Heterogeneous selective ensemble learning model for mill load parameters forecasting by using multiscale mechanical frequency spectrum

Zhuo Liu1 · Tianyou Chai1 · Jian Tang2 · Wen Yu3

Accepted: 25 April 2022 / Published online: 11 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
A ball mill is a heavy mechanical device and its safe operation affects the entire grinding process. Mill load is a key index in the optimum operation of the grinding process, but it cannot be measured directly. In industrial practice, operational experts normally estimate its value based on their experiences and the mechanical signals produced by the ball mill. In this paper, we proposed a heterogeneous selective ensemble method by using a multiscale mechanical frequency spectrum. The multicomponent adaptive decomposition algorithm is first used to decompose the original shell vibration and acoustic signals into sub-signals with different timescales. Then, selective ensemble (SEN) kernel projection to latent structure algorithm is used to model the spectral data of these sub-signals. Furthermore, the latent features of multiscale spectra are extracted to construct SEN models based on fuzzy inference. Finally, the two types of heterogeneous SEN models are fused by using information entropy. The main contribution of this study is that the proposed soft-sensing model has a dual-layer ensemble structure that can fuse multi-source information in different mechanical sub-signals with physical meaning. Moreover, the proposed model can simulate the fuzzy cognitive behavior of domain experts in the mineral grinding process. The effectiveness of the method is verified by the shell vibration and acoustical data of a laboratory-scale ball mill.

Keywords
Selective ensemble learning · Fuzzy system · Heterogeneous ensemble · Information entropy weighing · Mill load parameter forecasting

1 Introduction
Mill load is directly related to quality, efficiency, energy consumption, material consumption, and safe operation of the grinding process (Tang et al. 2018a). Overloading can cause splitting of ore, coarsening of mill output, blocking of the mill, and belly being full. Conversely, the low load can cause the “running with only ball,” which leads to energy and steel consumption, and may even cause damage to milling devices. All of these conditions can disrupt the safety of the production process. Thus, accurate forecasting of the mill load is one of the key factors in realizing the optimum operation of the grinding process (Zhou et al. 2009). As ball mills are a type of continuous rotatory closed mechanical device, the mill load (i.e., material, ball, and water loads inside the mill) cannot be measured directly. Moreover, the material and water loads are continuously changed with loop, and corrosion and wear mass of the ball load are not known. Therefore, estimating the mill load with an indirect approach is a necessary task.

In the actual mineral process, mechanical vibrations and acoustic signals generated during mill grinding are typically used by domain experts to estimate the mill load. The reason is that there is a non-deterministic nonlinear mapping relation between multi-source mill shell vibration/acoustic signals and the mill load. Therefore, operational experts in the mill grinding process can fuzzy estimate the mill load for their familiar specific mill based on the
"human brain model" by virtue of multi-source information on the industrial practice and experiential knowledge they have accumulated over the years. However, most of the former methods cannot simulate the fuzzy cognitive behavior of the domain experts. In nature, human ears function as a pair of adaptive band-pass filters (Rajaraman et al. 2013; Robinson et al. 2016). Normally, the reasoning recognition of experts through “hearing the sound” can be understood as a layer-by-layer cognition process that consists of frequency band selection, feature extraction, and knowledge rule-based reasoning (Tang et al. 2018a). However, this operating mode can be easily affected by the different experts’ experiences, their limited energy, and other subjective factors. Consequently, in practical milling process, the mill would operate under uneconomical conditions for a long time, thereby causing high energy consumption and low efficiency. Moreover, the reasoning recognition of experts in this field by hearing the sound cannot utilize mill shell vibration signals with high sensitivity and reliability. Thus, combining the different modeling methods is necessary to integrate the fuzzy cognition of domain experts and the data-fitting ability of the former methods.

Early studies based on mechanical signal were conducted in the mid-1990s (Zeng and Forssberg 1993) for soft-sensing models of pulp density (PD) and granularity inside the mill in which frequency spectrum sub-band features were used. Previous research indicated that mill acoustic spectrum contained more valuable information than the axis seat vibration spectrum. Then, soft-sensing models of three mill-load parameters (MLPs), namely, material-to-ball volume ratio (MBVR), PD, and ball-charge volume ratio (BCVR), are established based on external signals such as vibration, axis pressure, and mill current (Wang and Chen 2002; Li and Shao 2006). In practice, the BCVR inside the ball mill changes less within a short period and the grid ball mill may block within 60 s. Thus, charge volume ratio (BCVR) is used as an MLP to represent the volume of all loads inside the mill (Tang et al. 2012a). Furthermore, a soft-sensing MLP model based on feature selection and extraction, as well as model learning parameter combinational optimization, is proposed, which can overcome the high-dimensional collinearity of the shell vibration spectra (Tang et al. 2012b). However, the aforementioned soft-sensing models are all traditional single models, which cannot solve the problems of pattern recognition or regression modeling effectively in terms of small sample modeling, and have not fully utilized the rich information contained in multi-source and multiscale signals.

Ensemble learning can improve modeling performance and stability by fusing multiple different single sub-models. The generalization ability of an ensemble model should be balanced between the accuracy and diversity of ensemble sub-models (Granitto et al. 2005; Wang et al. 2014). A selective ensemble (SEN) model based on branch-and-bound (BB) and adaptive weighted fusion (AWF) algorithms was proposed by Tang et al. (Tang et al. 2013) to address problems such as the redundancy and complementarity of the shell vibration/acoustic frequency sub-bands spectrum, and limitation of the information contained by a single-sensor signal. Thus, an MLP model in terms of the selective fusion of single-scale spectrum feature subsets from multi-source signals is established. However, a mill grinding mechanism shows that the shell vibration/acoustic signals have non-stationary and multi-component characteristics. Thus, applying fast Fourier transform (FFT) to these original mechanical signals to gain a single-scale spectrum for modeling is unsuitable in theory (Lei et al. 2009). Thus, decomposing these original signals into stationary sub-signals is necessary.

Huang et al. (1998) and Wu and Huang (2009) proposed empirical mode decomposition (EMD) to effectively decompose the original non-stationary signal into stationary sub-signals with different timescales, i.e., intrinsic mode function (IMF). This method has also been extensively applied in the fault diagnosis of rotary machines (Rai and Mohanty 2007). To assess the mill load the signal was first detrended by a complete ensemble EMD (Twa et al. 2021). Aiming at the co-linearity and nonlinearity of high dimensional data, kernel projection to latent structure (KPLS) algorithm is suitable to address such issues (Nicolá et al. 2007; Bastien et al. 2015; Wang et al. 2015). Tang et al. introduced EMD, power spectral density (PSD), and PLS algorithm to analyze shell vibration (Tang et al. 2011) and established the MLP model with multiscale shell vibration spectrum features (Tang et al. 2012c). The change in IMF spectra under different grinding conditions is investigated (Zhao et al. 2012). Hilbert decomposition method can also decompose the vibration into multiscale sub-signals, and can explain the mapping relation between MLP and shell vibration from another perspective (Tang et al. 2015). However, the aforementioned models are all based on linear/nonlinear PLS algorithms. Recently, the MLP modeling method based on selective fusion multi-condition samples and multi-source features are constructed by using shell vibration and acoustic frequency spectra (Tang et al. 2018b). Dual-layer optimization strategy is employed to select the ensemble sub-models and learning parameters optimally (Tang et al. 2020). A method based on complete ensemble EMD with adaptive noise (CEEMDAN) is proposed to identify the mill load (Lya and Jc 2021). Selective ensemble modeling approach based on multi-modal feature subsets is proposed for mill load parameter forecasting (Liu et al. 2021) However, the ensemble sub-models are almost constructed by using the
KPLS algorithm. Thus, deficiencies in fitting modeling data with poor inference performance were observed. Moreover, the aforementioned models cannot simulate the intelligent cognition mechanism of industry experts based on auditory perception to estimate MLPs.

