A non-intrusive approach for fault detection and diagnosis of water distribution systems based on image sensors, audio sensors and an inspection robot

Ruikai He [a], Peng Xu [a,*], Zhibo Chen [a], Wei Luo [a], Zhineng Su [b], Jiong Mao [c]

[a] Department of Mechanical and Energy Engineering, Tongji University, Shanghai 201804, China
[b] Shanghai Hongqiao International Airport, Shanghai 201804, China
[c] Shanghai Dynawin Facility Management Co.,Ltd, Shanghai 201804, China

ARTICLE INFO

ABSTRACT

Fault diagnosis is important to maintain the normal operation of air-conditioning systems, reduce the energy consumption in buildings, and increase the service life of air-conditioning system equipment. We present a novel approach for fault detection and diagnosis system that relies on image and audio sensors and relevant algorithms.

This paper proposes a fault diagnosis algorithm based on a robot that can automatically capture audio and image signals from microphone arrays and cameras during inspection in a chiller room. It includes audio- and image-based fault diagnosis algorithms. The validity of the algorithm combined with sensors is verified using data from actual equipment in a chiller room.

The audio-based algorithm, which can monitor the abnormal sound of pumps to detect faults, utilizes Fourier transform, a finite impulse response digital filter, and an autoregressive integrated moving average model. We analyze the frequency domain of the pump signal and set the appropriate threshold to monitor abnormal signals based on the fitted model. Meanwhile, the image-based algorithms are divided into three sections to achieve three functions: 1) an AlexNet convolutional neural network is modified to classify the images of the chiller room equipment obtained by the visible light camera; 2) image morphology methods and trigonometric functions are used to read the dials' indicators acquired by the visible light camera; and 3) optical character recognition is used to obtain the highest temperature value in the infrared image of the pump captured by the infrared camera, which helps maintenance staff verify the operation of the pump and detect faults as soon as possible.

These diagnostic algorithms are non-intrusive, low cost, and easy to deploy. Combined with real-time data collection from the sensors on the robot, the algorithms can effectively improve the intelligence of the equipment room and allocate human resources more reasonably.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Background

The chiller room, which controls the temperature and humidity of the entire building, mainly includes core equipment such as chillers, chilled water pumps, cooling water pumps, water collectors, and manifolds, and some pipelines such as the chilled water and cooling water loops, which form an important part of a heating, ventilation, and air conditioning (HVAC) system Although these facilities provide thermal comfort to people, they are energy consuming when faults occur in improper routine operations and poor preventive maintenance of a HVAC system [1] Currently, with the restricted use of fossil fuels, fault detection and diagnostic (FDD) technologies for chiller room facilities play an important role in improving the energy efficiency of buildings. FDD approaches in HVAC area were categorized into three class: quantitative model-based method, qualitative model-based method and process history based method [1]. While these FDD approaches can diagnose specific problems (e.g., low delta-T syndrome of chilled water systems[2]), they all require large amounts of HVAC system operating data to analyze, and not all building automation systems (BASs)[3] have the ability to save such data in real. If the need to access these data could interfere with the normal operation of the equipment,

such as taking pressure points on pipes and stripping the insulation from pipes to measure temperature or pressure, it would be intrusive. In this study, we used images of equipment (visible images and infrared images) and sounds emitted by equipment to detect whether the equipment is operating abnormally without diagnosing the specific problems so as to schedule reasonably the working hours of operation and maintenance personnel.

1.2. Related work about equipment diagnostic methods

In addition to data collected by BASs, images and audio signals can be used to diagnose faults. Infrared thermography is a major focus in image-based because temperature is an important indicator of the structural health of equipment and components as malfunctioning machines often differ from well-working machines in their temperature distribution [4]. Bandyopadhyay et al. [5] developed a fault diagnosis combining an image processing method and the nearest neighbor algorithm to detect initial faults in insulated gate bipolar transistors of motor converters. Myrans et al. [5] developed a method for automatic, online, near real-time sewer fault detection using closed circuit television to extract still photos, and a random forest classifier with high accuracy using the hidden Markov model and order oblivious filter to reduce the false negative rate of detection. Guan Wang et al. [6] built a shallow convolutional neural network and implemented migration learning by fine-tuning the top layer of an already trained neural network (NN) to determine the disease severity in a small sample of fine-textured crops. They used multiple NNs and parameter combinations for training, and ultimately concluded that the VGG16 network using migration learning led to the most accurate predictions. This study uses an AlexNet [7] network and migration learning idea to identify photographs of equipment taken by robots during inspections in the machine room, as will be elaborated later in **2 Materials and Methods**. AlexNet and VGG16 are also both conventional neural networks, and both can be used for classification tasks. Adam Glowacz et al. [8] used infrared thermal images to diagnose a three-phase induction motor. The magenta image was extracted to obtain a fault feature vector, followed by backpropagation NN, nearest neighbor classifier, and K-means clustering methods for fault feature training and identification.

Overall, there are two main types of fault diagnosis using images at this stage. The first type of diagnosis applies traditional image processing techniques to visible or infrared images and combines them with feature extraction and classification methods to carry out effective fault identification. The other type uses large quantities of images to train deep NNs, and then utilizes the trained NN for fault classification to discriminate the fault.

Vilela et al. [9] stated that the acoustic properties of a machine usually change when a failure occurs. Sound can be used to analyze to operation condition of the fan coil in a HVAC system [10], the flight condition of a unmanned aerial vehicles (UAV)[11] and so on. Therefore, the sound of a machine carries information about its operation condition, and the information is useful in diagnosing. Extracting the acoustic characteristics of a machine is very useful in troubleshooting.

In terms of audio, there are two main types of machine sound information: 1) the vibration signal produced by the vibration of the machine when the machine is working, and 2) audio signals propagated by fluctuations in the surrounding air medium caused by machine vibration. The study of vibration signals mainly includes time domain, frequency domain, time frequency, and non-linear analysis [12]. The time domain analysis mainly includes time signature analysis (TCAs) that involve various factors such as frequency averages, standard deviations, and peak [13-16]. The frequency domain analysis considers the spectral characteristics, such as the frequency averages, standard deviations, and the

peak, energy, and spectral energy ratios, which are often used to troubleshoot pumps, motors, and gearboxes [13,17,18]. The short-time Fourier transform is a traditional time frequency analysis method[19,20] that can analyze unsteady signals. In the past 20 years, new techniques such as empirical modal decomposition [13,21], Hilbert–Huang transform [13,22], wavelet transform [23,24] and NN[25,26] have also been used in time frequency analysis. Audio signals can be converted to spectrograms and analyzed with convolutional neural networks (CNNs) [25] or recurrent neural networks (RNNs)[26]. Many advances methods using finite impulse response (FIR) have been used to monitor complex systems[27,28].Gino Iannace et al.[10] measured the noises emitted by fan coil in an open-plan office to detect its operating conditions. Correlation analysis was used to compare the average pectral levels in a 1/3 octave band. A recursive feature elimination and a decision tree were used to identify operating conditions.

Adam Glowacz et al. [29] identified single-phase induced motor bearing and stator faults based on acoustic signals. The significant frequency feature vector is extracted as a training fault feature, and finally, the fault feature classification identification is satisfactorily performed in the nearest neighbor classifier, nearest mean classifier, and Gaussian mixture model.

Glowacz [30] used audio signals for fault diagnosis analysis of single-phase induction motors. A motor in five operating states was studied, and audio signals in different states were classified using the nearest neighbor classifier.

Liu Mengyue [31] adopted vibration singals for fault diagnosis study of motor bearings. In order to extract fault features from complex signals, the fault extraction methods based on wavelet packet and ensemble empirical mode decomposition (EEMD) were studied separately. Multiclass relevance vector machine (M−RVM) was used to achieve the motor intelligent diagnosis of bearing faults. Li Ye [32] also used EEMD to handle noise in non-stationary and nonlinear vibration singals. Unlike Mengyue, improved local mean decomposition (ILMD) was proposed to extract fault features and incremental probabilistic neural network (IPNN) was used to diagnose bearing faults.

At this stage, the mainstream research method for audio signal fault diagnosis obtains the original signal, then processes the original signal and uses various mathematical methods to extract the transformed signal fault feature as a fault diagnosis data source. The diagnostic methods include not only simple traditional mathematical feature calculations but also machine learning models to train classifiers and identifiers. The overall process can be divided into two major parts: feature extraction and feature identification.

Whether it is image-based or audio-based diagnosis, the installation of sensors affects the normal operation of the equipment (e.g. motors, pumps) to a greater or lesser extent. Although the previous approach can diagnose specific problems within a particular facility, it is more demanding in terms of sensor location. The installation of sensors on or near the equipment can effectively reduce the interference of other sound sources on the fault diagnosis, but the fundamental issue is that it might affect the operation of the equipment. High demands are placed on the resolution and location of the image sensor to accurately determine the wear and tear of the equipment and other potential problems, which may be intrusive to equipment. In addition, the diagnostic methods based on these two different types of data are not sufficiently integrated. Conversely, we leverage the use of general images, infrared images, and audio data by combining them.

1.3. Outline of the non-intrusive FDD approach

In the FDD proposed in this paper, sensors that collect data for real-time diagnostics are installed on a inspection robot, which does not interfere with the operation of the equipment.. We use

a inspection robot for FDD in a chiller room. The cost of the inspection system is manageable, and there is no requirement for high-resolution image sensors. The quality of the existing BASs will not affect FDD. The proposed methodology contains an audio-based algorithm and an image-based algorithm, and we explain them in **2 Materials and Methods**. Relying on inspection robots to inspect the chiller room, sensors collect the operational audio signals and images of equipment through microphone, visible light camera, and infrared camera. The steps for FDD with a inspection robot: when the robot takes a picture of the equipment with the visible light camera module at the pre-set position, it identifies the type of equipment: 1) if it is a dial, the robot reads the dial and determine if the operating condition of pipeline is normal. 2) it is a pump, the robot calls the audio sensor so as to detect whether sound from the pump is normal or not, and calls the infrared camera module to read the temperature on the infrared picture to determine if the pump is overheated. The diagnostic idea is shown in Fig. 1.

