Analysis of Gas Leakage Early Warning System Based on Kalman Filter and Optimized BP Neural Network

GUOQUAN LIU1,2, ZHICHAO JIANG2, AND QI WANG2
1Jiangxi Engineering Research Center of Process and Equipment for New Energy, East China Institute of Technology, Nanchang 330013, China
2School of Mechanical and Electronic Engineering, East China University of Technology, Nanchang 330013, China
Corresponding author: Guoquan Liu (gqlecit@hotmail.com)

This work was supported in part by the Education Department of Jiangxi Province under Grant GJJ170484, in part by the Open Project Program of Jiangxi Engineering Research Center of Process and Equipment for New Energy, East China Institute of Technology, under Grant JXNE2020-01, in part by the Start-Up Foundation for Doctors of East China University of Technology under Grant DHBK2012201, in part by the Joint Fund of Science and Technology Department of Liaoning Province, in part by the State Key Laboratory of Robotics, China, under Grant 2020-KF-22-10, and in part by the National Natural Science Foundation of China under Grant 62063001.

ABSTRACT This paper proposes a method for gas leakage early warning system based on Kalman filter and back-propagation (BP) neural network to address the issue of inaccurate gas leakage detection and incapability of predicting concentration change of gas. First, Kalman filter is adopted to filter the noise from the gas concentration that is measured by a sensor. Then, predictions about the change of gas concentration are made using the BP neural network that is optimized by genetic algorithm. Next, the gas leakage early warning system, based on the proposed method, is designed. Last, to verify the effectiveness of the method proposed by simulation, methane, the main component of gas is chosen as an example. Also introduced in this paper are the determinant coefficient, mean absolute error, correlation coefficient and root-mean-square error—the four evaluation indicators methods to demonstrate the effectiveness and feasibility of the algorithm this paper proposed by comparing with Support Vector Machine (SVM), Long Short-term Memory (LSTM) and general Back Propagation Neural Network (BPNN). The best validation performance of BP neural network through simulation experiments and is 0.013518, and the probability of the relative error between the predicted value and the actual value within 10% is 0.7692. The proposed method can effectively improve the accuracy of gas concentration prediction as comparison results show, and it has advantage in fitting degree and error fluctuations.

INDEX TERMS Early warning system, gas leakage, Kalman filter, optimized BP neural network, genetic algorithm.

I. INTRODUCTION
Gas leakage accidents have become frequent occurrences with the popularization of piped gas and the expansion of gas supply systems. The reasons for gas leakage are many, including incorrect use of natural gas equipment by residents, and factors such as the open ad loose value. So, it may be because the apron connected to the gas stove is prone to aging and wear due to long-term contact with water and corrosive substances, or even due to poor quality. Serious accidents may also occur because the gas remaining in the pipe cannot be eliminated in time after each use of natural gas. Therefore, it is immensely significant to adopt a method of detection of the gas concentration in time to effectively prevent gas leakage.

So far, while aiming at more complete and accurate detection of gas leakages, many scholars have carried out research on the problem [1]–[5]. According to an earlier study [6], there are three different categories of leak detection methods, namely biological methods, hardware-based methods, and software-based methods. Among them, the hardware-based method uses external equipment like an infrared thermal imager, and a negative pressure detector to detect gas leakage signals. In [7]–[9], some scholars proposed the use of thermal imaging technology to detect gas leakage, analyze the heat emitted by the gas, and filter the image to accurately detect the gas leakage area in the image. This method can identify when the alarm that the gas concentration exceeds
the preset range, but the detection accuracy level is limited by the equipment, and the accuracy level is also not high. Therefore, research on software-based detection methods is now a high demand. To locate leakage events, pressure wave analysis is used in [10], and to reduce the false alarm rate, a multi-sensor pairing algorithm used. Sound wave detection technology is mainly used to detect pipeline gas leakage. First, to obtain sound signals, the sound field is established based on aeroacoustics, and simultaneously, to achieve the purpose of detecting leak signals, some methods such as wavelet transform are used to denoise the signal [11]–[13]. Results show that this method can not only effectively improve the detection ability of small leaks but also, at the same time, can reduce the false alarm rate; but this method is susceptible to external factors that reduce the accuracy of detection, such as temperature. The recognition accuracy of non-linear and unstable signals generated by gas leakage is not high, although the above method has the ability to identify gas leakage events.

The advancement of science and technology, especially with the emergence of machine learning and deep learning, has spurred the gas leak detection technology. Among them, the SVM model—the most widely used—mainly judges whether it is leaked, and its performance is mainly determined by kernel parameters and penalty factors; and the characteristic of this model is to solve the problem in a small sample size, but it is detected by SVM. The accuracy of the leakage, though, is not high [14]. In subsequent studies, therefore, fuzzy support vector machines have emerged. This method can improve the anti-noise ability of the support vector machine by assigning fuzzy membership values to each sample [15], as well as the detection accuracy. However, as this model is prone to local optimization problems, in such cases, particle swarm optimization (PSO) and other optimization algorithms are generally used to optimize the selection of SVM parameters and to reduce these problems. It is worth mentioning that SVM obviously has a better ability to deal with nonlinear problems when compared with the previous physical model-based detection techniques. However, the algorithms under deep learning framework such as neural BPNN and LSTM have better prediction and classification performance than SVM for the problem of predicting gas leakage [16]. Among them, BPNN is the most used prediction technique. BPNN and principal component analysis (PCA) is used to test acetone and ethanol [17]. BPNN was used to predict the chlorophyll content of rice leaves with the prediction error of 2.34 [18]. The prediction performance of BPNN was found very good, indicated by the above results. However, as a shallow structure algorithm, BPNN is limited in its ability to handle huge samples. In 1997, LSTM as a recurrent neural network was proposed by Sepp Hochreiter and Jürgen Schmidhuber. With better predictive performance, LSTM can solve the serious exploding/vanishing gradient problems [19]. An LSTM-based classifier was proposed in [20], which can handle the imbalanced and time-dependent data of the neural network, thereby improving the prediction accuracy.

This paper, based on the deep learning framework proposes to use the neural network (BP) based on Kalman filter (KF), and the optimization method using genetic algorithm (GEN) to achieve the prediction of leakage as the above research shows that the performance of the algorithm under the deep learning framework is better than other prediction algorithms. The performances of the four algorithms of KF-BP-GEN, SVM, BPNN and LSTM have been compared by determining the four parameters of coefficient, average absolute error, correlation coefficient and root mean square error.

