IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON LEARNING SYSTEMS BASED CONTROL AND OPTIMIZATION OF COMPLEX NONLINEAR SYSTEMS

Received January 4, 2018, accepted February 5, 2018, date of publication February 28, 2018, date of current version March 15, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2810266

Learning to Detect Local Overheating of the High-Power Microwave Heating Process With Deep Learning

KAI WANG, LONGKUN MA, QINGYU XIONG, SHAN LIANG, (Member, IEEE), GUOTAN SUN, XING YU, ZHENG YAO, AND TONG LIU

[1]Key Laboratory of Dependable Service Computing in Cyber Physical Society, MOE, Chongqing 400044, China
[2]College of Automation, Chongqing University, Chongqing 400044, China

Corresponding author: Kai Wang (akyle@163.com)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 106112017CDJXY170005, in part by the Chongqing Research Program of Basic Research and Frontier Technology under Grant cstc2017jcyjBX0025, in part by the National Natural Science Foundations of China under Grant 61771077, and in part by the Science and Technology Major Special Project of Guangxi under Grant GKAA17129002.

ABSTRACT As a new kind of heating technology, microwave heating could replace traditional heating methods, because it has the advantages of high efficiency, no secondary pollution, and rapid heating. But the microwave heating process, which involves complex coupling between time-varying electromagnetic field and thermal field, is extremely complicated. At this point, the heated medium may produce local overheating. Worse, it may cause unexpected safety accidents, such as burning and even explosion. However, the temperature variation during the period of microwave heating could barely be obtained. In order to solve the problem of local overheating, this paper proposes a deep learning algorithm based on multi-dimensional data to construct an anomaly detection model for detecting local overheating. The algorithm consists of convolutional neural networks (CNNs) and unsupervised learning method named isolation forest algorithm (IFA). First, CNNs is utilized to extract features of the data collected from a WXD15S microwave heating system. Then, IFA detects the local overheating. Compared with the algorithm with common model, experiment results show that the proposed algorithm owns better measurement performance and higher precision.

INDEX TERMS Microwave heating, local overheating, convolutional neural networks, isolation forest.

I. INTRODUCTION

Microwave heating is essentially based on the characteristics of microwave energy. Microwave with high energy, is capable of penetrating medium directly, which makes the molecules in the medium generate certain physical or chemical reactions [1]. Thus, temperature rises from the inside to the outside of the medium. As a new type of heating technology, microwave heating is widely applied in many fields, such as drying, minerals processing and waste disposal. Compared with traditional heating methods, it has the advantages of high efficiency, no secondary pollution and rapid heating [21], [23].

However, high-power microwave heating is a complex process, because there exist time-varying electromagnetic field and temperature field during the heating. Moreover, these two fields are strongly coupled [4], [22]. Meanwhile, when microwave heating is in progress, with temperature increasing, the dielectric coefficient and the thermal conductivity of medium heated are change significantly, which causes asymmetrical distribution of electromagnetic field. Thus, temperature distribution is uneven. Therefore, local overheating occurs frequently and it leads to unexpected safety accidents, for example, burning and even explosion [2], [3]. This problem has become major drawback for industrial applications of microwave heating.

To solve the problem mentioned above, firstly, power distribution should be obtained, then, temperature distribution could be controlled, finally, local overheating is settled and avoided. Lambert's law is one of common approaches to acquire the power distribution in practical industrial applications, such as material heating and drying process. However, it ignores the influence of the temperature directly related

FIGURE 1. The experiment environment of the WXD15S microwave heating system.

to the dielectric constant, which influences the accuracy of power distribution. So, the control of temperature distribution and local overheating is affected. From the aspect of data, this paper proposes a deep learning algorithm based on multi-dimensional data to construct an anomaly detection model for detecting local overheating. The data is collected from a WXD15S microwave heating system. Compared with the algorithm with common model, experiment results indicate that the proposed algorithm owns better measurement performance and higher precision.

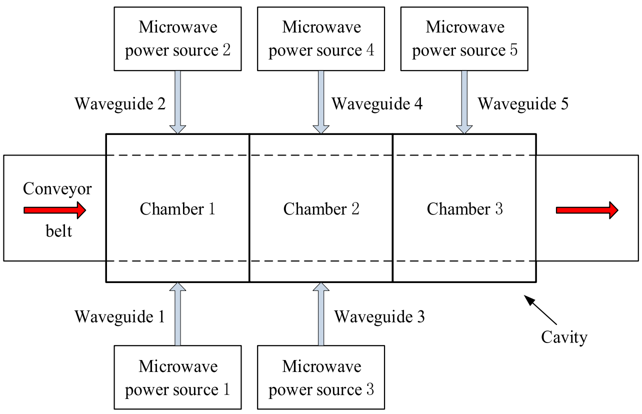


FIGURE 2. The plane diagram of the microwave heating system.

II. MATERIAL

A. EXPERIMENTAL ENVIRONMENT FOR DATA ACQUISITION

The WXD15S microwave heating system (Figure 1) is mainly divided into five microwave power sources (the power of each source is from 0KW to 3KW), detecting module measuring temperature and power with fiber sensors and power meters, waveguides, multimode resonant cavity with conveyor belt and PC. The cavity is made up of three continuous chambers (Figure 2) and each chamber uses an optical fiber temperature sensor which is inserted in the medium for temperature measurement.