2. Materials and methods

2.1. Hardware and software setup

The audio signals and images are the important basis for the non-instrusive FDD approach. We utilized the ReSpeaker Mic Array v2.0[33] as a sampling device for audio signals, which is a far-field sound capture array device that incorporates four Pulse Density Modulation (PDM) microphones to enhance the device's acoustic digital signal processing performance. Determination of sample frequency is discussed in **2.2.1 Selection of Sample Frequency**.

Pictures of equipment including temperature and pressure dials were taken in a real chiller room. We utilized RER-USB13MAF-V7S [34] as visible light camera module and FLIR A315[35] as infrared camera module. The details of the chiller room is in **3.1 The Details of the Chiller Room**.

We used Python3 and MATLAB to verify the algorithms in the proposed FDD approach.

2.2. Methods in the non-intrusive FDD approach

The proposed fault diagnosis algorithm is developed in terms of both sound and image, as shown in Fig. 2. The audio signal fault diagnosis algorithm is developed based on audio signals collected from different pumps in a LANDSEA chiller room. In **2.2.2 Frequency Domain Analysis**, Fourier Transform (FT)[12] is used to transform the time-domain signals into frequency domain signals, and FIR digital filters[12] are used to analyze the signals in three frequency bands: high-, medium-, and low-frequency domains. In **2.2.3 Time Series Model Fitting to Diagnose**, the extracted audio signals are modeled using an autoregressive integrated moving average (ARIMA)[36] model, and the fault diagnosis is performed using the threshold method based on the prediction confidence level. Image troubleshooting algorithms covers the classification of equipment in the freezer room using AlexNet[7] (**2.2.4 Equipment Images Classification**), the dial indicator reading based on morphological operations[37] (**2.2.5 Dial Indicator Reading**) and recognition of the infrared image temperature based on optical character recognition (OCR)[37] (**2.2.6 Infrared Images Recognition**). First, we classify the images of the facilities collected by the visible light camera sensor installed in the inspection robot. Then, for the images of pressure and temperature dials, the corresponding algorithm is used to judge whether the pipeline pressure and temperature are normal. Regarding the images of water pumps, OCR is used to identify the highest temperature indicator
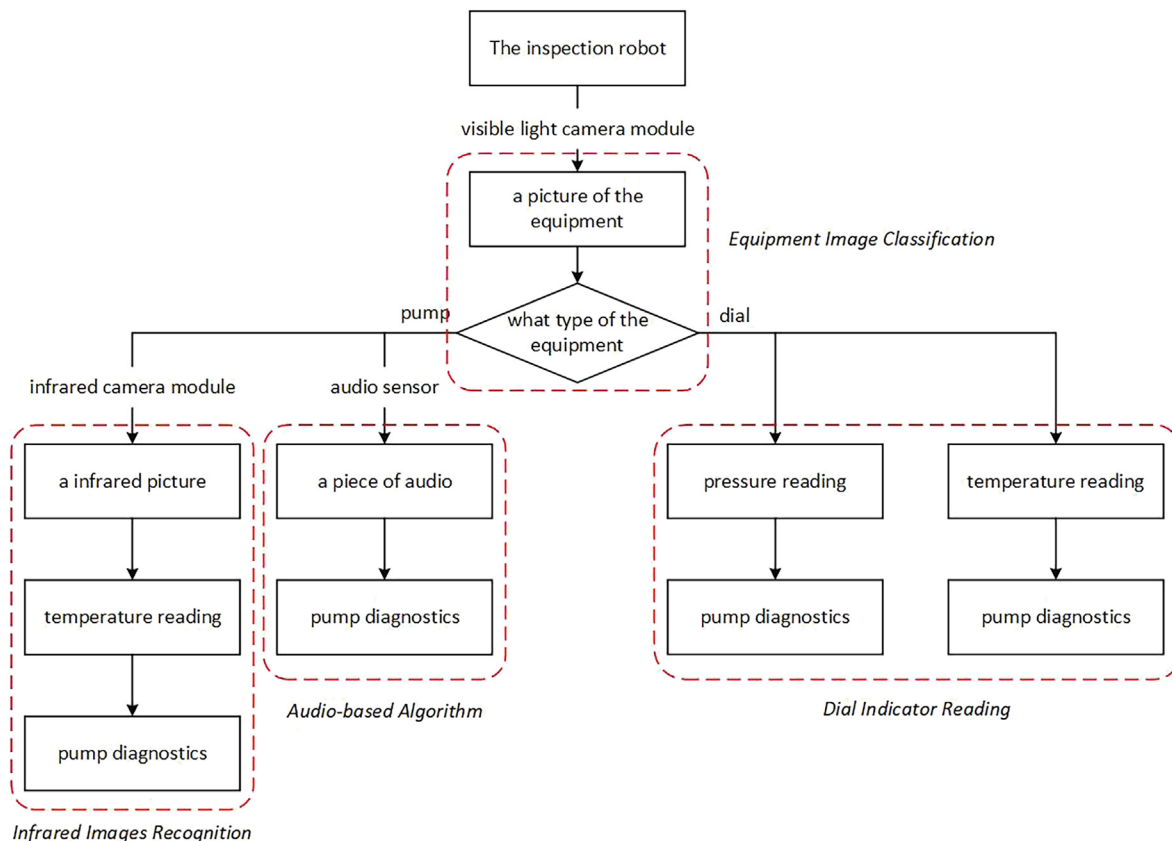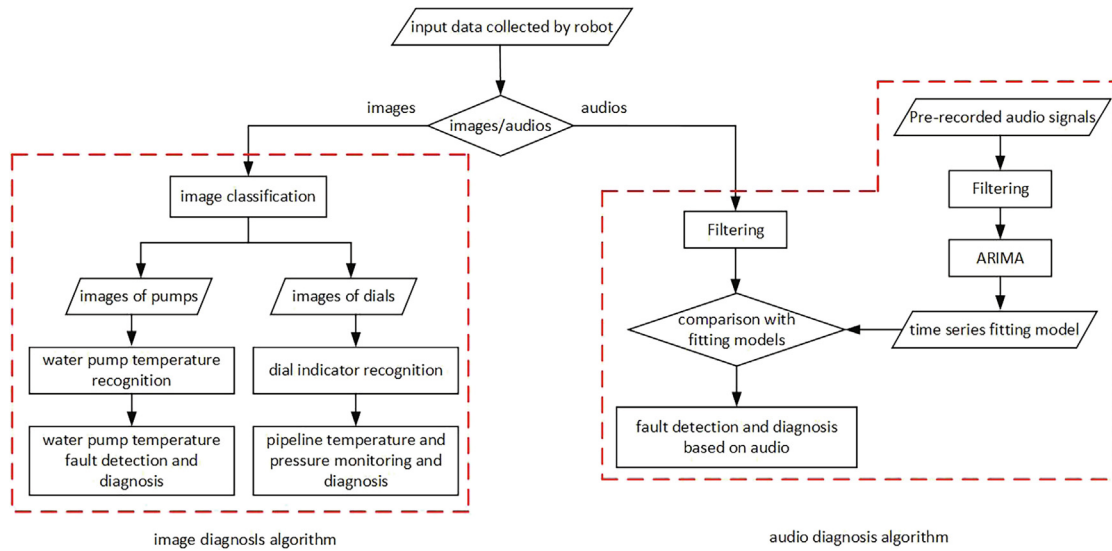

Fig. 1. General idea of the robot inspection.

Fig. 2. Overview of non-intrusive diagnostic algorithm.

of the infrared image, so as to judge whether the temperature of the water pump is abnormal.

2.2.1. Selection of sample frequency

In the audio-based fault diagnosis, the first step is to determine the audio sample frequency of the audio sensor. FT is a linear integral transform[12]. In practice, we sample discrete time signals in engineering situations. An infinitely long sequence must be intercepted into a finitely long sequence for discrete Fourier transform (DFT) [12] analysis and processing. Fast Fourier transform (FFT) [12] is a general term that designates a practical method for the scientifically efficient computation of DFT. FFT makes the analysis of actual signals more feasible. Therefore, we adopt different sample frequencies to collect audio signals, and then use Python to convert these by FFT. Eventually, we study the frequency domain range where the signal energy is mainly concentrated and ensure that there is no aliasing in the frequency domain of the collected signals. As described above, it is more reasonable to use the sample frequency for both microphone arrays and further analysis.

2.2.2. Frequency domain analysis

After determining the pump sound sample frequency, an FIR digital filter can be utilized to obtain the desired frequency domain to achieve signal separation for the further analysis. The original signal is filtered to obtain the required frequency components. According to the analysis of the pump audio signal in the frequency domain, we determined the audio-based diagnosis method. The FIR digital filter can be expressed as a Z-transformation:

$$H(z) = \sum_{n=0}^{N} h(n) Z^{-n} \tag{1}$$

where N is the number of filter steps designed, and $h(n)$ is a finite length series. The impulse response, i.e., the designed filter coefficients are

$$(n) = b_n, n = 0, 1, 2, \cdots N \tag{2}$$

We used MATLAB to design a GUI application for audio signal filtering (***3.3 Frequency Domain Characteristics of Water Pumps***). With this GUI, the original signal can be flexibly filtered into multiple signals according to its frequency components. The filtering results can be used to analyze the characteristics of pump audio

signals in different frequency domains, so as to determine whether the audio-based algorithm is developed from the time domain or the frequency domain.

2.2.3. Time series model fitting to diagnose

After the frequency domain analysis of the pump audio signal, the audio-based fault diagnosis algorithm was developed from the time domain analysis. We used the ARIMA[36] model to fit the pump high-, medium-, and low-frequency domain signals. If the original time series (audio signal) is non-smooth, the d-order difference operation is first performed to turn it into a smooth time series; for a smooth series, the ARMA (p,q) model is used directly. The ARIMA modeling flowchart is shown in Fig. 3.

The auto regression (AR) model describes the relationship between current and historical values and uses the variables' own historical time data to predict them:

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t \tag{3}$$

where x is the variable, t is the time, μ is a constant, ϕ_i is the autocorrelation coefficient, u_t is the error, and p is the number of the autoregressive terms. In particular, when $\phi_i < 0.05$, the model is inappropriate, and the premise of ARIMA modeling is that the data are smooth.

