A hybrid deep neural network based prediction of 300MW coal-fired boiler combustion operation condition

Journal:	*Science China Technological Sciences*
Manuscript ID	SCTS-2020-1458.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	14-Feb-2021
Complete List of Authors:	Han, Zhezhe; Southeast University, School of Energy and Environment Huang, Yizhi; Southeast University, School of Energy and Environment Li, Jian; Southeast University, School of Energy and Environment Zhang, Biao; Southeast University, School of Energy and Environment Hossain, Md. Moinul xu, chuanlong; Southeast University, School of Energy and Environment
Keywords:	Coal-fired power plant, Combustion operation condition prediction, Flame image, Convolutional sparse autoencoder, Least support vector machine
Speciality:	Environmental Science and Technology, Energy Sciences
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Figure 8.opj Figure 10.opj Figure 11.opj	

SCIENCE CHINA
Technological Sciences

• Article •

February 2021 Vol.xx No.2: xxx–xxx

doi: xxxxxxx

A hybrid deep neural network based prediction of 300MW coal-fired boiler combustion operation condition

HAN ZheZhe[1,3], HUANG YiZhi[1], LI Jian[1], ZHANG Biao[1],
Hossain Md. Moinul[2] & XU ChuanLong[1*]

[1] *Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education, School of Energy and Environment, Southeast University, Nanjing 210096, China;*

[2] *School of Engineering and Digital Arts, University of Kent, Canterbury, Kent CT2 7NT, UK;*

[3] *Lomon Billions Group Co., Ltd., Jiaozuo, Henan, 454191, China*

Received December 8, 2020, accepted February 8, 2021; published online xxx, 2021

In power generation industries, boilers are required to be operated under a range of different conditions to accommodate demands for fuel randomness and energy fluctuation. Reliable prediction of the combustion operation condition is crucial for an in-depth understanding of boiler performance and maintaining high combustion efficiency. However, it is difficult to establish an accurate prediction model based on traditional data-driven methods, which requires prior expert knowledge and a large number of labeled data. To overcome these limitations, a novel prediction method for the combustion operation condition based on flame imaging and hybrid deep neural network is proposed. The proposed hybrid model is a combination of convolutional sparse autoencoder (CSAE) and least support vector machine (LSSVM), i.e., CSAE-LSSVM, where the convolutional sparse autoencoder with deep architectures is utilized to extract the essential features of flame image, and then essential features are input into the least support vector machine for operation condition prediction. A comprehensive investigation of optimal hyper-parameter and dropout technique is carried out to improve the performance of the CSAE-LSSVM. The effectiveness of the proposed model is evaluated by 300MW tangential coal-fired boiler flame images. The prediction accuracy of the proposed hybrid model reaches 98.06%, and its prediction time is 3.06 millisecond/image. It is observed that the proposed model could present a superior performance in comparison to other existing neural network models.

coal-fired power plant, combustion operation condition prediction, flame image, convolutional sparse autoencoder, least support vector machine

*Corresponding author (email: chuanlongxu@seu.edu.cn)

Citation: Han Z Z, Huang Y Z, Li J, Zhang B, Hossain M M, Xu C L. A hybrid deep neural network based prediction of 300MW coal-fired boiler combustion operation condition. Sci China Tech Sci, 2021, xx: xxx, doi: xxx

1 Introduction

With the rapid development of industrialization and urbanization, energy demand and environmental problems have become increasingly prominent. Nowadays, renewable energy, such as wind energy and solar energy, has been widely applied in power system, but it also brings huge challenges for energy planning and electrical energy operation due to its stochastic and intermittent nature [1]. To absorb large-scale renewable energy for power generation, traditional coal-fired generator units are required to have the capacities of flexible operation and deep peak regulation. In this case, the utility boiler needs to operate under frequent and fast variable loads, easily causing unstable combustion conditions [2]. Unstable combustion will not only reduce energy utilization efficiency and increase pollutant emissions (such as NOx and SO2) but also cause furnace vibration and even safety accidents. Therefore, accurate and effective combustion operation condition monitoring is necessary, which is beneficial to increase boiler operating efficiency and reduce fuel consumption and emissions.

A great deal of effort has been devoted to developing intelligent methods for combustion operation condition monitoring [3, 4]. Compared with the traditional sensor methods (such as pressure/temperature sensors and flame detectors), flame imaging incorporating soft-computing technique [5] is considered to be a promising approach, which can provide more comprehensive measurement information, including temperature distribution, oscillation frequency, etc. Generally, these techniques involve two main stages, i.e., feature extraction and condition identification. Feature extraction methods extract low dimensional representative features of flame images [6]. For instance, Lin et al. [7] extracted first-, second-, third- and fourth-order statistic hue, saturation and intensity data of flame images to characterize the combustion operation conditions. However, these feature extraction methods have some deficits, such as (i) feature selection largely depends on the prior knowledge of expert experience, which is only suitable for specific diagnostic problems; (ii) poor generalization ability and (iii) lower prediction accuracy and difficulty to process high-dimensional data (such as high-resolution flame images). Therefore, the traditional feature extraction methods need further improvement for better generalization, big data processing ability, also to improve computational efficiency.

Recently, deep learning (DL) technique has attracted considerable attention in combustion diagnosis [8, 9]. The DL network learns discriminative features from flame images with the aid of multi-layer nonlinearity automatically, which not only overcomes the deficiency of inferior representative ability of

shallow models but also removes the tedious process of feature selection. For example, Abdurakipov et al. [10] established a convolutional neural network (CNN) to predict the combustion regimes in a swirling gas burner flame image. Wang et al. [11] proposed a modified CNN model to extract representative features of flame images and then predict the burning state. Although the prediction reliability of the combustion operation condition has significantly been improved by the DL network, the supervised learning (SL) network represented by CNN still has a thorny problem, that is, a large amount of labeled data is required during the training process whereas the scale and quality of labeled data directly determine the predictive performance. In practice, the formation of massive labeled data is challenging, which orders precise experimental devices and manual labeling [12]. This issue can only be addressed by an unsupervised learning network, which can be trained only using unlabeled data, thus reduce the demand for labeled data significantly. Akintayo et al. [13] adopted an autoencoder (AE) to detect the combustion operation condition using unlabeled laboratory-scale swirl combustor flame images. Qiu et al. [14] constructed a multi-layer AE to classify the combustion operation conditions of the pulverized coal furnace. In general, the unsupervised learning network usually has multi-hidden layers to extract higher-level data features. However, this multi-layer structure often brings training difficulties such as gradient vanishing or gradient explosion, resulting in difficulty to obtain the essential information of the input data. Therefore, an advanced feature extraction method is still worth exploring.

The combustion operation condition can be predicted by analyzing the extracted feature of each flame image and this is a classification process. Artificial neural network (ANN) and support vector machine (SVM) based techniques have been successfully applied in this process. For instance, Zhu et al. [15] introduced a three-layer ANN to classify combustion operation conditions of a supersonic combustor. However, ANN has some drawbacks, such as difficulty in determining the network structure (number of hidden layers and hidden neurons) and easy to cause overfitting. SVM has advantages compared to ANN, such as simple structure and strong generalization ability. Truong et al. [16] applied SVM for detecting flame conditions by analyzing the extracted parameters from the flame image. But the conventional SVM has high computational complexity and encounter overfitting [17]. The LSSVM converts inequality constraints of SVM's quadratic optimization problem into equality constraints in the solving process [18]. In this way, the LSSVM not only greatly simplifies the calculation process but also possesses good fitting accuracy and generalization performance.

This study presents a novel hybrid deep neural network (i.e., CASE-LSSVM) to predict combustion operation condition through flame images. In this model, the CASE with a deep structure is established for feature extraction from flame images. Then, the extracted features are analyzed by the LSSVM to predict the combustion operation condition. A detailed description of the hybrid model is discussed. Dropout training and performance evaluation metrics are also discussed and presented. The effectiveness of the proposed model is verified through 300MW tangential coal-fired boiler flame images. Experimental results and elaborates on the advantages of the proposed model compared with other benchmark models are also presented. Concluding remarks and future

direction of this study are given.

2 Proposed Hybrid Model

The technical strategy of the prediction model is shown in Fig. 1. The model consists of two stages: (1) model establishment, i.e., a combination of CSAE and LSSVM, and (2) model application. Detailed descriptions of the hybrid model can be found in Sections 2.1 and 2.2.

In the model establishment stage, flame images are acquired under different combustion operation conditions. Then, each flame image is resized to 128(H) × 128(V) and normalized to a range of 0 to 1. The preprocessed flame images are divided into the unlabeled and labeled dataset. The CSAE is established through unlabeled images. After that, the CSAE is used for feature extraction of labeled images. Finally, the LSSVM is established and trained using label information and extracted features.

In the model application stage, new unlabeled flame images are acquired and then performed data preprocessing, including image resize and normalization. After that, the CSAE-LSSVM model is used successively to perform feature extraction also to predict combustion operation condition without further training.

2.1 Convolutional sparse autoencoder for feature extraction

Autoencoder (AE) is an unsupervised neural network with three fully-connected layers, i.e., input layer (encoder), hidden layer (encoder vector) and output layer (decoder). The encoder is used to transform m-dimensional input sample x_i into n-dimensional encode vector h_i, typically $m>n$. Then, the decoder reconstructs the encode vector h_i to the m -dimensional output sample z_i. The AE performs training by minimizing the reconstruction error such as mean square error (MSE) between inputs and outputs, expressed as:

$$L_{MSE} = \frac{1}{M}\sum_{i=1}^{M}\left\|x_i - z_i\right\|^2, \quad (1)$$

An optimal weight matrix and bias vector can be obtained by utilizing a gradient descent algorithm. Once the AE training is completed, the extracted encode vector h_i can be regarded as a proper feature representation of the input sample.