This paper makes the following main contributions to the topic:

1. It presents the design of a gas leakage early warning system, that is based on STMicroelectronics 32-bit Series Microcontroller Chip (STM32). An alarm module and a gas meter are included in the system. The alarm module, combination with the gas meter control module and the exhaust module, can realize the monitoring of the gas leakage. Without any supervision, the gas meter can be turned off, and the exhaust fan can be turned on automatically in the place. By changing the detection sensor, the system can be used not only to detect the change of gas concentration; but also to detect the concentration of other general gases.

2. It uses the genetic algorithm for the error of BP neural network algorithm and updates its weights and thresholds to rectify and improve their iterative speed. BP neural network algorithm forecasts the change that there may be in the concentration of the gas in the air.

3. The algorithm proposed in this paper is compared with the support vector machine, LSTM network algorithm, and general feedforward neural network algorithm by introducing four evaluation coefficients of determination coefficient that are, average absolute error, correlation coefficient, and root mean square error. The comparison proves that the algorithm proposed in this paper has better performance in fitness and error fluctuation degree.

The rest of the structure this paper’s arrangement is as follows: the prediction technique of Kalman filtering algorithm and BP neural network has been explained in Section 2; Section 3 describes the design of gas leakage early warning system; the simulation and comparison to verify our method have been given in Section 4, and the conclusions and future works are given in Section 5.

II. PREDICTION TECHNIQUE

A. THE BASIC PRINCIPLE AND ALGORITHM FLOW OF BP NEURAL NETWORK

The thinking activities of neurons in the human brain are stimulated by an artificial neural network that is an algorithm. By learning existing knowledge and drawing certain judgment rules from it, an artificial neural network can identify, judge, and predict unknown things. The BP neural network [21], as one of the artificial neural networks, is a multi-layer forward neural network that can be composed of an input layer, a plurality of hidden layers, and an output layer.
While a plurality of neurons can be arranged in each layer, the neurons between layers adopt a full connection mode. One neuron has two states of excitation and inhibition in biological neurons. The normally inhibited neuron will superimpose the signals on the neuron after receiving various signals from the synapses of other neurons. The neuron will enter an excited state and transmit the signals to the next neuron when the superimposed intensity exceeds a certain threshold. The signals transmitted through neurons include excitation signals and excitation suppression signals, and these can control, as well as inhibit the excitation of neurons.

Figure 1 shows a typical structure of a BP neural network. $X^n_i$ is the input matrix; $I_n$ is the input layer; $H_l$ is the hidden layer; $\omega_{1ji}$ is the weight matrix of the input layer and hidden layer; $\omega_{2ij}$ is the weight matrix of the hidden layer and output layer; and $Y^m_j$ is the output of the output layer.

The characteristic of a biological neuron is stimulated in the superimposed intensity and outputs the added input signals in a set by the threshold value. Through a certain function expression, the transfer function outputs the partial derivative of error $E$ of the two through the learning of the input and output data.

The partial derivative of error $E$ to the weight from the input layer to the hidden layer is

$$
\frac{\partial E}{\partial \omega_{1ji}} = \sum_{k=1}^{n} \left( \delta_j^2 \omega_{2kj} \right) \cdot g'(net_k) x_i
$$

Therefore, the weight adjustment formula is

$$
\begin{align*}
\omega_{1ji}^{t+1} &= \omega_{1ji}^{t} + \Delta \omega_{1ji}^{t} \\
&= \omega_{1ji}^{t} - \eta_1 \frac{\partial E}{\partial \omega_{1ji}} \\
&= \omega_{1ji}^{t} - \eta_1 \delta_j^1 x_i,
\end{align*}
$$

The output of neurons in the output layer is

$$
net_k = z_k = g(\sum_{j=1}^{l} \omega_{kj}^2 O_j - \theta_k^1),
$$

where $z_k$ is the $k$-th neural of the output layer, $\omega_{kj}^2$ is the weight assigned from the hidden layer to the output layer; $\theta_k^1$ is the threshold from the hidden layer to the output layer; $O_j$ is the output of the hidden layer, that is, the input of the output layer; $k$ is the number of layers of the neural network where the current neuron is located; $j$ is the location of the neuron in the current layer and $g(\cdot)$ is the transfer function from hidden layer to the output layer.

The BP neural network can be trained by a series of input quantities and those can be determined under the current input quantities. The BP neural network can compare the predicted value with the theoretical output value and calculate the error of the two through the learning of the input and output data. If the error does not meet the set requirements, it is propagated to the input layer, by layer in the reverse direction through the output layer. Meanwhile the weight matrix and threshold matrix between layers are continuously updated until the errors of the predicted and the theoretical outputs meet the set requirements. BP neural network adopts Delta learning rule, that is, the gradient descent method to find the increment of the updated weight matrix.

The error between the predicted and the theoretical outputs is

$$
E = \frac{1}{2} \sum_{k=1}^{n} (y_k - z_k)^2.
$$

The partial derivative of error $E$ to the weight from the hidden layer to the output layer is

$$
\frac{\partial E}{\partial \omega_{2ij}} = \sum_{k=1}^{l} \left( \delta_j^2 \omega_{kj} \right) g'(net_k) O_i
$$

The partial derivative of error $E$ to the weight from the input layer to the hidden layer is

$$
\frac{\partial E}{\partial \omega_{1ji}} = \sum_{k=1}^{l} \left( \delta_j^1 \omega_{kj} \right) g'(net_k) x_i
$$

Therefore, the weight adjustment formula is

$$
\begin{align*}
\omega_{2ij}^{t+1} &= \omega_{2ij}^{t} + \Delta \omega_{2ij}^{t} \\
&= \omega_{2ij}^{t} - \eta_2 \frac{\partial E}{\partial \omega_{2ij}} \\
&= \omega_{2ij}^{t} - \eta_2 \delta_j^2 O_i,
\end{align*}
$$

where $\omega_{2ij}^{t}$ is the weight from the hidden layer to the output layer.
where η1 and η2 respect the learning step from the input layer to the output layer, as well as from the hidden layer to the output layer.