Moving average (MA) model: sometimes, in the AR model, u_t is a white noise, which is usually expressed as

$$u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{4}$$

where ε_t is random error, and q is the MA number of terms. When $u_t = \varepsilon_t$, u_t is white noise. Combining the AR model with the MA model yields Eq. (5), and each symbol has the same meaning as that in Eqs. (3) and (4).

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \# \tag{5}$$

In this study, the augmented Dickey-Fuller (ADF)[36] test was used to test the signal for smoothness. The p-order autoregressive process of the time series is constructed and the characteristic roots λ_i of its characteristic equations are found, where $i = 1, 2, \cdots, p$. If all the characteristic roots fall within the unit circle $|\lambda_i| < 1, i = 1, 2, 3, ..., p$, the time series is smooth. If a characteristic
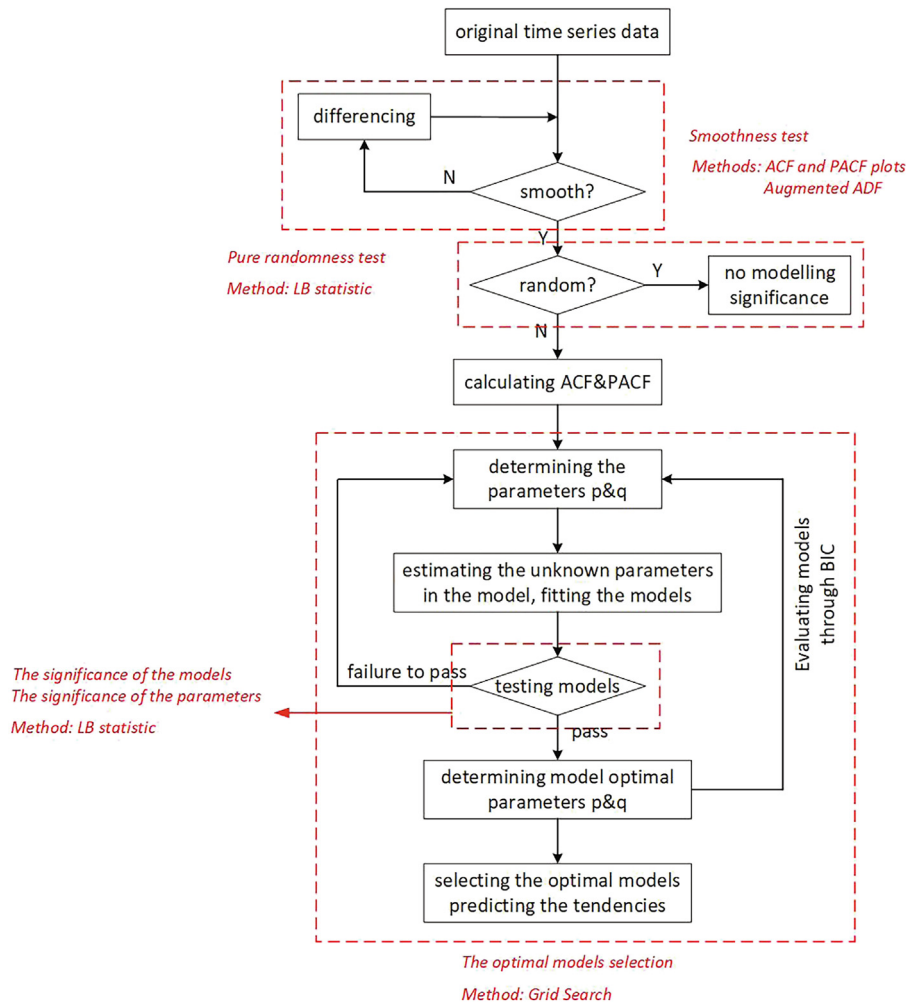
Fig. 3. ARIMA modeling flowchart.

root is the unit root, which means $|\lambda_i| = 1, \exists i \in 1, 2, 3, \cdots, p$, then the time series is not smooth. The hypothetical test condition is as follows:

Original hypothesis H_0: The time series is not smooth when at least one characteristic root is the unit root.

Alternative hypothesis H_1: All characteristic roots are within the unit circle, and thus the time series is smooth.

At some level of significance, the time series is a smooth series when it can significantly reject the original hypothesis H_0, which means that the time series is smooth.

To determine if the smooth time series has any value for further analysis, a pure randomness test needs to be performed to determine if it is a purely random series. As smooth time series generally have only short-term correlations, we only need to test whether there is a short-term correlation in the smooth time series to enable us to determine whether the series is a purely random sequence. The hypothetical test condition is as follows:

Original hypothesis H_0: Sequence values with a number of delay periods not greater than m periods are independent of each other.

Alternative hypothesis H_1: There is a correlation between sequence values for which the number of delay periods is not greater than m periods.

We use ρ to represent the autocorrelation coefficient and the subscript to represent the number of periods of delay; then, the hypothesis test condition can be expressed in Eqs. (6) and (7):

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0, \forall m \geqslant 1 \# \tag{6}$$

$$H_1 : \exists \rho_k \neq 0, \forall m \geqslant 1, k \leqslant m \# \tag{7}$$

The LB statistic is constructed to test this joint hypothesis.

$$LB = n(n+2) \sum_{k=1}^{m} \left(\frac{\hat{\rho}_k^2}{n-k} \right)^2 \tag{8}$$

The value of the LB statistic is calculated and its corresponding p-value is found. If the p-value is significantly greater than the set significance level α, then the original hypothesis H_0 cannot be rejected; conversely, if the p-value is less than the significance level α, then the original hypothesis H_0 can be rejected, thus determining that the sequence is not a purely random sequence.

After confirming the smooth non-pure random time series, the time series can be analyzed using the ARIMA model. The parameter d of the ARIMA model is confirmed by whether the original data is a smooth or differential smooth time series. p and q are determined by autocorrelation function (ACF) plots and partial autocorrelation function (PACF) plots, respectively. After determining the approximate range of the p and q values, the grid search algorithm is used to further confirm the p and q values using the Bayesian information criterion (BIC)[36]. The model that minimizes the BIC value is also the optimal model for this case.

$$BIC = -2ln(f(y, \theta_k)) + ln(n)K \tag{9}$$

where K is the number of model parameters, n is the number of samples, and $f(y, \theta_k)$ is the likelihood function[38], which is a function of the parameter θ and represents the likelihood of θ. $ln(n)K$

(penalty term) is effective in avoiding dimensional catastrophes when the number of dimensions is excessively large and the size of the training sample data is relatively small.

Finally, the validity of the model needs to be tested, which includes testing the significance of the model and the significance of the parameters. If the residual term after model fitting is a purely random series, the model is significantly valid, indicating that the fitted model adequately extracts valuable information from the original data. The method for the model significance test is the same as that for detecting whether a smooth time series is random or not.

The unknown parameters that correspond to key variables are needed for the detection, and these that fail to pass the significance test can be deleted because these parameters have little effect on the model. In the final fitted model, all parameters should significantly affect the model.

To validate the model, data equivalent to the length of the fitted data are predicted by the abovementioned ARIMA models. If the amplitudes of 5% of the sample points in an audio exceed the threshold (the upper and lower limits of the 95% confidence interval), the pump will be considered to be in abnormal operation during this period.

2.2.4. Equipment images classification

When the robot patrols the chiller room, it takes many pictures (of chillers, water pumps, temperature dials, and pressure dials) in the pre-set position by Light Detection and Ranging (LIDAR)[39], which is a method for measuring distances by illuminating the target with laser light and measuring the reflection with a sensor. When the equipment is identified accurately, further diagnosis makes sense. We adopt the modified AlexNet to recognize the visible light pictures taken by the robot. The feature extraction remains unchanged, and the part of the classification is rebuilt according to the requirements of the issues to be addressed: to change the output of the fully connected layer to 4 (means 4 types of equipment in **3.1 The Details of the Chiller Room**). In model training, only the parameters for the classification are updated. Pictures of the four types of equipment taken in the actual server room were used as training and test sets. The height and width of all images were adjusted to 224 * 224 and flipped horizontally with 50% probability to increase the number of equipment images in the training set.

To enhance the training effect of the model, we standardized each of the three color channels of the images: the mean and standard deviation of all pixels were determined separately for each color channel, and then each pixel of that color channel was calculated as a standardized pixel value according to Eq. (10):

$$Pixel_{out}[\text{channel}] = \frac{Pixel_{in}[channel] - Pixel_{mean}[channel]}{Pixel_{std}[channel]} \quad (10)$$

where the subscript *in* represents the pixel value before standardization, *out* represents the pixel value after standardization, *mean* represents the mean of the pixel value of the channel, and *std* represents the standard deviation of the pixel value of the channel.

2.2.5. Dial indicator reading

There are pressure dials and temperature dials on the pipes in the plant room, so reading the indications on the dials can be used to check the condition of pipes. To make the readings more accurate, standardized photography should be acquired before reading the dial. Standardized photography means that the photographs taken by the robot should be sharp and clear, and the dial to be read should be vertical and centered on the photograph. Fig. 4. illustrates the dial indicator reading process. Based on the above

assumptions, this part of the algorithm can be divided into three parts: 1) identify the position of the pointer in the image because the pointer is crucial to the reading, 2) calculate the angle between the pointer and the negative y-axis, which help us obtain the angle between the pointer and the zero mark, 3) convert it (the angle between the pointer and the zero mark) to the corresponding pressure or temperature value, thus realizing the function of reading the dial number automatically. Image scaling/rotating mainly involves adjusting the resolution of the image and changing its size without losing necessary information. When H/W of the input image is less than 1, it is designed to rotate 90° counterclockwise.

The grayscale images are binarized in R, G, and B channels after rotation and scaling, which is achieved by the maximum interclass variance (OTSU)[37], and then the logic AND operation is performed on the three binarized images. The purpose of this step is to initially separate the object from the background. After obtaining the binary image of the dial, we have to split the pointer from it. First, we take the complement set to the binary image so that only the dial is left in the entire image. Except for the white background in the image, the pointer section is the maximum connectivity domain, and thus we can extract the pointer from the image. For smoother pointer boundaries and higher quality image support for subsequent pointer angle algorithm design, we choose the closing operation[37] so that the tiny break in the tip of the pointer is partially bridged. The image is cropped so that the pointer fills the entire picture as much as possible. It is supposed that the binary image to be cropped is a matrix $M_{n \times m}$, where n represents the number of rows and m represents the number of columns.