The heated medium is water whose permittivity is relatively sensitive to temperature change [14]. The procedure (Figure 3) of the WXD15S microwave heating system is: firstly, when the system is closed, with the conveyor belt,

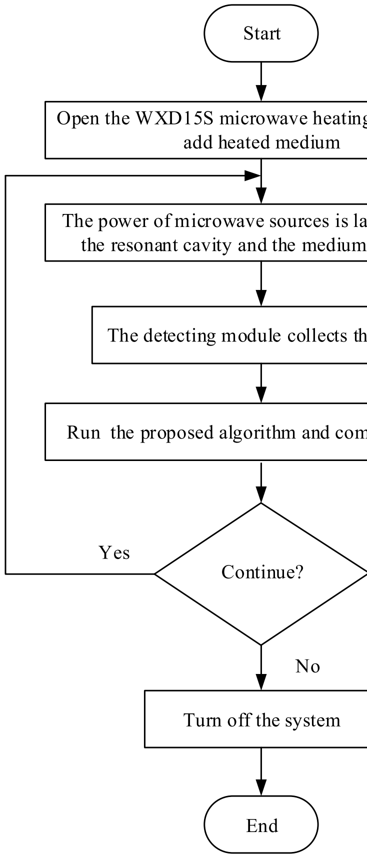


FIGURE 3. The procedure of experiment.

TABLE 1. The information of one set data.

Dimension	Initial temperature	Incident power	Reflected power
Value	20℃	700W	800W
Dimension	Temperature	Heating time	Dielectric constant
Value	30℃	1s	50-15j
Dimension	Thermal conductivity	Density	X-coordinate
Value	0.602W/(m· k)	1000kg/m^3	0.5m
Dimension	Y-coordinate		
Value	0.3m		

the heated medium in a glassware is placed to each chamber. Secondly, the power generated from the microwave power sources is launched into the resonant cavity by waveguides as soon as the system is open. Simultaneously, the medium is being heated. Thirdly, the detecting module collects the data of power, temperature and heating time. Fourthly, with the data, PC runs the proposed algorithm and computes result. Finally, The system and other equipment are turned off.

B. DATA DESCRIPTION

Besides the information of temperature and power, other elements of the multi-dimensional data are acquired. Table 1 represents the information of one set data in detail.

C. DATA PREPROCESSING AND FEATURE EXTRACTION

In order to improve the quality of data mining, more information statistics and feature engineering are applied to process the raw data to enrich the data features.

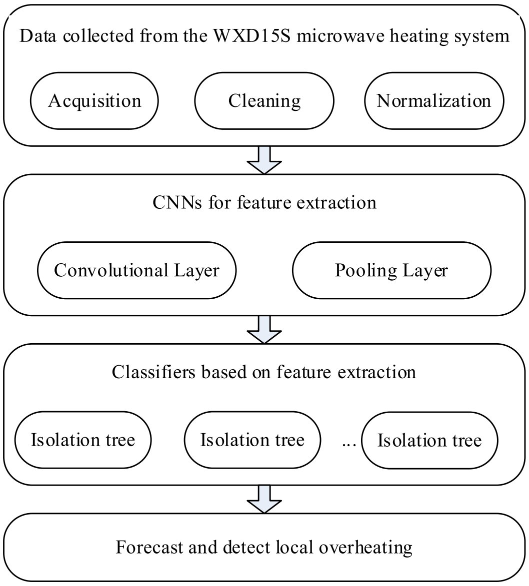


FIGURE 4. Process of data stream analysis.

In the context of detecting local overheating, feature extraction means transforming the processed data into depth features that can be used for the anomaly detection model to evaluate the anomaly message. Otherwise, feature extraction is a kind of effective expression for the original data and it can clearly convey the deep relationship of the data. Therefore, feature extraction in data mining is particularly prominent. This paper focuses on the combination of Auto-encoder and CNNs to achieve feature extraction. Figure 4 shows the process of data stream analysis.

D. SOFTWARE CONDITION

In this article, Anaconda software based on Python is used to build the anomaly detection model. In addition, the learning packages utilized to establish deep learning algorithms are keras, nolearn and scikit-learn.

III. METHODS

A. CNNs FOR EXTRACTING FEATURES FROM MULTIVARIATE TIME SERIES DATA

CNNs, recurrent neural networks (RNNs) and deep belief network (DBN) are deep learning algorithms which are widely applied in the field of voice-analysis and image recognition [6], [7], [10], [24], [25], [30]–[32]. For RNNs, there are many parameters to be trained and the process of training theses parameters are time-consuming. In addition, RNNs is hardly able to complete feature extraction [26], [27]. As for DBN, it is a kind of generation model [28] which means the classification accuracy is relatively low. However, compared with RNNs and DBN, CNNs with weights sharing, has few parameters trained and it takes less time to train parameters. Besides, the classification accuracy of CNNs is higher. More importantly, CNNs can fulfill feature extraction. Therefore, CNNs combing Auto-encoder is considered to extract features from multivariate time series data of microwave heating. To train the CNNs, denoising Auto-encoder, an unsupervised learning method, is applied. In the meantime, in order to capture a distributed representation of its leading factors of variation, denoising Auto-encoder transforms the information of the input signal, but without the linearity assumption of Principal Component Analysis (PCA) [29]. Then. the anomaly detection model could be built. Finally, local overheating is detected. .