Although the basic AE reconstructs the input information correctly, it is possible that the network simply copies the information from the input layer to the hidden layer. Then, the extracted features may be redundant and invalid. To avoid this, a sparse constraint is introduced to the AE, forming the sparse autoencoder (SAE). The principle of the SAE is to constrain the hidden-layer neurons in an inactive state, thereby extracting more representative features. If the activation value of the j^{th} hidden neuron is h_i^j ($i\in(1,M)$, $j\in(1,N)$), where M is the number of training samples; N is the number of neurons in the hidden layer. Then, the average activation value of the j^{th} hidden neuron can be calculated by:

$$p_j = \frac{1}{M}\sum_{i=1}^{M} h_i^j \quad (2)$$

where p_j is the average activation and the expected value is closer to zero. To achieve this goal, a penalty term $P_{penalty}$ is introduced to penalize p_j when it deviates from the sparse target p_{target}. The penalty term $P_{penalty}$ is defined as:

$$P_{penalty} = \sum_{j=1}^{N} KL\left(p_{target}\left\|p_j\right.\right) \quad (3)$$

where $KL(\cdot)$ is the Kullback–Leibler divergence (*KL* divergence) that can measure the difference between p_{target} and p_j, calculated by:

$$KL\left(p_{target}\left\|p_j\right.\right)=p_{target}\log\frac{p_{target}}{p_j}+\left(1-p_{target}\right)\log\frac{1-p_{target}}{1-p_j} \tag{4}$$

If $KL\left(p_{target}\left\|p_j\right.\right)=0$, $p_{target}=p_j$. Hence, the *KL* divergence is used as the loss function to achieve minimization, expressed as:

$$L_{Sparse}=\beta\sum_{j=1}^{N}KL\left(p_{target}\left\|p_j\right.\right) \tag{5}$$

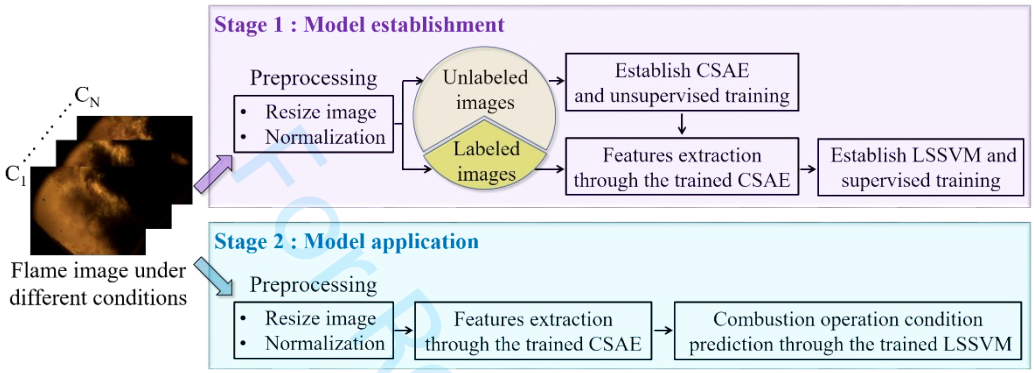


Figure 1 (Color online) Schematic diagram of the proposed hybrid model.

where β is the sparse rate. However, the fully-connected SAE suffers from training difficulties due to its numerous network parameters, especially in dealing with high-dimensional input data. To solve this issue, the convolution operation is utilized to replace the fully-connected operation and constructed convolutional sparse autoencoder (CSAE). In particular, the convolution operation has the characteristics of sparse connectivity and weight sharing, which can greatly reduce the number of network parameters. The architecture of the proposed CSAE is shown in Fig. 2.

In the CSAE, the flame image is sent to the input layer *x*, and then processed by five convolutional encoders, i.e., e_1, e_2, …, e_5 for feature extraction and five convolutional decoders, i.e., d_1, d_2, …, d_5 [refer to Fig. 2] for image reconstruction, and finally forming a reconstructed image in the output layer (*z*). The detailed configuration of the CSAE can be found in Table 1.

In the convolutional encoder, firstly, the input image is processed by the convolution layer [*C*(32@3×3+1)] of the encoder e_1, which has 32 filters, and each filter size is 3×3 and a stride of 1. After the convolution layer, the outputs are put through the activation function to improve the feature representation ability, such as rectified linear unit (ReLU) [$y(\lambda)$ = $max(0, \lambda)$, where λ represents hidden neuron]. Note that the Sigmoid function [$y(\lambda)$ = $1/1+\exp(-\lambda)$] is also used, i.e., d_5 operation (5th convolutional decoder). The purpose of using the ReLU function is to improve the training convergence speed, and the Sigmoid function is used to ensure the output range 0 to 1 [19]. Followed by the activation function, the pooling layer is adopted to reduce the dimensionality of the feature maps. The max-pooling

strategy [20] with a kernel size of 2×2 and a strider of 2 is used in this study. Finally, a feature map with a dimension of 64×64×16 can be extracted. After a series of similar encoder operations, the input image x with the dimension of 128×128×3 can be represented by the deep feature h with the dimension of 4×4×1.Afterwards, the extracted h is converted into a 16- dimensional vector to facilitate subsequent feature analysis.

In the convolutional decoder, the deep feature h is first sent into the upsampling layer [U(8×8)] of the decoder d_1 for dimensionality extension, and then successive processed by the convolution layer [C(4@3×3+1)] and ReLU activation function. After that, the feature map with the dimension of 8×8×4 is processed by a series of similar decoder operations for image reconstruction. Finally, the reconstructed image with the dimension of 128×128×3 is generated.

The CSAE training is only performed using unlabeled flame images, and the training target is to minimize the reconstruction error between the input image and the output image. After adding the sparse constraint in the CSAE training process, combining Eq. (1) and Eq. (5), the established CSAE loss function is expressed as:

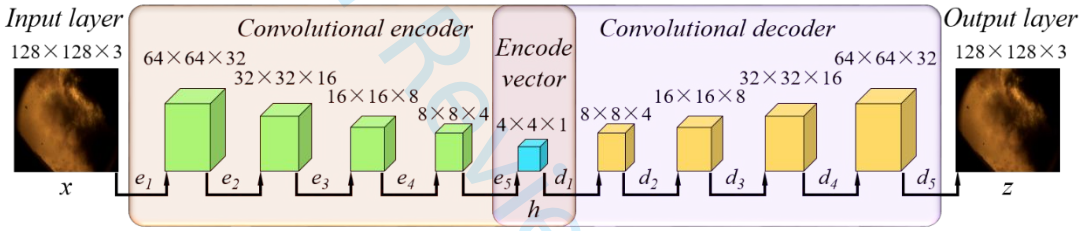


Figure 2 (Color online) The architecture of the convolutional sparse autoencoder.

Table 1 Structure and parameters of the CSAE

Items	Model type	Upsampling layer	Convolution layer	Activation function	Pooling layer	Output dimension
Input layer	/	/	/	/	/	128×128×3
Convolutional encoder	e_1	/	C(32@3×3+1)	*ReLU*	P(2×2+2)	64×64×32
	e_2	/	C(16@3×3+1)	*ReLU*	P(2×2+2)	32×32×16
	e_3	/	C(8@3×3+1)	*ReLU*	P(2×2+2)	16×16×8
	e_4	/	C(4@3×3+1)	*ReLU*	P(2×2+2)	8×8×4
	e_5	/	C(1@3×3+1)	*ReLU*	P(2×2+2)	4×4×1
Convolutional decoder	d_1	U(8×8)	C(4@3×3+1)	*ReLU*	/	8×8×4
	d_2	U(16×16)	C(8@3×3+1)	*ReLU*	/	16×16×8
	d_3	U(32×32)	C(16@3×3+1)	*ReLU*	/	32×32×16
	d_4	U(64×64)	C(32@3×3+1)	*ReLU*	/	64×64×32
	d_5	U(128×128)	C(32@3×3+1)	*Sigmoid*	/	128×128×3
Output layer	/	/	/	/	/	128×128×3

In Table 1, C represents convolution operation; P represents pooling operation; U represents upsampling operation; '/' represents no operation. $C(R@c\times c+g)$ represents the convolution layer with R filters. Each filter scans the input neurons with a fixed size of $c\times c$ and a stride of g. $P(p\times p+r)$ represents the pooling layer that condenses the feature map by selecting a maximum value with a kernel size of $p\times p$ and a step of r. $U(u\times u)$ represents the upsampling layer to extend the feature dimension to $u\times u$.

$$L_{CSAE} = L_{MSE} + L_{Sparse} \tag{6}$$

The L_{CSAE} is minimized iteratively by using the

adaptive moment estimation (Adam) algorithm [21]. The Adam algorithm has the advantages of high computational efficiency and low memory consumption.

Besides, overfitting is a common phenomenon in deep learning network training with a low training error, whereas high testing error. To avoid the occurrence of overfitting, the dropout technique [22] is utilized in the CSAE training. The main idea of the dropout technique is to randomly drop several neurons from the neural network with a probability of μ, thereby preventing an excessive co-adaptation between hidden neurons.

2.2 Least support vector machine for condition prediction

Once the CSAE training is completed, the deep features of the flame image are extracted by the trained convolutional encoder. These image features are then further used to train the LSSVM for combustion operation condition prediction. The LSSVM is an improved version of SVM, possessing a least-squares loss function and equality constraints [23]. It is a binary classifier, usually used for two-stage classification tasks. A brief review of the LSSVM algorithm for classification problems as following.

Assuming an observation dataset $\{X, Y\}$ of T samples, where $X=\{x_t\}_{t=1}^{T}$, $Y=\{y_t\}_{t=1}^{T}$, x_t denotes the t^{th} input sample, and y_t denotes the corresponding sample label, i.e., $y_t \in \{-1, +1\}$. Figure 3 presents a geometrical view of the LSSVM.