Algorithm: *image crop for pointer*

> **Input**: Matrix of the binary image to be cropped $M_{n \times m}$
> **Output**: boundary coordinates (row1,column1), (row1,column2), (row2,column1), (row2,column2)
> **for** i = 1 to m by 1 **do**
> **for** j = 1 to n by 1 **do**
> **if** $M[j,i]$ = 1 **then**
> row1 = i ; column1 = j
> break
> break
> **for** i = m to 1 by -1 **do**
> **for** j = n to 1 by -1 **do**
> **if** $M[j,i]$ = 1 **then**
> row2 = i ;column2 = j
> break
> break
> **return** row1;column1;row2;column2

Once the image cropping is complete, the tip of the pointer can be extracted from it, as the image of the pointer tip is more appropriate for calculating the angle due to the uniformity of the tip. We cropped the image of the pointer and assumed that the matrix of the image of the pointer is $M_{m' \times n'}$.

Algorithm: *image crop for tip of the pointer*

> **Input**: the matrix of the image of pointer $M_{m' \times n'}$
> **Output**: the matrix of the image cropped again $M_{m'' \times n''}$
> if $\frac{m'}{n'} > 1.2$ then
> do $m'' = \frac{1}{3} \times m'$; $n'' = n'$
> else
> do $n'' = \frac{1}{3} \times n'$; $m'' = m'$
> end if
> return m''; n''

Once an image of the part of the needle tip is available, the next step is to calculate the angle between the pointer and the y-axis. In the width direction, the midpoints of multiple pointer sections are
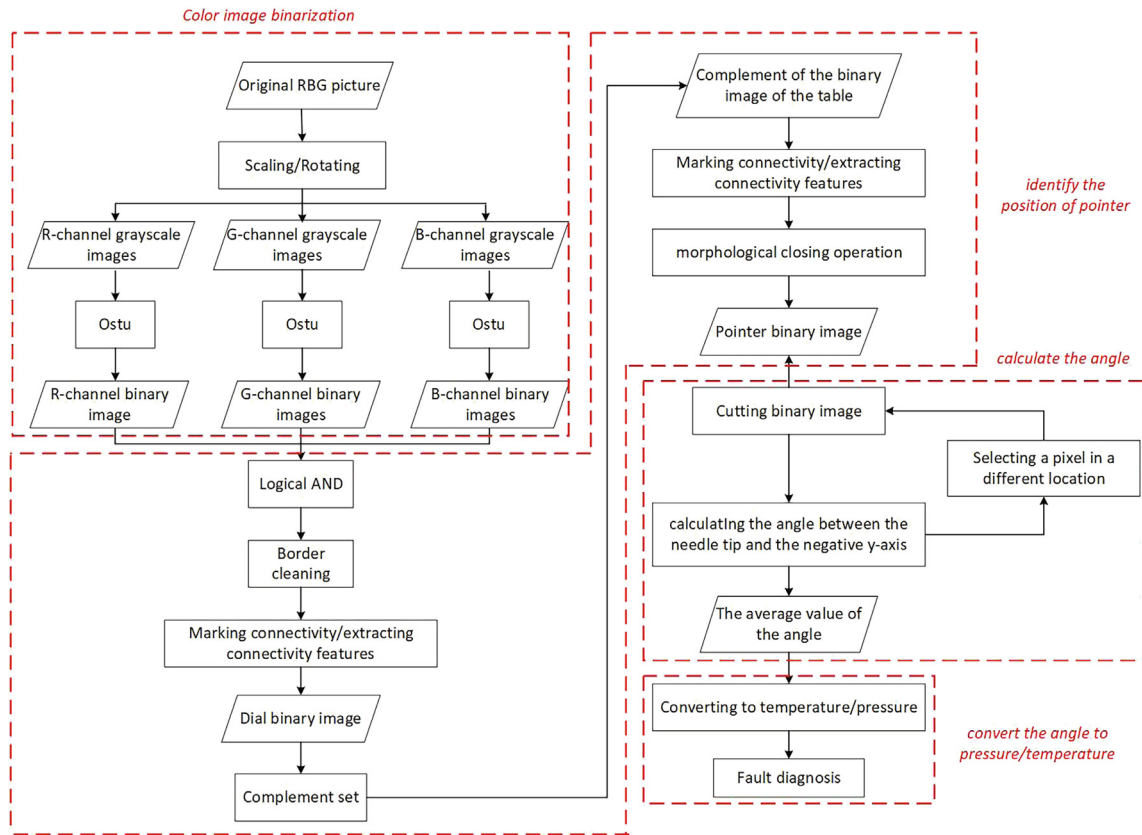
Fig. 4. Dial indicator reading process.

taken, and the angle is calculated using an arc-tangent trigonometric function. Therefore, to ensure the reliability and generalizability of the algorithm, the entire dial must be divided into different part. For pointers in different parts, the calculation method needs to be adjusted when calculating the angle. The partitioning of the dial is shown in Fig. 5. The blue box shows the image of the pointer. When $H/W > 1.2$, the pointer is located at part2 or part3. If $H/W < 1.2$, the pointer is in part1 or part4. $H/W = 1.2$ indicates that the angle between the tip of the pointer and the positive y-axis is 50°.
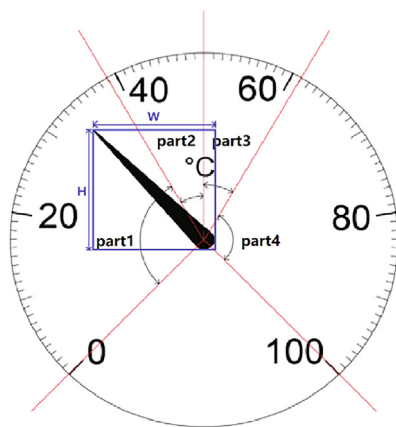
Fig. 5. Partitioning of the dial. The blue box shows the image of the pointer. When $H/W > 1.2$, the pointer is located at part2 or part3. If $H/W < 1.2$, the pointer is in part1 or part4. $H/W = 1.2$ indicates that the angle between the tip of the pointer and the positive y-axis is 50°. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Furthermore, we need to know whether the pointer falls in part2 or part3, and whether the pointer falls in part1 or part4. The judgment process is shown in Fig. 6. After initial partitioning, when the pointer falls in part1 or part4, at 1/8 and 7/8 of the width, the number of pixels is compared with a pixel value equal to 1. If the pointer lands on part2 or part3, a rotation of 90° counterclockwise is necessary before the abovementioned steps. When $w_1 = w_2$, the pointer is at 12o'clock.

Once the location of the pointer has been determined, the next step is to calculate the angle between the pointer and the negative y-axis in a counterclockwise direction. There are slight differences in the calculation in different parts. It is assumed that the size of the image of the tip is $p \times q$.

Calculation in part1: obtain two vertical coordinates $y_1, y_2, y_1 \neq y_2$ and $y_1 < y_2$ and find the corresponding horizontal coordinates x_1, x_2 of the midpoint. The angle is calculated using Eqs. (13 and 14):

$$angle^{'} = arctan\left(\frac{y_2 - y_1}{x_1 - x_2}\right), angle^{'} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \tag{13}$$

$$\begin{cases} angle = angle^{'}, angle^{'} \in (0, \frac{\pi}{2}) \\ angle = angle^{'} + \pi, angle^{'} \in (-\frac{\pi}{2}, 0) \end{cases} \tag{14}$$

Especially, when $x_1 = x_2$, angle = 90°.

Calculation in part2: First, rotate the image 90° counterclockwise, then repeat the steps in *calculation in part1*. The angle is calculated using Eq. (15):

$$angle = arctan\left(\frac{y_2 - y_1}{x_1 - x_2}\right) + \frac{\pi}{2}, angle \in \left(\frac{7\pi}{9}, \pi\right) \tag{15}$$

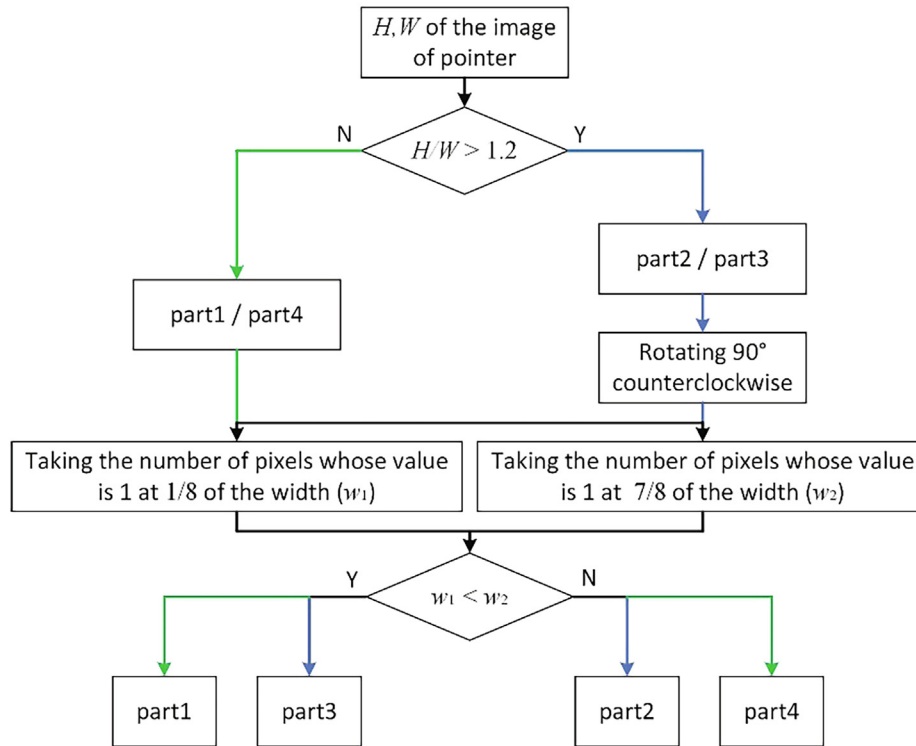Especially, when $x_1 = x_2$, angle = 180°.

