A Sintering State Recognition Framework to Integrate Prior Knowledge and Hidden Information considering Class Imbalance

Dingxiang Wang, Xiaogang Zhang, Hua Chen, Yicong Zhou, Fanyong Cheng

***Abstract*—Estimation of the sintering state has importance for clinker quality improvements and the safe operation of the rotary kiln. Class imbalanced thermal signals usually pose challenges in feature extraction and abnormal state recognition. In this paper, a novel framework that integrates prior knowledge and hidden information is developed for sintering state recognition in the class imbalance condition. For discriminative feature extraction of imbalanced data, a cascaded stack autoencoder (SAE) model is proposed to fuse our prior knowledge and hidden information. The model includes a feature extraction SAE and a deep fusion SAE: the former extracts hidden information from thermal signals, and the latter deeply fuses and compresses our prior knowledge and hidden information. For the class imbalance of sintering samples, we propose a data-dependent kernel modification optimal margin distribution machine (ddKMODM) as a sintering state recognition model. Modifying the original kernel function by a conformal function depending on the data distribution in kernel space, ddKMODM can change the local volume expansion coefficient of the feature space to eliminate the negative effects caused by imbalanced samples. Experiments on real data show that the proposed framework can balance the detection rate of each state in the class imbalanced condition, and its overall sintering state recognition accuracy exceeds 92%.**

***Index Terms*—Feature fusion, Imbalance data, Kernel modification, Sintering state recognition.**

I. Introduction

Rotary kilns are a type of large-scale thermal equipment that are extensively employed in metallurgical, chemical and cement industries for clinker production. To ensure the clinker quality and achieve optimal control of the kiln, the accurate recognition of sintering states is an urgent issue that need to be solved. However, as the kiln generally works in a normal state for safety and economic reasons in industrial fields, the quantity of process data that are collected in a normal state is expected to be larger than that of abnormal states. This class imbalance property always makes the extraction of discriminate features from process data more challenging [1], which further complicates the learning and discrimination process of the recognition model [2].

In practical applications, the data class imbalance always poses some challenges for pattern recognition tasks. In most cases, the pattern recognition effect mainly depended on the quality of feature extraction [3], the under-represented data of the minority class makes the capture of discriminative features for pattern classification difficult. Manual features based on expert experience are usually adopted in some simple pattern recognition tasks [3] but are not suitable for the fields with complex mechanisms and lack of prior knowledge. Many researchers deemed that by improving the objective function [4] of or introducing Laplacian regularization [1] in the deep neural networks (DNNs) model, the discriminative hidden information of imbalance data can be effectively extracted. In the case of a lack of expert experience, these methods may have substantial advantages. However, in some pattern recognition tasks, describing the patterns merely using deep information and disregarding useful prior knowledge accumulated in previous research is not prudent. Recently, the fusion of multi-features from different domains has been extensively utilized in the field of pattern recognition [5-7]. Many research results show that the fusion of prior knowledge and deep information may improve the separability of features and provide a better state description [5-6]. However, no unified framework exists for knowledge fusion.

On the other hand, the existence of class imbalance may cause failure of the classification model to learn the distribution of minority samples, which deteriorates the classification performance [7]. Standard classification models always aim to improve the overall classification accuracy. They usually focus on the over-represented majority samples, whereas the minority class is often the class of interest that needs more attention [2]. Various classification models have been proposed for some

This work was supported in part by the National Key R&D Program of China (No. 2018YFB1305900), in part by the National Natural Science Foundation of China (No.61674056 and 51475157), in part by the Natural Science Foundation of Hunan (No. 2018JJ2056), in part by the Science and Technology Development Fund, Macau SAR (File no. 189/2017/A3), and by University of Macau (File no. MYRG2018-00136-FST), and in part by the Research Foundation of Education Department of Anhui Province (KJ2019A0149). (Corresponding author: Xiaogang Zhang)

D. Wang, X. Zhang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: wangdx@hnu.edu.cn, zhangxg@hnu.edu.cn).

H. Chen is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: anneychen@126.com).

Y. Zhou is with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yicongzhou@um.edu.mo).

Fanyong Cheng is with the Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, School of Electrical Engineering, Anhui Polytechnic University, Wuhu, 241000, China (Email: b12090031@hnu.edu.cn)

fault diagnose and pattern recognition tasks in the class imbalance condition [7-9]. For example, [10] balances the number of samples by oversampling the minority class for fault diagnose of induction motors. [11] proposes an ensemble kernel extreme learning machine (ELM) for fault diagnosis of rotating machinery in the class imbalance condition. Although these methods have effectively improved the detection accuracy of the minority class in specific fields, they are not suitable for direct use as a sintering state recognition model.

For the application of sintering state recognition, the general frameworks of previous research extract useful information from process data to train the learning model, directly recognize the sintering states [12-14] or estimate other related indicators [15-17]. Similarly, these frameworks face difficulties in feature extraction and pattern classification due to data imbalances. The extracted useful information of process data refers to various manual [12-14,18] or hidden features [15-17] of flame images and thermal signals. These two kinds of features are obtained depending on prior knowledge and the DNN model. Usually only one kind of feature is applied to describe different sintering states. According to our experience, a single type of feature is insufficient for sintering state description since it always presents poor separability and may deteriorate the performance of subsequent classification models, especially in the class imbalance condition. In addition to the deficiency in feature extraction, to ensure the economic and safe operation of the rotary kiln, more attention should be paid to improve the detection rates of abnormal states. Unfortunately, in previous research, the standard classifiers for balance data, such as the support vector machine (SVM) [14] and ELM [13], are simply introduced as a sintering recognition model without improvements. Few studies consider the class imbalance issue in sintering state recognition systems and design a novel classifier for sintering state recognition in the class imbalance condition.

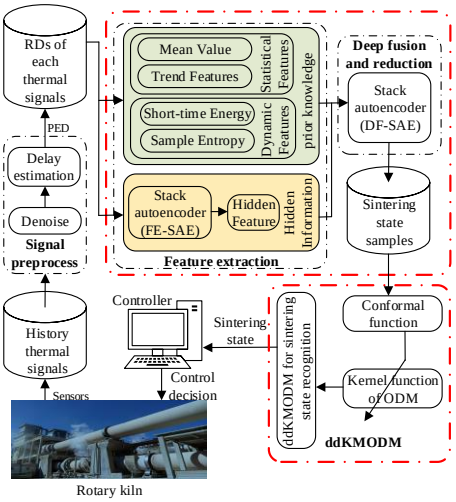


Fig. 1. Proposed integrated sintering state recognition framework.

To recognize the sintering state with higher and more balanced accuracy, this paper proposes a novel integrate framework based on prior knowledge and hidden information in the data imbalance condition. This framework, which is shown in Fig. 1, includes four modules: thermal signal preprocessing, feature extraction (FE), deep fusion (DF) and sintering state recognition. Please note that this paper does not aim to connect these modules in series but aims to develop a novel framework that integrates prior knowledge and hidden information for sintering state recognition in the class imbalanced condition. The main contribution of this paper is described as follows:

(1) A novel integrate sintering state recognition framework of the rotary kiln in the data imbalanced condition based on prior knowledge and hidden information is proposed. This is the first study to consider the class imbalance issue in a sintering state recognition framework.

(2) For the discriminative feature extraction of imbalanced data, a cascaded SAE model is proposed to integrate our prior knowledge and deep information of thermal signals. The model consists of two parts: FE-SAE and DF-SAE; the former extracts hidden information from thermal signals, whereas the latter deep fuses our prior knowledge and hidden information and compress their volume for the computation efficiency improvement of subsequent models. Using this model, the obtained fusion features can more comprehensively describe different sintering states.