Similarly, the threshold adjustment formula is obtained by using the partial derivative of the error to the threshold as follows:

\[
\begin{aligned}
\theta_j^1(t + 1) &= \theta_j^1(t) + \Delta \theta_j^1 \\
\theta_j^1(t) &= \theta_j^1(t) - \eta_1 \frac{\partial E}{\partial \theta_j^1} \\
\theta_j^2(t + 1) &= \theta_j^2(t) + \Delta \theta_j^2 \\
\theta_j^2(t) &= \theta_j^2(t) - \eta_2 \frac{\partial E}{\partial \theta_j^2}.
\end{aligned}
\]  

(8)

However, the general BP neural network will have a large number of iterations that will affect the training efficiency of the network, if the initial error is set improperly. We introduce a genetic algorithm to optimize the BP neural network to improve the training efficiency of the network. The minimum value of the error can be obtained by selecting the fitness function and calculating the same in each generation. In the end, calculating the best weights and thresholds provides feedback to BP neural networks.

**B. BP NEURAL NETWORK MODEL BASED ON Kalman FILTER**

Kalman filter [22] is an estimation theory in statistics that was proposed by the Hungarian mathematician Kalman. It is a recursive estimation algorithm with a minimum mean square error as an estimation criterion. Mainly using the state space model of signal and noise, it updates the estimation of state variables by using the estimation value at the previous time and the observation value at the current time to obtain the estimation value at the current time. Thereafter, the current state and the error between the predicted value and the real value are predicted according to the prediction equation.

The calculation step of the Kalman filter calculates the next prediction value at the observation time from the state equation according to the state estimation at the previous time. The optimal estimation is obtained when, according to the real-time observation value and the prior information at the current time, the correction value of the predicted value is calculated.

The data of training BP neural network has fluctuation in the form of Gaussian white noise interference. Inaccurate prediction results will be obtained if noise affects the stability of the data and makes the data abnormal. The Kalman filtering method filters the abnormal fluctuation in the original data to make it smooth. Then the Kalman filtering value is used as input to build a BP neural network model in the same as part A of Section 2 to predict the concentration of gas.

**C. SYSTEM MODEL**

This paper proposes the KF-BP-GEN algorithm model, based on the framework of deep learning. The model building process is given below:

*Step 1 (Build the General Model):* The change of gas concentration is predicted by building the leakage warning system model based on BP network model. The gas concentration of each of the four samples is in the same line as the input matrix; and these four samples can predict the concentration of the next sample. So, the model can be built as

\[
\tilde{x}_{t+4} = Z \\
= g\left(\sum_{j=1}^{l} \omega_{kj}^1 \left(\sum_{i=1}^{m} \omega_{ji}^0 \begin{bmatrix} x_i \\ x_i + 1 \\ x_i + 2 \\
\end{bmatrix} - \hat{x}_i^1\right) - \hat{\theta}_k^1\right) 
\]  

(9)

where Z is the output value of the output layer, \(x_i\) is the gas concentration input detected by sensor, and \(\tilde{x}_{t+4}\) is the prediction value though \(x_i, x_i+1, x_i+2\), and \(x_i+3\). Through this model, gas concentration \(x_i, x_i+1, x_i+2\), and \(x_i+3\) can predict gas concentration.

And the average relative error \(\sigma_i\) is

\[
\sigma_i = \frac{1}{m} \sum_{i=1}^{m} \frac{x_i - \tilde{x}_i}{x_i}.
\]  

(10)

*Step 2: Build the optimized system model based on Kalman filter and BP network*

The accuracy of the sensor’s measurement of concentration is affected by the Gaussian white noise. There are process and observation noises in the sensor’s measurement process. The value of the concentration of gas in the air, as detected by the sensor alone, is not the true value of the concentration at the current moment. Therefore, to effectively reduce the interference of noise and make the concentration of gas in the air detected by the gas sensor closer to the true value, Kalman filtering is adopted.

Gaussian distribution is satisfied, as an assumption, by both process noise \(w_t\) and observation noise \(v_t\).

That is,

\[
\begin{aligned}
w_t &\sim N(0, Q_t) \\
v_t &\sim N(0, R_t)
\end{aligned}
\]  

(11)

The system model is given by

\[
x_t = F_t x_{t-1} + B_t u_t + w_t,
\]  

(12)

where \(x_t\) is the true value of gas concentration in air at time \(t\); \(x_{t-1}\) is the true value of gas concentration in air at time \(t-1\); \(F_t\) is the state transition matrix; \(B_t\) is the input matrix, and \(u_t\) is the input value of the system at time \(t\). The system input value is set to 0, that is, \(u_t = 0\), in this detection process, so there is

\[
x_t = F_t x_{t-1} + w_t.
\]  

(13)

The following equality measurement model is given by

\[
z_t = H_t x_t + v_t,
\]  

(14)
where \( z_t \) is the sensor measurement value of gas concentration in the air at time \( t \) and \( H_t \) is the observation matrix.

The state variable prediction equation is

\[
\hat{x}_{t|t-1} = F_t \hat{x}_{t-1|t-1},
\]

(15)

where \( \hat{x}_{t|t-1} \) is the predicted value of concentration at time \( t-1 \), and \( \hat{x}_{t-1|t-1} \) is the updated value of concentration at time \( t-1 \), that is the real value.

Covariance equation between the predicted value and real value

\[
P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t,
\]

(16)

where \( P_{t|t-1} \) is the predicted value of the covariance of the error between the predicted value of the Kalman concentration at time \( t \) and the real value of the Kalman concentration at time \( t-1 \) is the covariance of the error between the updated predicted value of the Kalman concentration at time \( t-1 \) and the real value. \( P_{t-1|t-1} \) reflects the degree of correlation between the predicted value and the real values.

