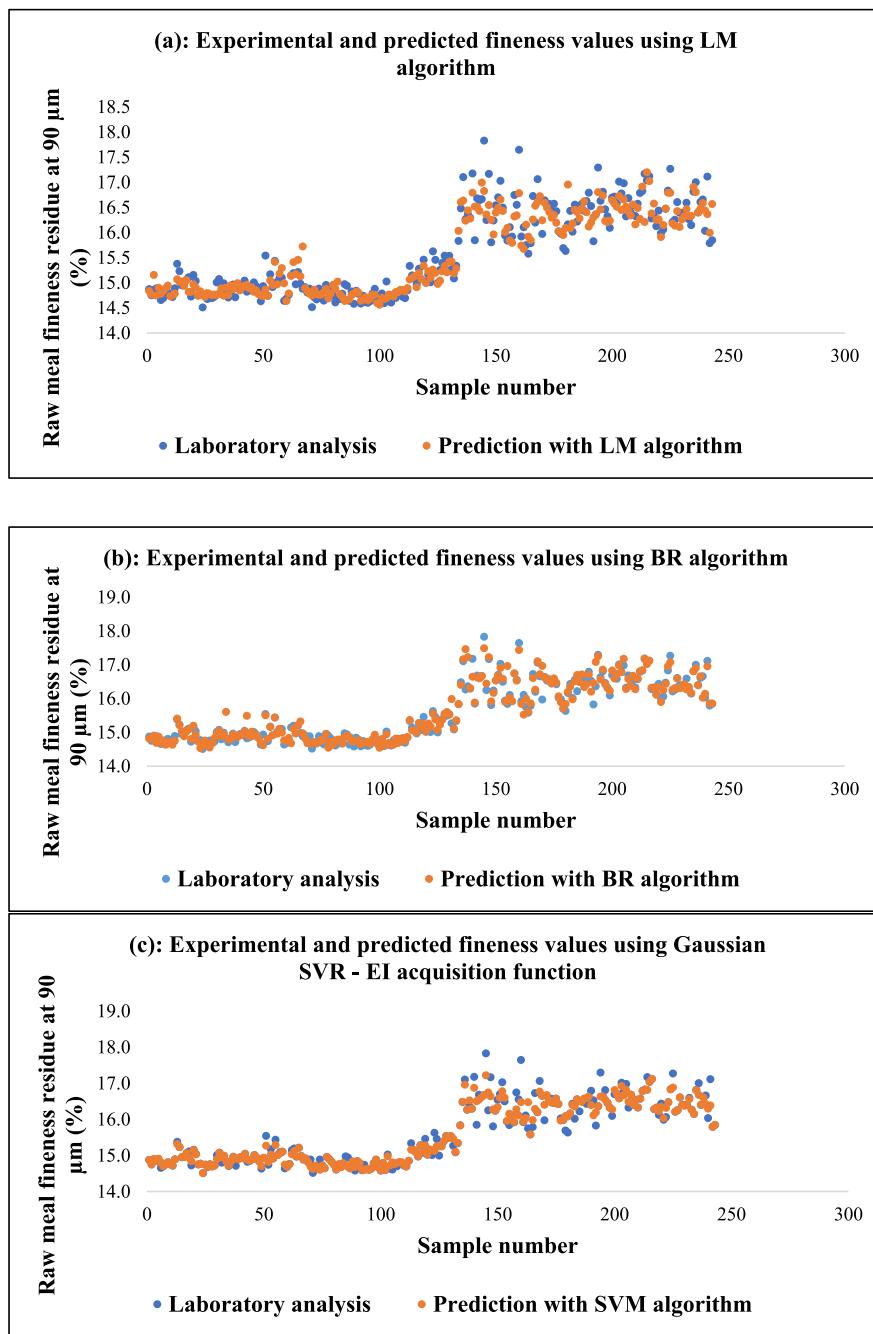
where,  $n$  is the input neuron.

#### 2.5. Evaluation of model performance

Model performance is evaluated by two main indices: mean square error (MSE) and coefficient of determination ( $R^2$ ) defined as follows:

$$\text{MSE} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (19)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (20)$$



**Fig. 8.** Raw meal fineness prediction during one year of VRM operation.

Where,  $y$  and  $\hat{y}$  are the  $i$ th experimental and predicted raw meal fineness values, respectively;  $\bar{y}$  and  $\bar{\hat{y}}$  are the mean values of experimental and predicted raw meal fineness, respectively. Both BR and LM algorithms use Jacobian for calculation, meaning that the performance is a mean or sum of squared errors. Thus, networks trained with these two algorithms should use the MSE performance function.