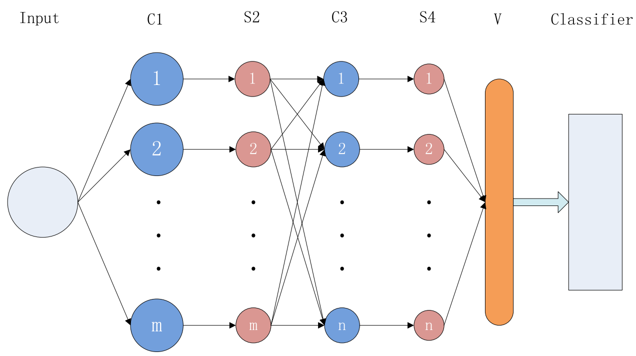


FIGURE 5. The structure of CNNs.

1) THE STRUCTURE OF CNNs

CNNs is a multilayer neural network, and each layer consists of a number of independent neurons [9]. Figure 5 shows the structure of CNNs. CNNs is mainly made up of convolutions with two layers (C1 and C3) and pooling with two layers (S2 and S4) [10]. In the process of feature extraction, firstly, the input data of microwave heating is convolved with m filters and m biases. Thus, m mapping data is generated in the C1 layer after convolution. Then the data is weighted, summed and added with biases through neurons. The processed data is conveyed to a Sigmoid function to obtain the S2 layer. Secondly, the S2 layer is filtered to get the C3 layer. Thirdly, with the same transformation trend, the S4 layer is generated from the C3 layer. Fourthly, the vector connected by the data of S4 layer is imported into the traditional neural network to gain the output. Finally, the output is applied in a error function to adjust weights and biases of the network.

The CNNs convolves the input signal with the kernel function [15], [16], and then from an activation function to obtain an output mapping [11], [17], [33]–[35]. In CNNs, the filters (the size is 2×1) with a stride of 2 down samples along time-axis are designed to process the data from the microwave heating system.

Equation (1) indicates the jth feature of each output in the lth layer x_j^l.

$$X_j^l = f\left(\sum\nolimits_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j\right) \tag{1}$$

where M_j is the jth combination of selected input feature, k^l_{ij} is the convolution kernel in the lth layer used for the connection between the ith input feature and the jth output feature, b_j is thejth bias of the output feature and $f(\cdot)$ is a activation function and $f(x) = \frac{1}{1+e^{-x}}$ or $f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

Equation (2) calculates the sensitivity δ^l_j of the jth feature in lth layer.

$$\delta^l_j = \delta^{l+1}_j W^{l+1}_j o f'\left(u^l\right) = \beta^{l+1}_j up\left(\delta^{l+1}_j\right) o f'\left(u^l\right) \tag{2}$$

Where $up(\cdot)$ represents an up-sampling operation, W^{l+1}_j is the weight of the jth feature in lth layer, $u^l = W^l x^{l-1} + b^l$ and o indicates multiplication.

The partial derivative $\frac{\partial E}{\partial b_j}$ of error cost function E for the bias b is obtained from Equation (3).

$$\begin{cases} E^n = \frac{1}{2}\sum_{k-1}^{c}\left(t^n_k - y^n_k\right)^2 \\ \frac{\partial E}{\partial b_j} = \sum_{u,v}(\delta^l_j)_{u,v} \end{cases} \tag{3}$$

where t^n_k and y^n_k are the label and the output of the kth sample in the nth dimension. c is the sample dimension and (u, v) represents the position of the elements in the matrix.

The partial derivative $\frac{\partial E}{\partial k^l_{ij}}$ of error cost function for the convolution kernel k is expressed in Equation (4).

$$\frac{\partial E}{\partial k^l_{ij}} = \sum_{u,v}(\delta^l_j)_{u,v}(p^{l-1}_i)_{uv} \tag{4}$$

where $(p^{l-1}_i)_{uv}$ is the patch of convolution between x^{l-1}_i and k_{ij}.

In the pooling layer, when feature mapping is generated, the pooling layer converges the continuous mapping feature of the convolution layer [18], [19], [36]–[38]. In this case, the layer could reduce the number of features and the size of the feature space. Thus, the number of parameters in the network could be reduced. Besides, the pooling layer could control the generation of the fitting. The max-pooling and the average-pooling are the two most common pooling ways. For each output feature x_j of the pooled layer, the x_j is obtained as follows:

The jth feature of each output in pooled layer is obtained in Equation (5).

$$x^l_j = f(\beta down\left(x^{l-1}_j\right) + b^i_j) \tag{5}$$

Where down represents down-sampling and β is a scalar parameter

Calculating the sensitivity of each layer:

$$\delta^l_j = f'\left(u^l_j\right) o\ conv2\left(\delta^{l+1}_j, rot180\left(k^{l+1}_j\right)', full'\right) \tag{6}$$

Where $conv2(\cdot)$ is a convolution function and rot represents the rotation of the parameters.

In pooling layer, The partial derivative $\frac{\partial E}{\partial b_j}$ of error cost function E for the bias b is indicated in Equation (7).

$$\frac{\partial E}{\partial b_j} = \sum_{u,v}(\delta^l_j)_{u,v} \tag{7}$$

In general, a simple CNNs is composed of a number of network layers connected. With a differentiable function, each layer of the network generates an activation value as the input of the next layer which pools the output of the previous layer. Thus, a complete framework of feature learning is formed. In order to deal with larger scale time-series data, in this paper, a network for the alternate structure of convolution layer and pool layer is designed.