Considering the predicted value at time \( t-1 \) and the difference between the measured and predicted values at time \( t-1 \), the predicted value at time \( t \) can be derived. Then, according to the Kalman gain and the observation matrix, the correlation coefficient between the predicted value and the true value at time \( t \) is updated. In summary, the predicted value of Kalman concentration at time \( t \) is as follows

\[
\begin{align*}
\{ \hat{x}_{t|t} &= \hat{x}_{t|t-1} + K_t (z_t - H_t \hat{x}_{t|t-1}) \\
P_{t|t} &= P_{t|t-1} - K_t H_t P_{t|t-1},
\end{align*}
\]

(17)

where \( K_t \) is the Kalman gain and

\[
K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t).
\]

(18)

So, taking \( \hat{x}_{t|t} \) as the input value, the model can be built as

\[
\hat{x}_{t+4|i+4} = g(\sum_{j=1}^{4} \alpha_{k,j} f(\sum_{i=1}^{m} \alpha_{k,j} \hat{x}_{i+1|i+1} + \hat{x}_{i+2|i+2} + \hat{x}_{i+3|i+3} - \theta_{k} - \theta_{i}))
\]

(19)

where \( \hat{x}_{t|i} \) is the concentration value by the Kalman filter and \( \hat{x}_{t+4|i+4} \) is the prediction value of gas concentration by optimized BP network.

III. DESIGN OF EARLY WARNING SYSTEM

A. THE OVERALL DESIGN OF THE SYSTEM

To realize the prediction of gas concentration and achieve accurate warning effect, the warning system adopts a combination of Kalman filtering and BP neural network algorithm. A certain probability of falling into a local minimum is the disadvantage of the BP neural network algorithm. We pass the original sample through the Kalman filtering and then put it into the BP neural network for offline training to get the weight value and threshold, in order to improve the generalization ability of the BP neural network. We use the trained neural network to predict the concentration value, and then judge the prediction result in practical applications. We take a series of control operations if the prediction result is greater than the set threshold.

The gas warning system we designed comprises a measurement and a control terminal. A gas concentration detection sensor, a wireless communication module and a microcontroller form the measurement terminal; it mainly realizes the collection of gas concentration data and the prediction of the concentration value. The data collected through Kalman filtering, uploaded into a trained neural network to predict the gas concentration. In the end, the microcontroller compares the predicted value with a preset threshold. The microcontroller controls the wireless communication module at the measurement end to send an early warning signal to the wireless communication module at the control terminal when the predicted value exceeds the threshold.

The control terminal, composed of a wireless communication module, a JK trigger, and a microcontroller, realizes the control of the on-off of the gas meter and the exhaust fan according to the early warning signal of the measurement end, in order to eliminate the risk of gas leakage. The microcontroller at the control terminal will output a pulse width modulation (PWM) pulse wave to the JK trigger to control the state of the JK trigger to close the gas meter and turn on the exhaust fan, when the control end receives the early warning signal sent by the measuring spot.

The **schematic diagram of the overall design** is given in Figure 2.

![Figure 2. Schematic diagram of the general principle of the gas system.](image)
process, the actual gas concentration value detection based on Kalman filter is completed.

The value filtered by the Kalman filter is taken as the training set to train the BP neural network. The genetic algorithm is used to optimize the training error of the BP neural network in the training process, to obtain the optimal weight and threshold value under the optimal error. At last, a complete BP neural network is built through the obtained optimal weight and threshold value. Figure 3 shows the algorithm flow chart.

According to Figure 3, the algorithm process is as follows.

**Step 1.** Establishing a system model, measurement model, the state variable prediction, and covariance equation between the predicted value and actual value equation, getting Eqs. (10)-(14).

**Step 2.** Calculating the Kalman update value of the current concentration through the update equation.

**Step 3.** Inputting the population size \( P \) and number of iterations \( I_{\text{max}} \).

**Step 4.** Initializing the connection weight between each input layer and the hidden layer and the connection weight between the hidden layer and the output layer.

**Step 5.** Determining the fitness function

\[
E = \frac{1}{2} \sum_{k=1}^{n} (y_k - z_k)^2 ,
\]

and calculating the value of each individual evaluation function and sorting them.

**Step 6.** Generating new individuals \( G'_i \) and \( G'_{i+1} \) by permutation and operating on individual \( G_i \) and \( G_{i+1} \) with probability \( P_m \).

**Step 7.** Mutating operations, using the probability \( P_m \) mutation to generate a new individual \( G'_i \).

**Step 8.** Inserting new individuals into population \( P \) and calculating their fitness function values.

**Step 9.** Judging whether the generation is under the max generation. If no, revert to Step 5. Otherwise getting the optimal solution as initial weight of BP neural network.

**Step 10.** Initializing weight of BP neural.

**Step 11.** Weighing input based on Eq. (2).

**Step 12.** Calculating errors and updating thresholds based on Eqs. (4)-(8).

**Step 13.** Judging whether the error meet the requirements. If yes, then for the output error, calculating the fitness function and then reverting to Step 4. Otherwise reverting to Step 11.

C. THE DESIGN OF SYSTEM IMPLEMENTATION CIRCUIT

The system design of the system implementation circuit includes a measuring spot and a control terminal. Their system circuit diagrams are shown in Figures 4 and 5, respectively. The measurement terminal consists of a 433M wireless communication module, a gas sensor, a microcontroller, and an LED display screen. The control terminal is made up of a 433M wireless communication module, a dual JK trigger, a gas meter module, and an exhaust fan module.

The gas meter control module is used to control the path of its power source, in turn, controls the value of the gas path, and the exhaust module is designed to control the working state of the exhaust fan.

The 433MHz wireless transceiver module in the gas meter control module adopts which adopts the SI4463 wireless transceiver chip introduced by Silicon Labs. The single-chip is internally integrated with a radio frequency transceiver. Therefore, the 433MHz wireless transceiver module has the characteristics of bidirectional communication with ultra-low power consumption. In the system, the wireless communication module sends an early warning.
signal when the gas concentration value exceeds a preset concentration threshold.

By controlling the state of the JK trigger, the wireless communication module at the control end receives the warning signal from the measurement end. After that signal, the microcontroller which is in the control end inputs PWM pulses to the CLK end of the JK trigger, in order to control the gas meter module and the exhaust fan module. The system can be reset by pressing the button connected to the clear end of the double JK flip-flop.

The microcontroller outputs a PWM wave to the JK trigger after receiving the early warning signal. Subsequently, the J terminal of the JK trigger is connected to a high level, the K terminal is connected to the ground, and the output of the JK trigger is connected to a switch configured NPN transistor, respectively. The Q terminal of the JK flip-flop will continuously output a low level when the CLK terminal of the JK flip-flop is activated by the PWM pulse, and the NPN tube changes from the previous on-state to the off-state. The collector-emitter is equivalent to an open circuit. As there are no current passes, the excitation current on the electromagnetic coil in the gas meter is zero. Then the magnetic piston is released and closed under the action of the spring, closing the gas passage. According to the description given above, the circuit diagram of the gas meter control module is shown in Figure 6.