Mean Absolute Error (MAE) and correlation coefficient ( $R$ ) will be also used to ensure accuracy of model performance, defined as follows:

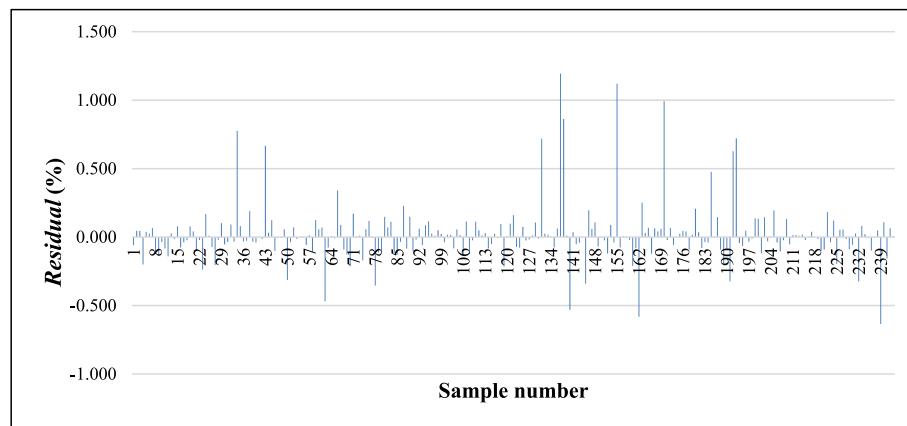
$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (21)$$

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (22)$$

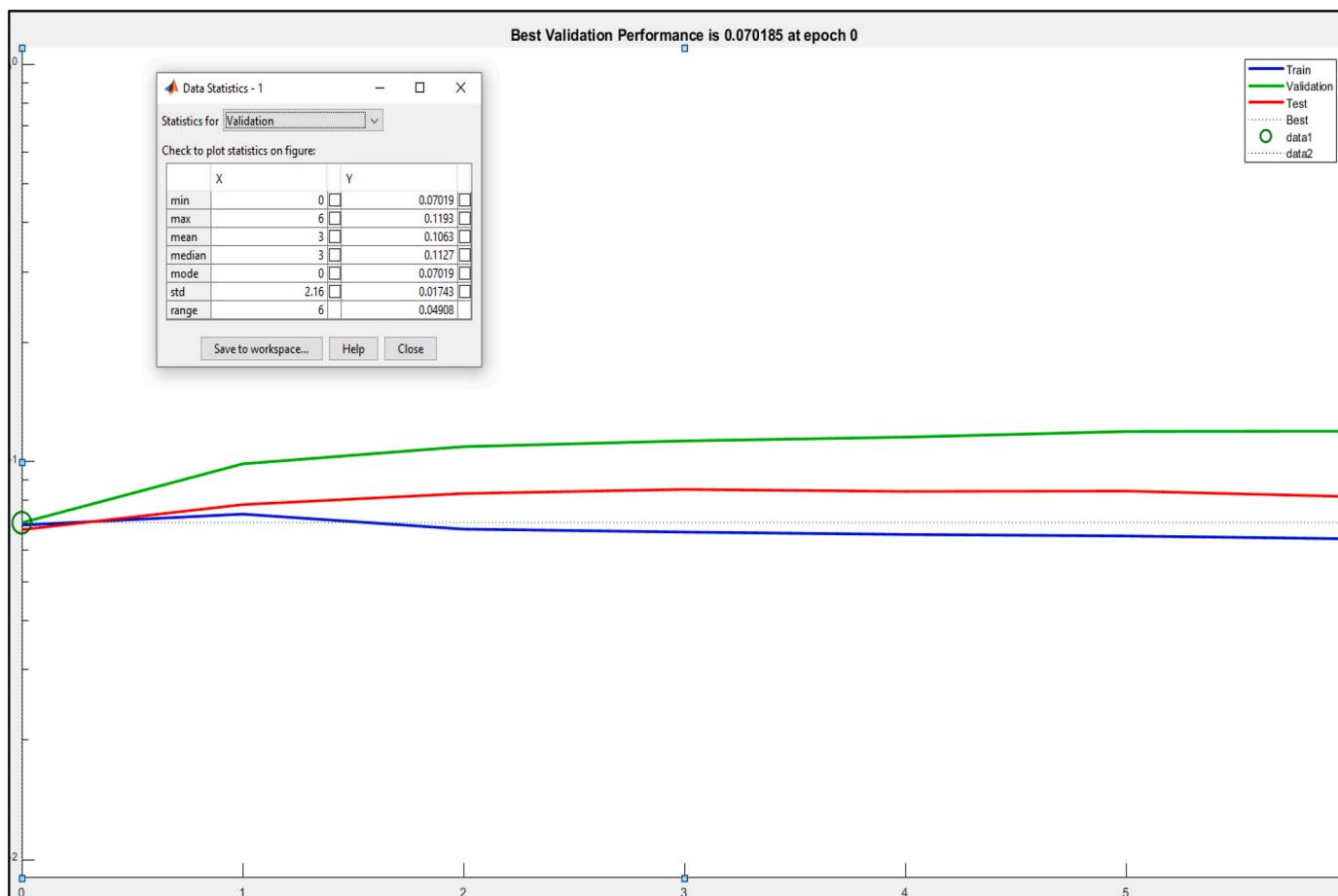
Training LM algorithm automatically stops when generalization stops improving, as indicated by an increase in the MSE of validation samples. On the other hand, training BR algorithm stops according to adaptive weight minimization (regularization).