The LSSVM aims to construct an optimal hyperplane as a linear decision boundary that separates samples into two categories with maximal margin. It mainly solves the following optimization problem:

$$\min_{w,b,\xi} J(w,b,\xi)=\frac{1}{2}w^T w+\frac{C}{2}\sum_{t=1}^{t=T}\xi_t^2 \tag{7}$$

subject to:

$$y_t\left[w^T\Phi(x_t)+b\right]=1-\xi_t \tag{8}$$

where $J(w, b, \xi)$ is the structural risk; w and b are weight vector and bias, respectively; $C (C>0)$ is regularization constant that balances the importance between the maximization of the margin width and the minimization of the training error; the slack variable ξ_t is the soft margin error; $\Phi(\cdot)$ denotes the nonlinear function mapping input data from the

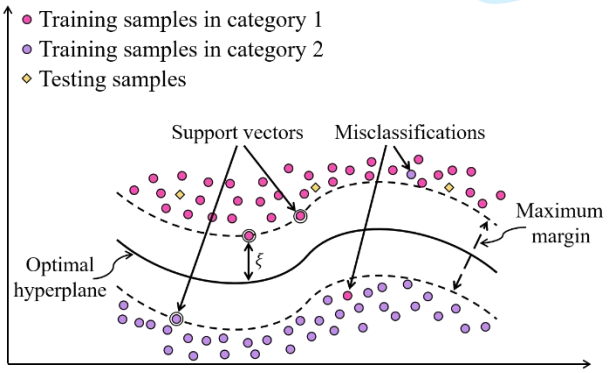


Figure 3 (Color online) Graphic representation of the LSSVM.

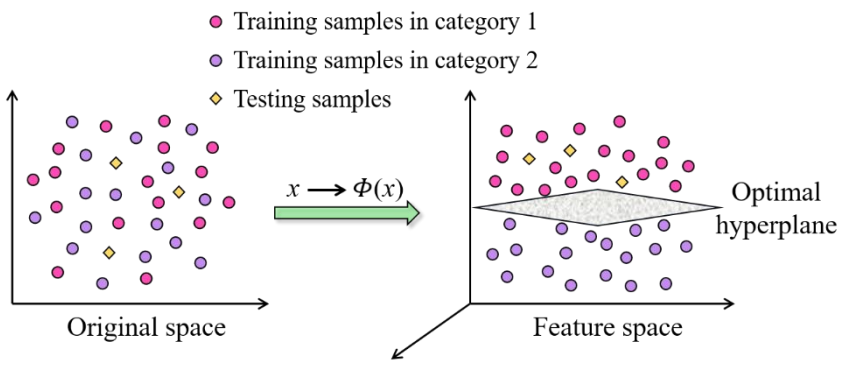


Figure 4 (Color online) Geometrical view of the nonlinear mapping.

original space into a high-dimensional feature space. This nonlinear mapping function is to solve the linear inseparable problem in the original space. As shown in Fig. 4, the training samples are inseparable in the original space (represented by 2-D space) but easy to separate in the high-dimensional feature space (represented by 3-D space).

By using the Lagrange method [24], the optimization problem Eq. (7) can be converted into a group of linear equations:

$$L(w,b,\xi,\alpha)=J(w,b,\xi)-\sum_{t=1}^{t=T}\alpha_t\left(w^T\Phi(x_t)+b-y_t+\xi_t\right) \tag{9}$$

where α_t(t=1,…,T) represents the Lagrange multiplier vector. According to the Karush–Kuhn–Tucker (KKT) condition [25], the optimal solution of Eq. (9) can be determined by:

$$\begin{cases}\dfrac{\partial L}{\partial w}=0\rightarrow w=\sum_{t=1}^{t=T}\alpha_t\Phi(x_t)\\ \dfrac{\partial L}{\partial b}=0\rightarrow\sum_{t=1}^{t=M}\alpha_t=0\\ \dfrac{\partial L}{\partial \xi_t}=0\rightarrow\alpha_t=C\xi_t\\ \dfrac{\partial L}{\partial \alpha_t}=0\rightarrow w^T\Phi(x_t)+b-y_t+\xi_t\end{cases} \tag{10}$$

By eliminating w, Eq. (10) can be transformed into the following matrix equation:

$$\begin{bmatrix}0 & \mathbf{1}^T\\ \mathbf{1} & \Omega+C^{-1}\boldsymbol{I}\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ \boldsymbol{y}\end{bmatrix} \tag{11}$$

where $\boldsymbol{I}$ is a $T\times T$ identity matrix; $\boldsymbol{y}=[y_1, y_2,\ldots, y_T]^T$ is the output matrix; $\boldsymbol{\alpha}=[\alpha_1, \alpha_2,\ldots, \alpha_T]^T$ is the Lagrange multipliers matrix; $\mathbf{1}=[1, 1,\ldots, 1]^T$ is the unit vector; Ω is the $T\times T$ kernel matrix and defined as:

$$\Omega=y_t y\Phi(x_t)^T\Phi(x)=y_t yK(x_t,x) \tag{12}$$

where x_t and x stand for different input points; $K(x_t, x)$ is the kernel function. It is worth noting that the kernel function defines the feature space of the input data, directly affecting the generalization performance of the LSSVM. Thus, selecting a proper kernel function is very critical. Through combining Eq. (11) and Eq. (12), $\boldsymbol{\alpha}$ and b can be obtained simultaneously. Then, the decision function of the LSSVM for data classification can be calculated as:

$$\psi(x)=sign\left[\sum_{t=1}^{t=T}y_t\alpha_t K(x_t,x)+b\right] \tag{13}$$

where *sign*(•) is the sign function, and the sign of the decision equation determined the category of the input data.

However, the combustion operation condition includes multiple categories. For this reason, the "one-against-one" approach [26] is considered for

performing the multi-class classification problem of the binary classifier LSSVM. In this approach, $n(n\text{-}1)/2$ classifiers are constructed based on every two categories, where n is the number of categories. A voting strategy is adopted to determine the final classification by the maximum number of votes. For example, combustion operation condition includes four categories: A, B, C and D. In the training process, six [4(4-1)/2] classifiers are constructed based on A *vs.* B, A *vs.* C, A *vs.* D, B *vs.* C, B *vs.* D, and C *vs.* D, respectively. In the testing process, these six classifiers perform estimation separately, and the final prediction result is determined according to the maximum number of votes.

2.3 Model evaluation criteria

To quantitatively evaluate the prediction performance of the proposed model, four metrics, including accuracy, precision, recall and F_1-score, are used. The formula for each evaluation metric is expressed as:

$$Accuracy(\varphi)=\frac{A}{B}\times 100\% \tag{14}$$

$$Precision(\varphi)=\frac{TP}{TP+FP} \tag{15}$$

$$Recall(r)=\frac{TP}{TP+FN} \tag{16}$$

$$F_1-score=2\times\frac{p\times r}{p+r} \tag{17}$$

where φ represents the prediction accuracy; A is the number of samples that are predicted correctly; B is the number of the testing samples; TP is the number of true positives; FP is the number of false positives; FN is the number of false negatives. The F_1-score is a harmonic mean of precision and recall, whose value ranges from 0 to 1. The higher value of these four evaluation metrics indicates better predictive performance.

3 Experiments on a 300MW coal-fired boiler

3.1 Experimental setup

To evaluate the effectiveness of the proposed hybrid model (CSAE-LSSVM) for predicting the combustion operation condition, experiments were conducted on a 300MW tangentially coal-fired boiler. This boiler adopted a direct-blowing pulverizing system and was equipped with five groups of roller mills (*A*–*E*). Only four groups (*A*–*D*) were utilized under the normal operation conditions, and the remaining *E* mill was usually on standby in case of accidents. The outlet of each mill is connected to four tilting burners, which are arranged at four corners of the boiler. The detailed structural design of the experimental boiler can be found in Ref. [27, 28].

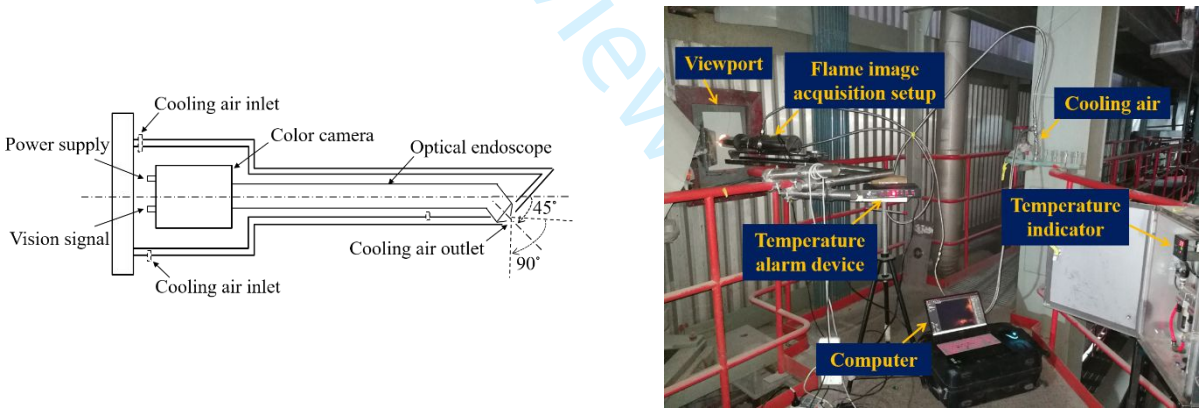
In this study, the combustion operation condition of the burner connected with mill *C* is selected as the research object. Figure 5 presents the structure and practical implementation of the flame imaging system. The flame imaging system is mainly composed of a color camera and an optical endoscope. The color camera (HIKVISION MV-CA003-50GC) is used to record the flame image with a resolution of 640(H) × 480(V) pixels at 20 f/s (frames per second). The optical endoscope can extend into the furnace and transmit the combustion flame to the camera through a set of lenses. It is equipped with a 90° angle of objective view lens, which provides a wider visual range and covers the main reaction area (i.e., root region) of the flame. Moreover, the front end of the endoscope has bent 45° to overcome the limitation of the installation position.