The principle of the exhaust fan module resembles that of the gas meter module. The closing and opening of the exhaust fan are controlled by the output state of the JK trigger. The exhaust fan is connected to a relay, and the state is normally open only when the JK trigger is receiving. After the PWM pulse wave, the relay and the exhaust fan are turned on. In order to better control the relay, this system adds an operational amplifier to amplify the control signal. The circuit diagram is as shown in Figure 7.

IV. SIMULATION AND RESULTS

A. METHANE CONCENTRATION EXPERIMENTAL COLLECTION

Figure 8 shows the plan diagram of an indoor natural ventilation environment. The door is closed, all the windows in a wall are open, and the leakage area is in the center of the natural wind field. Other objects in the environment are lower than the bottom edge of the windows. There is a methane concentration detector which is three meters away from the experimental area. The data is shown in Table 1.

| Number of samples | Concentration/% |
|-------------------|-----------------|
| 1                 | 0.06            |
| 2                 | 0.12            |
| 3                 | 0.13            |
| 4                 | 0.17            |
| 5                 | 0.39            |
| 6                 | 0.81            |
| 7                 | 2.20            |
| 8                 | 3.33            |
| 9                 | 4.85            |
| 10                | 7.03            |
| 11                | 8.98            |
| 12                | 11.90           |
| 13                | 13.88           |
| 14                | 16.34           |
| 15                | 18.76           |
| 16                | 20.21           |
| 17                | 26.73           |

In order to determine the location of the leak in space, 25 sets of beacon points are set within the experimental area. The DV-hop algorithm, which does not need to measure the distance is used to determine the location of the leak in the area. There are two main problems with the traditional DV-hop algorithm. One is that the error of the average distance per hop will increase as the number of node hops increases, and the other is that the unknown node only estimates its position relationship with the beacon point. By using the average distance per hop at the nearest beacon point, the average distance per hop is inaccurate.

Therefore, we introduce the genetic algorithm to optimize the selection of the average distance per hop. The adaptive function of the genetic algorithm can be derived by the reciprocal of the difference between actual distance and estimated distance. Finally, the average jump distance per hop with the smallest error in each generation population can be selected.

The adaptation function is as follows:

$$\frac{1}{\varepsilon} = \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 - AveD_i \cdot n}}$$
Finally, the location error of the leak can be used to analyze the accuracy of leak location.

The location error of the leak is

\[
\text{error} = \frac{\sqrt{(x - x_e)^2 + (y - y_e)^2}}{R} \times 100\%,
\]

where \((x, y)\) is the actual location of the leak; \((x_e, y_e)\) is the estimated location of the leak, and \(R\) is the communication radius.

**B. ALGORITHM SIMULATION**

According to the process established by the above model, two random variables \(w_t\) and \(v_t\) are generated using MATLAB. The random variables \(w_t\) and \(v_t\) contain 200 values and satisfy the unit Gaussian distribution with a mean value of 0, where \(w_t\) is the process noise and \(v_t\) is the observation noise. Set the true concentration value at time \(t = 0\) to 0, while the state transition matrix \(F_t\) takes the unit matrix. All the elements in the observation matrix \(H_t\) take 0.2.

As shown in Figure 9, and given to the process above, recursion is carried out to obtain the updated Kalman filter value, measured value and real value corresponding to 200 nodes.

In Figure 9, the red line represents the Kalman filter value, the green line represents the sensor measurement value, and...
the blue line represents the true concentration value. It is easy to detect a certain error between the measured value of the sensor and the real value, while the Kalman filter value tracks the real value well.

Then the true value of the gas concentration in air is the value created by the Kalman filter and will be put into the BP neural network to train.

In case of a leakage in a domestic gas pipeline, the methane concentration sensor is used to continuously collect the analog quantity of methane concentration in the 17 times equivalent (the volume proportion of methane in the air per unit volume) at an interval of 1s each. After the analog quantity of concentration is converted into digital quantity by A/D conversion inside the methane concentration sensor, the digital quantity is input into the STM32 chip through the serial port that is configured to the chip, and the binary quantity of methane concentration is converted into a decimal number in through software programming and then stored in an array. The methane concentration data, post the Kalman filtering processing on the collected data, are shown in Table 2.

The methane concentration data, post the Kalman filtering through software programming and then stored in an array.

TABLE 2. Data of measured methane concentration after kalman filter.

| Number of samples | Concentration/% |
|-------------------|-----------------|
| 1                 | 0.01            |
| 2                 | 0.14            |
| 3                 | 0.16            |
| 4                 | 0.20            |
| 5                 | 0.53            |
| 6                 | 0.90            |
| 7                 | 2.30            |
| 8                 | 3.47            |
| 9                 | 5.00            |
| 10                | 7.21            |
| 11                | 9.00            |
| 12                | 12.40           |
| 13                | 14.00           |
| 14                | 16.60           |
| 15                | 18.91           |
| 16                | 20.03           |
| 17                | 21.00           |

Subsequently, a $13 \times 4$ input matrix is constructed with the concentration values of all four consecutive times in Table 1 as the same row of the input matrix, while a $13 \times 1$ output matrix is constructed with the methane concentration levels corresponding to the next time of every four adjacent times.

The transpose $I$ of the input matrix is

$$
I = \begin{pmatrix}
0.12 & 0.14 & 0.16 & 0.2 \\
0.14 & 0.16 & 0.2 & 0.5 \\
0.16 & 0.2 & 0.5 & 0.9 \\
0.2 & 0.5 & 0.9 & 2.3 \\
0.5 & 0.9 & 2.3 & 3.4 \\
0.9 & 2.3 & 3.4 & 5 \\
2.3 & 3.4 & 5 & 7.2 \\
3.4 & 5 & 7.2 & 9 \\
5 & 7.2 & 9 & 12.4 \\
7.2 & 9 & 12.4 & 14 \\
9 & 12.4 & 14 & 16.6 \\
12.4 & 14 & 16.6 & 18.9 \\
14 & 16.6 & 18.9 & 20 \\
\end{pmatrix}^T
$$

Output matrix $T$ is

$$
T = \begin{bmatrix}
0.5 & 0.9 & 2.3 & 3.4 & 5 & 7.2 & 9 & 12.4 & 14 & 16.6 & 18.9 & 20 & 21 \\
\end{bmatrix}
$$

Let

$$
y = \frac{(y_{\text{max}} - y_{\text{min}})(x - x_{\text{min}})}{x_{\text{max}} - x_{\text{min}}} + y_{\text{min}}, \quad (25)
$$

where $y$ is the value after normalization; $x_{\text{max}}$ is the maximum value of elements in the same row of matrix $I$; and $x_{\text{min}}$ is the minimum value of elements in the same row of matrix $I$. The elements of each column in the matrix input between 0 and 1 are normalized by Eq. (25) to obtain an input normalization matrix $P$.