2) AUTO-ENCODER

The CNNs mentioned in the previous section is based on the supervised training process of tagged data [20]. It is usually necessary to connect a classifier at the end of the CNNs. Through the tagged data, the whole network could use the gradient-based learning method for global training, and the parameters could be corrected layer by layer. In this paper, the data collected from each node during the microwave heating process is unlabeled. Therefore, an Auto-encoder is connected to the final layer of the CNNs to modify the network parameters by the back-propagation algorithm.

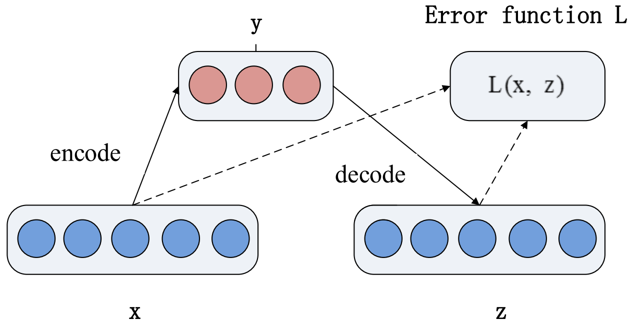


FIGURE 6. **The Auto-encode procedure.**

Auto-encoder is a neural network that tries to replicate the input signal. When an encoder handles the data, the corresponding code, which is a kind of input representation, is generated. To adjust the encoder and decoder parameters, the weights of the network are trained by the method of Back Propagation algorithm. Thus, the information between the code and the original input data is similar. Therefore, the code could express the original input data. Figure 6 shows the Auto-encoder procedure. x is a data set of microwave heating. y is encoded by x and z is decoded by y. L(x, z) is a error function connected with x and z.

y is represented as:

$$y = s(Wx + b) \tag{8}$$

where s is an Sigmoid function, $x = (x1, x2, x3, x4, x5, \ldots)$, W is weight and b is bias.

The hidden layer y reconstructs a signal z of the same shape as x with a decoder. z is shown as:

$$z = s(W'y + b') \tag{9}$$

where W' and b' are weight and bias, respectively.

L(x, z) is indicated as:

$$L(x, z) = \|x - z\|^2 \tag{10}$$

The cost function J (W, b) of Auto-encoder is defined in Equation (11).

$$\begin{cases} J(W, b) = A + B \\ A = \frac{1}{m}\sum_{i=1}^{m} J\left(W, b; x^{(I)}, y^{(i)}\right) \\ B = \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ji}^{(l)}\right)^2 \end{cases} \tag{11}$$

Where A is a mean square error term and B a regularization part (also called weight attenuation term) reducing the magnitude of the weights and prevent over-fitting.

Partial derivatives function is defined as:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b;\ x^{(i)}, y^{(i)})\right] + \lambda W_{ij}^{(l)} \tag{12}$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b) = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial b_i^{(l)}} J(W, b;\ x^{(i)}, y^{(i)}) \tag{13}$$

$W_{ij}^{(l)}$ and $b_i^{(l)}$ are iterated:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) \tag{14}$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b) \tag{15}$$

where α represents the learning rate.

When the iteration is finished, feature extraction could be completed. Thus, IFA is used to detect the local overheating based on feature extraction.

B. IFA FOR DETECTING LOCAL OVERHEATING

With the idea of stochastic forest, IFA, which is applied in the attack detection of network security, traffic anomalies and other analysis, could accomplish anomaly detection and outlier-based mining [12], [13]. Moreover, this algorithm could effectively and quickly process multi-dimensional and massive data. The procedure of IFA is:

Step 1: a data set with N samples and M dimensions (Figure 5) is extracted from the data of microwave heating with CNNs mentioned in Part III. A.

Step 2: ψ training samples are gained by uniformly sampling from N samples.

Step 3: a feature is randomly selected from the ψ training samples. Then, a value is randomly chosen from the feature and the ψ training samples are classified into two classes with binary partition based on the value.

Step 4: each class is continuously classified with Step 3 until that the class in the data is hardly divided or the depth of binary tree reaches $\log_2(\psi)$ [12].

Step 5: a binary tree is built based on ψ training samples.

In experiment, 500 binary trees are set to build according to Step 2- Step 5. Thus, prediction process could be conducted via inserting the test data into each of the binary trees. Then, the test data moves along the corresponding conditional branch until it reaches the leaf node. For each test data, the length h_i ($i = 1, 2, \ldots, 500$) of the ith path test data travels is recorded.

The mean length of the path $E(h_i)$ is defined as:

$$E(h(x)) = \frac{1}{L}\sum_{i=1}^{L} h_i(x) \tag{16}$$

where $L = 1, 2, \ldots 500$.

The abnormal scores S(n) for each test data is expressed as:

$$S(n) = 2^{\frac{E(h_i)}{c(n)}},\ S(n)\,\epsilon[0, 1] \tag{17}$$

$$c(n) = 2H(n-1) - (2(n-1)/n) \tag{18}$$

$$H(\cdot) = \ln(\cdot) + \xi \tag{19}$$

where c(n) is the mean length of binary tree, n is the number of samples and $\xi = 0.5772156649$ is Euler constant.

If S(n) is close to 0, the data is normal. But, if S(n) is close to 1, the data is abnormal, which means local overheating is detected.

IV. EXPERIMENTS AND DISCUSSION

In order to minimize the reconstruction error (RE) between the original data and the features extracted, and adjust the network weights during the training depth learning model, it is necessary to normalize the original data before it is transferred into the depth learning model. Equation (20) shows normalization process.

$$X^* = \frac{X - \min}{\max - \min} \tag{20}$$

where, in a column of data set, x, min and max indicate arbitrary value, minimum value and maximum value. X^* is the normalized result.