(3) Aimed at the class imbalance of sintering samples, a newly conformal transformation function that considers the data distribution in kernel space is designed to modify the kernel matrix of the ODM [19]. Thus, a novel data-dependent kernel modified ODM (ddKMODM) is proposed as a sintering state recognition model. By introducing this model, the higher overall recognition accuracy of the sintering state is achieved, and the detection rate of the abnormal state is also improved.

The remainder of this paper is organized as follows: Section II briefly introduces the details of thermal signals analysis. Details of FE and DF modules of imbalanced thermal signals based on prior knowledge and deep information is also given in this section. Section III describes the kernel modification methods and presents the ddKMODM model for sintering state recognition in the class imbalance condition. Section IV reports our experimental results and analysis. The conclusion of this work and further research directions are presented in section V.

II. Signal Analysis and Feature Extraction

In this section, the thermal signals of the kiln are briefly described and analysed, and then a cascaded SAE model is constructed for FE and DF of imbalanced thermal signals. The main structure of these modules is presented in Fig. 2.

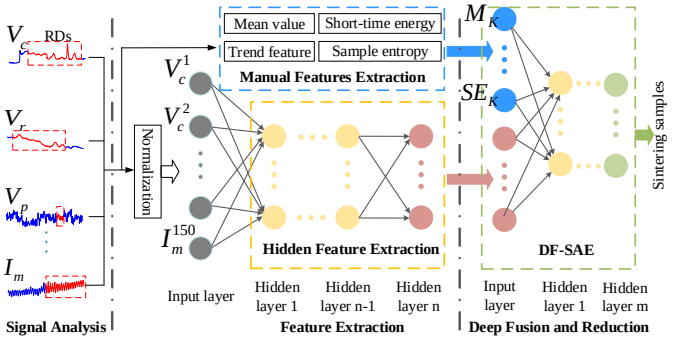


Fig. 2. Main structure of signal analysis, FE and DF modules.

A. Thermal Signals Analysis

Generally, three common sintering states of the kiln exist: normal, chilled and heated. The chilled state and heated state

are the two most important abnormal states, which indicate that the raw materials are calcined insufficient and excessive, respectively. The abnormal states usually cause low-quality clinker production and high economic losses. Considering the alumina rotary kiln as an example, in the chilled state, the raw materials are not melt sufficiently, which causes incomplete extraction of metal in subsequent smelting processes. In the heated state, the clinker becomes sticky and can easily agglomerate, which is not conducive to crushing and smelting. Simultaneously, the heated state may damage the refractory materials and increase the maintenance cost of a rotary kiln. Therefore, for the safe and efficient operations of a rotary kiln, accurately recognizing and avoiding these two abnormal states is important.

According to previous research, owing to disturbances, such as dust and smoke in the rotary kiln [14], the features of flame images may not adequately describe the sintering states. In comparison, thermal signals can more comprehensively reflect the sintering state of materials in the kiln. The thermal signals that are closely related to the sintering state include the coal feeding value (V_c), raw material flow (V_r), primary air (V_p), negative pressure (P_n), kiln head temperature (T_h), kiln tail temperature (T_t), main driven current (I_m) and flame temperature (T_f); their descriptions are presented in Table I. In the industrial field, T_f is measured by physical sensors, such as thermocouples and pyrometers, and is commonly applied as an important indicator to judge the sintering states. However, these sensors cannot provide sufficient measurements since they can only perform detection at single point or a small area of flame. In addition, recognize the sintering state by merely relying on T_f is not enough, as the sintering state is closely related to many other factors, such as the amount of material and the speed of the kiln. Therefore, in this paper, we attempt to extract the informatics features of thermal signals to recognize the sintering states.

TABLE I. THERMAL SIGNALS OF ROTARY KILN.

Signals	Description	Unit	PED (4s)	
			FRD	LRD
V_c	The feed rate of coal powder	t/h	15	223
V_r	The feed rate of material	t/h	70	276
V_p	The volume of Primary air	m^3/h	73	100
P_n	negative pressure of Kiln Tail	Pa	5	169
T_h	Temperature of gas in Kiln Head	ºC	3	184
T_t	Temperature of gas in Kiln Tail	ºC	10	125
I_m	Current of the main motor	ºC	43	81
T_f	Temperature of flame capture by colorimetric pyrometer	A	67	170

Before feature extraction, researching the characteristics of thermal signals to reduce data redundancy is necessary. Each thermal signal is considered to have a particular effect duration (PED) in the current sintering state, and the PED can be approximated by estimating the correlation between the thermal signal and T_f [15]. In this paper, a model-free Lipschitz method [20] is adopted to obtain the PED. This method can obtain the last relevant dynamic (LRD) and first relevant dynamic (FRD) of the SISO system with a thermal signal as input and T_f as output; its details are described in [20].

Since the thermal signals between the LRD and the FRD have a strong influence on T_f, the period between the LRD and the FRD of each thermal signal is defined as the PED, and the thermal data in PED are referred to as related dynamics (RDs) in this paper. The PED of each thermal signal is shown in Table I. Since the thermal signals that we collect have a 4-sec sampling time, the PEDs in Table I are very close to the values stated by experts. Consider V_c as an example; the PED means that once coal powder is sprayed into the kiln, the flame temperature is affected after one minute, and this effect can last approximately fifteen minutes.

B. Feature Extraction and Deep Fusion of Imbalance Thermal Signals

Aimed at the difficulty in the discriminative feature extraction of imbalanced data, in this section, both the prior knowledge and the hidden information are fully utilized with the expectation that the deep fusion of this knowledge can improve the separability of sintering samples and recognition accuracy of the sintering state.

1) Manual Feature Extraction Based on Prior Knowledge

According to our experience, the statistical and dynamic features of thermal signals are significant for the recognition of sintering states. In this section, the following four features of thermal signals in PED are extracted as manual features.

Mean Value

The mean value of the RDs of each thermal signal is highly correlated with the sintering states. For example, in the chilled state, the means of T_h and I_m are generally lower. Thus, the mean values of the RDs are calculated and represented as M_k.

Trend Feature

Since the sintering temperature of the kiln changes slowly, the energies of most thermal signals are concentrated in the low frequency part. Therefore, the trend of thermal signals is adopted as a feature of the sintering state. After linearly fitting the RDs of each signal channel, we can extract the optimal slope to represent the trend feature and denote it as A_k.

Short-time Energy

According to expert experience, significant differences exist in the stability of thermal signals in different sintering states. For example, in the chilled state, the fluctuation of T_h is more severe than that in other states. Thus, the short-time energies of the RDs, which are denoted as E_k, can be extracted as a feature for sintering state recognition.

Sample Entropy

Similarly, the complexity of the thermal signal is different in different sintering states, and the sample entropy of the RDs can be calculated to describe it. The sample entropy is sensitive to the changes in the complexity of data and can better distinguish different sintering states [14]; it is denoted as SE_k in this paper.

The manual feature of each sintering sample is composed of four features— M_k , A_k , E_k and SE_k —of the RDs of each thermal signal; it can be written as a feature vector with 32 dimensions $\{M_{V_c},...,M_{T_f},...,SE_{V_c},...,SE_{T_f}\}$.

2) Hidden Feature Extraction Based on SAE

Since the rotary kiln is a complex equipment with nonlinear, strong coupling, large delay and time-varying characteristics, according to previous research, the manual features based on prior knowledge are not enough to describe the information contained in thermal signals. The AE is one of the most extensively employed deep learning models that have achieved

excellent performance in many pattern recognition tasks [21-22]. The basic AE is a three-layer neural network, in which the output data are employed as the input. The dimension of the hidden layer is predominantly lower than that of the input layer or output layer, and the decoder uses these low-dimensional expressed features to reconstruct the input with the lowest reconstruction error [22]. By stacking multiple AEs, the SAE is able to adaptively capture the representative information from raw data via multiple nonlinear transformations and approximate, complex nonlinear functions with a small error.