### 3. Results and discussion

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.



**Fig. 9.** Residual distribution, for best neural network model using BR algorithm.



**Fig. 10.** MSE model performance for training, validation and testing datasets with only the 8 process input variables.

### 3.1. Support vector regression model performance

The analogical problem of choosing a neural model architecture is the problem of choosing a suitable kernel and AF in SVR. Therefore, performance of validation data was analyzed for each algorithm according to Gaussian, quadratic, and cubic kernels combined with EI, PI, and GP-LCB AFs. The models are trained using Regression Learner Toolbox in MATLAB, a 5-fold cross-validation is used during training to protect against overfitting by partitioning the dataset into folds, and calculate model performance on each fold of training datasets. The software computes different combinations of hyperparameter values by

using an optimization scheme that seeks to minimize the model MSE, and returns a model with the optimized hyperparameters. Table 4 shows model performance results of different SVR algorithms with corresponding best point and minimum error hyperparameters results. The optimal model performance is obtained with a Gaussian kernel and EI AF with a corresponding MSE of 0.114.

### 3.2. Neural network model performance

Matlab R2020b software was used for model development. Set of 243 data of each input is used to build the neural network model after data

preselection. 70% of data was used to train the model 15% for validation and the remaining 15% for testing the model. The optimal model is the one that produces the lowest error for the test or validation data, i.e. the best ability of generalization [21]. LM and BR algorithms were used for training the model. Table 5 shows simulation results of BR and LM algorithms with two types of activation functions defined in equations (17) and (18). Best model topology was analyzed according to the number of neurons from 1 to 20. Best model performance has the lowest MSE of 0.010 using BR algorithm, 10 neurons in the hidden layer, and hyperbolic tangent sigmoidal activation function. The activation function has a huge impact on model performance for both neural network models.

### 3.3. Model performance selection

Model performance was obtained by comparing predicted and experimental values of raw meal fineness. Fig. 7 represents this comparison for training, validation, testing and all model data. The overall model correlation coefficient obtained with BR neural network model is 97.51%. The overall model coefficient of determination is 96.98% with a MAE of 0.00016. Best MSE of training, validation and testing data are 0.0531, 0.0372 and 0.0095, respectively. Fig. 8 (a), (b) and (c) show simulation results of experimental and predicted raw meal fineness values for LM, BR and the retained Gaussian-EI AF SVR model, respectively, during one year of operation of the VRM and using all data of the model. It can be seen that LM and Gaussian SVR give slightly worse prediction results than BR algorithm, especially from data 50 to 70 and from 134 to 168. Therefore, BR model has a very accurate potential to predict raw meal fineness, and outperforms the other algorithms.

The performance of the best neural network model using BR algorithm was also evaluated by comparing the predicted values of raw meal fineness to the experimental one. This comparison is realized by plotting the residual error between experimental and predicted value in the function of sample number (Fig. 9). This figure shows that the points are distributed randomly around the zero axes, and the errors do not exceed the interval between -1.0 and 1.5.