The flame imaging system is mounted on the viewport located on the side of the burner and protected by cooling air that ensures that the camera operating temperature does not exceed 50°C. The temperature is monitored in real-time, with an over-temperature alarm function. The pressure of the cooling air is always higher than that of the boiler. Then, the cooling air can flow into the boiler through the front end of the optical endoscope, thus achieving a cooling effect. Also, the cooling air is conducive to keeping the monitoring windows clean for a long time running.

3.2 Flame image acquisition under different operation conditions

During the experiment, the coal-fired boiler was undergoing a peak shaving process. In this case, the boiler will continuously adjust the operational variables to meet the demand for energy fluctuations. Table 2 depicts the flame image dataset obtained from eight combustion operation conditions. These operation conditions are determined by the coal quantity and air volume. For each condition, 1000 RGB (Red, Green, and Blue) flame images were collected. Note that the coal quality is consistent in this experiment.

Examples of flame images under different operation conditions are shown in Fig. 6. Although the size, brightness and structure of the flame image are different, it is difficult to identify the condition based on these physical appearances accurately. In view of this, how to predict the operation condition effectively and accurately is desirable and yet challenging.



(a) Schematic diagram of the flame image imaging system. (b) Practical implementation.

Figure 5 (Color online) The physical implementation of the flame imaging system.

Table 2 Overview of the combustion operation condition used in this study

Condition	Coal quantity (t/h)	Air volume (km3/h)	Total images
C1	62.8	603.8	1000
C2	69.9	620.8	
C3	74.6	628.3	
C4	87.1	642.2	
C5	89.5	667.3	
C6	93.6	673.9	

C7	95.2	688.1
C8	96.5	695.9

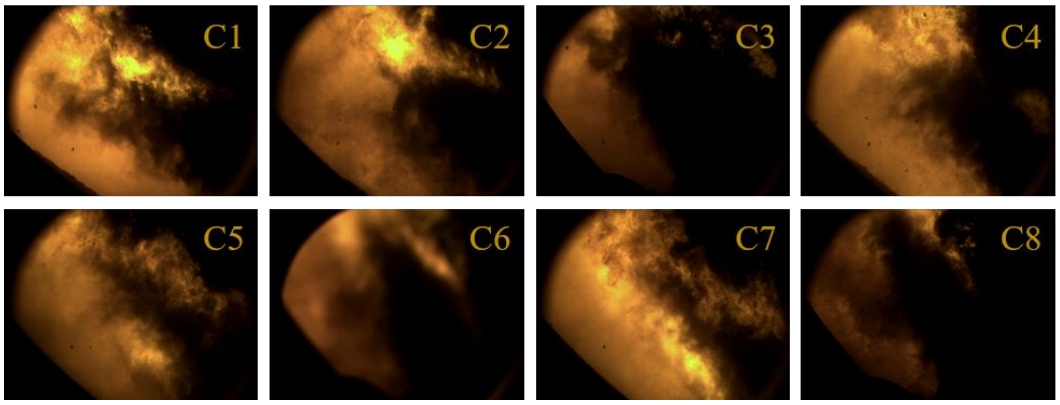


Figure 6 (Color online) Examples of flame images under eight combustion operation conditions.

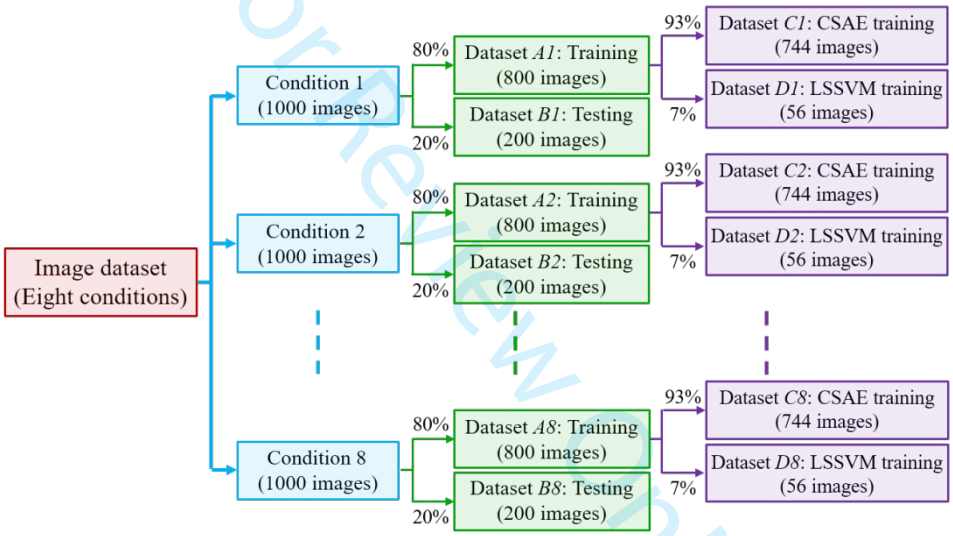


Figure 7 (Color online) Overview of the training and testing dataset.

3.3 Dataset and model establishment

To eliminate the influence of different image sizes and accelerate the convergence of the prediction model, all flame images are resized to 128(H) × 128(V) and normalized to 0 to 1 using the min-max scale [29]. Especially, the dimension of the input images is chosen empirically based on the trade-off between the prediction accuracy and prediction time of the model. Figure 7 illustrates the overall structure of the flame image dataset used for training and testing the prediction model. Take Condition 1 (containing 1000 images) as an example, 80% of the data is randomly selected to form training dataset *A1* (containing 800 images), and the remaining 20% is used as testing dataset *B1* (containing 200 images). Subsequently, 93% of the dataset *A1* is selected to form dataset *C1*(containing 720 images), and the rest of 7% forms dataset *D1*. Considering eight conditions, the dataset for unsupervised CSAE training contains 5952

unlabeled images, defined as dataset *C*; the dataset for supervised LSSVM training contains 448 labeled images, defined as dataset *D*; and the dataset for the CSAE-LSSVM testing contains 1600 labeled images, defined as dataset *B*.

From the dataset structure, a large amount of unlabeled data is used for feature extraction network CSAE training, and only a small amount of labeled data is used for condition identification network LSSVM training.

The CSAE-LSSVM training includes two independent steps, i.e., unsupervised CSAE training and supervised LSSVM training. In the CSAE training process, all the weight parameters are initialized by Gaussian distribution with a standard deviation of 0.02 and updated via back-propagation using the Adam algorithm [30]. The sparse penalty term is utilized to further improve CSAE feature learning ability, where the sparse rate β is set to be 0.4, and the sparse target p_{target} is set to be 0.08. Besides, the dropout technique is also introduced in the CSAE training with the *dropout rate* μ=0.2, which is determined via cross-validation compared to other rates such as 0.1, 0.3, 0.4 and 0.5. The LSSVM training is performed based on the CSAE feature and label information of the dataset *D*. Moreover, in comparison with different kernel functions such as linear and polynomial, the radial basis function (RBF) is selected as the kernel function.

All the calculations in this study were implemented in the Python programming language with an Intel i7-8700K CPU, 64 GB RAM and GeForce GTX 1080 Ti GPU. For each experimental result, ten trials are repeated to reduce particularity, contingency, and randomness.

4 Results and discussions

4.1 Image feature visualization

Once the CSAE training is completed, the CSAE is used to extract the flame image features. The extracted discriminative features can be visualized through the t-distributed stochastic neighbor embedding (t-SNE) technique [31]. The t-SNE method provides an effective solution for high-dimensional data visualization, which can convert the extracted 16-dimensional image features into a 2-dimensional (2-D) feature map with maximum preservation of data structure. Figure 8 shows the 2-D feature map of the flame image in testing dataset *B*.

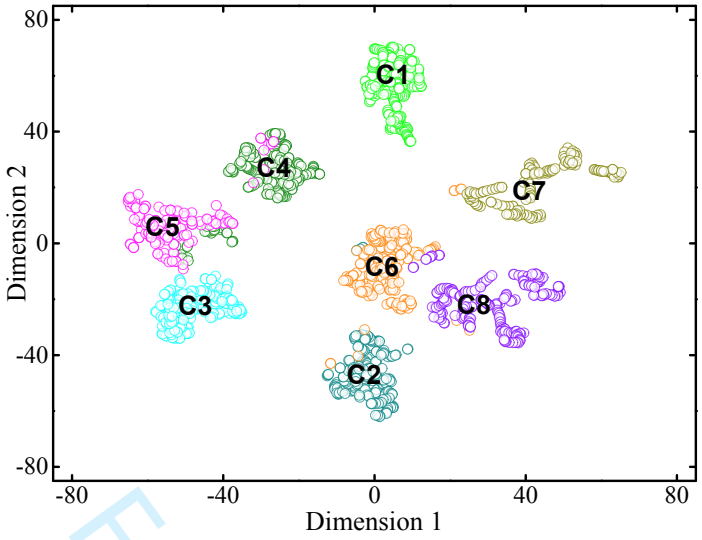


Figure 8 (Color online) Image feature visualization under eight combustion operation conditions.

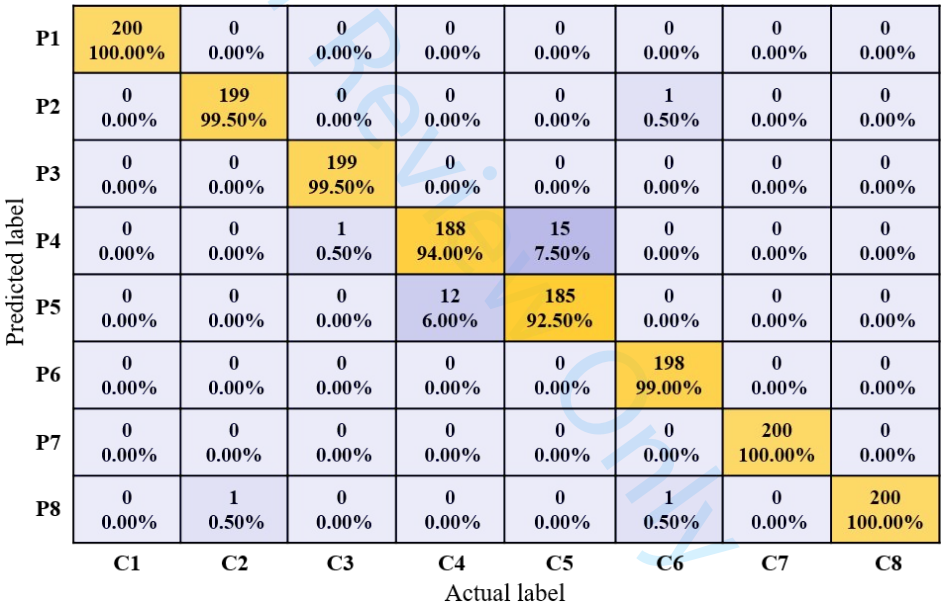


Figure 9 (Color online) Confusion matrix of classification results of the testing dataset.