In this paper, we construct an SAE model for the hidden feature extraction of the RDs. The input of this SAE model is the cascaded normalized RDs of each thermal signals, the dimension of it is fixed to 1050 according to the definition of PED in Table I. A schematic of adaptive feature extraction is shown in Fig. 2. Each layer of the SAE is a basic AE that can be pre-trained using the output of the previous layer as the input and output. In this way, all the hidden layers of the SAE can be pre-trained, and the obtained initial weights are suggested to be better than those that are randomly assigned [22]. After pre-training, the whole model can be fine-tuned using the thermal signals with state labels to improve the separability of hidden features. The outputs of the last hidden layer are hidden features that can be applied for sintering state recognition combined with manual features that we previously obtained. The hidden layer parameters of this FE-SAE model is optimized in experiment section.

3) Feature Fusion and Reduction

The useful prior knowledge and hidden information of original thermal signals can be represented by the manual and the previously extracted hidden features. How to further fuse these features and reduce their dimensions to guarantee efficient learning and high accuracy recognition is a key issue. Some feature fusion models can be adopted to project the high-level representation of features to a low-dimension space. The most commonly employed models include principal component analysis (PCA) [23], t-Distributed Stochastic Neighbour Embedding (t-SNE) [24] and linear discriminant analysis (LDA) [25]. PCA and t-SNE are unsupervised methods, and the computational complexity of t-SNE is too high, which is not suitable for sintering state recognition tasks. LDA aims to maximize the between-class covariance while minimizing the within-class covariance in a supervised way. However, the target dimension of LDA is limited by the number of the class; finding the optimal dimension of features expression may not be necessary. In addition to the previously stated methods, some DNNs, such as the restricted Boltzmann machine (RBM) [6] and AE [26], are popular tools for deep fusion and reduction of features.

In this paper, we adopt a SAE network to realize deep fusion and dimensionality reduction of the extracted sintering state features. By introducing a Softmax classifier to fine-tune the DF-SAE model, the obtained low-dimension features will be highly separable and suitable for sintering state recognition in the next step. The hyperparameter of this DF-SAE network will be optimized in the experiment section, and the performance of the SAE and other methods in feature fusion and reduction will be compared in the experimental part.

III. DDKMODM FOR IMBALANCE CLASSIFICATION

In the class imbalance condition, a novel classification model, which is based on the kernel modification method, is proposed in this section to balance the detection rate of each sintering state while improving the overall recognition accuracy.

A. Class Imbalance Classification

The most extensively employed strategy for the class imbalance issue is to balance the influence of each class by adopting various oversampling [27] or under-sampling [28] techniques to generate balanced training data. However, the performance of these methods depends on the structure and distribution of training data, the resampling is unavoidable introduce noise (over-sampling) or eliminate useful information (under-sampling) and cause performance degradation of the classifier. The most common cost-sensitive methods increase the misclassification penalty for minority samples [29-30] or the samples are difficult to classify [8]. The cost-sensitive methods are minimize the misclassification costs and cannot change the spatial distribution of the training data. According to the KKT conditions, some studies indicate that increasing the misclassification penalty may not affect the Lagrange multiplier α, which determines the weight of the influence of the training data on the classifier.

In addition to the two categories of methods, another way exists to eliminate the negative effects of class imbalance. The kernel function has a key role in a classifier since it can map the linear inseparable training data in the input space into a high dimensional kernel space, in which the training data are considered to be linear separable. Some research revealed that the Riemannian metric of the feature space and the training data distribution can be changed via modifying the kernel function [9, 31]. These kernel modification technologies were originally proposed to improve the generalization performance of a classifier [32]. By constructing a conformal function and performing conformal transformation of the original kernel function, the local volume expansion coefficient (VEC) of the area near the initial separator in kernel space can be amplified. In this way, the separability of the training data can be improved. Furthermore, when the imbalance rate (IR) is introduced in the conformal function [7], the area of different classes would have different VEC assignments. For instance, by magnifying the VEC of the majority class, the influence of different classes on the learned separator can be balanced effectively [7].

In general, most kernel modification methods follow these steps: 1) apply the standard classifier to training data; 2) construct the conformal function according to the results obtained in step 1, such as the margin of training data and support vectors; and 3) perform the conformal transformation of the original kernel function using the obtained new kernel to reclassify the training data. The support vectors in step 2 are determined by the margin of data. Thus, most of these methods can be categorized as margin-based kernel modification methods. The main disadvantage of this type of method is that the margin of training samples can only be obtained by complex algorithm iterations and it cannot adequately reflect the spatial distribution of the training samples.

B. ddKMODM

Consider that a conventional classifier, such as the SVM, is aimed at optimizing the minimum margin to learn the separator [33] and generally yields poor generalization performance. In this paper, the multi-class ODM (mcODM) is adopted as a robust sintering state recognition model. Compared with the SVM, the mcODM optimizes not only the minimum margin but also the margin distribution; details are provided in [19].

For the linear inseparable situation, the mcODM introduces a kernel function to map the data into a high-dimensional feature space. The kernel function can be denoted as $K(x_i, x_j) = \varphi(x_i)\varphi(x_j)$. Using the mapping function φ, the training data in the input space are embedded into a curved Riemannian manifold in the feature space. The Riemannian metric is defined as [32]:

$$g_{ij}(x) = \left(\frac{\partial^2 K(x, x')}{\partial x_i \partial x_j} \right)_{x=x'} \tag{1}$$

The VEC reflects the volume expansion rate and can be written as $\sqrt{\det(\mathbf{g}(\mathbf{x}))}$, and the dominant eigenvector of the Riemannian metric matrix $\mathbf{g}(\mathbf{x})$ refers to the principal spread direction (PSD). For classification tasks, the generalization performance of the classifier and data separability can be improved if the VEC is magnified in the heterogeneous region and reduced in the homogeneous region. The VEC induced by the extensively employed kernels is constant [9]. The parameters adjustment of these kernels has very limited ability to improve the classification performance. The conformal transformation of the kernel function is an effective way to modify the kernel function as expected; it can be formulated as:

$$\tilde{K}(x, x') = D(x)D(x')K(x, x') \tag{2}$$

The conformal function $D(x)$ is the key factor that determines whether the kernel modification is a success. The modified kernel can induce a local change of the VEC to finely adjust the data distribution in the kernel space for accurate classification. Many conformal functions have been proposed to handle the imbalance data [7,31-32]. However, most of these functions depend on the margin obtained by a pre-trained standard classifier. The pre-trained classifier generally produces high computation complexity and its results cannot adequately reflect the real data distribution. Thus, a novel conformal function that depends on the data distribution in the kernel space is proposed in this work.