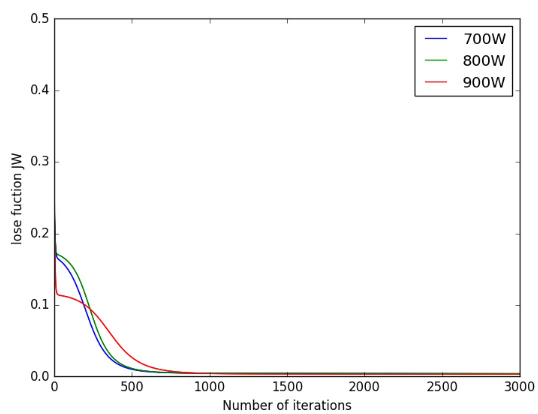


FIGURE 7. The number of iterations of the model.

When the power of five microwave sources is 700W, 800W and 900W, figure 7 represents the results of RE. With the increase of iteration number, RE gradually tends to zero

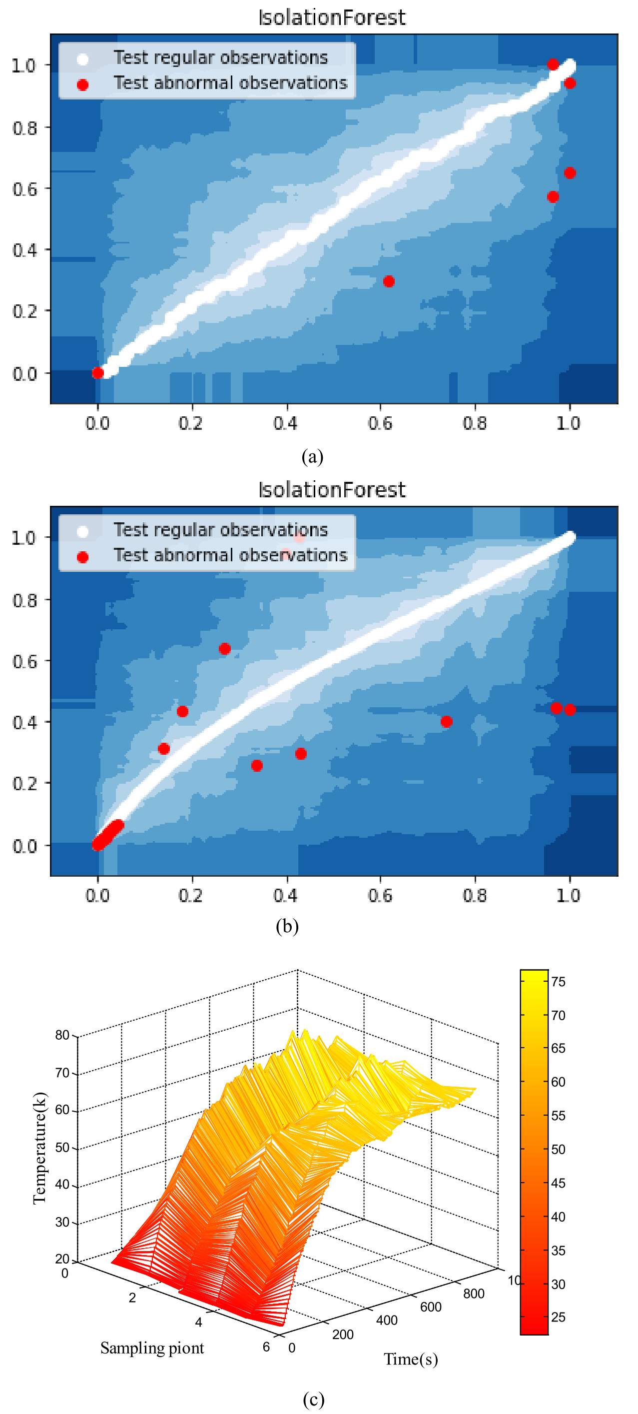


FIGURE 8. **The result of local overheating detection when the power is 700W: (a) Without CNNs; (b) With CNNs; (c) Temperature field distribution of samples.**