The fuzzy system provides an effective means of complicated industrial modeling with mechanism complexity, strong coupling, uncertainty, and other comprehensive characteristics. For dry mill load detection, a patent for the load detection of a mill that integrated shell vibration and acoustic is applied (Si et al. 2007). A cloud model based on experimental ball mill axis vibration signal is built to infer and measure mill loads (Yan et al. 2014). For industrial wet mill during mineral grinding, rule-based reasoning is used to monitor and control mill overload status based on the current and process variables of mills (Zhou and Chai 2008). Furthermore, data fusion and case-based reasoning algorithm are combined to estimate mill loads (Bai and Chai 2009). However, these methods cannot use mill shell vibration signals with high sensitivity and reliability. Thus, a SEN fuzzy model based on multiscale frequency spectrum features must be constructed. In terms of simulating expert cognition behavior, this model has good inference, but its fitting ability to the modeling data is weak.

Therefore, we proposed a novel heterogeneous SEN (HSEN) modeling strategy by using SEN KPLS and SEN fuzzy models. The contribution of the proposed strategy includes the following: (1) combining two types of heterogeneous models with the error information entropy to obtain the new HSEN model, (2) compensating the cognitive behavior of the domain experts in the practical process with a latent feature-based fuzzy model, and (3) selectively fusing with multiscale mechanical frequency spectrum and the multiple experts’ knowledge. To show the advantages of the novel soft-sensing method, we used the shell vibration and acoustical data of a laboratory-scale ball mill to verify it.

2 Mill load in grinding process

The objective of the grinding process is to crush the broken ore through the ball mill into a qualified mineral pulp. It provides selected materials for subsequent sorting operations. Closed grinding circuit (GC) of two stages is widely used in mineral grinding in China. The MLP of Stage I is examined and shown in Fig. 1.

Figure 1 shows that the fresh ore is fed to the conveyor belt by the feeder, and then transported to the wet pre-concentration, which enters Stage I of GC. The wet pre-concentration selects the useful ore and discards the tailings. Then, the recycled slurry from the hydrocyclone is mixed with the useful core. The steel balls and the mill water are added periodically. Thereafter, the mixture enters the ball mill. The rotating ball mill drives the steel ball to impact and crush the ore. The slurry drops down into the sump and is mixed with sump water pumped into the hydrocyclone. The hydrocyclone divides the slurry into overflow with fine particle size and coarse grit. The former part is transported to stage II of GC. The latter is the recycle to the wet preconcentration.

Grinding process is a “bottleneck” operation in mineral processing. The mill load cannot be measured directly. In real application, the operation experts identify the familiar load and internal parameters of a specific mill by their own listening experience. It is shown that the human ear is essentially a set of adaptive band-pass filters. From a certain point of view, the process of expert “listening” reasoning and recognition can be understood as a layer-by-layer cognitive process of human ear band-pass filtering ability, human brain feature extraction ability, and expert experience rule reasoning ability.

The acoustic signals used by operation experts are adaptively decomposed into different sub-signals through the human ear band-pass filter. After feature extraction, the fuzzy values (such as high, low, and moderate) of a certain band signal are obtained. The fuzzy identification of mill load is realized based on the knowledge of different experts. Obviously, this is a selective information fusion process based on human expert experience for uncertain reasoning, but the differences in expert experience and their limited energy causes difficulty in ensuring that the mill can be operated in the optimal load status for a long time. Thus, one focus of this paper is how to simulate the identification process of operational experts based on existing technology.

The vibration signal has higher sensitivity and reliability, but the operation experts in the industrial field cannot use the signal directly and effectively with human ears. Therefore, with the modern signal analysis technology and modeling technology, the vibration and acoustic signals of the ball mill can be effectively fused to realize the simulation of operation expert cognitive process.

3 Heterogeneous SEN-based MLP forecasting

A new heterogeneous SEN (HSEN)-based modeling strategy is proposed by using multiscale mechanical frequency spectrum for soft-sensing MLP, which is shown in Fig. 2. This strategy includes two modules, namely, multiscale frequency spectral transformation and HSEN forecasting. The latter consists of a frequency spectrum-based SEN KPLS model, a latent feature-based SEN fuzzy model, and error information entropy-based combination sub-modules.
In Fig. 2, superscripts t and f represent the time domain and frequency domain, respectively; and subscripts V and A represent the shell vibration and acoustical signals. An explanation of the nomenclature used in this study is given in Table 1.

The functions of different modules are described as follows:

1. Multiscale frequency spectral transformation module: the adaptive decomposition of shell vibration and acoustic signals is conducted to simulate the band-pass filtering function of the human ears for multicomponent mechanical signals, and then transform them into multiscale frequency spectra. EEMD algorithm is used to adaptively decompose the original vibration and acoustic signals into stationary sub-signals (IMFs) with different timescales. Although FFT is not suitable for decomposing non-stationary and multicomponent original signals, it is good at transforming the IMFs into frequency spectra.

2. HSEN forecasting model module: frequency spectrum-based SEN KPLS model sub-module establishes candidate sub-models based on different frequency spectra by using the KPLS algorithm, and sub-models are selected and combined to obtain SEN KPLS models. The latent feature-based SEN fuzzy model sub-module extracts the latent features of multiscale spectra, establish candidate fuzzy sub-models based on these latent features, and select and combine fuzzy sub-models to obtain SEN fuzzy models. Finally, the error information entropy-based combination sub-module fuses the aforementioned SEN model based on the information entropy of prediction errors.

3.1 Multiscale frequency spectral transformation

The EMD algorithm lacks theoretical foundation and has an end effect, and has difficulty determining the decomposition termination criterion. Losing the physical meaning of the IMF sub-signals because of mode mixing effect is a prominent problem. Ensemble EMD overcomes this problem through auxiliary noise analysis technology by using two parameters, namely, additional noise $A_{noise}$ and ensemble number $M$. Their relationship can be described as follows (Tang et al. 2018b):

$$\ln e_{EEMD} + \frac{A_{noise}}{2} \ln M = 0$$

where $e_{EEMD}$ is the error between the original signal and corresponding IMFs.
EEMD can be described as follows: (1) $M$ and $A_{\text{noise}}$ are initialized, (2) $A_{\text{noise}}$ is added to the original signal, (3) the EMD of the new signal for $M$ times is applied, (4) the average EMD value for $M$ times is calculated as the final EEMD result.

The EEMD result of the shell vibration signal can be expressed as

$$x^t_V = \sum_{j=1}^{J_t} x^t_{\text{VEEMD}j} + r^t_{\text{VEEMD}j},$$  \hspace{1cm} (2)$$

The relationship between EEMD and EMD can be expressed as

$$\begin{align*}
x^t_{\text{VEEMD}j} &= \frac{1}{M} \sum_{m=1}^{M} (x^t_{\text{EMD}j})_m, \\
r^t_{\text{VEEMD}j} &= \frac{1}{M} \sum_{m=1}^{M} (r^t_{\text{EMD}j})_m,
\end{align*}$$  \hspace{1cm} (3)$$

where $x^t_{\text{VEEMD}j}$ is $j$th IMF of $m$th EMD, and $r^t_{\text{VEEMD}j}$ is the residual error after decomposition.

The decomposition of the mill shell vibration and acoustic signals can be expressed as

These decomposed signals are sorted according to their frequency, that is, from high to low. Frequency domain analysis is required because extracting the valuable information within the time domain is difficult. Thus, each IMF is transformed into the frequency domain through FFT. The relationship between time and frequency domain sub-signals can be expressed as

$$X^t_V = \{(X^t_V)_n\}_{n=1}^{N} \Rightarrow \left\{X^t_{\text{VEEMD}j} = \{(X^t_{\text{VEEMD}j})_n\}_{n=1}^{N}\right\},$$  \hspace{1cm} (4)$$

$$X^t_A = \{(X^t_A)_n\}_{n=1}^{N} \Rightarrow \left\{X^t_{\text{AEEMD}j} = \{(X^t_{\text{AEEMD}j})_n\}_{n=1}^{N}\right\},$$  \hspace{1cm} (5)$$
where $J = J_V + J_A$ is the number of multiscale spectra of vibration and acoustic signals after the combination.