While normalizing the elements in the matrix output between zero and one using Eq. (25), the output normalized matrix $n$ is obtained as follows:

$$
n = \begin{bmatrix}
0.0195 & 0.0878 & 0.1415 & 0.2195 & 0.3268 & 0.4146 \\
0.5805 & 0.6585 & 0.7854 & 0.8976 & 0.9512 & 1.0000 \\
\end{bmatrix}
$$

First, a three-layered BP neural network composed of an input layer, an implicit layer and an output layer is constructed; second, the number of neurons in the implicit layer is set to 9, the number of learning times is set to 1000, the learning rate is set to 0.01, and the error is set to $1 \times 10^{-4}$; and last, the neural network is constructed. While $P$ is taken as the input matrix of the neural network, $n$ is taken as the output normalization matrix of the neural network to train the neural network.

The weight matrix $\omega^1$ from the input layer to the hidden layer is obtained as

$$
\omega^1 = \begin{bmatrix}
1.9398 & -3.7176 & -0.0142 \\
4.2480 & 0.5783 & -0.1295 \\
-3.6515 & 2.0901 & 0.0007 \\
-3.5131 & -3.6659 & -0.5777 \\
-1.1475 & 4.3702 & -0.2519 \\
3.8766 & -3.1815 & 0.2570 \\
4.0843 & -1.7303 & 0.4695 \\
3.6927 & -2.1500 & 0.0108 \\
1.2494 & -4.0156 & -0.0105 \\
\end{bmatrix}
$$
The threshold matrix $\theta^1$ from the input layer to the hidden layer is

$$\theta^1 = \begin{bmatrix}
-4.2142 \\
-3.2795 \\
2.1007 \\
0.4723 \\
-0.2519 \\
1.3070 \\
2.5695 \\
3.1608 \\
4.1895 \\
\end{bmatrix}$$

The threshold matrix $\omega^2$ from the hidden layer to the output layer is

$$\omega^2 = [-0.0701 \ 1.2050 \ 0.6709 \ -1.2813 \ -1.2802 \ 
-1.8689 \ 1.0459 \ 0.4969 \ 0.0679]$$

The weight $\theta^2$ from the hidden layer to the output layer is

$$\theta^2 = -0.2968$$

After using the genetic algorithm, the weights and thresholds are rectified. The relation between generation time and the error of the genetic algorithm is depicted in Figure 10.

From Figure 10, it can be seen that with the increase of generation, the error continuously decreases. While comparing the error of general BP networks with the BP networks optimized by genetic algorithm, it is found that the genetic algorithm can get the minimum value of error by genetic iterations. Every time it will get the optimal individual characteristic according to the fitness function Eq. (20). This error is smaller than that in the BP networks. Figure 11 shows the convergence of trained BP neural network.

As seen from Figure 11, in the training process of BP neural network, the error meets the set error standard after 11 iterations, and the trained BP neural network is convergent; the minimum error is 0.013518.

The trained BP neural network is used to predict the concentration of methane in the air at the next moment of four consecutive moments, and it is compared with the methane concentration that is actually measured by the sensor. The curves of the predicted concentration and the actual values are shown in Figure 12.
four consecutive moments, and is compared with the methane concentration actually measured by the sensor. The curves of the predicted concentration and the actual values are shown in the following Figure 12.

Among them, the green curve is the predicted value of methane concentration in air by BP neural network, and the red curve is the analog quantity of methane concentration in air measured by the methane concentration sensor.

It is easy to find that the predicted value of BP neural network is quite consistent with the actual value of concentration, and 10 of the 13 data predicted by BP neural network are very consistent. This shows that the change of methane concentration can well be predicted by the BP neural network.

C. SYSTEM OVERALL SIMULATION

The simulation model of the early warning system through MATLAB that includes the control signal wireless transmission part and JK trigger control part at the control signal receiving end, according to the design content of the early warning system in Section 3, is shown in Figure 13. The rectangular is produced by a controller that is used to control the JK trigger control part to action. The sine wave input is the signal carrier that carries the target rectangular wave signal. The high-frequency signal is sampled by the trigger subsystem after passing the bandpass filter and the low pass filter. The target signal can be restored by the use of inverse Fourier transform. The control signal generated by STM32 is transmitted by the wireless transmitting device through carrier modulation technology in the simulation, and the control signal is restored at the wireless receiving device through band-pass filtering, low pass filtering, and demodulation processes. The high level of the control signal at the receiving end represents triggering the emergency treatment, and the low level represents resetting the state of the gas meter and the exhaust fan part to relieve the current emergency state and restore the normal use in the simulation process. The JK trigger will control to cut the gas meter’s power supply circuit and open the exhaust equipment after receiving the control signal.

In Figure 14, X1 is the value before the signal passes through the band pass filter and X2 is the value of the signal after passing through the band pass filter. The waveform of X1 still retains the general shape of sine wave in form, the amplitude of carrier signal X2 is lower than X1, and there is no general shape of the sine wave, but there is still a series of high frequency noises in X2.
lost the high frequency signals, and the clutters of X3 have been filtered.

Figure 16 is a control signal after demodulation that is the same PWM pulse waves as the input signal.

This paper simulated the control reset terminal, using a step signal and two-state switches. When the input step signal is greater than 0, the switch state will remain unchanged, and the actual state of the gas meter and exhaust fan will remain unchanged, that is, the gas meter is in a closed state. The exhaust fan is turned on in this state. On the contrary, when the control signal is less than 0, the switch state is reversed to open the gas meter and close the exhaust fan; the simulation results are shown in Figures 17 and 18, respectively.