First, a common kernel is used to map the original data into a kernel space, in which the data are approximate linear separable. Considering that the data with a small distance from heterogeneous data in the kernel space have a high probability of being located in overlap regions, we can design the conformal function depending on the average distance between the target data and the heterogeneous data:

$$D(x) = e^{-N\frac{1}{n-n_c}\sum_{x_i \notin c}\|\varphi(x)-\varphi(x_i)\|^2}\Big|_{x \in c}, \tag{3}$$

where x_i denotes the *i-th* data, $x(i)$ is the *i-th* feature of *x*, and n_c and n represent the number of data of class *c* and the total. *N* is the parameter that reflects the IR of the training data, which can be formulated as

$$N = \left(\frac{\max(n_c)}{n_c} \right)^{\rho}, \tag{4}$$

where $\max(n_c)$ represents the largest class size; and ρ is a parameter that controls the spatial expansion ratio of different regions, which can be selected from the set [0, 0.1, …, 2] according to our experience. The average distance between data x in class c and its heterogeneous data in the kernel space can be calculated by the kernel trick:

$$\sum_{x_i \notin c}\|\varphi(x)-\varphi(x_i)\|^2 = K(x,x) - 2\sum_{x_i \notin c}K(x,x_i) + \sum_{x_i, x_l \notin c}K(x_i,x_l). \tag{5}$$

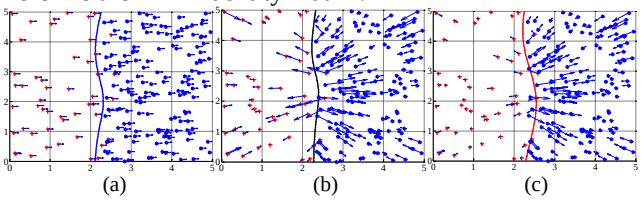
Let us denote $\mathbf{u}(x) = (D_1(x), D_2(x), ..., D_d(x))^T$, where $D_i(x) = \dfrac{\partial D(x)}{\partial x(i)}$. When the original kernel function is selected as $e^{-\gamma_0\|x-x'\|}$, according to [9], the VEC and PSD of the new kernel modified by $D(x)$ can be written as

$$\mathbf{g}(\mathbf{x}) = \mathbf{u}(\mathbf{x})\mathbf{u}(\mathbf{x})^T + 2\gamma_o D^2(\mathbf{x})\mathbf{I}, \tag{6}$$

$$\mathbf{VEC} = \sqrt{\det(\mathbf{g}(\mathbf{x}))}\sqrt{(2\gamma_o D^2(x))^{n-1}(2\gamma_o D^2(x) + \|\mathbf{u}(x)\|^2)}, \tag{7}$$

$$\mathbf{PSD} = sign(D(x)) \cdot \mathbf{u}(x), \tag{8}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix.



(a) (b) (c)

Fig. 3 Obtained separator and local change in VEC deduced by different RBF kernels. (a) RBF kernel; (b) modified kernel without considering imbalanced data; (c) modified kernel considering imbalanced data

Fig. 3 shows the learned separator and local change in the VEC deduced by different RBF kernels. As shown in Fig. 3 (a), the conventional ODM with RBF kernel changes the VEC in a constant way. The learned separator is skewed toward the minority class and produces a relatively low detection rate for the minority class. However, in Fig. 3 (b) and (c), when the proposed $D(x)$ is adopted to modify the RBF kernel, the VEC deduced by the newly obtained kernel is magnified along the class boundary, and the PSD vectors point to the class centre. Thus, the spatial resolution of the heterogeneous region increases, and the homogeneous region is decreased to favour the fine classification. For the application of the ODM in the class imbalance condition, a smaller VEC of the minority samples always pushes the learned separator toward the majority class as the classifier tends to optimize the margin distribution while minimizing the penalty of misclassification. As shown in Fig. 3 (c), by introducing the parameter N to reflect the IR of the training data in the conformal function, the learned separator is moved toward the majority class.

The pseudo-codes of the ddKMODM are given in Algorithm 1. First, the average distance between the training sample and its heterogeneous data in the original kernel space is obtained. $D(x)$ of each training sample can be achieved using "(3)", and the original kernel matrix $\mathbf{K}$ can be adjusted. The new kernel matrix $\tilde{\mathbf{K}}$ is employed to learn the final classifier, which produces the ddKMODM.

Algorithm 1 ddKMODM

1: Input: training set $\mathbf{X}_{train}$ with c classes; original kernel matrix $\mathbf{K}$ of ODM.

2: Output: new kernel matrix $\tilde{\mathbf{K}}$; output classifier ddKMODM.

3: Function:

ODMTrain($\mathbf{X}_{train}$, $\mathbf{K}$) : train the ODM model using $\mathbf{X}_{train}$ and $\mathbf{K}$.

4: **Begin**

5: **for** each training data $\mathbf{x}$ of class c **do**

6: Compute average distance between $\mathbf{x}$ and its heterogeneous data in kernel space.

7: Compute parameter N for class c using "(4)".

8: Compute conformal function using "(3)" .

9: **end for**

10: **for** each k_{ij} in $\mathbf{K}$ **do**

$$\tilde{\mathbf{K}} \Leftarrow \tilde{k}_{ij} = D(\mathbf{x}_i) D(\mathbf{x}_j) k_{ij} .$$

11: **end for**

12: Training ODM using new kernel matrix $\tilde{\mathbf{K}}$

$$\text{ddKMODM} \Leftarrow \text{ODMTrain}(\mathbf{X}_{train}, \tilde{\mathbf{K}}) ;$$

13: **Return** ddKMODM

14: **End**

IV. Experiment Analysis

To validate the performance of the proposed sintering state recognition framework, an online sintering state recognition system has been developed for the #3 kiln of the ZhongZhou Aluminum Corporation in China to assist kiln operators in state recognition and control decision-making. The training of this version of the system is offline. In the process of online application, only the estimation results of the sintering state are provided, the parameters of the model obtained by offline training will not be changed dynamically.

TABLE II. Sintering Samples Extracted From Three Rotary Kilns

	Rotary kiln			Total
	#1	#2	#3	
chilled	432	775	302	1509
heated	469	648	297	1414
normal	3249	3755	1751	8755
Total	4150	5178	2350	11678

The cascaded SAE model shown in Fig. 2 is performed by 500 epochs, and its batch size is set to 1000. The initial learning rate is set to 0.01 based on the trials between 0.001 and 0.02. The RELU function is selected as the activation function of the SAE, and the parameter optimization for this model is performed by an Adam optimizer. The sintering samples are extracted without overlap and marked by three kiln experts. In our system, the classification model is trained and tested with the samples extracted from the history thermal signals with 4-sec intervals of three rotary kilns. The information of these samples is listed in Table II.

TABLE III. Baseline Recognition Model

Balance model		mcSVM [33], mcODM [19], Softmax.
Imbalance model	Resampling based	WKSMOTE [27], RBO [28].
	Cost sensitive based	AWENSVM [29], BadaCost [30], FocalNN [8]
	Kernel based	KMODM [7]

All the experiments described here use cross-validation to ensure the generalization performance of the model. Parameters optimization of the FE-SAE and DF-SAE use 10-fold cross-validation, while other experiments use nested cross-validation. When performing nested cross-validation, samples of two kilns were selected as the training data and that of another kiln were selected as the test data. For the model tuning, the 10-fold cross-validation method is still used to optimize the model parameters in the inner loop.

The F1-score and G-mean are adopted as metrics to assess the recognition performance of the proposed framework in this paper. For performance comparison with the proposed ddKMODM, the nine classification models shown in Table III were selected as the baseline model. Among them, mcSVM, mcODM and Softmax are commonly employed for balance classification tasks, and the remaining models are improved for imbalanced classification.

A. Parameters Optimization of FE-SAE and DF-SAE

TABLE IV. Experimental Result Comparison With Different Hidden Layer Parameters of FE- SAE

Input size	# hidden layer	# hidden nodes	F1 score
1050	2	500;50	81.78
	2	500;100	81.85
	2	500;200	82.09
	3	500;100;50	83.25
	3	525;263;50	83.19
	3	**525;263;100**	**83.67**
	4	525;263;130;25	82.98
	4	525;263;130;50	83.07
	5	525;263;130;50;20	83.13
	5	525;263;130;50;30	82.60

The structure of the FE-SAE and DF-SAE models have importance in the extraction of effective features and achieve high performance of the sintering state recognition framework. Therefore, some experiments are employed to optimize the hidden layer parameters of these two models.