It can be seen that the image features in the same conditions are clustered automatically, while the image features in different conditions are separated well. This result shows that even though the original flame images are difficult to distinguish intuitively, their features have a good separability in the feature space. Notably, some mixing points are appeared and mainly concentrated in combustion operation conditions 4 and 5. This confusion reflects some similarities between flame images, especially from adjacent operation conditions. Nevertheless, most feature points can be distinguished, indicating the effectiveness of CSAE in image feature extraction.

4.2 Prediction accuracy

After the CSAE-LSSVM training, the dataset *B* (containing 200 labeled images per condition) is used to examine its prediction performance. As shown in Fig. 9, the confusion matrix summarizes the classification results of the eight combustion operation conditions. In the confusion matrix, columns represent the actual labels, and rows represent the predicted labels. Diagonal cells show the number and accuracy of correctly estimated samples, while off-diagonal cells display the number of misclassified samples.

It can be seen that conditions 1, 7 and 8 can be fully identified with the accuracy of 100%. Especially, the prediction accuracy of condition 5 is the lowest with the accuracy of 92.50%, where only 185 images are correctly classified, and the remaining images are mainly misclassified into condition 4. Followed by condition 4, the accuracy is only 94.00%, showing the appearance of confusion with condition 5. This result is consistent with that given in Fig. 8, that is, misclassification is prone to occur under condition 4 and condition 5. Because of this, it can be inferred that if the flame images of different conditions cannot be completely separated in the feature space, it will inevitably affect the prediction accuracy of the subsequent classifiers. This suggesting that the representative features of flame image play a decisive role in high-precision combustion operation condition prediction. Although some misclassifications occur, most of the testing images can be accurately classified, and the average prediction accuracy has reached 98.06% [refer to Table 3]. This sufficiently confirms that the proposed CSAE-LSSVM performed well in predicting the combustion operation condition.

4.3 Comparative study and performance analysis

The performance of the hybrid model (CSAE-LSSVM) is verified under different hyper-parameter, such as *dropout rate* and kernel functions. Furthermore, to verify the superiority of the hybrid model, a comprehensive comparative study is performed with other typical networks, including different feature extraction networks and condition prediction models.

4.3.1 Effect of different dropout rates and kernel functions

To tackle the overfitting problem, the dropout technique is involved in the CSAE training process. Since the *dropout rate* is a vital hyper-parameter, it is meaningful to study its impact on prediction performance. Besides, the kernel function has a significant influence on the generalization ability of the LSSVM. Thus, the LSSVM performance under three different kernel functions is also investigated. The three kernel functions are specifically expressed as:

(1) Linear function (LF): $K\left(x_t, x\right) = x_t \cdot x$

(2) Polynomial function (PF): $K\left(x_t, x\right) = \left(x_t \cdot x + z\right)^v$

(3) Radial basis function (RBF):

$$K\left(x_t, x\right) = exp\left(-\frac{\left\|x_i - x\right\|^2}{2\sigma^2}\right)$$

where z, v and σ are parameters of the kernel function. Figure 10 presents the prediction results of the CSAE-LSSVM, where the *dropout rate* is changed from 0 to 0.5 with a step size of 0.1.

The result demonstrates that, with the increase of the *dropout rate*, the prediction accuracy of these three networks has experienced a process of first increasing and then decreasing. The best performance was obtained at a *dropout rate* close to 0.2. This verifies that the reasonable dropout method can improve the CSAE performance, but a high

dropout rate may lose important neurons, resulting in the degradation of classification performance. The result also indicates that the CSAE-LSSVM with PF kernel function can achieve the highest prediction accuracy in all cases, so it is preferred.

4.3.2 Effect of different feature extraction networks

To verify the superiority of the CSAE feature method, a comparative study is carried out with other feature extraction methods, such as empirical formula (EF), principal component analysis (PCA) and convolutional autoencoder (CAE). The EF method uses the pre-defined empirical formula to derive various statistical features from the flame image. In this study, seven different statistical features were considered, including geometric parameters (ignition point and ignition area) and luminous parameters (luminous region, brightness, non-uniformity, mean intensity and flame area). The PCA is an effective data dimensionality reduction method that can discover the principal variables of the flame images. For the CAE, its structure is the same as the CSAE, except that sparse constraint is not considered. For a fair comparison, different types of features are input to the same LSSVM to perform prediction. Table 3 summarizes the prediction accuracy, F_f-score and prediction time of the neural network established based on different feature extraction methods.

The EF-based feature extraction method is at a relatively low level in terms of prediction accuracy and prediction

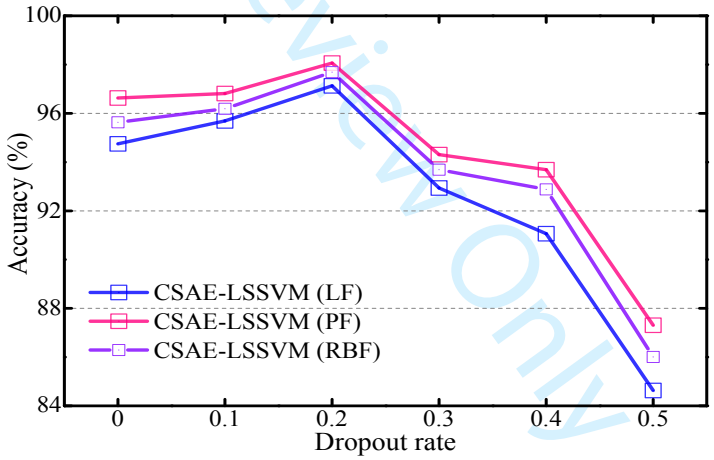


Figure 10 (Color online) Performance testing of the CSAE-LSSVM under different dropout rates and kernel functions.

Table 3 Prediction network performance based on different feature extraction methods

Prediction network	Accuracy (%)	F_f-score	Prediction time (ms/image)
EF-LSSVM	63.06%	0.63	33.93
PCA-LSSVM	91.75%	0.91	1.15
CAE-LSSVM	96.94%	0.97	2.48
CSAE-LSSVM	**98.06%**	**0.98**	**3.06**

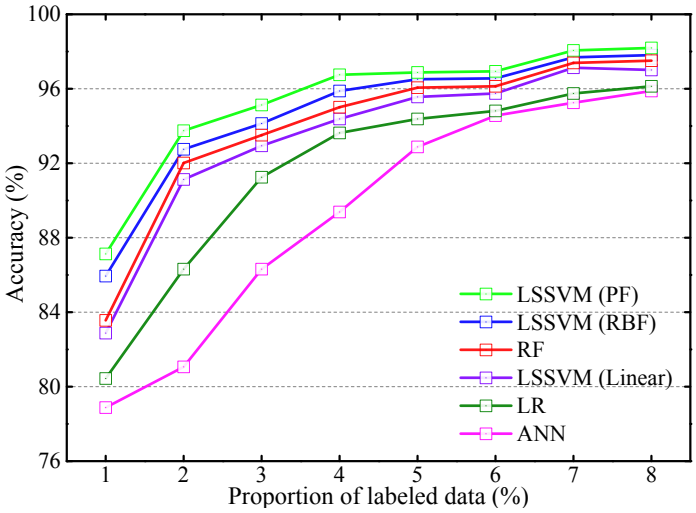


Figure 11 (Color online) Prediction accuracy of the classifiers under different proportions of labeled images.

time. More importantly, these pre-defined formulas are often accompanied by a complicated feature selection process, which relies heavily on prior expert knowledge. The result also indicates that, although PCA-LSSVM achieves the fastest prediction time (i.e., 1.15 ms/image), its accuracy is lower than the CSAE-LSSVM. This is mainly because PCA is a simple linear transformation method that cannot extract the essential data features.Compared with CAE-LSSVM, CSAE-LSSVM achieves a more satisfactory performance with an accuracy of 98.06% and an F_1-score of 0.98, although its consumption time is slightly higher (i.e., 3.06 ms/image) due to the increased sparse constraint. However, from a practical engineering perspective, the prediction time of the CSAE-LSSVM is acceptable, satisfying the requirement of real-time processing data acquired by the color camera with a frame rate of 20 f/s. It can be concluded that the CSAE-LSSVM is better than the shallow PCA-LSSVM, indicating that the hybrid deep learning network can extract more representative flame image features. The CSAE-LSSVM is superior to the CAE-LSSVM, suggesting that the sparse constraint can further enhance feature learning ability. In summary, the CSAE is more recommended as an optimum feature extraction network for flame images.

4.4.3 Effect of different classifiers and proportions of labeled data

After extracting the representative features of the flame image, it is critical to select an appropriate classifier, which is beneficial to the accurate analysis of the features. A trial is carried out to compare the LSSVM performance with other typical classifiers, such as ANN, random forest (RF), and logistic regression (LR). Meanwhile, the prediction of the LSSVM under three different kernel functions is also considered. Typically, due to the different structures of the neural network classifiers, the demand for labeled data is also different. It is desirable to adopt an optimal classifier that can provide better accuracy with a minimum labeled data, which will help reduce the burden of preparing high-quality and large-scale labeled data. Therefore, the proportion of labeled data has also been carefully investigated. In this study, the proportion of labeled data refers to the proportion of dataset *D* to dataset *A*, as shown in Fig. 7. Figure 11

shows the prediction accuracy of different classifiers based on CSAE features, where the proportion of labeled data varies from 1% to 8% with a step size of 1%.