To double check the impact of material quality characteristics, model performance was evaluated using the retained BR model with only 8 input variables  $x_1, x_2, \dots, x_8$ . With the same modeling procedure using only the 8 process input variables, best MSE of training, validation and testing data are: 0.06391, 0.07019 and 0.06741, respectively. Fig. 10 shows MSE model performance for training, validation and testing datasets with only the 8 process input variables. The MSE performance for validation data drops down from 0.03720 to 0.07019 when model is built without material quality characteristics. Therefore, it is undeniable that the overall model performance is better when material quality characteristics are considered in raw meal fineness soft sensing.

## 4. Conclusions

In spite of the level of theoretical knowledge achieved on grinding systems, the experimental return on experience is necessary to enhance our process knowledge because of the nonlinearities and the dynamic aspect of the industrial model. This is illustrated by an industrial application of finesse soft sensing in a vertical roller mill using robust BR algorithm that outperforms LM and SVR models. A particular attention was given to variables selection in order to deal with raw material heterogeneity that impacts raw meal fineness. It has been proved that the overall model performance is better when material quality characteristics are considered as input variables in raw meal fineness soft sensing.

Monitoring the raw mill using predicted soft sensor fineness is a very useful tool for operators to achieve a better and stable performance. Operators can act on separator speed to adjust predicted fineness in real time to reduce fineness fluctuations, and get a better product quality and mill stability.

Material stability is the main issue when dealing with cement grinding plants. Thus, the soft sensor approach used in this study is also available for cement grinding plants, where clinker and additives are usually imported from different clinker kilns. Therefore, applications of this study can be extended to all cement grinding plants using the same modelization strategy and adapting it to cement grinding plants constraints.

## Credit author statement

**Fahd Belmajdoub:** Investigation, Methodology, Formal analysis, Writing original draft, review & editing. **Souad Abderafi:** Conceptualization, Supervision and Revising the manuscript critically for important intellectual content, review & editing.

## Funding

This research received no external funding.

## Declaration of competing interest

There are no conflicts to declare. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Data availability

Data will be made available on request.

## Acknowledgments

The authors gratefully acknowledge support provided by Moroccan cement plant CIMAT.

## References

- [1] K. Sankar, P. Kale, R.A. Somani, et al., Online control and optimization of the grinding process in ball mills, *ZKG Int.* 55 (2002) 92–99.
- [2] G. Hefer, T. Kopetzky, Tilo, Expert system for optimization of cement mills, *ZKG Int.* 64 (2011) 48–53.
- [3] A. Bhaukam, S. Banerjee, J. Sil, Designing of intelligent expert control system using PetriNet for grinding mill operation, *WSEAS Trans. Inf. Sci. Appl.* 2 (2005) 360–365.
- [4] X. Chen, J. Zhai, Q. Li, S. Fei, Fuzzy logic based on-line efficiency optimization control of a ball mill grinding circuit, in: Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), IEEE, 2007, pp. 575–580, <https://doi.org/10.1109/FSKD.2007.329>.
- [5] X.S. Chen, Q. Li, S.M. Fei, Supervisory expert control for ball mill grinding circuits, *Expert Syst. Appl.* 34 (2008) 1877–1885, <https://doi.org/10.1016/j.eswa.2007.02.0132>.
- [6] S.P. Kavitha, M. Guruprasath, A.K. Tangirala, Developing a soft sensor for fineness in a cement ball mill, *IFAC Proc. Vol.* 47 (2014) 1019–1025, <https://doi.org/10.3182/20140313-3-IN-3024.00211>.
- [7] D. Stanišić, N. Jorgovanović, N. Popov, V. Čongradac, Soft sensor for real-time cement fineness estimation, *ISA Trans.* 55 (2015) 250–259, <https://doi.org/10.1016/j.isatra.2014.09.019>.
- [8] K.D. Morgenerie, L. Runge, M. Wächter, J. Vogel, B. Seller, Neuro-Fuzzy-Regelung einer Zementmühle, *ZKG Int.* 55 (2002) 72–80.
- [9] F. Belmajdoub, S. Abderafi, Specific electricity consumption optimization of raw grinding workshop in a Moroccan cement plant, in: 2018 6th International Renewable and Sustainable Energy Conference (IRSEC), IEEE, 2018, pp. 1–6, <https://doi.org/10.1109/IRSEC.2018.8702965>.
- [10] A.K. Pani, H.K. Mohanta, Soft sensing of particle size in a grinding process: application of support vector regression, fuzzy inference and adaptive neuro fuzzy inference techniques for online monitoring of cement fineness, *Powder Technol.* 264 (2014) 484–497, <https://doi.org/10.1016/j.powtec.2014.05.051>.
- [11] A.K. Pani, H.K. Mohanta, Online monitoring and control of particle size in the grinding process using least square support vector regression and resilient back propagation neural network, *ISA Trans.* 56 (2015) 206–221, <https://doi.org/10.1016/j.isatra.2014.11.011>.
- [12] Y. Bao, Y. Zhu, W. Du, W. Zhong, F. Qian, A distributed PCA-TSS based soft sensor for raw meal fineness in VRM system, *Control Eng. Pract.* 90 (2019) 38–49, <https://doi.org/10.1016/j.conengprac.2019.06.009>.