TABLE V. Experimental Result Comparison With Different Hidden Layer Parameters of DF-SAE

Input size	# hidden layer	# hidden nodes	F1 score	
			training	testing
132	1	10	85.75	86.17
	1	8	86.28	85.74
	1	5	85.49	85.09
	1	3	86.05	85.26
	2	66;10	88.13	87.60
	2	66;8	87.78	88.05
	2	66;5	87.97	87.62
	3	66;33;10	88.54	88.43
	3	66;33;8	88.95	87.83
	3	**66;33;5**	**89.27**	**88.79**
	3	66;33;3	87.90	87.31

For the FE-SAE model, the number of input nodes is fixed to 1050, which is equal to the number of total RDs. The model that can offer hidden features with the best recognition accuracy is optimal, and then the hidden layer parameters can be selected by feeding the obtained hidden feature to a Softmax classifier. To demonstrate the influence of the number of hidden layers and hidden nodes for the FE-SAE module, the recognition accuracies with different hidden layer parameters are listed in Table IV. According to these results, three hidden layers with 525-263-100 nodes are employed as the FE-SAE model.

The obtained hidden features are combined with manual features by a cascade pattern and are sent to the DF-SAE for deep fusion and reduction. Thus, the input size of the DF-SAE model is determined by the dimensions of the manual and hidden features; it can be fixed to 132 in this paper. The way of the hidden layer parameters optimization is similar to that of the FE-SAE model. The output size of this module determines the

dimensions of the sintering samples. The F1 scores of different model parameters are recorded in Table V. The F1 score, which corresponds to different dimensions of sintering samples, is shown in Fig. 4. According to these results, the SAE, which includes three hidden layers with 66-33-5 nodes, is selected as the DF-SAE.

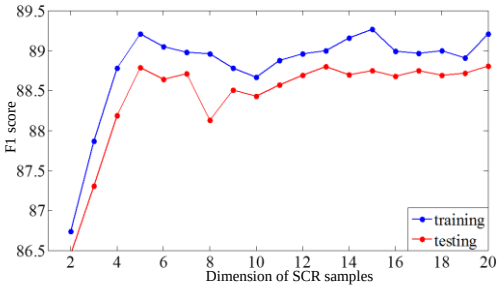


Fig. 4. Selection of the dimension of sintering samples

In this part, several popular DF models, including PCA, LDA and RBM, are selected to compare with the SAE regarding the performance in feature fusion and reduction. Fig. 5 shows the distribution of the F1 score attained by different DF models integrated with the ten recognition models shown in Table III; each model is tested 10 times. The boxes illustrate the distribution range of the F1 score (circles) between the first quartile and third quartile. The red solid lines indicate the median value; the red cross symbols represent the outliers; and the dashed lines represent the outlier range. These DF models are ranked as SAE, RBM, LDA and PCA according to the average of the F1 score. Although the RBM obtains an average value that is close to that of the SAE, it has more outlier points as the RBM cannot easily learn an accurate probability distribution of training data in the class imbalance condition. The feature dimension obtained by LDA is limited by the number of class. Obtaining a suitable feature description dimension may not be possible since the class number of the sintering state is too small. PCA is an unsupervised model, and the separability of the obtained features cannot be guaranteed; thus, it produces the largest box and also has the maximum number of outliers, i.e., the most unstable recognition performance.

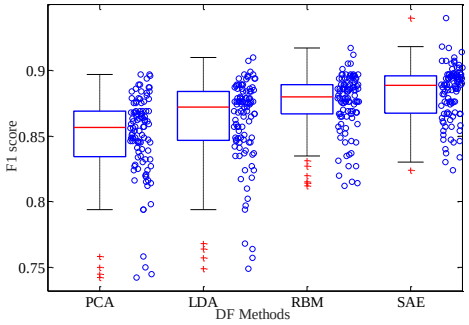


Fig. 5. Distributions of F1 score attained by different DF models.

B. Parameter Analysis of ddKMODM Model

In the previous experiments, the optimal parameters of the FE-SAE and DF-SAE models have been determined. For the optimization of parameter ρ in the ddKMODM model, in this part, the extracted sintering samples are employed to compare the recognition accuracy of different ρ. The training and testing accuracy of each sintering state with different ρ are shown in Fig. 6.

The accuracy of abnormal states generally increases with an increase in parameter ρ, whereas the accuracy of the normal state changes in the opposite direction. The main factor of this phenomenon is that when ρ = 0, the ddKMODM does not consider the class imbalance issue, which produces lower recognition accuracies of abnormal sintering states and lower overall F1 scores. With an increase in ρ, the VEC of the majority class increases, while that of the minority class decreases. This result causes the learned seperator to move toward majority class, which improves the recognition accuracy of the abnormal states. From the results displayed in Fig. 6, we observe that when ρ = 1.3, the ddKMODM can obtain the highest F1 score, at which time, the recognition accuracy of all sintering states are relatively balanced. As a result, ρ is fixed to 1.3 in the following experiments.

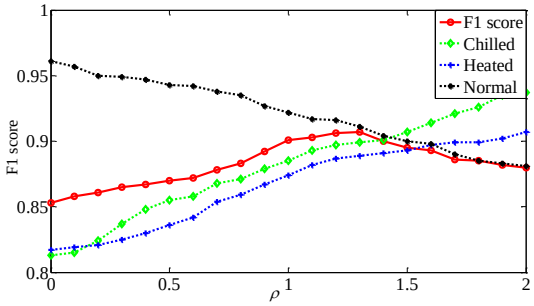


Fig. 6. Recognition accuracy of each sintering state with different ρ

C. Analysis of Sintering State Recognition Results

In this section, a series of experiments are carried out to analyse the performance of the proposed framework in various aspects, such as recognition accuracy, robustness and computational complexity.

1) Comparison of Recognition Results

Table VI summarizes the average recognition accuracy of each sintering condition obtained by different recognition models, and the overall F1-score and G-mean score using manual (#M), hidden (#H) and fusion (#F) features are reported. As shown in the table, the recognition accuracy of each single feature is always lower than that of the fusion feature. For example, the F1 score of FocalNN using fusion features reaches 90.59%, while that of the manual feature and hidden feature is at least 4% lower. These results reveal that the proposed cascaded SAE model can well integrate our prior knowledge and hidden information of imbalanced thermal signals and the obtained fusion features can describe the characteristics of thermal signals more comprehensively and accurately than a single feature.

In addition, due to the class imbalance of the training samples, the mcSVM, mcODM and Softmax obtain a seriously imbalanced recognition accuracy of each state, and the detection rates of the normal condition are approximately 8% higher than those of the abnormal sintering states. However, the ddKMODM and other recognition models which consider the imbalanced distribution of the training data, can significantly improve the recognition accuracy of two abnormal conditions and obtain a more balanced detection rate. The FocalNN and ddKMODM achieve an overall accuracy of more than 90%, and the ddKMODM achieves more than 92% recognition accuracy. This finding verifies the effectiveness of the

proposed feature extraction method and ddKMODM-based sintering state recognition framework.