### 3.2 HSEN forecasting model

The root mean square relative error (RMSRE) of the HSEN-based forecasting model can be expressed as

$$
e_{\text{main}} = \sqrt{\frac{1}{k} \sum_{i=1}^{k} \left( \frac{y^i - \hat{y}^i}{y^i} \right)^2}$$

$$= \sqrt{\frac{1}{k} \sum_{i=1}^{k} \left( \frac{y^i \left( \sum_{j=1}^{J} \left[ w_{\text{fuzzy}} \cdot f_{\text{fuzzy}}^{\text{sel}}(\cdot)_{j=1} + w_{\text{latent}} \cdot f_{\text{latent}}^{\text{sel}}(\cdot)_{j=1} \right] \right) + \left( \sum_{j=1}^{J} \left[ w_{\text{fuzzy}} \cdot f_{\text{fuzzy}}^{\text{sel}}(\cdot)_{j=1} + w_{\text{latent}} \cdot f_{\text{latent}}^{\text{sel}}(\cdot)_{j=1} \right] \right) y^i}{y^i}}$$

where $k$ is the number of samples, $w_{\text{fuzzy}}$ and $w_{\text{latent}}$ are the weighting coefficients of fuzzy inference SEN model and latent structure SEN model, $f_{\text{fuzzy}}^{\text{sel}}(\cdot)_{j=1}$ and $f_{\text{latent}}^{\text{sel}}(\cdot)_{j=1}$ are the $j_{\text{sel}}$th fuzzy inference SEN sub-model and $j_{\text{lat}}$th latent structure SEN model, $w_{\text{fuzzy}}$ and $w_{\text{latent}}$ are the weighting coefficients of the corresponding SEN sub-model, and $J_{\text{fuzzy}}$ and $J_{\text{latent}}$ are the ensemble size of the fuzzy reasoning SEN model and latent structure SEN model.

The preceding HSEN model includes a dual-layer SEN. For the first layer, we should determine the number of sub-models $(J_{\text{fuzzy}}^{\text{sel}})$, where $(f_{\text{fuzzy}}^{\text{sel}}(\cdot)_{j=1})$ and $(f_{\text{latent}}^{\text{sel}}(\cdot)_{j=1})$ and their weighting coefficients $(w_{\text{fuzzy}}^{\text{sel}})$ and $(w_{\text{latent}}^{\text{sel}})$ should be selected optimally, and the $w_{\text{fuzzy}}^{\text{entropy}}$ and $w_{\text{latent}}^{\text{entropy}}$ in the second layer should be determined.

Therefore, by maximizing the optimization objective, the solution process can be expressed as the following optimization problems:

### Table 1 Nomenclature

| Symbol | Description |
|--------|-------------|
| $x_j$  | $j$th vibration sub-signal in frequency domain |
| $y^j_{\text{latent}}$ | Output of candidate KPLS sub-models based on $j$th frequency spectrum |
| $y^j_{\text{fuzzy}}$ | Output of candidate fuzzy inference sub-models based on latent features of $j$th frequency spectrum |
| $y_{\text{latent}}$ | Output of SEN latent structural mapping model |
| $\hat{y}$ | Output of obtained HSEN-based soft-sensing model |
The heterogeneous selective ensemble learning model for mill load parameters forecasting by using KPLS sub-models. Taking the aforementioned method are used to establish here, learning parameters for various candidate sub-models to simplify performance for SEN models, we select the same learning reasoning model. To ensure better diversity and modeling learning parameters of the latent structure model and fuzzy forecasting problem is transformed into optimizing the algorithm as

\[
\begin{aligned}
\text{max} & \quad e_{\text{main}} = \theta_{\text{th}} - \sqrt{\frac{1}{k} \sum_{j=1}^{k} \left( y_j - \left( w_{\text{Fuzzy entropy}} \sum_{j_{\text{sel}}=1}^{J} w_{j_{\text{sel}}} f_{\text{Fuzzy}}(\cdot)_{j_{\text{sel}}} + w_{\text{Latent entropy}} \sum_{j_{\text{sel}}=1}^{J} w_{j_{\text{sel}}} f_{\text{Latent}}(\cdot)_{j_{\text{sel}}} \right) \right)^2
\end{aligned}
\]

\[
\sum_{j_{\text{sel}}=1}^{J} w_{j_{\text{sel}}} = 1, \quad 0 \leq w_{j_{\text{sel}}} \leq 1, \quad 1 < j_{\text{sel}} < J_{\text{sel}} < J,
\]

\[
\sum_{j_{\text{sel}}=1}^{J} w_{j_{\text{latent}}} = 1, \quad 0 \leq w_{j_{\text{latent}}} \leq 1, \quad 1 < j_{\text{sel}} < J_{\text{sel}} < J
\]

s.t.

\[
w_{\text{Fuzzy entropy}} + w_{\text{Latent entropy}} = 1
\]

\[
\{f_{\text{Fuzzy}}(\cdot)_{j_{\text{sel}}}\}_{j_{\text{sel}}=1}^{J} \in \{f_{\text{can}}(\cdot)\}_{j=1}^{J}
\]

\[
\{f_{\text{Latent}}(\cdot)_{j_{\text{sel}}}\}_{j_{\text{sel}}=1}^{J} \in \{f_{\text{can}}(\cdot)\}_{j=1}^{J}
\]

From the perspective of global optimization, the HSEN forecasting problem is transformed into optimizing the learning parameters of the latent structure model and fuzzy reasoning model. To ensure better diversity and modeling performance for SEN models, we select the same learning parameters for various candidate sub-models to simplify the selection process of these model learning parameters.

### 3.2.1 Frequency spectrum-based SEN KPLS model

Here, J multiscale frequency spectra acquired through the aforementioned method are used to establish J candidate KPLS sub-models. Taking jth spectrum \((X_j)_{l=1}^{k}\) as an example, we use the following to realize nonlinear mapping:

\[
K^j_{\text{Ker}} = \Phi((x_j)_l^T \Phi((x_j)_m), \quad l, m = 1, 2, \ldots, k
\]

(10)

where Ker is the kernel parameter.

The \(K^j_{\text{Ker}}\) is obtained through the centralization of kernel matrix \(K^j_{\text{Ker}}\) by using the following formula:

\[
K^j_{\text{Ker}} = (I - \frac{1}{k} 1_k 1_k^T) K^j_{\text{Ker}} (I - \frac{1}{k} 1_k 1_k^T)
\]

(11)

where I is a k-dimensional unit matrix, and 1_k is a vector with a value of 1 and length of k.

The output of a candidate KPLS sub-model based on spectrum \(X_j\) can be expressed on the basis of the KPLS algorithm as

\[
y^\text{latent}_j = \tilde{K}^j_{\text{Ker}} U_j (K^j_{\text{Ker}} K^j_{\text{Ker}} U_j)^{-1} T^T y
\]

(12)

where \(T_j\) and \(U_j\) are the latent score matrices of the input and output data obtained on the basis of the KPLS algorithm.

The test samples are calibrated by using the following formula,

\[
\tilde{K}_{ij} = (K_{ij} - \frac{1}{k} 1_{k_t} 1_{k_t}^T K^j_{\text{Ker}}) (I - \frac{1}{k} 1_k 1_k^T)
\]

(13)

where \(K_{ij}\) is the kernel matrix of the test samples, \(K_{ij} = K_j((X_j)_l, (X_j)_m)\) and \((X_j)_m\) are the training data, \(k_t\) is the number of testing samples, and 1_{k_t} is a vector with a value of 1 and length of \(k_t\).