Fig. 6. Determination of the place where the pointer falls.

Calculation in part3: The steps are the same as that in *the calculation in part2*. The angle is calculated using Eq. (16):

$$angle = arctan(\frac{y_2 - y_1}{x_1 - x_2}) + \frac{3\pi}{2}, angle \in \left(\pi, \frac{11\pi}{9}\right) \qquad (16)$$

Especially, when $x_1 = x_2$, angle = 180°.

Calculation in part4: the same procedure as that in *the calculation in part1*. The angle is calculated using Eqs. (17 and 18):

$$angle' = arctan(\frac{y_2 - y_1}{x_1 - x_2}), angle' \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \qquad (17)$$

$$\begin{cases} angle = angle' + \pi, angle' \in (0, \frac{\pi}{2}) \\ angle = angle' + 2\pi, angle' \in (-\frac{\pi}{2}, 0) \end{cases} \qquad (18)$$

Especially, when $x_1 = x_2$, angle = 270°.

To ensure accuracy, it is advisable to take multiple sets of points to calculate the angle repeatedly. In the wide direction, a set of points $y_1, y_2 (y_1 < y_2)$ is picked up by a certain step and the above calculation steps are repeated. We take the 0.4 and 0.6 percentile angle values and subsequently average them, which better prevent outliers. Ultimately, we have to translate the angle into the corresponding temperature and pressure.

The conversion angle into temperature/pressure is different for different dials, and thus the temperature dials and pressure dials (Fig. 7) in the LANDSEA chiller room are used as research objects in this study. Taking two anchor points, one anchor point corresponds to 50°C at 12o'clock on the dial, and the other anchor point corresponds to 0°C. The angle between the two temperatures was 42°. Given that the actual shooting is not perfect, the angle ranges from 42° to 46°, and 44° (average value) is taken as the final angle result and the correspondence relation is approximately 1.136 ℃/°. The conversion from angle to pressure is the same as that for temperature: an anchor is taken as 0.8 MPa at 12o'clock and the other corresponds to 0 MPa. The angle between these is an average of 130°, and the corresponding relationship is approximately 6.154

× 10–3 MPa/°. Eventually, we can obtain the actual indication on the dials.

2.2.6. Infrared images recognition

The temperature of the water pump during operation is also a key indicator of whether the pump is operating properly. Dial reading is a good method to monitor the temperatures and pressures of pipelines, but this technique is impractical for pump temperature monitoring. For pump temperature monitoring, this study proposes another method, which involves recognizing infrared images. Reading the number in the upper right corner of the infrared image helps the staff to know the highest temperature of the pump. Fig. 8 illustrates the process of the infrared image.

In the infrared image taken by the FLIR infrared camera mounted on the inspection robot, the maximum temperature is displayed in a fixed position whose coordinates are positioned from 279 to 315 on the x-axis and from 5 to 25 on the y-axis, assuming that the lower left corner of the image is the origin of the Cartesian coordinate system. The cropped color image is grayed and subsequently binarized using the OTSU[37]. To ensure recognition accuracy, the image needs to be enlarged and the enlargement scale is determined by grid search. In the search, the optimal scale factor is designed to be chosen from 1 to 5 in increments of 0.1, and then we record the accuracy of identifying 100 different infrared images at different scales. The results of the grid search are shown in Fig. 9.

3. Results

3.1. The Details of the chiller room

The proposed algorithm was validated in a real chiller room in Shanghai. The equipment in the chiller room includes 6 screw heat pump units, 2 screw chiller units, and several vertical and horizontal pumps. There are also water system pipelines, which are

Fig. 7. Temperature dial (left) and pressure dial (right).
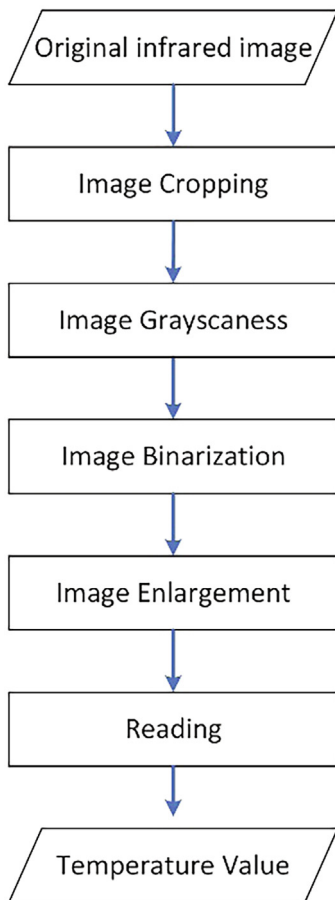

Fig. 8. Process of infrared image recognition.

equipped with dials to show the pressure or temperature of the medium in the pipes (Fig. 10.).

3.2. Determination of sample frequency

We used the radio signals of hot water pump #1 from a real chiller room to analyze the pump radio. This audio signals include the sound emitted by the pump itself when it was operating and the ambient sound around it. Three different sampling frequencies, 44100, 22050, and 11025 Hz, were used in the sampling process (20 s). An FFT is performed on the signal (0.1 s) using Python. It is more reasonable to determine which sampling frequency is used for subsequent audio signal analysis by observing the results of the Fourier transform. Fig. 11 shows the results of the FFT. Finally, we chose 11025 Hz as the optimal sampling frequency because spectral overlap was small at the sampling frequency of 11025 Hz and filtering did not have a significant impact on the 0–3000 Hz band.

3.3. Frequency domain characteristics of water pumps

We used MATLAB to design a GUI application to filter the audio signals (Fig. 12). We choose the fir2 function of the FIR digital filter in MATLAB to design the filter. In particular, when using the fir2 function to design the filter, we should truncate the filter coefficients obtained, and the truncation length is the design order. After truncation, we can obtain the FIR. In the fir2 filter, the frequency break points corresponding to the 18 sliders are set, and frequency points $\in [0, 1]$, which represent the percentage of the actual frequency domain from 0 to 5512.5 Hz. It should be noted that the frequency break point corresponding to the filter frequency value is the product of 5512.5 Hz and the break point value. The imported original data comprise the pump audio signal of 30 s.

We filtered the audio signal from the hot water secondary pump # 1 in the LANDSEA chiller room. The low-, medium-, and high-frequency signals are separated, where the higher amplitude part is worthy of further study because this is the main information from the equipment itself, and the lower amplitude part is mainly from various other sources, such as other facilities. The low-frequency domain ranges from 0 to 535 Hz. We moved the sliders to 1 before 0.097 in the filter GUI. The frequency of the medium-frequency domain signal ranges from 625 to 1333 Hz when all the sliders are moved from 0.1134 to 0.2418 in the GUI of the filter to 1. The frequency rate of the high-frequency domain signal ranges from 1666 to 4000 Hz, and we pushed all the sliders between 0.3023 and 0.7256 of the filter GUI to 1.

In addition, we filtered specific frequency bands, which were the peak amplitude of the low-frequency domain signal, the peak amplitude of the medium-frequency domain signal, and the peak amplitude of the high-frequency domain signal. The low-frequency peak amplitude signal and the high-frequency peak amplitude signal are displayed in Fig. 13(d)(e). Moving the sliders
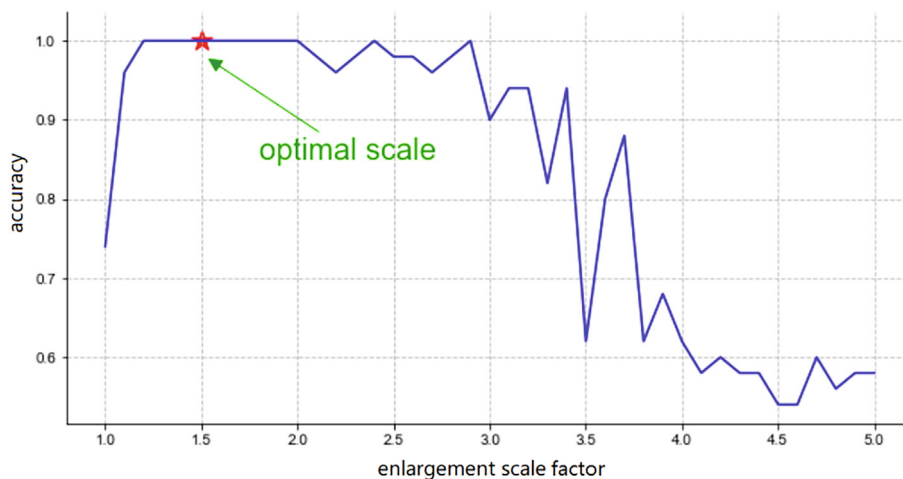
Fig. 9. Relationship between enlargement scale and recognition accuracy. When the enlargement scale is getting larger and larger, the recognition accuracy of OCR is decreasing generally. When the enlargement scale takes 1.2 to 2.0, the recognition accuracy reaches 1.0, so we choose 1.5 as the optimal enlargement scale according to our experience.

chiller compressors   dials   vertical pumps   horizontal pumps
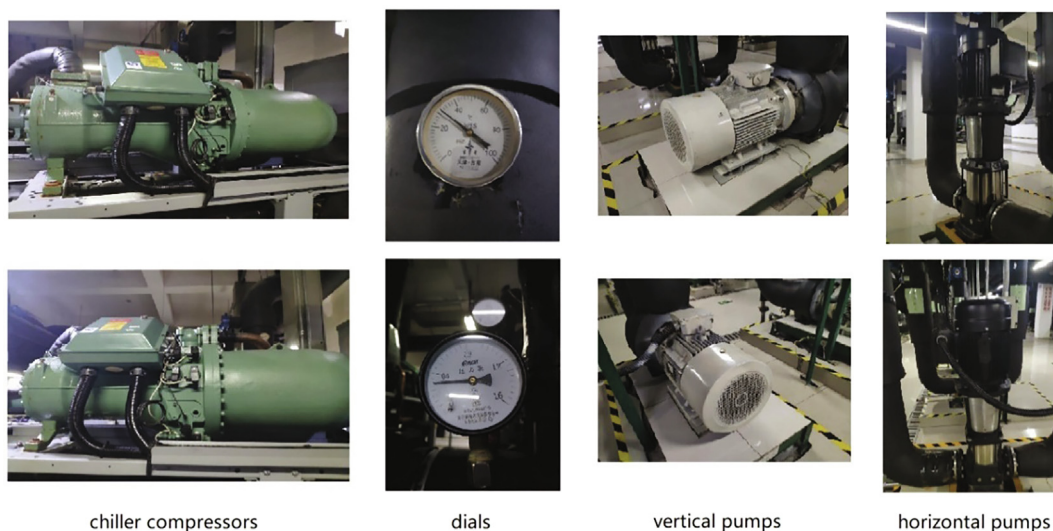
Fig. 10. The equipment in the real chiller room.

to 1 before 0.0227 yielded the peak amplitude frequency domain of the low-frequency signal (Fig. 13(d)), with a similar process for the analysis of medium- and high-frequency signals. Moving the sliders from 0.3023 and 0.3628 to 1 provided the frequency of the peak portion of the high-frequency signal (Fig. 13(e)).