D. MODEL ANALYSIS AND DISCUSSION

We focus on the comparison of SVM, BPNN, LSTM algorithm in order to demonstrate the effectiveness of KF-BP-GEN NNS algorithm as proposed. This section introduces four criteria, that is, determinant coefficient, mean absolute error, correlation coefficient, and root-mean-square error. The criteria are expressed by the following equation:

\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,
\]  

(26)

where \(n\) is the amount of sample, \(y_i\) is the \(i\)-th sample, and \(\bar{y}\) is the average.

\[
SS_{\text{tot}} = \sum_{i=1}^{n} (y_i - \bar{y})^2,
\]  

(27)

where \(SS_{\text{tot}}\) is the total sum of squares, \(y_i\) is the \(i\)th sample, and \(\bar{y}\) is the average.

\[
SS_{\text{reg}} = \sum_{i=1}^{n} (f_i - \bar{y})^2,
\]  

(28)

where \(SS_{\text{reg}}\) is regression sum of squares, \(f_i\) is the prediction of the \(i\)th sample, and \(\bar{y}\) is the average.

\[
SS_{\text{res}} = \sum_{i=1}^{n} (y_i - f_i)^2 = \sum_{i=1}^{n} e_i^2,
\]  

(29)
where $SS_{res}$ is residual sum of squares, $y_i$ is the $i$th sample, $f_i$ is the prediction of the $i$th sample, and $e_i$ is the error of $y_i$ and $f_i$.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

where $R^2$ is the determinant coefficient.

$$\delta = \frac{\sum_{i=1}^{n} e_i}{n},$$

where, $\delta$ is mean absolute error, $e_i$ is the error of $y_i$ and $f_i$.

$$r(f_i, y_i) = \frac{\text{Cov}(f_i, y_i)}{\sqrt{\text{Var}(f_i) \text{Var}(y_i)}},$$

where, $f_i$ is the prediction of the $i$-th sample, $y_i$ is the $i$th sample, and $r$ is the correlation coefficient.

$$S^2 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2},$$

where, $S^2$ is root mean square error, $e_i$ is the error of $y_i$ and $f_i$.

And we use the four criteria to evaluate the KF-BP-Gen NNS algorithm. This paper proposed and compared with the SVM, BPNN and LSTM algorithm.

The results of the four criteria are given in the Table 3.

### TABLE 3. The four criteria of SVM, BPNN, LSTM and KF-BP-GEN NNS.

| Algorithm   | $R^2$ | $\delta$ | $r$  | $S^2$ |
|-------------|-------|----------|------|-------|
| SVM         | 0.8959| 1.2714   | 0.5886| 2.9762|
| BPNN        | 0.8763| 1.2674   | 0.6794| 2.9441|
| LSTM        | 0.9087| 1.2494   | 0.7142| 2.9258|
| KF-BP-GEN   | 0.9185| 1.2531   | 0.6975| 2.9089|

We can give a comparison of the four algorithms through in Table 3. It is known that the KF-BP-GEN NNS as proposed in this paper, has the best goodness of fit when considering the determinant coefficient $R^2$. This is obviously superior to the general BPNN. Although the average absolute error $\delta$ of KF-BP-GEN NNS, as proposed in this paper, is not the best, it is only inferior to the LSTM and superior to the SVM and BPNN. Considering the correlation coefficient $r$, the best is the LSTM, but the KF-BP-GEN NNS this paper proposed is better than the SVM and BPNN, and the SVM is the worst. Comparing the root-mean-square error $S^2$ in Table 3, the best is KF-BP-GEN NNS and the worst is SVM.

In sum, the KF-BP-GEN NNS this paper proposed has the best goodness of fit among of the SVM, BPNN, LSTM and KF-BP-GEN NNS; and the KF-BP-GEN NNS this paper proposed also has the good degree of closeness between the predicted and the measured values, and the absolute error of predicted and measured values has the minimum degree of fluctuation.

The gas leakage early warning system based on Kalman filter and BP neural network can not only detect and predict the gas concentration, but it can also detect and predict any other general gas. Kalman filter algorithm and BP neural network algorithm can be used to effectively predict the change of gas concentration in the air. The gas safety early warning system based on the above algorithm can make a timely detection of the gas concentration information in the air and make emergency treatment in case of gas leakage. By synthesizing the predicted value and the measured value of the sensor after knowing the measured value of the sensor in the next time period, the main purpose of its prediction is to obtain a Kalman value closer to the real value. Therefore, the accuracy correction of the measured value of the sensor is Kalman filtering. The BP neural network can learn the methane concentration value in the air, find out the change rule of gas concentration, and use this rule to predict the change of gas concentration. The use of the Kalman value obtained by Kalman filtering is the training sample of BP neural network. The local convergence of the BP neural network can be reduced by the combination of Kalman filter and BP neural network. In addition, there is the problem of too many iterations of errors in BP neural networks. In this case, to reduce the learning time of BP neural network, genetic algorithm can be used to optimize the weights and thresholds of BP neural network in the process of error iteration.

**Remark 1:** The KF-BP-GEN NNS as proposed in this paper has an advantage in fitting degree and error fluctuation.

**Remark 2:** The KF-BP-GEN NNS as proposed in this paper has good portability because it can not only be applied to detect and predict the gas concentration, but also detect and predict any other general gas.

### V. CONCLUSION AND FUTURE WORKS

This paper proposes, the gas leakage early warning system based on Kalman filter and BP neural network. The KF-BP-GEN algorithm model is designed and simulated. To improve the dynamic performance of the BP neural networks and enhance the accuracy of the prediction of gas leakage by the BP neural networks, the algorithm model consists of Kalman filter, Genetic algorithm, and BP neural networks is established. To compare our algorithm with SVM, LSTM and BPNN, four indicators is introduced, which are determinant coefficient, mean absolute error, correlation coefficient and root-mean-square error. The simulation and comparison results show that the KF-BP-GEN algorithm this paper proposed, has the better fitting degree and the smaller error fluctuation. And, by replacing the sensor, the gas leakage early warning system also can detect and predict any other general gas.

In the future, we will plan for the adoption of more advanced detection systems to measure gas leakage, and use deep learning algorithms to analyze and deal with gas leakage of methane, ethane, propane, and butane.