The candidate sub-model output of testing sample \((X_j)_l\) can be expressed as

\[
y^\text{latent}_{ij} = \tilde{K}_{ij} U_j (T_{ij}^T K^j_{\text{Ker}} U_j)^{-1} T_{ij}^T y_i
\]

(14)

The number of latent variables, which refer to the number of layers of the KPLS model denoted as \(h\), should be determined.

The establishment of the jth candidate KPLS model can be expressed as

\[
\{((X_j, y_j)_{l=1}^{k})_{h=1}^{k} \}^{\text{KPLS}}_{\text{Ker}, h} \longrightarrow f^\text{can}_{\text{KPLS}}(\cdot)_j
\]

(15)

The set of all J candidate KPLS sub-models can be expressed as
\[ S_{\text{KPLS}}^{\text{Can}} = \{ f_{\text{KPLS}}^{\text{Can}}(\cdot) \}_{j=1}^{J} \]  

where \( S_{\text{KPLS}}^{\text{Can}} \) is the set of all candidate KPLS sub-models. The BB-based SEN is used to select and combine the ensemble KPLS sub-models. The candidate KPLS sub-models and KPLS sub-models are expressed as \( \{ f_{\text{KPLS}}(\cdot) \}_{j=1}^{\text{pl latent}} \), and the relationship between ensemble KPLS sub-models and candidate KPLS sub-models is expressed as,

\[ S_{\text{KPLS}}^{\text{Sel}} = \{ f_{\text{KPLS}}(\cdot) \}_{j=1}^{\text{sel}} \subseteq S_{\text{KPLS}}^{\text{Can}} \quad j_{\text{sel}} \leq J \]  

where \( S_{\text{KPLS}}^{\text{Sel}} \) is the set of ensemble KPLS sub-models, \( j_{\text{sel}} = 1, 2, \ldots, \text{pl latent} \), and \( j_{\text{sel}} \) is the ensemble size of SEN KPLS models.

The AWF algorithm is used to calculate the weighting coefficients of ensemble KPLS sub-models by using the following formula:

\[ w_{j_{\text{sel}}}^{\text{latent}} = 1 / \left( (\sigma_{j_{\text{sel}}})^2 \sum_{j_{\text{sel}}=1}^{\text{pl latent}} 1 / (\sigma_{j_{\text{sel}}})^2 \right) \]  

where \( \sum_{j_{\text{sel}}=1}^{\text{pl latent}} w_{j_{\text{sel}}}^{\text{latent}} = 1 \) and \( 0 \leq w_{j_{\text{sel}}}^{\text{latent}} \leq 1 \), and \( w_{j_{\text{sel}}}^{\text{latent}} \) is the weighting coefficient that corresponds to the \( j_{\text{sel}} \)-th ensemble KPLS sub-model, \( \sigma_{j_{\text{sel}}} \) is the standard deviation of the sub-models’ output value \( \{ (y_{j_{\text{sel}}}^{\text{latent}}) \}_{j=1}^{J} \), and \( k \) is the modeling sample number.

The output value \( \hat{y}_{\text{latent}} \) of the SEN KPLS model is calculated by using the following formula:

\[ \hat{y}_{\text{latent}} = \sum_{j_{\text{sel}}=1}^{\text{pl latent}} w_{j_{\text{sel}}}^{\text{latent}} f_{\text{KPLS}}(\cdot)_{j_{\text{sel}}} = \sum_{j_{\text{sel}}=1}^{\text{pl latent}} w_{j_{\text{sel}}}^{\text{latent}} y_{\text{latent}} \]  

where \( y_{\text{latent}}^{j_{\text{sel}}} \) is the output of the \( j_{\text{sel}} \)-th KPLS ensemble sub-model.

The establishment of the aforementioned SEN KPLS model can be expressed as

\[ \{ f_{\text{KPLS}}^{\text{Can}}(\cdot) \}_{j=1}^{J} \xrightarrow{}\text{BBSEN} \{ f_{\text{KPLS}}^{\text{Can}}(\cdot) \}_{j=1}^{\text{pl latent}} \xrightarrow{}\text{BBSEN} \{ f_{\text{KPLS}}^{\text{Sel}}(\cdot) \}_{j_{\text{sel}}=1}^{\text{pl latent}} \Rightarrow \]  

\[ \hat{y}_{\text{latent}} = \sum_{j_{\text{sel}}=1}^{\text{pl latent}} w_{j_{\text{sel}}}^{\text{latent}} y_{\text{latent}} \]  

3.2.2 Latent feature-based SEN fuzzy model

The input of the SEN fuzzy model consists of latent features of multiscale frequency spectra. Here, the same number of latent variables is selected for each frequency spectrum, and the number is denoted as \( h' \). The latent feature extracted from the \( j \)-th spectrum \( \{ (x_{j}) \}_{l=1}^{k} \) is expressed on the basis of the latent feature extraction method (Tang et al. 2014) as follows:

\[ z_{j} = [z_{j1}, \ldots, z_{jh'}] \]  

The latent feature subset extracted from all multiscale frequency spectra is denoted as \( \{ z_{j} \}_{j=1}^{J} \). The extracted latent features are used to construct the candidate fuzzy sub-models, and the \( j \)-th candidate fuzzy sub-model can be established as

\[ \{ (z_{j1}, \ldots, z_{jh'}, y) \}_{j=1}^{k} \xrightarrow{L} f_{\text{Fuzzy}}(\cdot)_{j} \]  

where \( L \) is the clustering threshold set during the establishment of fuzzy models.

The set of all \( J \) candidate fuzzy sub-models can be expressed as

\[ S_{\text{Fuzzy}}^{\text{Can}} = \{ f_{\text{Fuzzy}}^{\text{Can}}(\cdot) \}_{j=1}^{J} \]  

where \( S_{\text{Fuzzy}}^{\text{Can}} \) is the set of all candidate fuzzy sub-models. Here, all the selected ensemble fuzzy sub-models are expressed as \( \{ f_{\text{Fuzzy}}^{\text{Sel}}(\cdot) \}_{j_{\text{sel}}=1}^{\text{fuzzy}} \), and the relationship between ensemble fuzzy models and candidate fuzzy sub-models can be expressed as

\[ S_{\text{Fuzzy}}^{\text{Sel}} = \{ f_{\text{Fuzzy}}^{\text{Sel}}(\cdot) \}_{j_{\text{sel}}=1}^{\text{fuzzy}} \subseteq S_{\text{Fuzzy}}^{\text{Can}} \quad j_{\text{sel}} \leq J \]  

where \( S_{\text{Fuzzy}}^{\text{Sel}} \) is the set of ensemble fuzzy sub-models, \( j_{\text{sel}} = 1, 2, \ldots, j_{\text{sel}} \), and \( j_{\text{sel}} \) is the ensemble size of the SEN fuzzy model.

The AWF algorithm is used again to calculate the weighting coefficients of ensemble fuzzy sub-models by using the following formula:

\[ w_{j_{\text{sel}}}^{\text{fuzzy}} = 1 / \left( (\sigma_{j_{\text{sel}}})^2 \sum_{j_{\text{sel}}=1}^{\text{fuzzy}} 1 / (\sigma_{j_{\text{sel}}})^2 \right) \]  

where \( \sum_{j_{\text{sel}}=1}^{\text{fuzzy}} w_{j_{\text{sel}}}^{\text{fuzzy}} = 1 \) and \( 0 \leq w_{j_{\text{sel}}}^{\text{fuzzy}} \leq 1 \), \( w_{j_{\text{sel}}}^{\text{fuzzy}} \) is the weighting coefficient that corresponds to the \( j_{\text{sel}} \)-th ensemble fuzzy sub-model, \( \sigma_{j_{\text{sel}}} \) is the standard deviation of the output values \( \{ y_{j_{\text{sel}}}^{\text{k}} \}_{l=1}^{k} \) of fuzzy inference ensemble sub-models, and \( k \) is the modeling sample number.
Heterogeneous selective ensemble learning model for mill load parameters forecasting by using...