It can be observed that, in 0.1 s audio signal, the extent of oscillation gradually becomes stronger from the low- to high-frequency domain signals. At the same time, they are not of the same amplitude, reflecting the different weights of each part of the signal in the original signal, and the medium-frequency domain signal accounts for a relatively larger proportion. After extracting the low-, mid-, and high-frequency peak amplitude signals, the frequency range of the signal becomes narrower, with fewer frequency components, and therefore the signal oscillations in the time domain become weaker and more regular. In general, 1) the frequency domain of the original pump signal after FFT is complex, with high amplitude in some frequency domains and low amplitude in most frequency domains. It can be inferred that the higher amplitude, to some extent, can reflect the characteristics of the equipment. 2) The collected audio signals, including signals from the machine itself and ambient mixed signals, are complex and

irregular in components. The source of the interfering signal is difficult to control, and thus it cannot be conclusively assumed to be a useless signal component.

3.4. Pump fault detection from audio signals

The signal frequency component is complex and confusing; therefore, signal analysis using the time domain is more appropriate to address the issues in this study. We proposed to model each of the three frequency domains of audio signals using ARIMA for pump diagnosis. The following experiment was implemented in Python. Before applying the ARIMA model to a pump audio, preprocessing of the signal is required, as follows.

1) The signal is verified for smoothness. For the original data, the first and last 1 s data are deleted, leaving the middle 28 s of signal data, of which the first 25 s are used as training data, after 3 s used as test data. Because too many data points (due to high sample frequency) are not suitable for real-time inspection and fault diagnosis, sample frequency reduction is necessary, and the data are taken in a fixed step and the sampling frequency is reduced
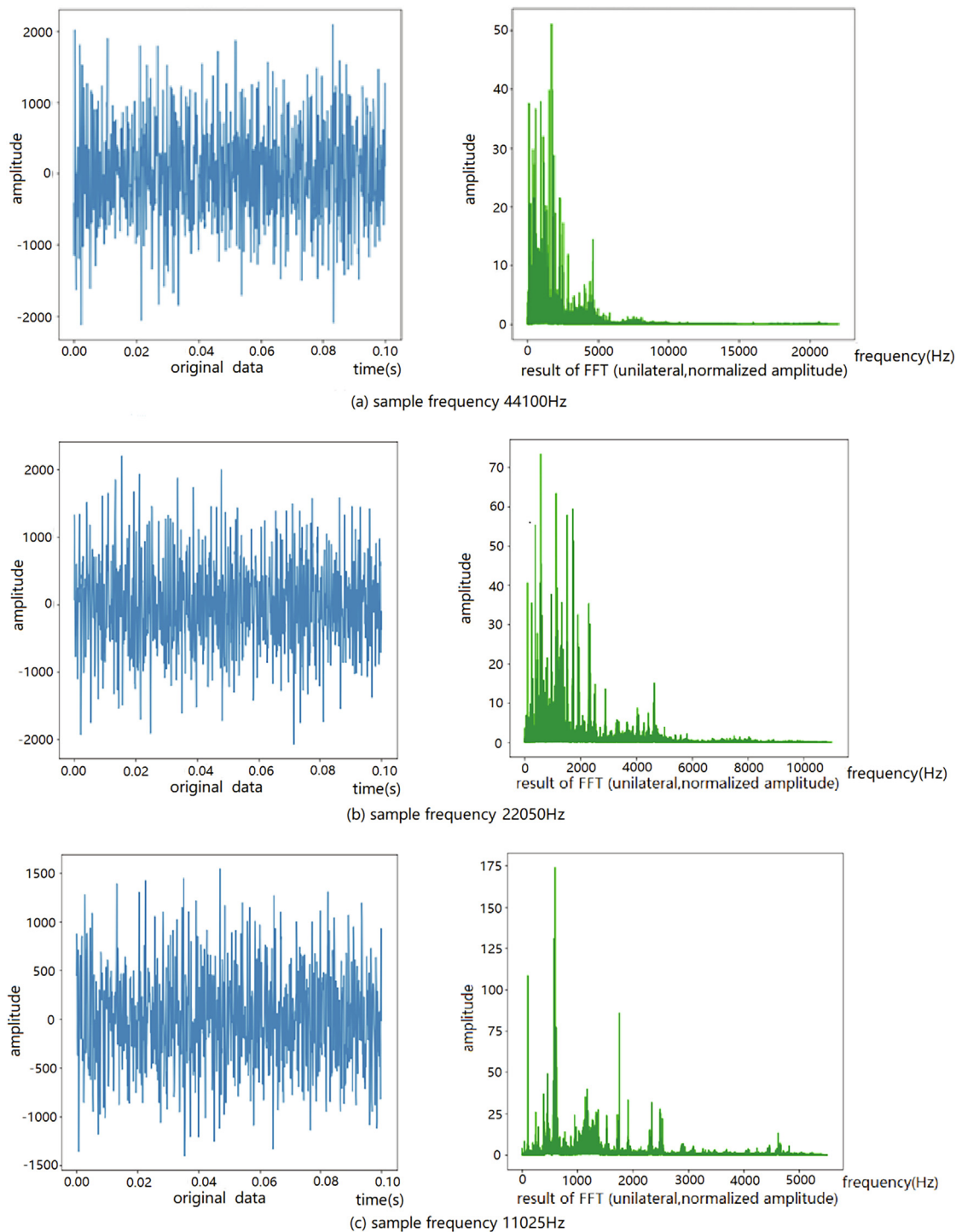
Fig. 11. Results of FFT (sampling frequency 44100, 22050, 11025 Hz). (a–c) show the FFT results after continuous sampling of the sound from the same pump using 44100 Hz, 22050 Hz and 11025 Hz, respectively. We found that the signal energy was mainly concentrated within 3000 Hz, so we were mainly concerned with the 0–3000 Hz band.

to 256 Hz. A smoothing check is performed on the training data and test data.

By analyzing the ACF and PACF plots (Fig. 14), it can be inferred that the original data are closer to a smooth time series. When detecting the smoothness of a low-frequency signal using the unit root, $p_{train\_low} = 2.7 \times 10^{-11} \ll 0.1$, $p_{test\_low} = 2.6 \times 10^{-25} \ll 0.1$. The original hypothesis H_0 can be rejected, and the training and test data can be considered as a smooth time series. In a similar way, $p_{train\_medium} = 5.9 \times 10^{-19} \ll 0.1$, $p_{test\_medium} = 2.3 \times 10^{-7} \ll 0.1$, $p_{train\_high} = 1.5 \times 10^{-12} \ll 0.1$, and $p_{test\_high} = 8.6 \times 10^{-11} \ll 0.1$. The

medium- and high-frequency domain signals are also smooth. The ACF and PACF plots of the medium- and high-frequency domains are shown in Fig. 14(b,c), which have a similar pattern to the low-frequency signal. The original data are a time-smoothed series.

2) Detecting that the time series to be fitted is not a purely random sequence ensures that the time series is of analytical value. Combined with the ACF and PACF plots of the signal, the number of delay periods for correlation is small, barely more than 10 periods, and thus we selected a delay of 12 periods when calculating
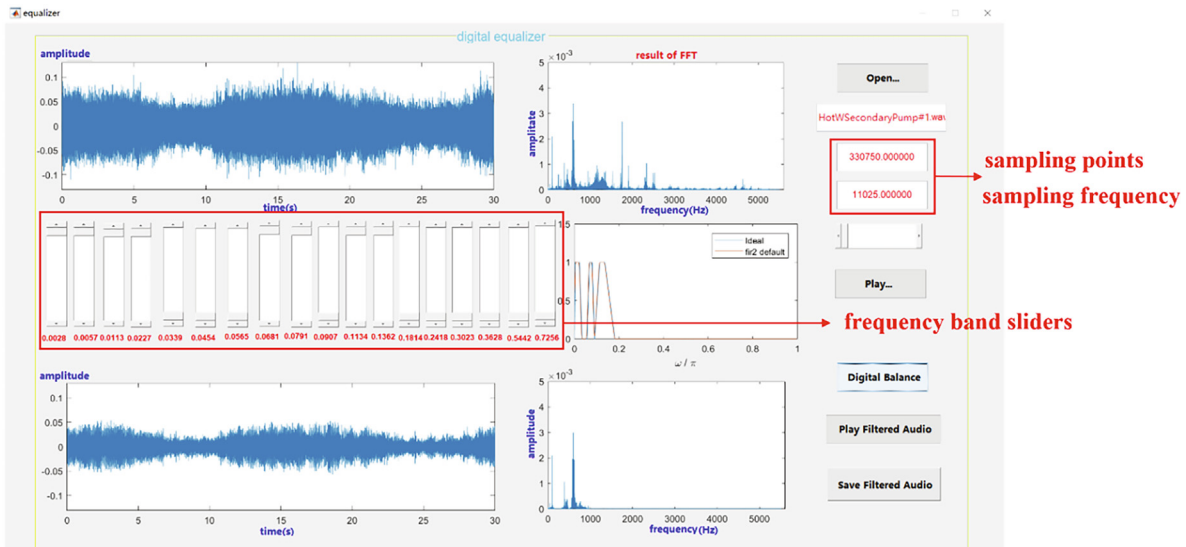
Fig. 12. Interface of the designed GUI application.

the LB statistic. Here, we set the significance level of the test at 1%. The LB statistics and p-value results of the training data for the signals in the three frequency domains are presented in Table 1.