The LSSVM with the PF kernel function always maintains the highest prediction accuracy in all cases. As the proportion increases from 1% to 7%, the accuracy of all classifiers increases rapidly. Although only 6% of the labeled images are available, the minimum prediction accuracy reaches 95% (provided by the ANN), which meets the accuracy requirements of actual engineering [5]. With the further increase of the proportion, the prediction accuracy improves slightly but stabilized after 7%. The obtained results not only demonstrate the LSSVM with PF kernel function achieves the best performance in predicting the combustion operation condition but also reflects that 7% of the label data can provide a high-precision prediction. Therefore, it can be determined that the proposed prediction model is suitable for prediction tasks with limited labeled data.

5 Conclusion

A novel hybrid deep neural network model (CASE-LSSVM) is proposed to predict the combustion operation condition. The CASE is used to feature extraction of flame images, and the LSSVM is utilized to predict the combustion operation condition based on the extracted flame features. The proposed model overcomes the shortcomings of the traditional techniques where prior expert knowledge and massive labeled data are required. Experiments were carried out on a 300MW tangential coal-fired boiler and collected flame images under different operation conditions and to evaluate the prediction model. The main conclusions of this study can be drawn as follows:

(a) Flame structure changes with the combustion operation conditions. Although flame images in different conditions are difficult to distinguish based on their physical appearance, the extracted features can be separated accurately. For the examined conditions, the established CASE-LSSVM provided 98.06% prediction accuracy and 3.06 ms/image prediction time, superior to other prediction models.

(b) The CASE can automatically extract the representative features of the flame image in an unsupervised manner. In comparison with other feature extraction methods, the CSAE not only achieve good prediction accuracy but also revoke the tedious process of feature selection. More importantly, practically proved that representative flame features are the key factor in obtaining satisfactory prediction performance of combustion operation condition.

(c) This study conducts a detailed exploration, including the sparse constraint method, reasonable selection of kernel function, and the optimum *dropout rate*. Through the cross-validation method, it is suggested that the polynomial function is more suitable with a *dropout rate* of 0.2.

(d) Experimental results demonstrate that the proposed model performs well in dealing with the problem of limited labeled data, which significantly reduces the demand for image labels. Consequently, the proposed model has a great application prospect in combustion operation condition prediction.

Overall, the proposed model can easily be applied to other combustion processes such as heavy-oil, biomass co-combustion, etc. The future work will be focused on tailoring the proposed model for predicting combustion operation conditions in more combustion

facilities.

This work was supported by the National Natural Science Foundation of China [grant number 51976038], the Natural Science Foundation of Jiangsu Province, China for Young Scholars [grant number BK20190366], and the China Scholarship Council [grant number 202006090164].

1 Morrone P, Algieri A, Castiglione T. Hybridisation of biomass and concentrated solar power systems in transcritical organic Rankine cycles: A micro combined heat and power application. Energy Convers Manage, 2019, 180: 757-768

2 Li X, Huang Y, Zhao D, et al. Stability study of a nonlinear thermoacoustic combustor: Effects of time delay, acoustic loss and combustion-flow interaction index. Appl Energy, 2017, 199: 217-224

3 González-Cencerrado A, Gil A, Peña B. Characterization of PF flames under different swirl conditions based on visualization systems. Fuel, 2013, 113: 798-809

4 Golgiyaz S, Talu M, Onat C. Artificial neural network regression model to predict flue gas temperature and emissions with the spectral norm of flame image. Fuel, 2019, 255: 115827

5 Bai X, Lu G, Hossain M, et al. Multi-mode combustion process monitoring on a pulverised fuel combustion test facility based on flame imaging and random weight network techniques. Fuel, 2017, 202: 656-664

6 Sun D, Lu G, Zhou H, et al. Quantitative Assessment of Flame Stability Through Image Processing and Spectral Analysis. IEEE Trans on Instrum Meas, 2015, 64(12): 3323-3333

7 Lin B, Jørgensen S. Soft sensor design by multivariate fusion of image features and process measurements. J Process Control, 2011, 21(4): 547-553

8 Chen H, Yan T, Zhang X. Burning Condition Recognition of Rotary Kiln Based on Spatiotemporal Features of Flame Video. Energy, 2020, 211: 118656

9 Liu Y, Fan Y, Chen J. Flame Images for Oxygen Content Prediction of Combustion Systems Using DBN. Energy Fuels, 2017, 31(8): 8776-8783

10 Abdurakipov S, Gobyzov O, Tokarev M, et al. Combustion Regime Monitoring by Flame Imaging and Machine Learning. Optoelectron Instrument Proc, 2018, 54(5): 513-519

11 Wang Z, Song C, Chen T. Deep Learning based Monitoring of Furnace Combustion State and Measurement of Heat Release Rate. Energy, 2017, 131:106-112

12 Lu C, Wang Z, Zhou B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Adv Eng Inform, 2017, 32: 139-151

13 Akintayo A, Lore K, Sarkar S, et al. Early Detection of Combustion Instabilities using Deep Convolutional Selective Autoencoders on Hi-speed Flame Video. Int J Prognostics Health Monitor, 2016

14 Qiu T, Liu M, Zhou G, et al. An Unsupervised Classification Method for Flame Image of Pulverized Coal Combustion Based on Convolutional Auto-Encoder and Hidden Markov Model. Energies, 2019, 12(13): 2585

15 Zhu X, Cai Z, Wu J, et al. Convolutional neural network based combustion mode classification for condition monitoring in the supersonic combustor. Acta Astronautica, 2019, 159: 349-357

16 Truong T, Kim J. Fire flame detection in video sequences using multi-stage pattern recognition techniques. Eng Appl Artif Intell, 2012, 25(7): 1365-1372

17 Kim D, Lee H, Cho S. Response modeling with support vector regression. Expert Syst Appl, 2008, 34(2): 1102-1108

18 Han H, Cui X, Fan Y, et al. Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features. Appl Therm Eng, 2019, 154:540-547

19 Jiao J, Zhao M, Lin J, et al. A Multivariate Encoder Information based Convolutional Neural Network for Intelligent Fault Diagnosis of Planetary Gearboxes. Knowl Based Syst, 2018, 160: 237-250

20 Han Z, Li J, Zhang B, et al. Prediction of combustion state through a semi-supervised learning model and flame imaging. Fuel, 2020, 289(4): 119745

21 Fei Z, Wu Z, Xiao Y, et al. A new short-arc fitting method with high precision using Adam optimization algorithm. Optik, 2020, 212: 164788

22 Sun W, Shao S, Zhao R, et al. A Sparse Auto-encoder-Based Deep Neural Network Approach for Induction Motor Faults Classification. Measurement, 2016, 89:171-178

23 Houthuys L, Langone R, Suykens J. Multi-View Least Squares Support Vector Machines Classification. Neurocomputing, 2018, 282: 79-88

24 Tan Z, De G, Li M, et al. Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. J

Clean Prod, 2020, 248: 119252

25 Liu C, Tang L, Liu J. Least Squares Support Vector Machine with Self-organizing Multiple Kernel Learning and Sparsity. Neurocomputing, 2018, 331: 493-504

26 Hsu C, Lin C. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Networks, 2002, 13(2): 415-425

27 Gubba S, Ingham D, Larsen K, et al. Numerical modelling of the co-firing of pulverised coal and straw in a 300 MWe tangentially fired boiler. Fuel Process Technol, 2012, 104: 181-188

28 Wang X, Tan H, Niu Y, et al. Experimental investigation on biomass co-firing in a 300MW pulverized coal-fired utility furnace in China. Proc Combust Inst, 2011, 33(2): 2725-2733

29 Han Z, Hossain M, Wang Y, et al. Combustion stability monitoring through flame imaging and stacked sparse autoencoder based deep neural network. Appl Energy, 2019, 259: 114159

30 Wan K, Hartl S, Vervisch L, et al. Combustion regime identification from machine learning trained by Raman/Rayleigh line measurements. Combust. Flame, 2020, 219: 268-274