TABLE VI. AVERAGE RECOGNITION ACCURACY (ACCURACY\STANDARD DEVIATION) OF DIFFERENT RECOGNITION MODELS USING DIFFERENT FEATURES; THE BEST RESULTS ARE SHOWN IN BOLD.

		mcSVM	mcODM	Softmax	WKSMOTE	RBO	BAdaCost	KMODM	AWENSVM	FocalNN	ddKMODM
#M	heated	81.04\2.5	81.73\1.8	82.62\2.2	83.65\2.7	83.85\2.6	83.73\2.4	84.13\2.8	84.01\2.5	84.24\3.7	**85.05\2.5**
	chilled	82.49\1.8	82.71\2.4	83.43\2.5	83.54\3.2	83.40\3.1	83.27\2.7	83.81\3.0	83.56\3.9	84.07\2.3	**84.36\2.1**
	normal	**91.14\2.6**	91.05\2.1	90.10\2.2	89.67\2.3	90.45\3.9	89.41\2.5	88.65\2.9	90.06\2.7	88.64\3.2	89.87\2.9
	F1	83.58\2.4	83.16\2.6	83.49\2.5	84.44\2.9	83.88\3.2	84.63\2.3	84.70\3.1	84.57\2.9	84.83\3.1	**85.88\2.7**
	G-mean	82.80\2.3	82.53\2.4	83.27\2.7	83.13\2.6	84.16\3.0	83.40\2.4	85.17\2.8	84.29\3.1	85.09\2.8	**86.19\2.5**
#H	heated	83.03\2.7	82.15\2.7	83.75\2.9	84.72\3.3	85.35\3.3	84.87\3.4	84.75\3.1	84.88\2.6	85.34\2.8	**87.45\3.0**
	chilled	82.47\2.2	83.44\2.6	83.10\2.6	84.16\2.8	83.92\3.7	84.55\2.9	86.01\2.4	85.66\3.3	**86.51\3.3**	85.94\2.5
	normal	90.43\2.4	89.72\3.0	**90.44\1.5**	89.76\2.9	89.79\2.8	87.93\3.2	88.56\2.8	88.82\2.9	88.31\31	88.46\2.2
	F1	83.64\3.4	84.05\2.7	84.22\2.5	85.01\3.0	85.21\3.3	85.14\2.6	86.07\2.7	85.50\3.1	86.79\2.9	**87.37\2.4**
	G-mean	83.42\3.0	84.18\2.5	85.03\2.9	85.49\2.7	85.20\3.0	85.09\3.1	85.71\2.6	85.27\3.3	86.03\3.1	**86.51\2.5**
#F	heated	84.89\3.1	85.94\2.1	85.61\2.1	86.52\2.5	87.06\3.4	86.95\2.7	88.04\3.4	87.92\2.6	88.77\2.7	**90.80\3.2**
	chilled	86.01\3.3	86.51\2.7	85.97\2.5	86.72\3.0	87.28\3.3	87.88\2.8	87.82\3.5	86.80\3.1	89.88\3.2	**91.64\3.5**
	normal	**95.25\2.7**	93.93\2.9	94.74\2.0	93.51\2.9	92.70\2.5	92.85\2.4	91.79\2.5	92.57\2.5	91.73\2.9	93.76\2.7
	F1	86.27\2.7	87.16\2.3	85.95\2.5	88.45\2.6	88.79\2.9	89.13\2.6	89.59\3.3	89.74\2.9	90.59\2.9	**92.69\3.1**
	G-mean	86.91\3.0	87.80\2.7	85.86\3.1	88.10\2.8	87.96\3.1	88.32\2.4	89.10\3.0	88.67\2.8	90.47\2.7	**92.40\3.0**

2) Noise Sensitive Analysis

The thermal signals collected from the rotary kiln are often disturbed by noise, which is usually introduced by the sensor or electromagnetic environment. In this section, white noise stress tests were carried out to demonstrate the robustness and effectiveness of the employed methods with different signal-to-noise ratio (SNR) values, where the average F1 score of the 30 tests were considered to be the final result, as shown in Fig. 7.

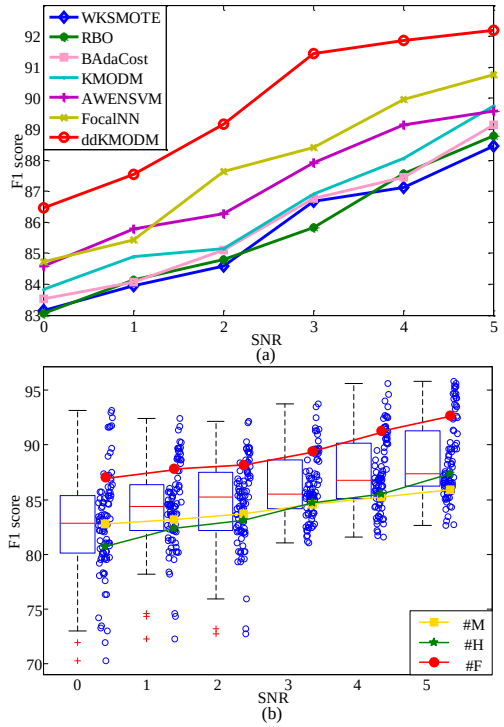


Fig. 7. (a) Average F1 score of different imbalance models using fusion feature over different SNR. (b) F1 score of ddKMODM using different features over different SNR.

As shown in Fig. 7(a), the additional noise with different SNRs in the thermal signal have a negative impact on the sintering state recognition accuracy of different recognition models. However, the noise of the same SNR has a similar effect on the performance of different models, and the proposed ddKMODM can achieve the highest recognition accuracy with different SNRs. As shown in Fig. 7(b), the mean value of the F1 score obtained by different features are connected by a line for a better comparison. The larger is the slope of the line, the more sensitive the feature is to the noise. In this figure, with an increase in the noise ratio, the recognition accuracy obtained by the manual features decreases less, while the hidden features are larger. As a result, the manual features extracted based on prior knowledge have strong anti-noise abilities, and the hidden features are more sensitive to noise interference. Regarding the benefits to the deep fusion of these features, the noise sensitivity of the fusion feature is in the middle.

3) Imbalance Rate Sensitive Analysis

To analyse the impact of the IR of the sintering samples on the performance of the recognition model, in this section, the heated samples are treated as the minority class, and the normal and chilled samples are treated as the majority class. By randomly undersampling the minority or majority samples, nine data sets with different IRs from 1.6 to 16.7 are obtained without introducing noise data. Fig. 8 shows the average F1 scores of different imbalanced models for 10 trials on these data sets.

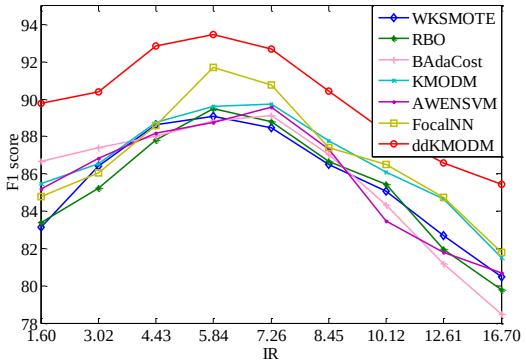


Fig. 8. Average F1 score of different imbalanced models over different IRs.

A comparison of these results reveals that the increase in the IR has the greatest impact on the performance of the model based on resampling technology, such as WKSMOTE and RBO. The oversampling technology that is employed is dependent on the data distribution of the minority class, while the data sets with greater IR are obtained by randomly deleting minority samples, which is detrimental to the implementation

of oversampling. The cost-sensitive methods assign different penalty factors to samples of different class according to the IRs, which enable better robustness to the change in the IR than the resampling-based method. The proposed ddKMODM method adjusts the VEC of the kernel space by considering the distribution and IR of the training data, which changes the distribution of the samples in the kernel space and eliminates the negative effects of the imbalanced data. Deleting a part of the samples that are not located at the class boundary have a minimal effect on the VEC of the kernel space; thus, the model is the most robust model regarding the change in the IR.

4) Significant Analysis

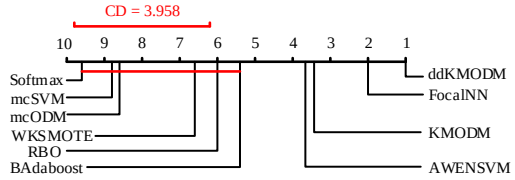


Fig. 9. CD diagram of the post hoc Friedman test.

Based on the reported results of different recognition models, Fig. 9 shows the critical difference (CD) diagram of the post hoc Friedman test with the significance level set to 0.05. The ddKMODM has a CD with the six models on the left, which means that the ddKMODM obviously outperforms them. On the other hand, although the ddKMODM only slightly outperforms the FocalNN, KMODM and AWENSVM, its overall performance is still better than that of the other models.