The BBSEN algorithm was used to select and combine the ensemble fuzzy sub-models. Its output value \( \hat{y}_{\text{fuzzy}} \) is calculated as

\[
\hat{y}_{\text{fuzzy}} = \sum_{j_{\text{sel}}=1}^{n_{\text{sel}}} w_{j_{\text{sel}}} \hat{y}_{\text{fuzzy}}^{j_{\text{sel}}} = \sum_{j_{\text{sel}}=1}^{n_{\text{sel}}} w_{j_{\text{sel}}} \hat{y}_{j_{\text{sel}}}^{\text{fuzzy}}
\]

(26)

where \( \hat{y}_{j_{\text{sel}}}^{\text{fuzzy}} \) is the output of the \( j_{\text{sel}} \)th ensemble fuzzy sub-model.

The establishment of the aforementioned SEN fuzzy model can be expressed as

\[
\begin{align*}
\{f_{\text{can}}^{\text{valid}}(\cdot)\}_{j=1}^{J} & \quad \{x_{\text{valid}}, y_{\text{valid}}\} \\
\{f_{\text{valid}}^{\text{valid}}(\cdot)\}_{l=1}^{L} & \quad \{f_{\text{can}}^{\text{valid}}(\cdot)\}_{j=1}^{J} \\
\{w_{\text{can}}^{\text{valid}}(\cdot)\}_{l=1}^{L} & \quad \{f_{\text{valid}}^{\text{valid}}(\cdot)\}_{l=1}^{L} \\
\Rightarrow \hat{y}_{\text{fuzzy}} = \sum_{l=1}^{L} w_{l} \hat{y}_{l}^{\text{fuzzy}}
\end{align*}
\]

(27)

### 3.2.3 Error information entropy-based combination

The two types of heterogeneous models can be fused based on information entropy. The weighting coefficients of the two types of heterogeneous SEN models are determined on the basis of the output values of the training data (Tang et al. 2012a).

The relative error of the predicted output of the \( j_{\text{Entropy}} \)th ensemble SEN sub-model at each time \( l \) is calculated as

\[
\epsilon_{\text{entropy}}^{j_{\text{Entropy}}} = \begin{cases} 
\frac{|y_l - \hat{y}_{\text{Entropy}}^{j_{\text{Entropy}}}|}{y_l} & 0 \leq \frac{|y_l - \hat{y}_{\text{Entropy}}^{j_{\text{Entropy}}}|}{y_l} < 1 \\
1 & \frac{|y_l - \hat{y}_{\text{Entropy}}^{j_{\text{Entropy}}}|}{y_l} \geq 1
\end{cases}
\]

(28)

where \( \hat{y}_{\text{Entropy}}^{j_{\text{Entropy}}} \) is the output value of the modeling sample at time \( l \) using the \( j_{\text{Entropy}} \)th SEN sub-model. In this study, weighing \( j_{\text{Entropy}} = 1, 2 \), which represents the latent structure SEN model and fuzzy reasoning SEN model, respectively, i.e., \( \hat{y}_{\text{Entropy}}^{j_{\text{Entropy}}} = \hat{y}_{\text{latent}}^{j_{\text{Entropy}}} \) or \( \hat{y}_{\text{Entropy}}^{j_{\text{Entropy}}} = \hat{y}_{\text{fuzzy}}^{j_{\text{Entropy}}} \) with \( j_{\text{Entropy}} = 1 \) or 2.

The proportion \( p_{\text{fuzzy}}^{j_{\text{Entropy}}} \) of the relative error of the predicted output of the \( j_{\text{Entropy}} \)th ensemble SEN sub-model is calculated as

\[
p_{\text{fuzzy}}^{j_{\text{Entropy}}} = \frac{\epsilon_{\text{entropy}}^{j_{\text{Entropy}}}}{k} \left( \epsilon_{\text{entropy}}^{j_{\text{Entropy}}} \right)
\]

(29)

Entrophy \( E_{\text{entropy}}^{j_{\text{Entropy}}} \) of the relative error of the predicted output of the \( j_{\text{Entropy}} \)th ensemble SEN sub-model is calculated as

\[
E_{\text{entropy}}^{j_{\text{Entropy}}} = \frac{1}{\ln k} \sum_{l=1}^{k} p_{\text{fuzzy}}^{j_{\text{Entropy}}} \ln p_{\text{fuzzy}}^{j_{\text{Entropy}}}
\]

(30)

The weighting coefficient of the \( j_{\text{Entropy}} \)th ensemble SEN sub-model is calculated as

\[
W_{j_{\text{Entropy}}} = \frac{1}{j_{\text{Entropy}} - 1} \left( 1 - \frac{E_{\text{entropy}}^{j_{\text{Entropy}}}}{\sum_{j_{\text{Entropy}}=1}^{J_{\text{Entropy}}} (1 - E_{\text{entropy}}^{j_{\text{Entropy}}})} \right)
\]

(31)

where \( \sum_{j_{\text{Entropy}}=1}^{J_{\text{Entropy}}} W_{j_{\text{Entropy}}} = 1 \), and \( J_{\text{Entropy}} \) is the number of ensemble SEN sub-models.

In this study, \( J_{\text{Entropy}} = 2 \), which indicates that the following corresponding relationships exist between the SEN KPLS and SEN fuzzy models using the aforementioned weighting algorithm:

\[
\begin{align*}
W_{\text{Entropy}}^{\text{latent}} &= W_{1_{\text{Entropy}}} \\
W_{\text{Entropy}}^{\text{fuzzy}} &= W_{2_{\text{Entropy}}}
\end{align*}
\]

(32)

\[
\begin{align*}
\hat{y}_{\text{Entropy}}^{\text{entropy}} &= \hat{y}_{\text{latent}}^{j_{\text{Entropy}}} \\
\hat{y}_{\text{Entropy}}^{\text{fuzzy}} &= \hat{y}_{\text{fuzzy}}^{j_{\text{Entropy}}}
\end{align*}
\]

(33)

where \( \hat{y}_{\text{latent}} = \{y_{\text{latent}}^{i}\}_{i=1}^{k} \), and \( \hat{y}_{\text{fuzzy}} = \{y_{\text{fuzzy}}^{i}\}_{i=1}^{k} \).

The output of the HSEN model can be expressed as

\[
\hat{y} = w_{\text{entropy}} \hat{y}_{\text{fuzzy}} + w_{\text{entropy}} \hat{y}_{\text{latent}}
\]

(34)

### 4 Experimental validation

#### 4.1 Data description

The experiments were conducted on a XMQL 420 × 450 ball mill, where the outer diameter and length of its shell were 460 mm. The mill was driven by a 2.12 kw three-phase motor, with 80 kg maximum steel ball loading capacity, 10 kg/h power grinding capacity, and 57 r/min revolving speed. An opening, which was used to add the steel ball, mill, and water load, was in the middle of the mill. The materials used in this experiment included copper ores with diameter less than 6 mm and density of 4.2 t/m³. Steel balls with diameters of 30, 20, and 15 mm were used as grinding media, and the matching ratio was 3:4:3.

The data acquisition system was installed on the mill shell and mainly consisted of acceleration sensor and DSP equipment. In this study, the acquisition frequencies of mill shell vibration and acoustic signals were 51,200 Hz and 8,000 Hz, respectively.
Due to the limited modeling samples, the data are partitioned as the training and testing ones, which are the same as those reported by Tang et al. (2018b) and Tang et al. (2016). In their works, bootstrap algorithm is used to address the sample partition problem. The results are almost the same as those of the training and testing methods.