31 Laurens V, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res, 2008, 9(2605): 2579-2605

Fig. 1

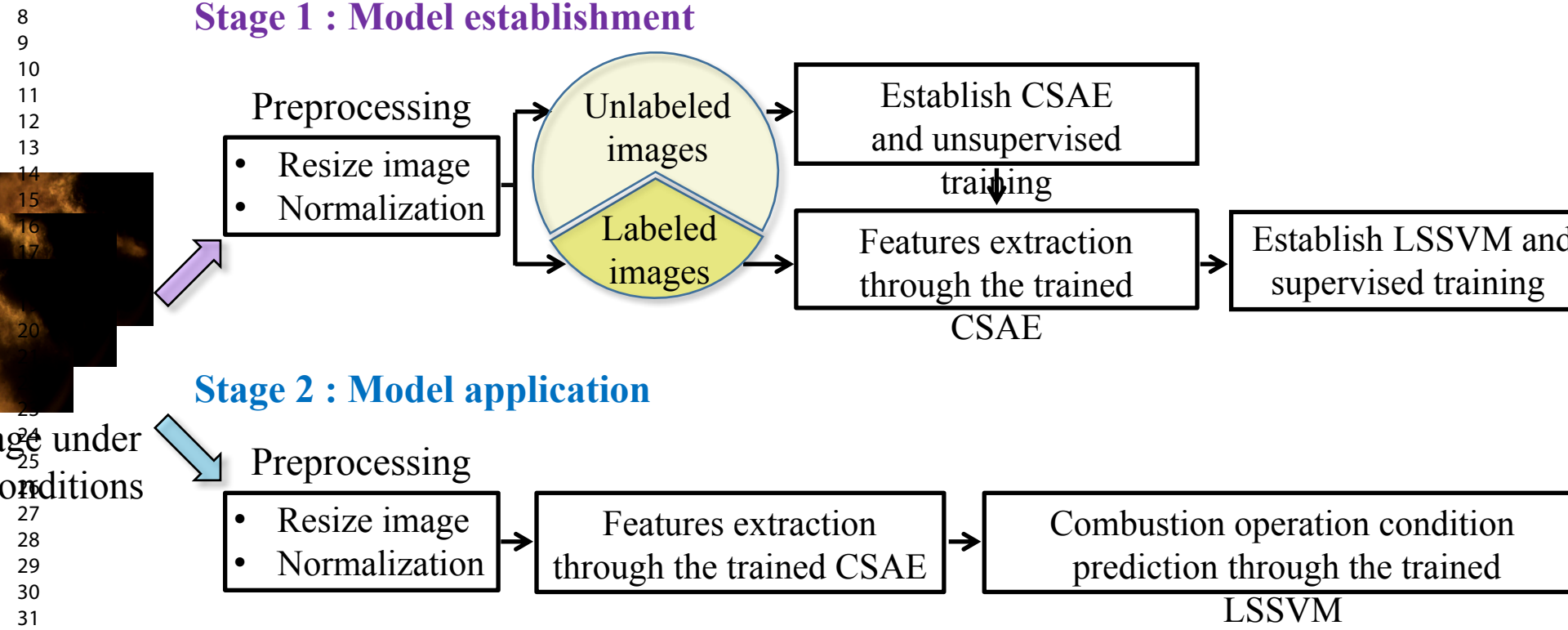

Stage 1 : Model establishment
Preprocessing
Resize image
Normalization
Unlabeled images
Labeled images
Establish CSAE and unsupervised training
Features extraction through the trained CSAE
Establish LSSVM and supervised training
Stage 2 : Model application
Preprocessing
Resize image
Normalization
Features extraction through the trained CSAE
Combustion operation condition prediction through the trained LSSVM

Fig. 2

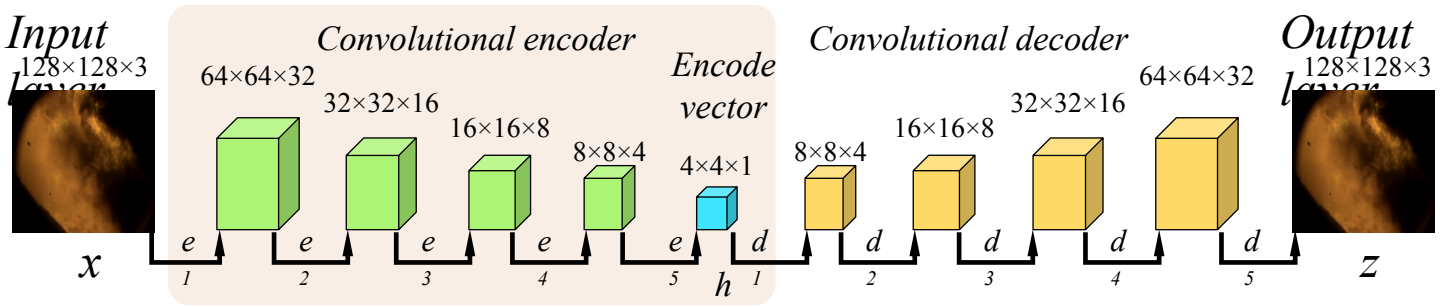

Input
128×128×3
Convolutional encoder
64×64×32
32×32×16
16×16×8
8×8×4
4×4×1
Encode vector
Convolutional decoder
8×8×4
16×16×8
32×32×16
64×64×32
Output
128×128×3
e
d
x
h
z

Fig. 3

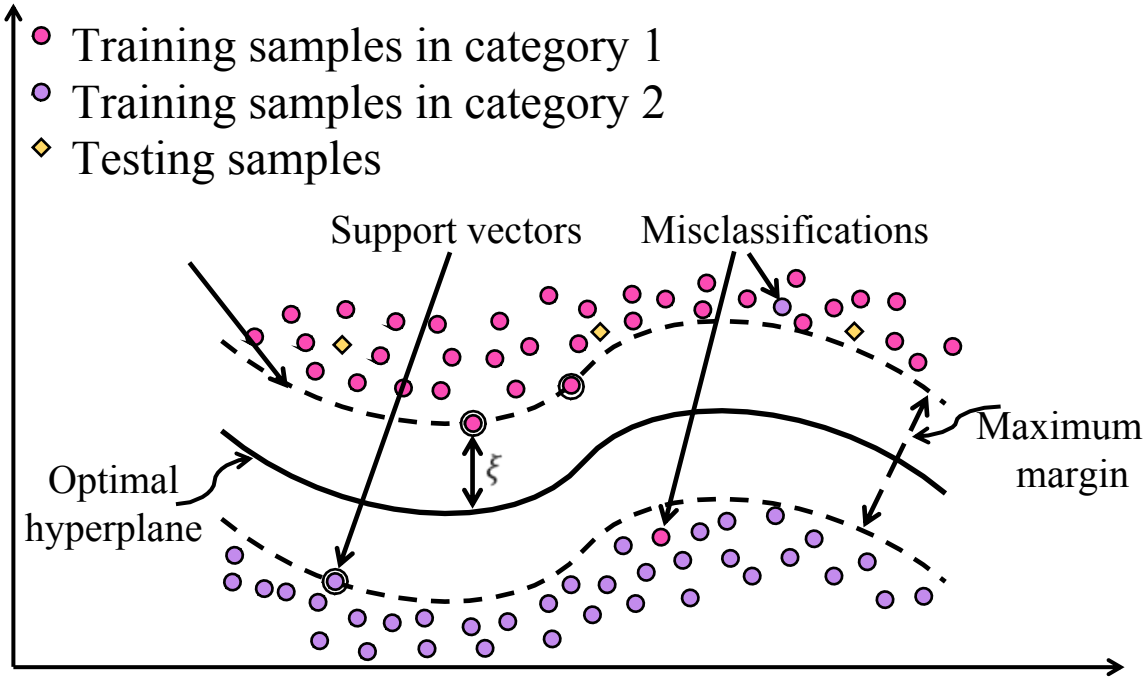

Training samples in category 1
Training samples in category 2
Testing samples
Support vectors
Misclassifications
Maximum
margin
Optimal
hyperplane
ξ

Fig. 4

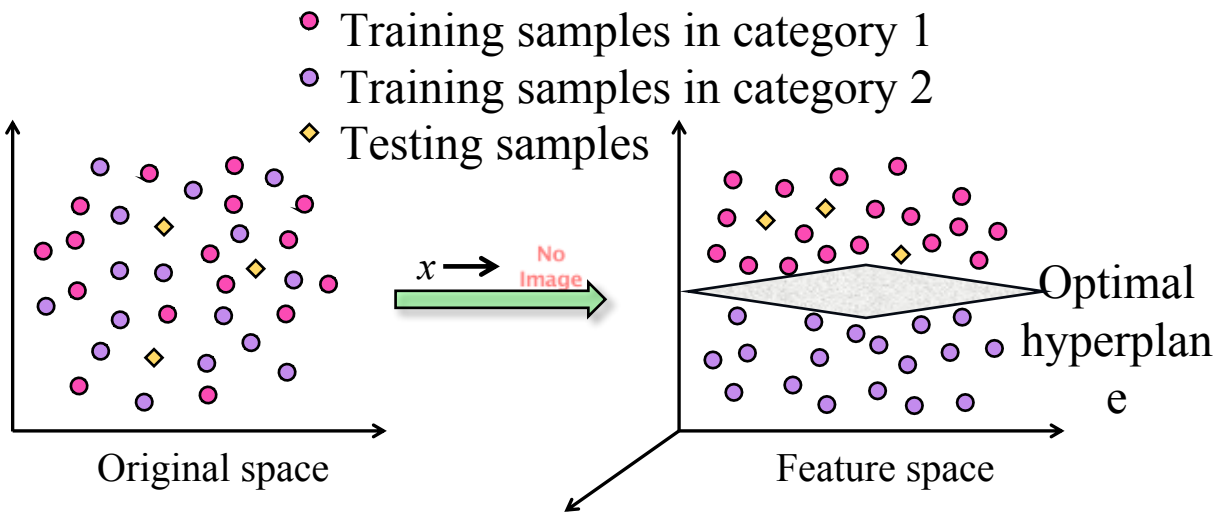

Training samples in category 1
Training samples in category 2
Testing samples
x
No Image
Optimal hyperplan e
Original space
Feature space

Fig. 5 (a)

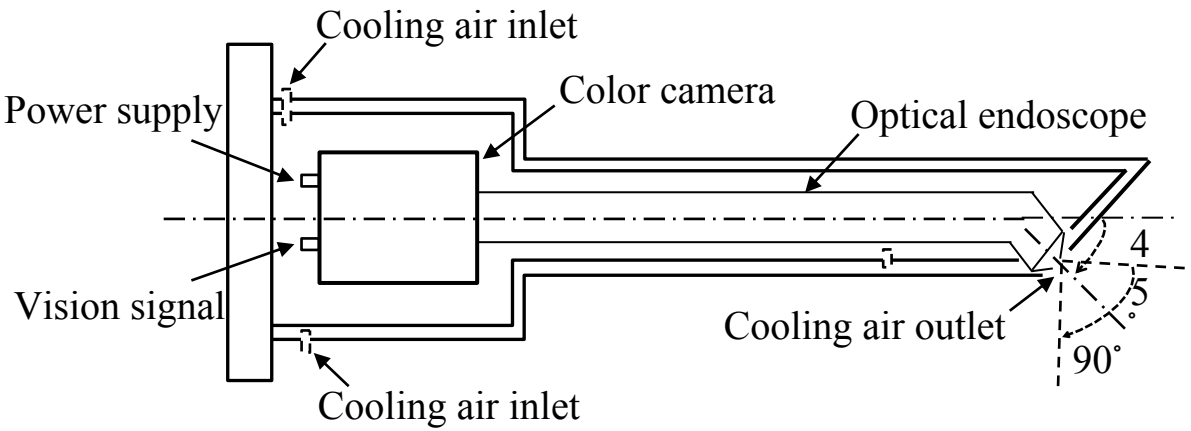

Cooling air inlet
Power supply
Color camera
Optical endoscope
Vision signal
Cooling air outlet
Cooling air inlet
4
5°
90°