5) Low Operation Risk Analysis

TABLE VII CONFUSION MATRIX OF DDKMODM AND MCODM. THE DATA IN CONFUSION MATRIX ARE THE AVERAGE OF 30 EXPERIMENTAL RESULTS.

		mcODM			ddKMODM		
		chilled	heated	normal	chilled	heated	normal
#M	chilled	248 82.1%	20 6.6%	34 11.3%	257 85.0%	12 3.9%	33 11.1%
	heated	19 4.4%	248 83.5%	30 10.1%	12 4.0%	260 87.6%	25 8.4%
	normal	105 6.0%	53 3.0%	1593 91.0%	93 5.3%	82 4.7%	1576 90%
#H	chilled	255 84.4%	18 6.0%	29 9.6%	263 87%	11 3.7%	28 9.3%
	heated	21 7.0%	248 83.5%	28 9.5%	16 5.4%	255 85.9%	26 8.7%
	normal	82 4.7%	58 3.3%	1611 92.0%	90 5.1%	94 5.4%	1567 89.5%
#F	chilled	260 86.1%	13 4.3%	29 9.6%	**277** **91.7%**	**6** **2.0%**	19 6.3%
	heated	13 4.4%	259 87.2%	25 8.4%	**4** **1.3%**	**278** **93.6%**	15 5.1%
	normal	54 3.0%	34 2.0%	**1663** **95.0%**	56 3.2%	58 3.3%	1637 93.5%

The confusion matrixes of the ddKMODM and mcODM using different features are shown in Table VII. Adopting our fusion features, the ddKMODM achieves a relatively balanced recognition accuracy of 91.7%, 93.6% and 93.5% for the three different sintering states. Furthermore, as the operations that correspond to the two abnormal sintering states are quite different, the misjudgement between the chilled state and the heated state may cause undesirable consequences. For instance, our model misjudges the current heated state as being chilled, and subsequent improper operation may cause a higher degree of overheating of the rotary kiln and increase its operation risk. The higher is the misclassification rate between two abnormal states, the higher is the operational risks caused by the recognition system.

The confusion matrix indicates that the ddKMODM not only achieves a more balanced detection rate of each class but also a misclassification rate of 1.6% between two abnormal sintering states. This result is approximately one third of that of the mcODM (4.4%). The ddKMODM significantly improves the recognition rate of the abnormal states and reduces the operational risk deduced by abnormal sintering state misclassification.

6) CPU Time Cost

Fig. 10 shows the CPU time cost comparison of training different recognition models using different features. These results can reflect the computational complexity of different models to a certain extent. All experiments are performed with Matlab2017a on the workstation with 2×3.3GHz CPUs and 4 GB main memory. As shown in the figure, since the hidden feature has the largest dimension, it consumes the most training time, and the fusion feature has the lowest dimension after dimensionality reduction; thus, it can substantially decrease the training time. Furthermore, a comparison of the time cost of different recognition models reveals that the proposed ddKMODM is more computationally complex than the balance classification model, as its additional processes, such as conformal function calculation and kernel matrix modification, increase the computational complexity. Compared with other imbalanced models, especially WKSMOTE, RBO and BAdacost, its advantage of time consumption is obvious. Considering that the change in the sintering state in the kiln is usually slow, the time cost of the ddKMODM can satisfy the requirement of the sintering state recognition system.

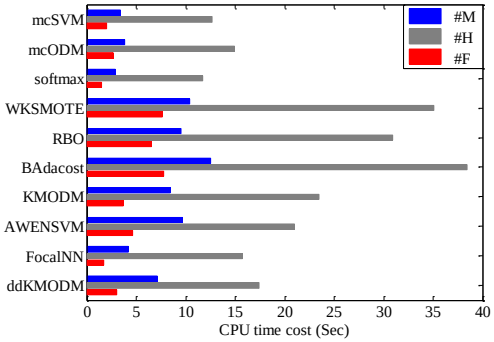


Fig. 10. CPU time cost of each recognition model

D. Industrial application

The online sintering state recognition system has been installed on the #3 kiln and is successfully operating. Fig. 11 shows the installation locations of the sensors, which collect various thermal signals, and the central control room.

Fig. 11. Installation locations of sensors (left) and the central control room (right).

To reduce the complexity, in our experiments, the kiln operator merely adjusted V_C ($\pm$ 0.2 each time) to control the sintering state, and the operation interval was 2 minutes. Before using our system, the operators manually judge the sintering states by comprehensively analysing the thermal signals and increase or decrease V_C accordingly. After installing our system, the operators employ the same strategy to control the kiln based on the sintering state recognition results that are automatically provided by the system every 2 minutes.

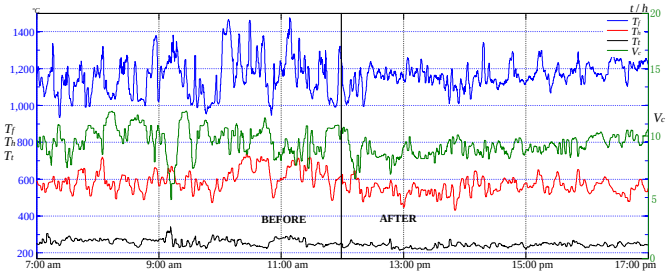


Fig. 12. Fluctuations of thermal signals before and after using our system.

Fig. 12 shows the fluctuations of four main thermal signals before and after using our system. These thermal signals include V_c, T_f, T_h, and T_t. In general, the smoothness of these signals indicates whether the control of the rotary kiln is stable. As shown in Fig. 8, after using our system, the curves of these signals are smoother, and the abnormal sintering states are significantly less than previous states. The improvement is mainly attributed to the notion that our system can accurately predict the sintering state of the rotary kiln to ensure that the corresponding operation can be performed in advance to reduce the occurrence of abnormal states.

V. Conclusion and Prospect

This paper proposes an efficient integrated framework for sintering state recognition of the rotary kiln in the class imbalance condition. For distinguishable features extraction of imbalanced thermal signals, a cascaded SAE model is proposed to extract hidden information and deep fuse it with our prior knowledge. The obtained fusion features have lower dimensional expressions and are more separable than single features; thus, they can more comprehensively describe different sintering states. For the recognition model improvements in the class imbalance condition, a novel conformal function that depends on the average distance between the target data and heterogeneous data is designed to modify the kernel function of mcODM, and a newly sintering state recognition model, which is named ddKMODM, is proposed. The novelty of the ddKMODM is to change the VEC of the feature space to eliminate the skewness of the learned separator caused by class imbalanced training data.

The achieved results show that the proposed integrated framework efficiently recognizes the sintering states in the class imbalance condition and reduces the operation risk caused by misclassification between two abnormal states. The obtained sintering state recognition results can provide a decision basis for the subsequent control system, which is a direction of our future research. The training of the proposed ddKMODM is offline. The algorithm for training this model online is a challenge that we are currently conquering, which is very important for the development of a sintering state recognition system and even the intelligent control system of a rotary kiln.

References

[1] Zhao, Xiaoli, Minping Jia, and Mingyao Lin. "Deep Laplacian Auto-encoder and its Application into Imbalanced Fault Diagnosis of Rotating Machinery." *Measurement* (2019): 107320.

[2] Hu, Zhixin, and P. Jiang. "An Imbalance Modified Deep Neural Network with Dynamical Incremental Learning for Chemical Fault Diagnosis." *IEEE Transactions on Industrial Electronics* (2018):540-550.

[3] Lei, Yaguo, et al. "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data." *IEEE Transactions on Industrial Electronics* (2016):3137-3147.

[4] Chen, Longting, et al. "Learning deep representation of imbalanced SCADA data for fault detection of wind turbines." *Measurement* 139 (2019): 370-379.

[5] R. Razavi Far *et al*., "Information Fusion and Semi-Supervised Deep Learning Scheme for Diagnosing Gear Faults in Induction Machine Systems," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 8, pp. 6331-6342, Aug. 2019.