4.2 Experimental results

4.2.1 Multiscale frequency spectrum transformation results

The parameters $A_{\text{noise}} = 0.1$ and $M = 10$ are selected based on the work of Tang et al. (2018b). The original shell vibration and acoustic signals of the four rotating periods of the mill are decomposed into time-domain sub-signals with different timescales based on EEMD. These IMFs

![Graph](image-url)
with different timescales are transformed into multiscale frequency spectra by FFT. The spectra of the former eight shell vibration sub-signals (VIMF) and acoustic sub-signals (AIMF) are shown in Fig. 3.

The figure indicates that these multiscale sub-signals, which have high-dimensional features, are sorted from high to low frequency.

Table 2 Statistical table of ensemble sub-models of MBVR SEN KPLS model

| No | Ensemble size | Sub-model No | Testing RMSRE |
|----|---------------|--------------|---------------|
| 1  | 2             | 2 3          | 0.4458        |
| 2  | 4             | 6 4 2 3      | 0.4497        |
| 3  | 8             | 9 7 5 1 6 4 2 3 | 0.4643       |
| 4  | 5             | 1 6 4 2 3    | 0.4752        |
| 5  | 7             | 7 5 1 6 4 2 3 | 0.4843        |
| 6  | 3             | 4 2 3        | 0.4852        |
| 7  | 6             | 5 1 6 4 2 3  | 0.4892        |
| 8  | 9             | 10 9 7 5 1 6 4 2 3 | 0.4921      |
| 9  | 10            | 18 10 9 7 5 1 6 4 2 3 | 0.4955      |

Bold values indicate the best one

Table 3 Statistical table of ensemble sub-models of PD SEN KPLS model

| No | Ensemble size | Sub-model No | Testing RMSRE |
|----|---------------|--------------|---------------|
| 1  | 2             | 2 3          | 0.6162        |
| 2  | 3             | 6 2 3        | 0.6223        |
| 3  | 4             | 1 6 2 3      | 0.6252        |
| 4  | 5             | 4 1 6 2 3    | 0.6289        |
| 5  | 6             | 8 4 1 6 2 3  | 0.6346        |
| 6  | 7             | 1 4 8 4 1 6 2 3 | 0.6402      |
| 7  | 8             | 5 1 4 8 4 1 6 2 3 | 0.6457      |
| 8  | 9             | 16 5 1 4 8 4 1 6 2 3 | 0.6522     |
| 9  | 10            | 11 16 5 1 4 8 4 1 6 2 3 | 0.6593     |

The figure indicates that these multiscale sub-signals, which have high-dimensional features, are sorted from high to low frequency.
4.2.2 Frequency spectrum-based SEN KPLS model results

The number of kernel latent values (KLVs) of the KPLS model determines the model structure and generalization performance. The relationships of KLV number with MBVR, PD, and CVR soft-sensing models are shown in Figs. 4, 5 and 6.

According to Figs. 4, 5 and 6, the number of KLVs of the MBVR, PD and CVR KPLS model is 3, 8 and 6, respectively.

RBF kernel function and kernel parameters are selected. The relationship between the kernel parameter value and the RMSRE of the SEN KPLS model is shown in Figs. 7, 8 and 9.

The learning parameters of MBVR, PD, and CVR soft-sensing models with SEN KPLS algorithm are determined based on the aforementioned results, and the corresponding ensemble sub-models are shown in Tables 2, 3 and 4. The multiscale spectra that correspond to the ensemble sub-models numbered as 1–10 are VIMF1–VIMF10, and those that correspond to the ensemble sub-models numbered as 11–20 are AIMF1–AIMF10.

| No | Ensemble size | Sub-model No | Testing RMSRE |
|----|---------------|--------------|---------------|
| 1  | 2             | 1 3          | 0.3078        |
| 2  | 3             | 2 1 3        | 0.3104        |
| 3  | 6             | 9 6 4 2 1 3  | 0.3192        |
| 4  | 5             | 6 4 2 1 3    | 0.3318        |
| 5  | 4             | 4 2 1 3      | 0.3329        |
| 6  | 7             | 5 9 6 4 2 1 3| 0.3330        |
| 7  | 8             | 20 5 9 6 4 2 1 3 | 0.3353 |
| 8  | 9             | 7 20 5 9 6 4 2 1 3 | 0.3421 |
| 9  | 10            | 14 7 20 5 9 6 4 2 1 3 | 0.3477 |

Bold values indicate the best one
The aforementioned results indicate that the ensemble sub-models selected by the SEN KPLS model mainly come from shell vibration signals.

4.2.3 Latent feature-based SEN fuzzy model results

The relationships between the clustering threshold and the RMSRE of the SEN fuzzy models are shown in Figs. 10, 11 and 12.

Table 5: Statistical results of ensemble sub-models selected by MBVR SEN fuzzy model

| No | Ensemble size | Sub-model No       | Testing RMSRE |
|----|---------------|-------------------|---------------|
| 1  | 4             | 5 16 8 11         | 0.5037        |
| 2  | 2             | 8 11              | 0.5625        |
| 3  | 16            | 6 15 2 14 7 1 10 20 17 19 18 13 5 16 8 11 | 0.6080        |
| 4  | 3             | 16 8 11           | 0.6124        |
| 5  | 5             | 13 5 16 8 11      | 0.6161        |
| 6  | 12            | 7 1 10 20 17 19 18 13 5 16 8 11 | 0.6171        |
| 7  | 17            | 12 6 15 2 14 7 1 10 20 17 19 18 13 5 16 8 11 | 0.6221        |
| 8  | 14            | 2 14 7 1 10 20 17 19 18 13 5 16 8 11 | 0.6264        |
| 9  | 19            | 4 3 9 12 6 2 14 7 1 10 20 17 19 18 13 5 16 8 11 | 0.6281        |

Bold values indicate the best one

The relationships between kernel parameters and RMSRE of MBVR SEN fuzzy inference model

Fig. 13

The relationships between kernel parameters and RMSRE of PD SEN fuzzy inference model

Fig. 14

The relationships between kernel parameters and RMSRE of CVR SEN fuzzy model

Fig. 15
According to Figs. 10, 11 and 12, the clustering threshold value of the MBVR, PD, and CVR KPLS models is 0.5, 0.3, and 0.25, respectively.

The kernel parameters with RBF kernel are selected for feature extraction. The relationships between the kernel parameter values and RMSRE of the SEN fuzzy model are shown in Figs. 13, 14 and 15.

The learning parameters of MBVR, PD, and CVR soft-sensing models with the SEN fuzzy algorithm based on the aforementioned results are determined, and the corresponding ensemble fuzzy sub-models are shown in Tables 5, 6 and 7.

Tables 5, 6 and 7 show that half of the ensemble sub-models selected by the fuzzy SEN model come from shell vibration signals, and the other come from acoustic signals.

### 4.2.4 Error information entropy-based combination results

The weighing coefficients of the SEN KPLS and SEN fuzzy models for the MBVR model are 0.6148 and 0.3851,

![Fig. 16](image_url)  
**Fig. 16** RMSRE comparison of different models for MBVR HSEN model

According to Figs. 10, 11 and 12, the clustering threshold value of the MBVR, PD, and CVR KPLS models is 0.5, 0.3, and 0.25, respectively.

The kernel parameters with RBF kernel are selected for feature extraction. The relationships between the kernel parameter values and RMSRE of the SEN fuzzy model are shown in Figs. 13, 14 and 15.

The learning parameters of MBVR, PD, and CVR soft-sensing models with the SEN fuzzy algorithm based on the aforementioned results are determined, and the corresponding ensemble fuzzy sub-models are shown in Tables 5, 6 and 7.

Tables 5, 6 and 7 show that half of the ensemble sub-models selected by the fuzzy SEN model come from shell vibration signals, and the other come from acoustic signals.

### 4.2.4 Error information entropy-based combination results

The weighing coefficients of the SEN KPLS and SEN fuzzy models for the MBVR model are 0.6148 and 0.3851,
respectively, indicating that the contribution rate of the KPLS model is approximately twice that of the fuzzy model. The prediction errors of different models are shown in Fig. 16.