### REFERENCES

[1] C.-C. Ting and C.-C. Chen, “Detection of gas leakage using microcolor Schlieren technique,” Measurement, vol. 46, no. 8, pp. 2467–2472, Oct. 2013.
[2] T. H. Mujawar, V. D. Bachewar, M. S. Kasbe, A. D. Shaligram, and L. P. Deshmukh, “Development of wireless sensor network system for LPG gas leakage detection system,” Int. J. Sci. Res., vol. 6, no. 4, pp. 558–563, Apr. 2015.

[3] A. Varma, S. Prabhakar, and K. Jayavel, “Gas leakage detection and smart alerting and prediction using IoT,” in Proc. 2nd Int. Conf. Comput. Commun. Technol. (ICICT), Chennai, India, Feb. 2017, pp. 327–333.

[4] K. D. Romanak, P. C. Bennett, C. Yang, and S. D. Hovorka, “Process-based approach to CO2 leakage detection by vadose zone gas monitoring at geologic CO2 storage sites,” Geophys. Res. Lett., vol. 39, no. 15, pp. 1649–1653, Aug. 2012.

[5] B. D. Jolhe, P. A. Potdukhe, and N. S. Gawai, “Automatic LPG booking, leakage detection and real time gas measurement monitoring system,” Int. J. Eng. Res. Technol., vol. 2, no. 4, pp. 1192–1195, Apr. 2013.

[6] J. Zhang, “Designing a cost-effective and reliable pipeline leak-detection system,” Pipes Pipelines Int., vol. 42, no. 1, pp. 20–26, Nov. 1997.

[7] M. S. Jadin and K. H. Ghazali, “Gas leakage detection using thermal imaging technique,” in Proc. UKSim-AMSS 16th Int. Conf. Comput. Modelling Simulation, Cambridge, U.K., Mar. 2014, pp. 302–306.

[8] K. Jach, M. Kopica, Z. Mierczyk, P. Pregowski, and M. Strzelec, “Active thermal imaging of natural gas leakages,” in Proc. Pacific Rim Conf. Lasers Electro-Opt., Seoul, South Korea, 1999, p. 581.

[9] Z. Xu, W. Jin, L. Li, W. Xia, and Q. Chao, “Research progress on passive infrared imaging detection technology and system performance evaluation of natural gas leakage,” Infr. Laser Eng., vol. 48, no. S2, pp. 47–59, Jun. 2019.

[10] C. H. Tian, J. C. Yan, J. Huang, Y. Wang, D.-S. Kim, and T. Yi, “Negative pressure wave based pipeline leak detection: Challenges and algorithms,” in Proc. IEEE Int. Conf. Service Oper. Logistics, Inform., Suzhou, China, Jul. 2012, pp. 372–376.

[11] Q. Xu, L. Zhang, and W. Liang, “Acoustic detection technology for gas pipeline leakage,” Process Saf. Environ. Protection, vol. 91, no. 4, pp. 253–261, Jul. 2013.

[12] W. Liang, L. Zhang, Q. Xu, and C. Yan, “Gas pipeline leakage detection based on acoustic technology,” Eng. Failure Anal., vol. 31, pp. 1–7, Jul. 2013.

[13] L. Meng, L. Yuxing, W. Wuchang, and F. Juntao, “Experimental study on leak detection and location for gas pipeline based on acoustic method,” J. Loss Prevention Process Industries, vol. 25, no. 1, pp. 90–102, Jan. 2012.

[14] J. Feng, H. Zhang, and D. Liu, “Applications of fuzzy decision-making in pipeline leak localization,” in Proc. IEEE Int. Conf. Fuzzy Syst., Budapest, Hungary, Jul. 2004, pp. 599–603.

[15] J. Wang and Y.-I. Lin, “A fuzzy multicriteria group decision making approach to select configuration items for software development,” Fuzzy Sets Syst., vol. 134, no. 3, pp. 343–363, Mar. 2003.

[16] F. Qian, L. Chen, J. Li, C. Ding, X. Chen, and J. Wang, “Direct prediction of the toxic gas diffusion rule in a real environment based on LSTM,” Int. J. Environ. Res. Public Health, vol. 16, no. 12, pp. 21–33, Jun. 2019.

[17] H. Xiao, J. Xu, B. Tang, and Z. Zhang, “Classification and concentration prediction of combustible gas based on BPNN and PCA,” in Proc. 2nd Workshop Adv. Res. Technol. Ind. Appl. (WARTIA), 2016, pp. 1561–1568.

[18] P. Wang, F. Huang, X. Liu, and L. Zhao, “Estimating chlorophyll content of rice under soil contamination stress using BPNN,” in Proc. 6th Int. Conf. Natural Comput., Yantai, China, 2010, pp. 1649–1653.

[19] J. Q. Wang, Y. Du, and J. Wang, “LSTM based long-term energy consumption prediction with periodicity,” Energy, vol. 197, no. 11, pp. 71–97, Apr. 2020.

[20] P. Xu, R. Du, and Z. Zhang, “Predicting pipeline leakage in petrochemical system through GAN and LSTM,” Knowl.-Based Syst., vol. 175, pp. 50–61, Jul. 2019.

[21] W. Liang, G. Wang, X. Ning, J. Zhang, Y. Li, C. Jiang, and N. Zhang, “Application of BP neural network to the prediction of coal ash melting characteristic temperature,” Fuel, vol. 260, pp. 1–8, Jan. 2020.

[22] S. Greš, M. Döhler, P. Andersen, and L. Mevel, “Kalman filter-based subspace identification for operational modal analysis under unmeasured periodic excitation,” Mech. Syst. Signal Process., vol. 146, pp. 1–24, Jan. 2021.

**GUOQUAN LIU** was born in 1982. He received the B.S. degree from Zhengzhou University, in 2005, the M.S. degree from the Sichuan University of Science and Engineering, in 2008, and the Ph.D. degree from Chongqing University, in 2011. His research interests include neural networks, stability analysis, nonlinear systems, and embedded system and its applications.

**ZHICHAO JIANG** was born in Jingdezhen, Jiangxi, China, in 2000. She is currently pursuing the bachelor’s degree with the East China University of Technology, China. Her research interests include electronic communications and neural networks.

**QI WANG** was born in Yingtian, Jiangxi, China, in 2000. He is currently pursuing the bachelor’s degree with the East China University of Technology, China. His research interests include intelligent control and neural networks.