- [13] R. Lange, T. Lange, T.L. van Zyl, Predicting particle fineness in a cement mill, in: 2020 IEEE 23rd International Conference on Information Fusion (FUSION), IEEE, 2020, pp. 1–8, <https://doi.org/10.23919/FUSION45008.2020.9190236>.
- [14] G. Liu, Z. Ouyang, X. Hao, X. Shi, L. Zheng, Y. Zhao, Prediction of raw meal fineness in the grinding process of cement raw material: a two-dimensional convolutional neural network prediction method, Proc. IME J. Syst. Control Eng. 235 (2021) 823–838, <https://doi.org/10.1177/0959651820965447>.
- [15] H. Fernandes, A. Halim, W. Wahab, Modeling vertical roller mill raw meal residue by implementing neural network, in: 2019 IEEE International Conference on Innovative Research and Development (ICIRD), IEEE, 2019, pp. 1–6, <https://doi.org/10.1109/ICIRD47319.2019.9074747>.
- [16] F. Burden, D. Winkler, Bayesian regularization of neural networks, Artif. Neural Netw. (2008) 23–42, [https://doi.org/10.1007/978-1-60327-101-1\\_3](https://doi.org/10.1007/978-1-60327-101-1_3).
- [17] Y. Zhou, X. Zhao, C. Jiang, S. Liu, Z. Han, G. Wang, Permeability prediction of multi-stage tight gas sandstones based on Bayesian regularization neural network, Mar. Petrol. Geol. 133 (2021), 105320, <https://doi.org/10.1016/j.marpgeo.2021.105320>.
- [18] Z. Yue, Z. Songzheng, L. Tianshi, Bayesian regularization BP Neural Network model for predicting oil-gas drilling cost, in: 2011 International Conference on Business Management and Electronic Information, IEEE, 2011, pp. 483–487, <https://doi.org/10.1109/ICBME11.5917952>.
- [19] K. Hirschen, M. Schäfer, Bayesian regularization neural networks for optimizing fluid flow processes, Comput. Methods Appl. Mech. Eng. 195 (2006) 481–500, <https://doi.org/10.1016/j.cma.2005.01.015>.
- [20] D.J.C. MacKay, Maximum entropy connections: neural networks, in: Maximum Entropy and Bayesian Methods, Springer, Dordrecht, 1991, pp. 237–244, [https://doi.org/10.1007/978-94-011-3460-6\\_22](https://doi.org/10.1007/978-94-011-3460-6_22).
- [21] F.D. Foresee, M.T. Hagan, Gauss-Newton approximation to Bayesian learning, in: Proceedings of International Conference on Neural Networks (ICNN'97), IEEE, 1997, pp. 1930–1935, <https://doi.org/10.1109/ICNN.1997.614194>.
- [22] Y. Jbari, S. Abderafi, Liquid Density Prediction of Ethanol/Water, Using Artificial Neural Network, 2021, <https://doi.org/10.33263/BRIAIC14.56255637>.
- [23] F. El Hamdani, S. Vaudreuil, S. Abderafi, T. Bounahmidi, Determination of design parameters to minimize LCOE, for a 1 MW<sub>e</sub> CSP plant in different sites, Renew. Energy 169 (2021) 1013–1025, <https://doi.org/10.1016/j.renene.2021.01.060>.
- [24] H. Okut, X.L. Wu, G.J. Rosa, S. Bauck, B.W. Woodward, R.D. Schnabel, D. Gianola, Predicting expected progeny difference for marbling score in Angus cattle using artificial neural networks and Bayesian regression models, Genet. Sel. Evol. 45 (2013) 1–13, <https://doi.org/10.1186/1297-9686-45-34>.
- [25] H.H.M. Darweesh, Influence of sun flower stalk ash (SFSA) on the behavior of Portland cement pastes, Res. Eng. 8 (2020), 100171, <https://doi.org/10.1016/j.rineng.2020.100171>.
- [26] R. Elmrabet, Y. Hmidani, R. Mariouch, A. Elharfi, M.S. Elyoubi, Influence of raw meal composition on clinker reactivity and cement properties, Mater. Today Proc. 45 (2021) 7680–7684, <https://doi.org/10.1016/j.matpr.2021.03.178>.
- [27] P.D. Tennis, M.D.A. Thomas, W.J. Weiss, State-of-the-Art Report on Use of Limestone in Cements at Levels of up to 15%. PCA R&D SN3148, Portland Cement Association, Skokie, IL, 2011.
- [28] J.W. Osborne, A. Overbay, The power of outliers (and why researchers should always check for them), Practical Assess. Res. Eval. 9 (2004) 6, <https://doi.org/10.7275/qf69-7k43>.
- [29] L. Fortuna, S. Graziani, A. Rizzo, M.G. Xibilia, Soft Sensors for Monitoring and Control of Industrial Processes, Springer, London, UK, 2007.