Fig. 5 (b)

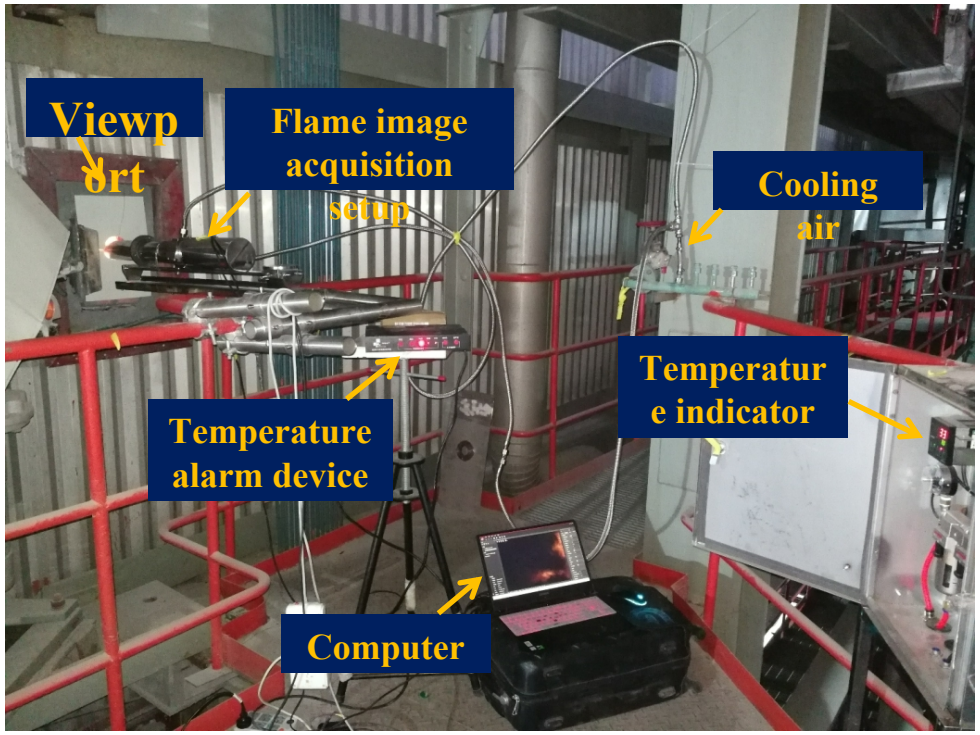

Viewport
Flame image acquisition setup
Cooling air
Temperature indicator
Temperature alarm device
Computer

Fig. 6

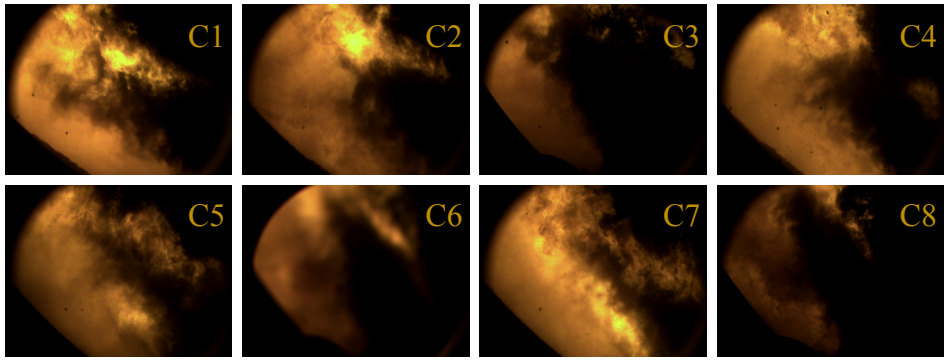


Fig. 7

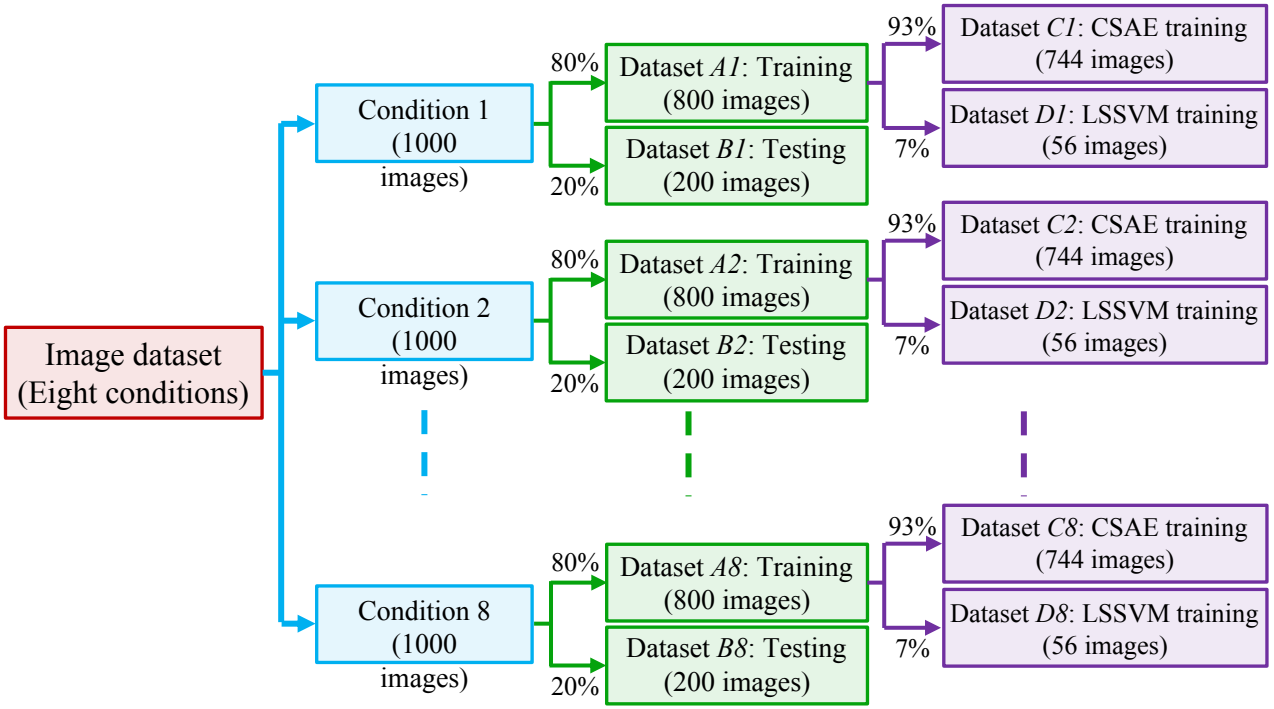
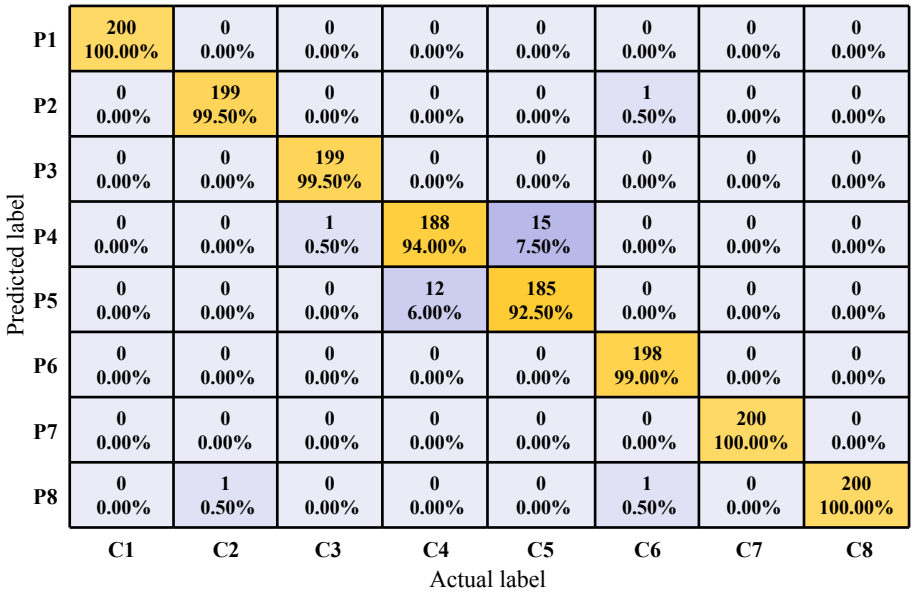


Image dataset
(Eight conditions)
Condition 1
(1000
images)
Condition 2
(1000
images)
Condition 8
(1000
images)
80%
20%
Dataset A1: Training
(800 images)
Dataset B1: Testing
(200 images)
Dataset A2: Training
(800 images)
Dataset B2: Testing
(200 images)
Dataset A8: Training
(800 images)
Dataset B8: Testing
(200 images)
93%
7%
Dataset C1: CSAE training
(744 images)
Dataset D1: LSSVM training
(56 images)
Dataset C2: CSAE training
(744 images)
Dataset D2: LSSVM training
(56 images)
Dataset C8: CSAE training
(744 images)
Dataset D8: LSSVM training
(56 images)

Fig. 9


Predicted label
P1
P2
P3
P4
P5
P6
P7
P8
200 100.00%
0 0.00%
199 99.50%
1 0.50%
188 94.00%
15 7.50%
12 6.00%
185 92.50%
198 99.00%
C1
C2
C3
C4
C5
C6
C7
C8
Actual label