As shown in Fig. 16, the training error of the SEN KPLS model is much smaller than that of the testing data, indicating that overfitting exists during the training. The training error of the SEN fuzzy model is one-third of the testing error, indicating that overfitting is reduced and verifying the feasibility of the modeling based on the SEN fuzzy method. The training data error of the HSEN model is between that of the SEN KPLS model and that of the SEN fuzzy model. Meanwhile, the HSEN model has the minimum testing data error, thereby verifying that the fusion of the SEN KPLS and SEN inference models has improved the generalization performance.

For the PD model, the weighting coefficients of the SEN KPLS model and SEN fuzzy model are 0.4605 and 0.5395, respectively, indicating that the contribution rate of the SEN KPLS model is similar to that of the SEN fuzzy model. The prediction errors of various models are shown in Fig. 17.

In Fig. 17, the training error of the SEN KPLS model is smaller than that of the testing data, which is approximately 1/1000 of the testing training error, indicating that overfitting exists during the training of the SEN KPLS model. The training error of the SN fuzzy model is only one-third of the testing error. This finding indicates that the overfitting is reduced and verifies the feasibility of the modeling process based on the SEN fuzzy method. The training data error of the HSEN model is only half of the testing data error, and its training and testing precisions are improved compared with the SEN fuzzy and SEN KPLS models. These results indicate that the fusion of the two models has improved the generalization performance.

4.3 Comparative results

4.3.1 HSEN model and its sub-models

Considering the trade-off difficulty between the diversity and precisions of the sub-models, the proposed HSEN models select the model learning parameters through global optimization. In terms of ensemble models, the proposed MLP soft-sensing model is a double-layer ensemble with dual-layer sub-models. The first-layer sub-models refer to the SEN fuzzy and SEN KPLS models. The second-layer sub-models refer to the fuzzy and KPLS sub-models. The statistical results of different MLP models obtained by the proposed method are shown in Table 8.

Table 8 shows that:

1) The HSEN model has satisfactory generalization performance and fuses two heterogeneous SEN sub-models effectively. The modeling performance of the SEN KPLS model is stronger than that of the SEN fuzzy model, which is based on a previous analysis.

2) In the first-layer sub-models, for MBVR, the contribution rate of the SEN KPLS model is higher than that of the SEN fuzzy model. However, for the PD and CVR, the contribution rates are similar between two types of first-layer SEN sub-models.

3) For the contribution rates of the second-layer sub-models, three different MLP models select the

| MBVR  | {5, 16, 8,11} | 0.5037 | {2,3} | 0.3851 | 0.6148 | 0.08822 |
|-------|--------------|--------|-------|--------|--------|---------|
| PD    | {14, 5, 11, 4,8} | 0.6162 | {2,3} | 0.5395 | 0.4605 | 0.1355 |
| CVR   | {12, 6, 18, 5} | 0.5049 | {1,3} | 0.5598 | 0.4401 | 0.1697 |
| Average error | # | 0.5416 | # | 0.3666 | # | 0.1311 |

In Fig. 18, the training error of the SEN KPLS model is smaller than that of the testing data, indicating that overfitting exists during the training of the SEN KPLS model. The training error of the SEN fuzzy model is only one-third of the test error, indicating that overfitting declines to a certain degree and verifying the feasibility of the modeling process based on the SEN fuzzy method. The training data error of the HSEN model is only half of the testing data error, and its training and testing precisions are improved compared with the SEN fuzzy and SEN KPLS models. These results indicate that the fusion of the two models has improved the generalization performance.
frequency spectra of VIMF5 and VIMF3, and all come from shell vibration signals. The number of fuzzy second-layer sub-models selected by the first-layer sub-model (SEN fuzzy model) is large and contains both shell vibration and acoustic multiscale spectra, which immensely improve the precision of a HSEN model.

4) For the prediction accuracy of the MLP model, the MBVR model has minimum prediction error, indicating that this method is suitable for estimating MBVR, which aligns with the cognitive process of industrial operation experts. However, deep coupling mechanism remains to be elucidated.

4.3.2 HSEN model

The proposed soft-sensing method is compared with single-scale spectral feature extraction and selection method (Tang et al. 2012d), linear PLS model-based multiscale ensemble modeling method (Zhao et al. 2012), nonlinear KPLS model-based multiscale SEN modeling method (Tang et al. 2014), latent features and adaptive genetic algorithm-based single model method (Liu et al. 2014), selective fusion multi-condition samples and multi-source features method (Tang et al. 2018b), and dual-layer optimized selective information fusion using multi-source frequency spectrum features method (Tang et al. 2020). The results of the comparison are reported in Table 9.

Table 9 Test results of different soft measurement methods of MLPs (RMSREs)

| Method                                      | MBVR  | PD    | CVR   | Mean value | Remark          |
|---------------------------------------------|-------|-------|-------|------------|-----------------|
| Ref (Tang et al. 2012d)                     | 0.2711| 0.1158| 0.1368| 0.1745     | [FFT,V]         |
| Ref (Zhao et al. 2012)                      | 0.5454| 0.3074| 0.2527| 0.3685     | [EMD, V]        |
| Ref (Tang et al. 2014)                      | 0.3173| 0.1876| 0.1932| 0.2327     | [EMD,V,A]       |
| Ref (Liu et al. 2014)                       | 0.3620| 0.1346| 0.08992| 0.1955     | [EMD,PCA,GA optimization,V] |
| Ref (Tang et al. 2018b)                     | 0.1649| 0.07004| 0.1112| 0.1153     | [FFT/KPLS,GASEN,V,A] |
| Ref (Tang et al. 2020)                      | 0.1892| 0.08191| 0.1136| 0.1282     | [HVD,EMD,EEMD,GA optimization,V,A] |
| The method                                  | 0.08822| 0.1355| 0.1697| 0.1311     | [EEMD,V,A]      |

Bold values indicate the best one

1) The proposed method has the best average modeling performance for MBVR, and its RMSRE is 0.08822. One reason is that EEMD improves the adaptive decomposition precision of the shell vibration and acoustic signals. Another reason is that the HSEN modeling strategy that fuses the two heterogeneous SEN models can simulate and compensate the MLP estimation mechanism of operation experts more effectively compared with the previous method. Moreover, in practice, most of the operational experts can estimate MBVR for their familiar ball mill by hearing the acoustical signal. That is, the proposed method immensely improves MBVR in terms of modeling performance, which is in-line with the fuzzy estimation of the MBVR method in industrial practice.

2) The traditional single model (Liu et al. 2014) based on the shell vibration signal has the best modeling performance for CVR and its RMSER is 0.08992, which uses GA to jointly select the input latent features and learning parameters. The traditional single model established by Tang et al. (2012d) provides poor physical explanation.

3) The dual-layer SEN in terms of selective fusion multi-condition samples and multi-source features based on the single-scale frequency spectrum has the best modeling performance for PD and its RMSER is 0.07004. By using three multi-component signal decomposition approaches, Tang et al. obtained the secondary prediction performance of PD (Tang et al. 2020). Thus, different MLPs require various soft-sensing modeling strategies related to the production mechanism of MLPs on shell vibration/acoustical signals, and deep cognition remains to be elucidated.

4) This study and the method presented by Tang et al. (2018b, 2020) have greater complexity than the single-scale frequency spectrum ones in terms of mechanical signal decomposition. Based on this study as an example, for the signal decomposition in the model training phase, the calculation quantity of EEMD is $M$ times of EMD and $J*M$ times of FFT, where $M$ is the EMD execution times, and $J$ is the number of multiscale sub-signals decomposed by EMD. For the multiscale spectral feature selection, the proposed method has not conducted feature selection compared with the works of Tang et al.
(2012d, 2014). In the establishment phase of the soft-sensing model, the proposed HSEN model has double layers that are more complex than those of other single models and single-layer SEN modeling methods.