- [30] L.H. Chiang, R.J. Pell, M.B. Seasholtz, Exploring process data with the use of robust outlier detection algorithms, J. Process Control 13 (2003) 437–449, [https://doi.org/10.1016/S0959-1524\(02\)00068-9](https://doi.org/10.1016/S0959-1524(02)00068-9).
- [31] R.K. Pearson, Outliers in process modeling and identification, IEEE Trans. Control Syst. Technol. 10 (2002) 55–63, <https://doi.org/10.1109/87.974338>.
- [32] S. Seo, A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets, University of Pittsburgh, 2006. Thèse de doctorat.
- [33] P.J. Rousseeuw, M. Hubert, Robust statistics for outlier detection, Wiley interdiscipl. Rev.: Data Min. Knowl. Discov. 1 (2011) 73–79, <https://doi.org/10.1002/widm.2>.
- [34] A.K. Pani, H.K. Mohanta, Online monitoring of cement clinker quality using multivariate statistics and Takagi-Sugeno fuzzy-inference technique, Control Eng. Pract. 57 (2016) 1–17, <https://doi.org/10.1016/j.conengprac.2016.08.011>.
- [35] J.S. Zeng, C.H. Gao, Improvement of identification of blast furnace ironmaking process by outlier detection and missing value imputation, J. Process Control 19 (2009) 1519–1528, <https://doi.org/10.1016/j.jprocont.2009.07.006>.
- [36] W.J. Egan, S.L. Morgan, Outlier detection in multivariate analytical chemical data, Anal. Chem. 70 (1998) 2372–2379, <https://doi.org/10.1021/ac970763d>.
- [37] R.G. Garrett, The chi-square plot: a tool for multivariate outlier recognition, J. Geochem. Explor. 32 (1989) 319–341, [https://doi.org/10.1016/0375-6742\(89\)90071-X](https://doi.org/10.1016/0375-6742(89)90071-X).
- [38] P. Filzmoser, A multivariate outlier detection method, in: S. Aivazian, P. Filzmoser, Yu Kharin (Eds.), Proceedings of the Seventh International Conference on Computer Data Analysis and Modeling vol. 1, Belarusian State University, Minsk, 2004, pp. 18–22.
- [39] Á. López-Oriona, J.A. Vilar, Outlier detection for multivariate time series: a functional data approach, Knowl. Base Syst. 233 (2021), 107527, <https://doi.org/10.1016/j.knosys.2021.107527>.
- [40] E. Cabana, R.E. Lillo, H. Laniado, Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators, Stat. Pap. 62 (2021) 1583–1609, <https://doi.org/10.1007/s00362-019-01148-1>.
- [41] SAS Institute Inc., Cary, NC, USA. JMP®. Version 14.2.0.
- [42] K. Varmuza, P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, first ed., Taylor & Francis Group, 2016 <https://doi.org/10.1201/9781420059496>. Boca Raton CRC press.
- [43] R.L. Mason, J.C. Young, *Multivariate Statistical Process Control with Industrial Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 2002. SIAM. (Chapter 5).
- [44] M. Boulvin, A.V. Wouwer, R. Lepore, C. Renotte, M. Remy, Modeling and control of cement grinding processes, IEEE Trans. Control Syst. Technol. 11 (2003) 715–725, <https://doi.org/10.1109/TCST.2003.816406>.
- [45] D. Altun, H. Benzer, N. Aydogan, C. Gerold, Operational parameters affecting the vertical roller mill performance, Miner. Eng. 103 (2017) 67–71, <https://doi.org/10.1016/j.mine.2016.08.015>.
- [46] R.K. Dash, T.N. Nguyen, K. Cengiz, A. Sharma, Fine-tuned support vector regression model for stock predictions, Neural Comput. Appl. (2021) 1–15, <https://doi.org/10.1007/s00521-021-05842-w>.
- [47] A.S. Noushabadi, A. Dashti, F. Ahmadijokani, J. Hu, A.H. Mohammadi, Estimation of higher heating values (HHVs) of biomass fuels based on ultimate analysis using machine learning techniques and improved equation, Renew. Energy (2021) 550–562, <https://doi.org/10.1016/j.renene.2021.07.003>.
- [48] Z. Zhang, W.C. Hong, Application of variational mode decomposition and chaotic grey wolf optimizer with support vector regression for forecasting electric loads, Knowl. Base Syst. 228 (2021), 107297, <https://doi.org/10.1016/j.knosys.2021.107297>.
- [49] R. Nyirandayisabhe, H. Li, Q. Dong, T. Hakuzwyezu, F. Nkinahamira, Automatic pavement damage predictions using various machine learning algorithms: evaluation and comparison, Res. Eng. 16 (2022), 100657, <https://doi.org/10.1016/j.rineng.2022.100657>.