For low-frequency signals, the p-values of the training data LB statistic are far less than 0.01 after order 3; for medium-frequency signals, the p-values of the training data LB statistic are far less than 0.01 after order 2; for high-frequency signals, all the p-values of the training data LB statistic are far less than 0.01 after order 1. The statistical test results of the signals in all the three frequency domains indicate that the p-value tends to decrease as the number of delay periods increases. The results of the above analysis indicate that the training data are a non-pure random smooth series, which is valuable for time series analysis. The ARIMA model was then used to analyze the time series.

The three parameters *d, p,* and *q* are crucial in determining the ARIMA model. By observing the ACF and PACF plots in Fig. 14, it can be observed that the data have a weak short-term correlation, which is barely more than a 5th order delay, and thus the ranges of *p* and *q* are not too wide for a reasonable model. Therefore, the grid search selects a range of 0–5 for *p* and 0–5 for *q*. According to the BIC criterion, the ultimate optimal model should be the model that meets the BIC value minimization in the grid search. The final grid search results are listed in Table 2.

Finally, the model and parameters were tested for significance. The ACF and PACF plots of the residuals after fitting the ARMA models with the signal training data of the three frequency domains are shown in Fig. 15.

The residuals of the models were then tested for pure randomness based on the LB statistic, and the results are presented in Table 3. The LB statistic of the residuals of the model and their p-values are much greater than 0.01, and therefore the original hypothesis H_0 that the residuals of the low-frequency signal fitting model are purely random sequences cannot be rejected.

The fit statistics of the ARMA model for the low-frequency domain signal training data are listed in Table 4. From the column of $P > |t|$, except for the first parameter of *q*, which almost satisfies the 5% confidence level significance test, all the parameters pass the significance test; thus, they can be considered as valuable parameters, and the selection of the whole parameter set is reasonable. In a similar way, in the fit statistic (Table 4) of the ARMA model for the medium-frequency domain signal training data, except for the first and third parameter of *q*, which are some way

from meeting the 5% confidence level significance test, all the parameters can be considered as valuable parameters. In the fit statistic (Table 4) of the ARMA model for the high-frequency domain signal training data, all five parameters passed the significance test and could be considered valuable parameters, and the entire set of parameters was reasonably selected.

Based on the final fit (Figure 16), the upper and lower 95% confidence intervals of the fitted data were used as thresholds to detect whether the signal was unusual due to the high uncertainty of the signal itself. The threshold determines the range limit for the fluctuation of the normal audio signal, which can detect some abnormal sounds such as whistling and other sharp noises[40]. Fig. 17 shows the result of backward prediction of data of equal length using the ARMA model. For other pumps, the audio signals can also be divided into three frequency domains to establish the ARIMA model using thresholds to determine whether the audio signal is unusual or not.

### 3.5. Equipment identification

According to **2.2.4 Equipment Images Classification**, AlexNet is retained using Pytorch. The training set contains 403 pictures of equipment in the chiller room, including horizontal pumps, vertical pumps, dials, and compressors, and it is verified with 20 homologous pictures. The model was trained on the GPU for 20 rounds, and all the images were trained by the model in each round. The initial learning rate λ was set to 0.005, the learning rate decay was set to 10% every 5 rounds, the weight decay coefficient was set to 0.0001, and the momentum parameter was set to 0.9. The final training and validation losses were both reduced to zero, and the training and validation accuracy reached 100%. The results of the model classification of new images are shown in Fig. 18.

### 3.6. Temperature and pressure monitoring of pipelines

When the robot recognizes the photo of the dial during inspection, it can read the dial indicator using the corresponding algorithm(**2.2.5 Dial Indicator Reading**). Fig. 17 shows the process of reading the dial indication by processing the dial image with the proposed algorithm. From the final results, the result of the automatic reading is similar to that of the manual reading. Different pressure and temperature thresholds are set depending on the pipe
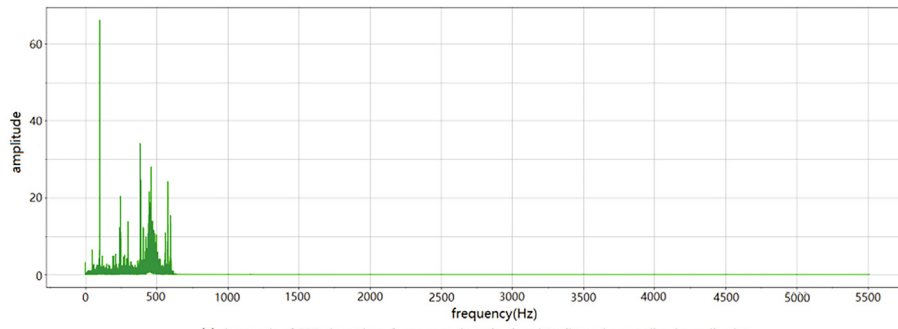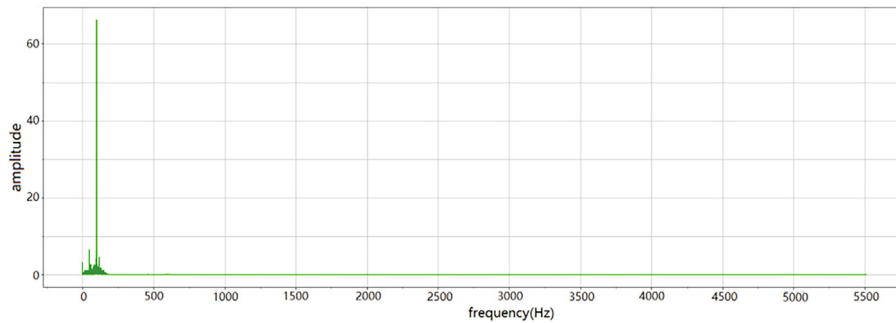
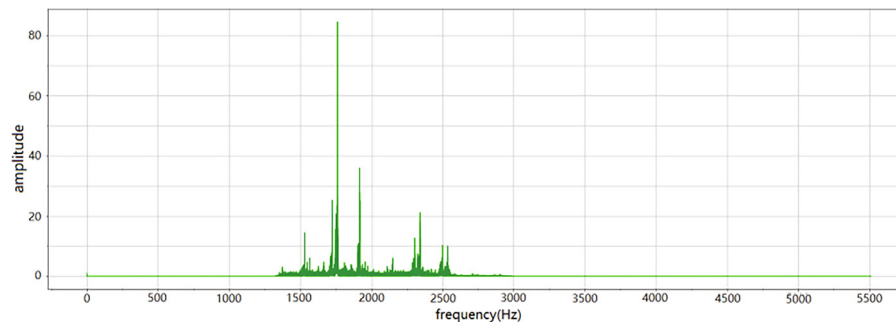(a) the result of FFT about low-frequency domain signal (unilateral,normalized amplitude)
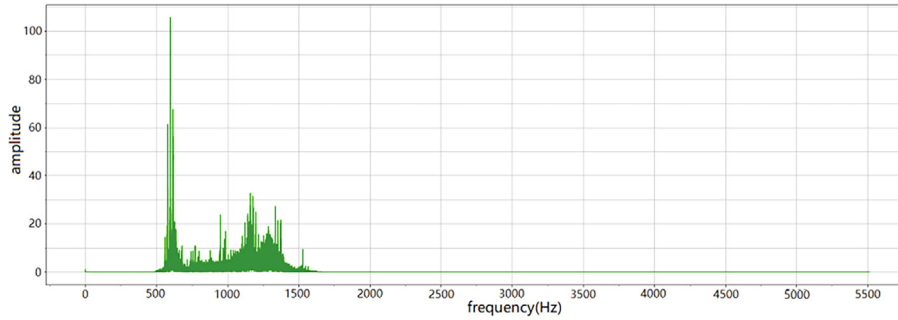
(b) the result of FFT about medium-frequency domain signal (unilateral,normalized amplitude)
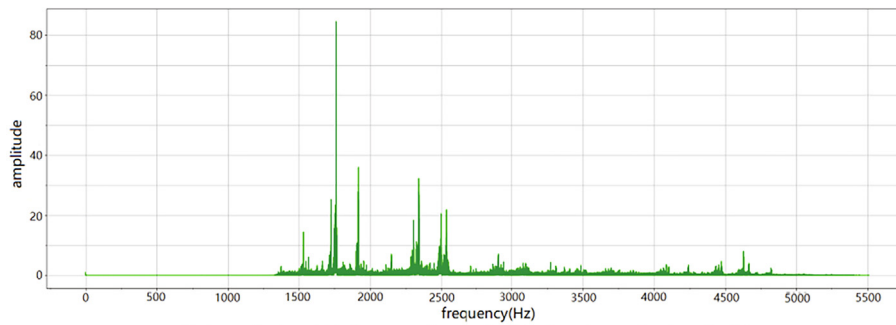
(c) the result about high-frequency domain signal (unilateral,normalized amplitude)

(d) the result of FFT about low-frequency peak amplitude signal (unilateral,normalized amplitude)