5) Among all the methods, only the proposed method uses the fuzzy inference mechanism. Thus, the proposed method has advantages in terms of the simulation and compensation of the cognition mechanism of operation experts.

5 Conclusions

In this study, we successfully combine two types of heterogeneous models, namely, frequency spectrum-based KPLS and latent feature-based fuzzy models, by using the prediction error information entropy. This soft-sensing method can simulate and compensate the fuzzy cognition behavior of industrial experts in the mineral grinding process. This novel model is a dual-layer ensemble construction strategy that can fuse multi-source information in various grinding conditions.

The simulation results are from a small sample of experimental data on a laboratory-scale ball mill. The experiment results show that the new model performs better than the other algorithms.

Our future work will apply the industrial ball-mill data closing to actual working conditions. Moreover, intelligent optimization algorithms will be used to select the learning parameters of the HSEN strategy.

Acknowledgements This study was funded by the Natural Science Foundation of China (Grant Nos. 62073006, 61973226, 61703089 and 61673097) and the Beijing Key Laboratory of Process Automation in Mining & Metallurgy (BGRIMM-KZSKL-2020-02).

Funding The authors have not disclosed any funding.

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Bai R, Chai TY (2009) Optimal control of ball mill loads based on data fusion and case reasoning. J Chem Ind Eng (china) 60(7):1746–1751

Bastien P, Bertrand F, Meyer N, Mauny-Bertrand M (2015) Deviance residuals-based sparsepls and sparse kernel pls regression for censored data. Bioinformatics 31(3):397–404

Granitto PM, Verdes PF, Ceccatto HA (2005) Neural network ensembles: evaluation of aggregation algorithms. Artif Intell 163(2):139–162

Huang NE, Shen Z, Long SR (1998) The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc Royal Soc London A 454:903–995

Lei YG, He ZJ, Zi YY (2009) Application of the EEMD method to rotor fault diagnosis of rotating machinery. Mech Syst Signal Process 23(4):1327–1338

Li Y, Shao C (2006) Application of a novel grey relation analysis algorithm to soft sensor [J]. Acta Automatica Sinica 32(2):311–317

Liu Z, Chai TY, Yu W, Tang J (2014) Multi-frequency signal modeling using empirical mode decomposition and PCA with application to mill load. Neurocomputing 169:392–402

Liu Z, Tang J, Chai TY, Yu W (2021) Selective ensemble modeling approach for mill load parameter forecasting based on multi-modal feature sub-sets [J]. Acta Automatica Sinica 47(8):11

Lya B, Jc C (2021) A method to identify wet ball mill’s load based on CEEMDAN, RCMDE and SRNN classification [J]. Miner. Eng. 165

Nicolai BM, Theron KI, Lammertyn J (2007) Kernel pls regression on wavelet transformed nir spectra for prediction of sugar content of apple. Chem Intel Lab Syst 85(2):243–252

Rai VK, Mohanty AR (2007) Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform. Mech Syst Signal Process 21(6):2007–2165

Rajaraman K, Mhatre N, Jain M, Postles M, Balakrishnan R, Robert D (2013) Low-pass filters and differential tympanal tuning in a paleotropical bushcricket with an unusually low frequency call. J Exp Biol 216(5):777–787

Robinson SR, Thompson S, Allen JB (2016) Effects of negative middle ear pressure on wideband acoustic immittance in normal-hearing adults. Ear Hear 37(4):452–464

Si GQ, Cao H, Zhang YB, Jia LX (2007) Barrel-type steel ball load detection method and device based on multi-sensor fusion, Xi’an, Xi’an Jiaotong University, Shanxi

Tang J, Zhao LJ, Yue H, Yu W, Chai TY (2011) Vibration analysis based on empirical mode decomposition and partial least squares. Procedia Eng 16:646–652

Tang J, Cai TY, Zhao LJ, Yue H, Zheng XP (2012a) Ball mill load parameters ensemble modeling method during the ore grinding process based on vibration spectra. Control Theory Appl 29(2):183–191

Tang J, Zhao LJ, Yu W, Yue H, Chai TY (2012b) Soft sensor for parameters of mill load based on multi-spectral segments PLS sub-models and on-line adaptive weighted fusion algorithm. Neurocomputing 78(1):38–47

Tang J, Zhao LJ, Jia J, Chai TY, Yu W (2012c) Selective ensemble modeling parameters of mill load based on shell vibration signal. Lect Notes Comput Sci 7367:489–497

Tang J, Chai TY, Yu W, Zhao LJ (2013) Modeling load parameters of ball mill in grinding process based on selective ensemble multisensor information. IEEE Trans Autom Sci Eng 10(3):726–740

Tang J, Chai TY, Yu W, Zhao LJ (2013) Modeling load parameters of ball mill in grinding process based on selective ensemble learning algorithm. Acta Automatica Sinica 40(9):1853–1866

Tang J, Chai TY, Cong QM, Liu Z, Yu W (2014) Soft sensor approach for modeling mill load parameters based on EMD and selective ensemble learning algorithm. Acta Automatica Sinica 40(9):1853–1866

Tang J, Chai TY, Cong QM, Liu Z, Yu W (2015) Modeling mill load parameters based on selective fusion of multi-scale shell vibration frequency spectra. Control Theory Appl 12:1582–1591

Tang J, Chai TY, Yu W, Liu Z, Zhou XJ (2016) A Comparative study that measures ball mill load parameters through different single-scale and multi-scale frequency spectra-based approaches [J]. IEEE Trans Industr Inf 12(6):2008–2019
Tang J, Qiao JF, Liu Z, Zhou XJ, Yu G, Zhao JJ (2018a) Mechanism characteristic analysis and soft measuring method review for ball mill load based on mechanical vibration and acoustic signals in the grinding process. Miner Eng 128:294–311
Tang J, Qiao JF, Wu ZW, Chai TY, Zhang J, Yu W (2018b) Vibration and acoustic frequency spectra for industrial process modeling using selective fusion multi-condition samples and multi-source features. Mech Syst Signal Process 99:142–168
Tang J, Qiao JF, Liu Z, Sheng N, Yu W, Yu G (2020) Dual-layer optimized selective information fusion using multi-source multi-component mechanical signals for mill load parameters forecasting. Mech Syst Signal Process 135(1):106371
Tang J, Chai TY, Wen Y, Zhao LJ (2012d) Feature extraction and selection based on vibration spectrum with application to estimating the load parameters of ball mill in grinding process. Control Eng Pract 20:991–1004
Twa B, Wz A, Rx A, et al (2021) Assessing load in ball mill using instrumented grinding media. Miner Eng 173
Wang ZH, Chen BC (2002) Tri-factor testing of the internal parameters of ball mill [J]. Metal Mine 307(1):32–35
Wang XZ, Xing HJ, Li Y et al (2014) A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning. IEEE Trans Fuzzy Syst 23(5):1–1
Wang M, Yan G, Fei A (2015) Kernel pls based prediction model construction and simulation on theoretical cases. Neurocomputing 165:389–394
Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 1(1):1–41
Yan GW, Gong XX, Li GY (2014) Soft measurement of ball mill loads based on vibration signal and cloud reasoning. Control Decis 29(6):1109–1115
Zeng Y, Forssberg E (1993) Monitoring grinding parameters by signal measurements for an industrial ball mill. Int J Miner Process 40(1):1–16
Zhao LJ, Tang J, Zheng WR (2012) Ensemble modeling of mill load based on empirical mode decomposition and partial least squares. J Theor Appl Inf Technol 45:179–191
Zhou P, Chai TY (2008) Intelligent monitoring and control of mill loads during ore grinding process. Control Theory Appl 25(66):1095–1099
Zhou P, Chai TY, Wang H (2009) Intelligent optimal-setting control for grinding circuits of mineral processing. IEEE Trans Autom Sci Eng 6(4):730–743

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.