- [50] P. Sultana, A.K. Dey, D. Kumar, Empirical approach for prediction of bearing pressure of spread footings on clayey soil using artificial intelligence (AI) techniques, Res. Eng. 15 (2022), 100489, <https://doi.org/10.1016/j.rineng.2022.100489>.
- [51] M.S. Jassim, G. Coskuner, M. Zontul, Comparative performance analysis of support vector regression and artificial neural network for prediction of municipal solid waste generation, Waste Manag. Res. 40 (2022) 195–204, <https://doi.org/10.1177/0734242X211008526>.
- [52] M. Faldziński, P. Fiszeder, W. Orzeszko, Forecasting volatility of energy commodities: comparison of GARCH models with support vector regression, Energies 14 (2021) 6, <https://doi.org/10.3390/en14010006>.
- [53] N. Vapnik Vladimir, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- [54] R. Ghawi, J. Pfleffer, Efficient hyperparameter tuning with grid search for text categorization using kNN approach with BM25 similarity, Open Comp. Sci. 9 (2019) 160–180.
- [55] R.G. Mantovani, A.L. Rossi, J. Vanschoren, B. Bischi, A.C. De Carvalho, Effectiveness of random search in SVM hyper-parameter tuning, in: 2015 International Joint Conference on Neural Networks (IJCNN), Ieee, 2015, pp. 1–8, <https://doi.org/10.1109/IJCNN.2015.7280664>.
- [56] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, Adv. Neural Inf. Process. Syst. 25 (2012).
- [57] R.P. Adams, A Tutorial on Bayesian Optimization for Machine Learning, Harvard University, 2014.
- [58] M.S. Alam, N. Sultana, S.Z. Hossain, Bayesian optimization algorithm-based support vector regression analysis for estimation of shear capacity of FRP reinforced concrete members, Appl. Soft Comput. 105 (2021), 107281, <https://doi.org/10.1016/j.asoc.2021.107281>.
- [59] H.J. Kushner, A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise, 1964. <https://doi:10.1115/1.3653121>.
- [60] A.H. Victoria, G. Maragatham, Automatic tuning of hyperparameters using Bayesian optimization, Evolving Syst. 12 (2021) 217–223, <https://doi.org/10.1007/s12530-020-09345-2>.
- [61] N. Srinivas, A. Krause, S.M. Kakade, M. Seeger, Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design, 2009, [https://doi.org/10.1109/TIT.2011.2182033 arXiv preprint arXiv:0912.3995](https://doi.org/10.1109/TIT.2011.2182033).
- [62] D.R. Jones, A taxonomy of global optimization methods based on response surfaces, J. Global Optim. 21 (2001) 345–383, <https://doi.org/10.1023/A:101277102557>.
- [63] M.A. Gelbart, J. Snoek, R.P. Adams, Bayesian Optimization with Unknown Constraints, 2014 [arXiv preprint arXiv:1403.5607](https://arxiv.org/abs/1403.5607).
- [64] Y. Jbari, S. Abderafi, Parametric study to enhance performance of wastewater treatment process, by reverse osmosis-photovoltaic system, Appl. Water Sci. 10 (2020) 1–14, <https://doi.org/10.1007/s13201-020-01301-4>.
- [65] M.A.W. Saduf, Comparative study of back propagation learning algorithms for neural networks, Int. J. Adv. Res. Comput. Sci. Software Eng. 3 (2013) 1151–1156.
- [66] Y. Ji, Z. Kang, X. Liu, The data filtering based multiple-stage Levenberg–Marquardt algorithm for Hammerstein nonlinear systems, Int. J. Robust Nonlinear Control 31 (2021) 7007–7025, <https://doi.org/10.1002/rnc.5675>.
- [67] Z. Sabir, M.R. Ali, M.A.Z. Raja, M. Shoaib, R.A.S. Núñez, R. Sadat, Computational intelligence approach using Levenberg–Marquardt backpropagation neural networks to solve the fourth-order nonlinear system of Emden–Fowler model, Eng. Comput. (2021) 1–17, <https://doi.org/10.1007/s00366-021-01427-2>.
- [68] S. Gouravaraju, J. Narayan, R.A. Sauer, S.S. Gautam, A Bayesian regularization-backpropagation neural network model for peeling computations, J. Adhes. (2021) 1–24, <https://doi.org/10.1080/00218464.2021.2001335>.
- [69] L.M. Saini, Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks, Elec. Power Syst. Res. 78 (2008) 1302–1310, <https://doi.org/10.1016/j.epsl.2007.11.003>.