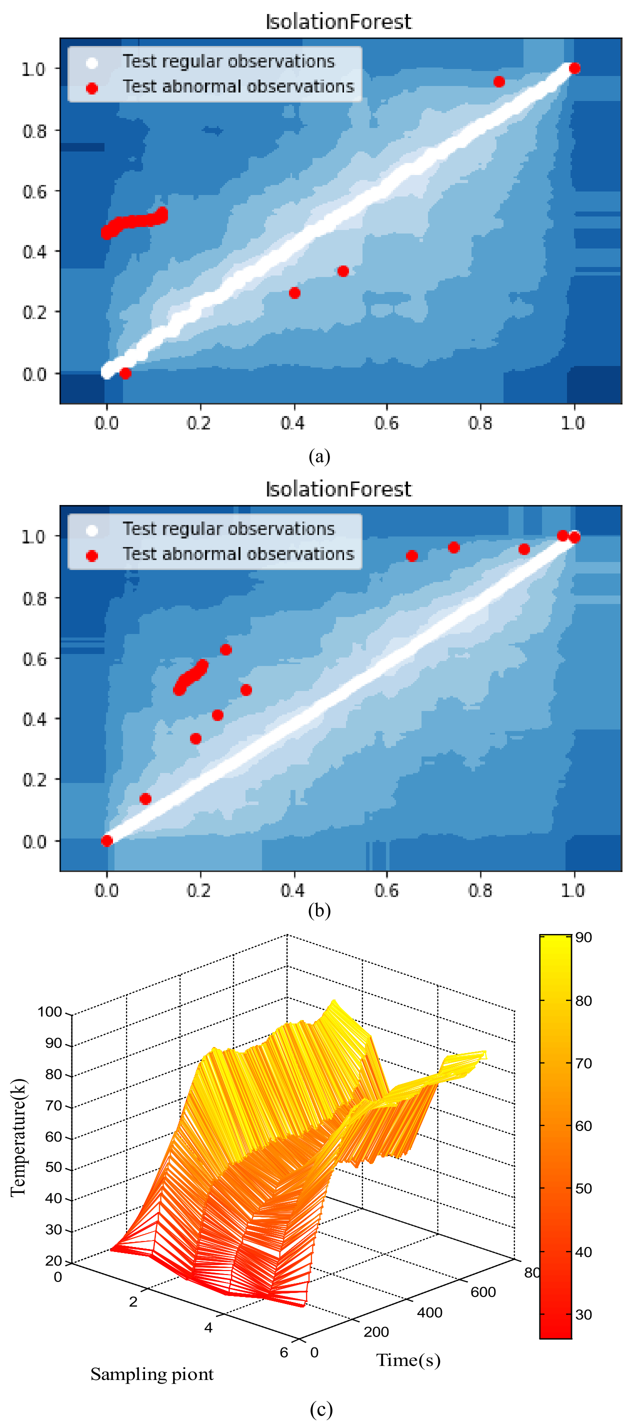


FIGURE 9. **The result of local overheating detection when the power is 800W: (a) Without CNNs; (b) With CNNs; (c) Temperature field distribution of samples.**

which means the extracted features basically express the original data.

When the power is set to 700W, 800W and 900W, three test data sets are gained. Figure 8-10 present the results of local heating detection with CNNs and without CNNs based on the test data sets. The red points are abnormal. Meanwhile, the white points are normal.

Area Under Curve (AUC) is used to estimate the algorithm accuracy. Table 2 indicates the calculation result of AUC.

From table 2, when the power is 700W, without and with CNNs, most of the abnormal points in the testing data set are basically detected and the corresponding AUC values are relatively small and close. As the power is increasing, without

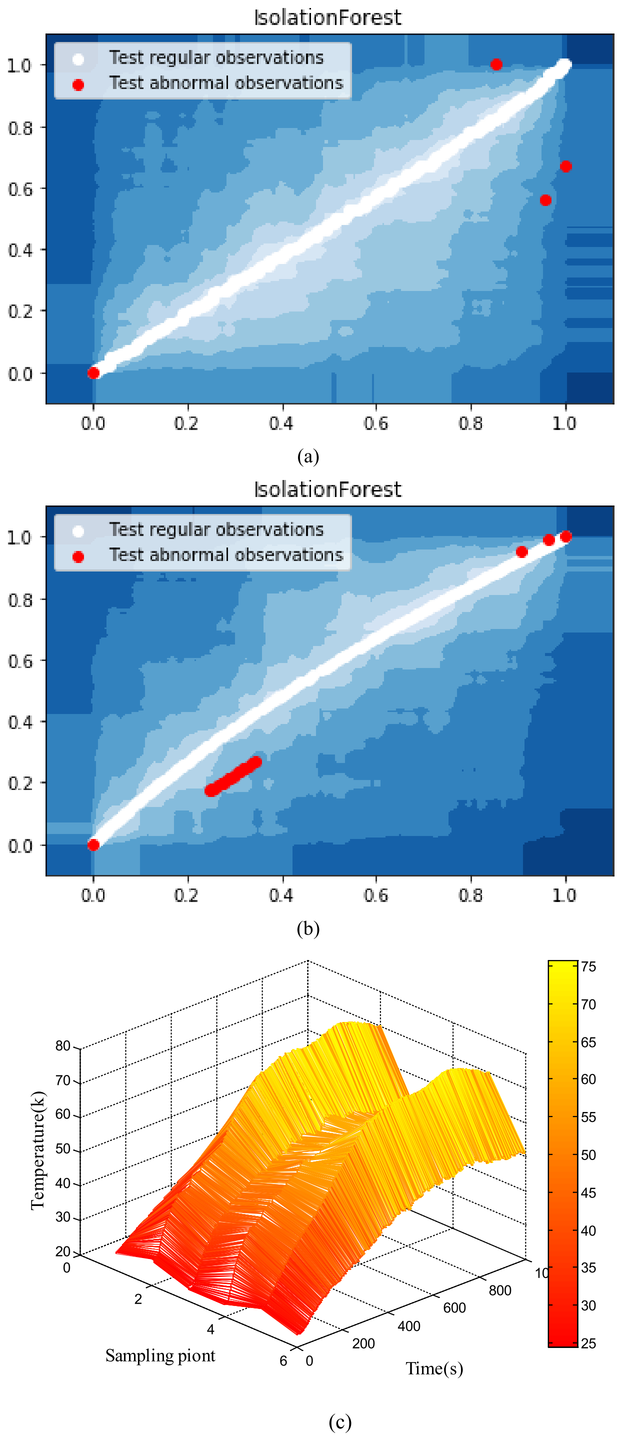


FIGURE 10. **The result of local overheating detection when the power is 900W: (a) Without CNNs; (b) With CNNs; (c) Temperature field distribution of samples.**

CNNs, the AUC value is reducing. However, with CNNs, the AUC value is increasing.