[6] Han Guanghui, et al. "Hybrid resampling and multi-feature fusion for automatic recognition of cavity imaging sign in lung CT." *Future Generation Computer Systems*, 99(2019):558-570.

[7] ZhangXiaogang, WangDingxiang, et al. "Kernel modified optimal margin distribution machine for imbalanced data classification," *Pattern Recognition Letters*, 2019, 125(1) :325-332

[8] T. Lin, P. Goyal, riyal, R. Girshick, K. He and P. Dollar. "Focal loss for dense object detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2018, pp. 2999-3007

[9] Xiong, Huilin, *et al.* "Learning the Conformal Transformation Kernel for Image Recognition," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 28, no. 1, pp. 149-163, Jan. 2017.

[10] Razavi-Far, Roozbeh, Maryam Farajzadeh-Zanjani, and Mehrdad Saif. "An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors." *IEEE Transactions on Industrial Informatics* 13.6 (2017): 2758-2769.

[11] Pang, Shan, et al. "Fault diagnosis of rotating machinery with ensemble kernel extreme learning machine based on fused multi-domain features." *ISA transactions* (2019).

[12] W. Li, D. Wang, T. Chai, "Flame Image-Based Burning State Recognition for Sintering Process of Rotary Kiln Using Heterogeneous Features and Fuzzy Integral," *IEEE Trans. Ind. Inform*, vol. 8, no.4, pp. 780-790, 2012.

[13] Li W, Wang D, Chai T. Burning state recognition of rotary kiln using ELMs with heterogeneous features. *Neurocomputing*, 2013; 102:144–53.

[14] Chen, Hua, et al. "Recognition of the Temperature Condition of a Rotary Kiln Using Dynamic Features of a Series of Blurry Flame Images," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 148-157, 2015.

[15] Bakdi, Azzeddine, A. Kouadri, and A. Bensmail. "Fault detection and diagnosis in a cement rotary kiln using PCA with EWMA-based adaptive threshold monitoring scheme," *Control Engineering Practice*, 2017(66): 64-75.

[16] Makaremi, Iman, et al. "Identification and Abnormal Condition Detection of a Cement Rotary Kiln," *IFAC Proceedings*, 2008,41(2):7233-7238.

[17] Xiaogang Zhang, Lei Zhang, Hua Chen, et al. "Prediction of coal feeding during sintering in a rotary kiln based on statistical learning in the phase space," *ISA Transactions*, vol. 83, pp. 248–260, 2018.

[18] Zhang, Xiaogang, and Jingjing Zhao. "Prediction model for rotary kiln coal feed based on hybrid SVM." *Procedia Engineering*, 15 (2011): 681-687.

[19] Teng Zhang, Zhi-Hua Zhou. "Multi-Class Optimal Margin Distribution Machine." *Proceedings of the 34th International Conference on Machine Learning, PMLR*, 70(2017) 4063-4071.

[20] Makaremi, Iman, A. Fatehi, and B. N. Araabi. "Lipschitz Numbers: A Medium for Delay Estimation." *Ifac Proceedings*, 2008,41(2): 7468-7473.

[21] Ma, Suliang, et al. "High-Voltage Circuit Breaker Fault Diagnosis Using a Hybrid Feature Transformation Approach Based on Random Forest and Stacked Auto-Encoder." *IEEE Transactions on Industrial Electronics* (2018):9777-9788.

[22] G.E.Hinton, S.Osindero, Y.Teh, "A fast learning algorithm for deep belief nets. " *NeuralComput*, 18(2006)1527-1554.

[23] Qiao J, Chai T. "Soft measurement model and its application in raw meal calcination process. " *J Process Control*, 2012; 22:344-51.

[24] der Maaten, L. V., & Hinton, G. (2008). "Visualizing data using t-sne. " *Journal of Machine Learning Research*, 9, 11.

[25] Songcan, and D. Li. "Modified linear discriminant analysis." *Pattern Recognition*, 38.3(2005):441-443.
[26] Shahroudy, Amir, et al. "Deep multimodal feature analysis for action recognition in rgb+d videos." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40.5 (2017): 1045-1058.
[27] Mathew J, Pang C K, Luo M, *et al*. "Classification of Imbalanced Data by Oversampling in Kernel Space of Support Vector Machines," *IEEE Transactions on Neural Networks & Learning Systems*, 2017, 99(1): 1-12.
[28] Michał Koziarski. "Radial-Based Undersampling for imbalanced data classification", Pattern Recognition, Volume 102, 2020, 107262.
[29] Kai Qi, et al. "A new adaptive weighted imbalanced data classifier via improved support vector machines with high-dimension nature", Knowledge-Based Systems, Vol. 185, 2019, 104933.
[30] Fernández-Baldera, Antonio, Buenaposada, José M., and L. Baumela. "BAdaCost: Multi-class Boosting with Costs," *Pattern Recognition*, vol. 79, pp. 467-479, Jul. 2018.
[31] Maratea A, Petrosino A, Manzo M. "Adjusted F-measure and kernel scaling for imbalanced data learning," *Information Sciences*, vol. 257, no. 2, pp. 331-341, Jan. 2014.
[32] Williams, Peter, *et al*. "A Geometrical Method to Improve Performance of the Support Vector Machine," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 942-947, May. 2007.
[33] Crammer, Koby, and Y. Singer. "On the Learnability and Design of Output Codes for Multiclass Problems," *Machine Learning*, vol. 47, no. 2-3, pp. 201-233, May. 2002.

Dingxiang Wang received the B.S. degree from Hunan Institute of Engineering, China, and M.S. degrees from Changsha University of Science & Technology, China. He is currently a Ph.D. candidate of Control Science and Engineering of Hunan University, Changsha, China.

His research interests include machine learning, pattern recognition and data mining for industrial process applications.

Xiaogang Zhang (Corresponding Author) received the Bachelor's, Master's and Ph.D. degrees in Control theory and Engineering from Hunan University in 1996, 1999 and 2003 respectively. From 2015 to 2016, he was a Visiting Scholar with the Department of Computer Science, Brandeis University, Boston, USA.

He has been a Full Professor with the College of Electrical and Information Engineering, Hunan University, Changsha, China. His research interests include pattern recognition and data mining for industrial control systems. He has applied his expertise extensively in industrial practice, particularly in metallurgical process control and industrial robots.

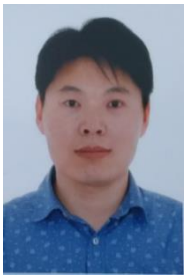
Hua Chen received the B.S., M.S. and Ph.D. degrees in control science and engineering from Hunan University, Changsha, China in 1996,2002 and 2014, respectively.

Now she is an Associate Professor with the College of Computer Science and Electronic Engineering, Hunan University. Her research interests include image and video processing, machine learning, and computer vision.

Yicong Zhou (M'07-SM'14) received the B.S. degree from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees from Tufts University, Massachusetts, USA, all in electrical engineering.

He is currently an Associate Professor and Director of the Vision and Image Processing Laboratory in the Department of Computer and Information Science at University of Macau. His research interests include image processing, computer vision, machine learning, and multimedia security.

Dr. Zhou is a senior member of the International Society for Optical Engineering (SPIE). He was a recipient of the Third Price of Macau Natural Science Award in 2014. He is a Co-Chair of Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society. He serves as an Associate Editor for *IEEE Transactions on Neutral Networks and Learning Systems*, *IEEE Transactions on Circuits and Systems for Video Technology*, *IEEE Transactions on Geoscience and Remote Sensing*, and four other journals.



Fanyong Cheng received the PhD degree in Control Science and Engineering from the Hunan University, China. His current research interests include Machine Learning, Pattern Recognition and Image Processing, Fault Detection and Diagnose.