- [70] S. Garg, Forecasting of gold prices using Bayesian regularization neural network, in: Nanoelectronics, Circuits and Communication Systems, Springer, Singapore, 2021, pp. 147–153, [https://doi.org/10.1007/978-981-15-7486-3\\_15](https://doi.org/10.1007/978-981-15-7486-3_15).
- [71] P. Lauret, E. Fock, R.N. Randrianarivony, J.F. Manicom-Ramsamy, Bayesian neural network approach to short time load forecasting, *Energy Convers. Manag.* 49 (2008) 1156–1166, <https://doi.org/10.1016/j.enconman.2007.09.009>.
- [72] J.L. Ticknor, A Bayesian regularized artificial neural network for stock market forecasting, *Expert Syst. Appl.* 40 (2013) 5501–5506, <https://doi.org/10.1016/j.eswa.2013.04.013>.
- [73] M. Kayri, Predictive abilities of bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: a comparative empirical study on social data, *Math. Comput. Appl.* 21 (2016) 20, <https://doi.org/10.3390/mca21020020>.
- [74] D.J. MacKay, Bayesian interpolation, *Neural Comput.* 4 (1992) 415–447, <https://doi.org/10.1162/neco.1992.4.3.415>.
- [75] F. Dan Forsee, M.T. Hagan, Gauss–Newton approximation to Bayesian learning, in: Proceedings of International Conference on Neural Networks (ICNN'97), IEEE, 1997, pp. 1930–1935, <https://doi.org/10.1109/ICNN.1997.614194>.
- [76] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE Trans. Neural Network.* 5 (1994) 989–993, <https://doi.org/10.1109/72.329697>.
- [77] Z. Yue, Z. Songzheng, L. Tianshi, Bayesian regularization BP Neural Network model for predicting oil-gas drilling cost, in: 2011 International Conference on Business Management and Electronic Information, IEEE, 2011, pp. 483–487, <https://doi.org/10.1109/ICBMEI.2011.5917952>.
- [78] F. Burden, D. Winkler, Bayesian regularization of neural networks, *Artif. Neural Netw.* (2008) 23–42, [https://doi.org/10.1007/978-1-60327-101-1\\_3](https://doi.org/10.1007/978-1-60327-101-1_3).
- [79] Y. Zhang, P. Tiño, A. Leonardis, K. Tang, A survey on neural network interpretability, *IEEE Tran. Emerg. Top. Comput. Intell.* (2021), <https://doi.org/10.1109/TETCI.2021.3100641>.
- [80] V.R. Radhakrishnan, Soft sensors for the cement industry–neural network models for kiln and grinding mill, *Me Three Number T* (2003) 29.