Therefore, experiment results show that the proposed algorithm owns better measurement performance and higher precision, which means the data collected from microwave heating is analyzed and local overheating could be detected.

TABLE 2. **The calculation result of AUC.**

Power of microwave sources	Without CNNs	With CNNs
700W	0.76	0.81
800W	0.73	0.84
900W	0.53	0.78

V. FUTURE RESEARCH

Future research should focus on adjusting the network of CNNs to improve the measurement accuracy and applying the proposed algorithm to assist the construction of control strategy in microwave heating system. More detection equipment, such humidity sensor, should be installed in the system to collect stable and substantial data. In addition, a method for temperature field measurement during microwave heating could be considered, because common temperature measurement methods are inappropriate in the environment of microwave. Thus, according to the results of temperature field measurement and CNNs, the power of microwave sources may be controlled immediately and the problem of local overheating is solved effectively.

REFERENCES

[1] X. Shen, Q. Xiong, W. Shi, K. Wang, and G. Lai, "Temperature distribution monitoring using ultrasonic thermometry based on Markov radial basis function approximation and singular values decomposition," *Math. Problems Eng.*, vol. 2014, Apr. 2014, Art. no. 835619.

[2] Z. Y. Li, R. F. Wang, and T. Kudra, "Uniformity issue in microwave drying," *Drying Technol.*, vol. 29, no. 6, pp. 652–660, 2010.

[3] X. Zhao, K. Huang, L. Yan, and Y. Yao, "A preliminary study on numerical simulation of microwave heating process for chemical reaction and discussion of hotspot and thermal runaway phenomenon," *Sci. China G, Phys., Mech. Astron.*, vol. 52, no. 4, pp. 551–562, 2009.

[4] H. D. Kimrey, Jr., "Microwave heating system with enhanced temperature control," U.S. Patent 9 357 590 B2, May 31, 2016.

[5] B. I. Cirstea and L. Likforman-Sulem, "Improving a deep convolutional neural network architecture for character recognition," *Electron. Imag.*, vol. 17, pp. 1–17, Feb. 2016.

[6] W. Dai, C. Dai, S. Qu, J. Li, and S. Das, "Very deep convolutional neural networks for raw waveforms," in *Proc. ICASSP*, Mar. 2017, pp. 421–425.

[7] D. Eigen, R. Fergus, J. Rolfe, and Y. LeCun, "Understanding deep architectures using a recursive convolutional network," *Comput. Sci.*, Dec. 2013.

[8] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Int. Conf. Artif. Neural Netw.*, 2011, pp. 52–59.

[9] P. Dighe, G. Luyet, A. Asaei, and H. Bourlard, "Exploiting low-dimensional structures to enhance DNN based acoustic modeling in speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 5690–5694.

[10] B. Lee, E. Erdenee, S. Jin, and P. K. Rhee, "Efficient object detection using convolutional neural network-based hierarchical feature modeling," *Signal Image Video Process.*, vol. 10, no. 8, pp. 1503–1510, 2016.

[11] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit number recognition from street view imagery using deep convolutional neural networks," *Comput. Sci.*, Dec. 2013.

[12] W. R. Chen, Y.-H. Yun, M. Wen, H.-M. Lu, Z.-M. Zhang, and Y.-Z. Liang, "Representative subset selection and outlier detection via isolation forest," *Anal. Methods*, vol. 8, no. 39, pp. 7225–7231, 2016.

[13] L. Sun, S. Versteeg, S. Boztas, and A. Rao. (2016). "Detecting anomalous user behavior using an extended isolation forest algorithm: An enterprise case study." [Online]. Available: https://arxiv.org/abs/1609.06676

[14] X. Liao, G. Chen, Q. Liu, J. Chen, and J. Peng, "Investigation on drying characteristics of high titanium slag using microwave heating," *Green Process. Synthesis*, vol. 6, no. 1, pp. 103–112, 2017.

[15] A. Mahendran and A. Vedaldi, "Visualizing deep convolutional neural networks using natural pre-images," *Int. J. Comput. Vis.*, vol. 5, no. 4, pp. 233–255, 2015.
[16] T. N. Sainath *et al.*, "Improvements to deep convolutional neural networks for LVCSR," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Dec. 2013, pp. 315–320.
[17] J.-T. Huang, J. Li, and Y. Gong, "An analysis of convolutional neural networks for speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 4989–4993.
[18] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," *Comput. Sci.*, Jan. 2013.
[19] T. Wiatowski and H. Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1845–1866, Mar. 2018.
[20] T. Salimans. (2016). "A structured variational auto-encoder for learning deep hierarchies of sparse features." [Online]. Available: https://arxiv.org/abs/1602.08734
[21] S. Xin, L. Jiannan, X. Qingyu, and Y. Yupeng, "Research on the uniformity evaluation model in microwave heating," *Chin. J. Sci. Instrum.*, vol. 9, p. 003, Sep. 2014.
[22] J. Li, Q. Xiong, K. Wang, X. Shi, S. Liang, and M. Gao, "Temperature control during microwave heating process by sliding mode neural network," *Drying Technol.*, vol. 34, no. 2, pp. 215–226, 2016
[23] A. Putranto and X. D. Chen, "Microwave drying at various conditions modeled using the reaction engineering approach," *Drying Technol.*, vol. 34, no. 14, pp. 1654–1663, 2016.
[24] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.
[25] C. Du and S. Gao, "Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network," *IEEE Access*, vol. 5, pp. 15750–15761, 2017.
[26] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.*, vol. 3, no. 3, pp. 127–149, 2009.
[27] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[28] R. Salakhutdinov and I. Murray, "On the quantitative analysis of deep belief networks," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 872–879.
[29] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 20–33, May 2013.
[30] L. Wu, Y. Gao, J. Liu, and H. Li, "Event-triggered sliding mode control of stochastic systems via output feedback," *Automatica*, vol. 82, pp. 79–92, Aug. 2017.
[31] Y. Zhao, Y. Shen, A. Bernard, C. Cachard, and H. Liebgott, "Evaluation and comparison of current biopsy needle localization and tracking methods using 3D ultrasound," *Ultrasonics*, vol. 73, pp. 206–220, Jan. 2017.
[32] J. Liu, C. Wu, Z. Wang, and L. Wu, "Reliable filter design for sensor networks using type-2 fuzzy framework," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1742–1752, Aug. 2017.
[33] J. Liu, S. Vazquez, L. Wu, A. Marquez, H. Gao, and L. G. Franquelo, "Extended state observer-based sliding-mode control for three-phase power converters," *IEEE Trans. Ind. Electron.*, vol. 64, no. 1, pp. 22–31, Jan. 2017.
[34] H. Wang, W. Liu, J. Qiu, and P. X. Liu, "Adaptive fuzzy decentralized control for a class of strong interconnected nonlinear systems with unmodeled dynamics," *IEEE Trans. Fuzzy Syst.*, to be published.
[35] H. Wang, P. X. Liu, and P. Shi, "Observer-based fuzzy adaptive output-feedback control of stochastic nonlinear multiple time-delay systems," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2568–2578, Sep. 2017.
[36] H. Wang, P. X. Liu, and S. Liu, "Adaptive neural synchronization control for bilateral teleoperation systems with time delay and backlash-like hysteresis," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3018–3026, Oct. 2017.
[37] D. Xu, Y. Shi, and Z. Ji, "Model free adaptive discrete-time integral sliding mode constrained control for autonomous 4WMV parking systems," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 834–843, Jan. 2018.
[38] D. Xu, J. Liu, X.-G. Yan, and W. Yan, "A novel adaptive neural network constrained control for multi-area interconnected power system with hybrid energy storage," *IEEE Trans. Ind. Electron.*, to be published.

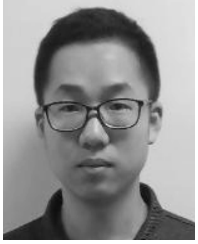
KAI WANG received the B.S. degree in electronics and information engineering from Sichuan University, Chengdu, China, in 2000, and the Ph.D. degree in control theory and control engineering from Chongqing University, Chongqing, China, in 2009. He was a Visiting Scholar with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA, in 2014. He is currently an Associate Professor with the Department of Automation, Chongqing University, Chongqing, China. His current interests include machine learning, deep learning, and the application of big data.

LONGKUN MA is currently pursuing the master's degree with the Department of Control Engineering, University of Chongqing, under the supervision of Prof. K. Wang. His research interests mainly include machine learning and data mining.

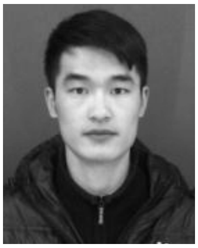
QINGYU XIONG received the B.Sc. and M.Sc. degrees from Chongqing University in 1986 and 1991, respectively, and the Ph.D. degree from Kyushu University, Japan, in 2002. He is currently a Professor and a Supervisor for the Ph.D. student with Chongqing University. His main research interests include intelligent system and intelligent computation, and pervasive computation and embedded system.

SHAN LIANG (M'08) received the M.Sc. degree in control science and engineering from the College of Automation, Chongqing University, Chongqing, China, in 1995, and the Ph.D. degree from the Department of Mechanical Systems Engineering, Kumamoto University, Kumamoto, Japan, in 2004. His current research interests include numerical modeling, electromagnetic theory, nonlinear systems, adaptive control, and sensor networks. He is currently a Member of the IEEE Control Systems and the Society of Instrument and Control Engineers, Japan. He is also a Co-Editor of the *International Journal of Sensing, Computing and Control.*

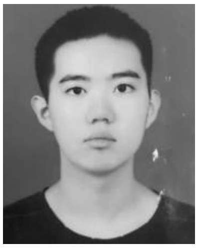
GUOTAN SUN is currently pursuing the master's degree with the Department of Control Engineering, University of Chongqing, under the supervision of Prof. K. Wang. His research interests mainly include machine learning and data mining.



XING YU is currently pursuing the master's degree in control engineering with Chongqing University, under the supervision of Prof. K. Wang whose research interests mainly include machine learning and data mining.



ZHENG YAO is currently pursuing the master's degree in control engineering with Chongqing University under the supervision of Prof. K. Wang. His current research interests include machine learning and data mining.



TONG LIU is currently pursuing the master's degree in control science and engineering with Chongqing University, Chongqing, China, under the supervision of Prof. S. Liang. His current research interests include the application of control theory in microwave heating systems.