(e) the result of FFT about high-frequency peak amplitude signal (unilateral,normalized amplitude)
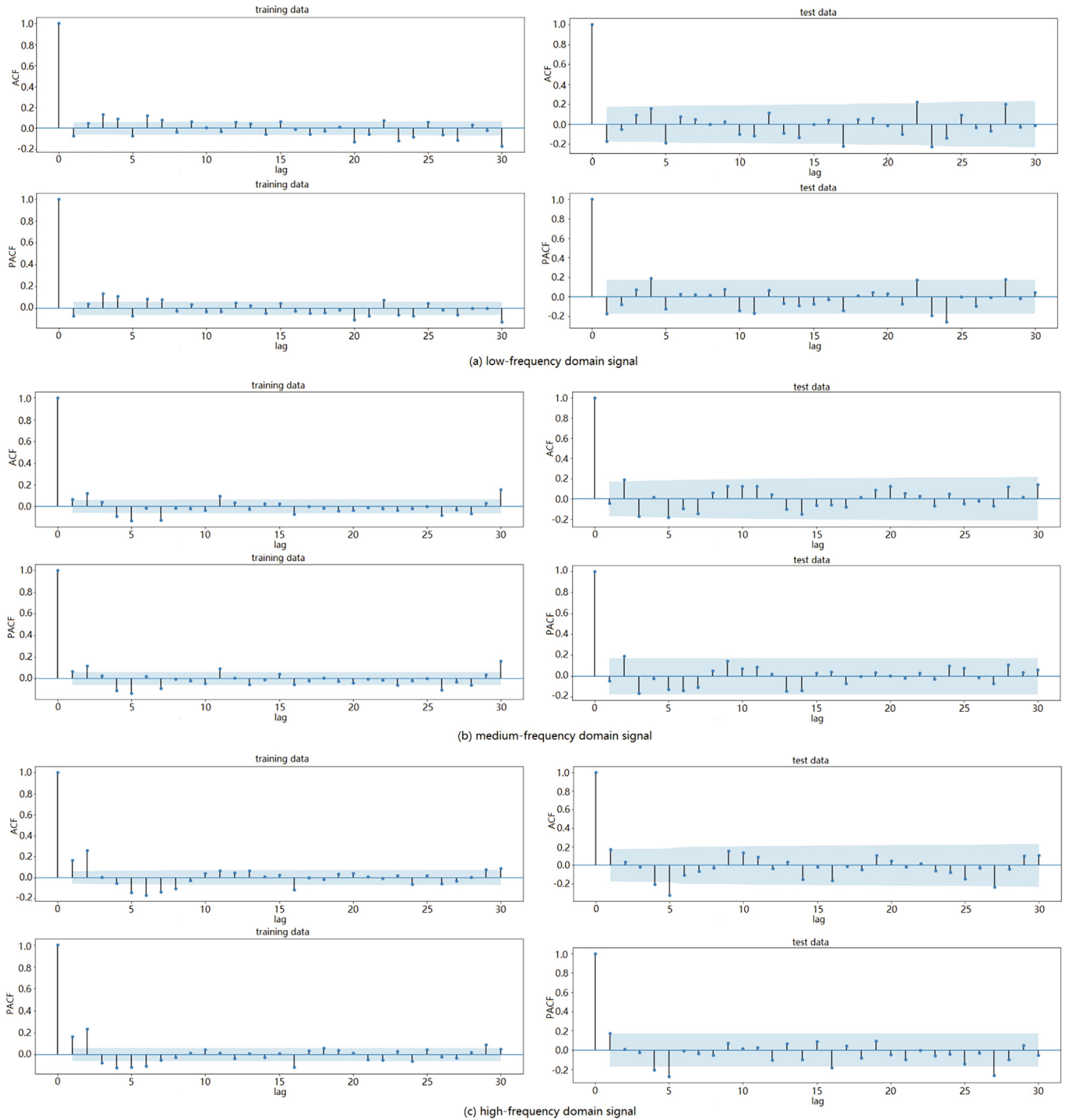
Fig. 14. ACF and PACF plots of three different frequency domain signals. ACF and PACF plots of training data in the three frequency domains for 1–25 s are shown on the left and the right side shows ACF and PACF plots of the test data in the three frequency domains for 25–28 s, which shows that both the training and test data are not highly correlated after a 1st order delay.

type to indicate whether the medium in the pipe is in a normal state or not.

When the robot recognizes the photo of the dial during inspection, it can read the dial indicator using the corresponding algo-

rithm. Fig. 19 shows the process of reading the dial indication by processing the dial image with the proposed algorithm. From the final results, the result of the automatic reading is similar to that of the manual reading. Different pressure and temperature thresh-

Fig. 13. Results of filtering. (a) shows the 0–535 Hz (low frequency band) frequency domain characteristics, (b) shows the 625–1333 Hz (medium frequency band) frequency domain characteristics, and (c) shows the 1666–4000 Hz (high frequency band) frequency domain characteristics. It can be seen that the peak amplitudes in the three frequency domains are comparable. (d) shows the peak amplitude in the low-frequency domain in the 0–125 Hz band. (e).

Table 1
LB statistics and p-value results of the training data for the signals.

	Low-frequency domain		Medium-frequency domain		High-frequency domain	
	LB-value	p-value	LB-value	p-value	LB-value	p-value
1	5.147	2.328e−02	4.700	3.016e−02	29.857	4.652e−08
2	7.537	2.308e−02	20.876	2.930e−05	100.000	1.929e−22
3	25.847	1.026e−05	22.582	4.935e−05	100.022	1.537e−21
4	34.719	5.307e−07	31.515	2.403e−06	103.381	1.874e−21
5	39.983	1.505e−07	50.251	1.231e−09	124.730	3.122e−25
6	56.141	2.726e−10	50.503	3.726e−09	155.773	4.650e−31
7	63.348	3.226e−11	68.234	3.354e−12	176.775	9.343e−35
8	64.737	5.441e−11	68.438	1.005e−11	188.760	1.484e−36
9	69.035	2.353e−11	68.961	2.433e−11	189.553	5.107e−36
10	69.097	6.622e−11	70.490	3.565e−11	191.219	1.090e−35
11	69.979	1.233e−10	80.266	1.311e−12	195.603	6.070e−36
12	73.528	6.966e−11	81.427	2.204e−12	197.542	1.048e−35

Table 2
Ultimate optimal model for the three frequency domain signals.

	p	q	d	Result
Low-frequency signal	4	3	0	ARMA(4,3)
Medium-frequency signal	5	0	0	ARMA(5,0)
High-frequency signal	3	2	0	ARMA(3,2)

and skimage. The actual image processing flow and results are shown in Fig. 20. The proposed algorithm can accurately read the temperature on an infrared image. It is determined whether the pump is in an overheated state based on the identified maximum temperature of the pump, thus achieving the purpose of fault diagnosis. The alternating current windings of the water pump motors
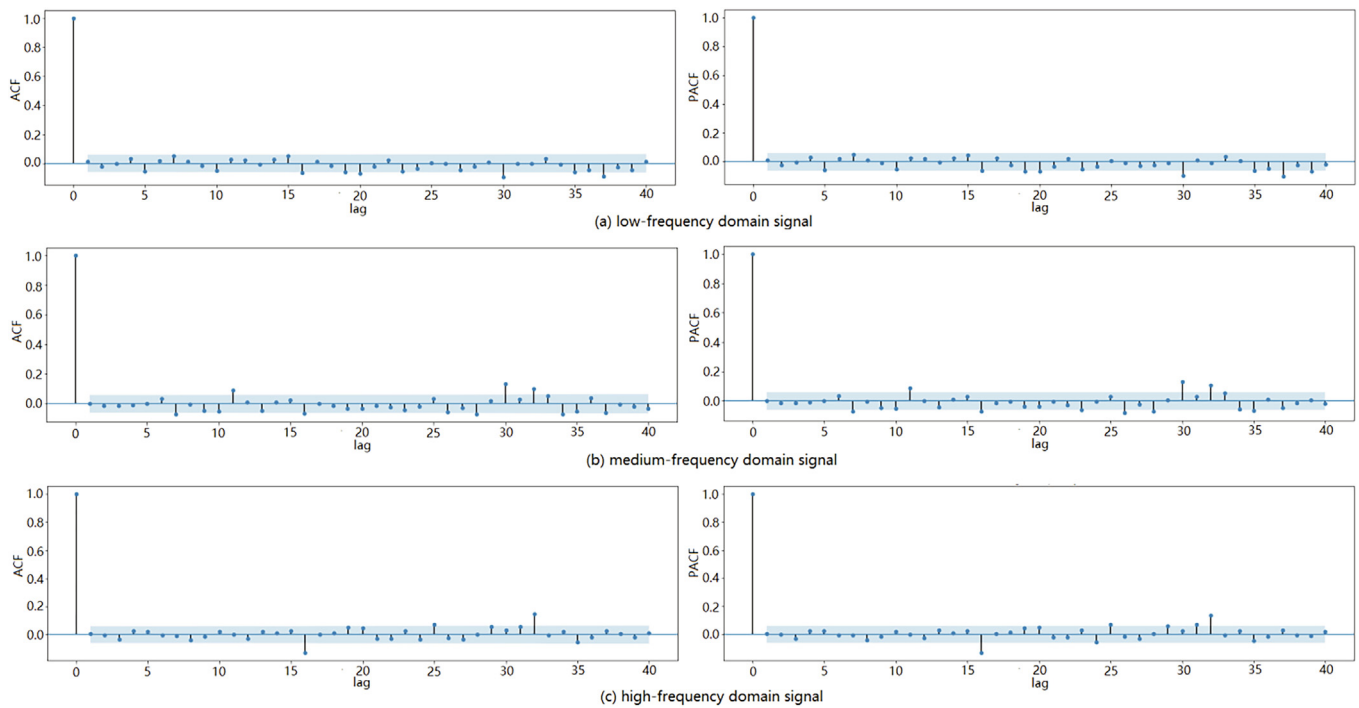
Fig. 15. ACF and PACF plots of the residuals after fitting. We can observe that the models are truncated in the first order. Because of the smooth time series with short-term correlation, it can be concluded that the residuals are white noise sequences, and the model extracts sufficient information.

olds are set depending on the pipe type to indicate whether the medium in the pipe is in a normal state or not. By reading the dial indicator, sensor fault and thermodynamic fault can be detected. For example, the temperature dial on the chilled water pipeline shows 30°C three times in a row(15 min interval).

3.7. Equipment temperature monitoring

During the robot inspection in the chiller room, the temperature reading of the infrared image of the pump(**2.2.6 Infrared Image Recognition**) is achieved with the help of Python's pytesseract

in chiller rooms are heat exchanged with air, and their temperature limit is 110°C[41]. The proposed non-intrusive approach cannot measure the maximum temperature of the internal part of a motor. The temperatures of the water pumps in the experimental chiller room are not more than 65°C. In order to reduce false alarm, the temperature threshold in the room is set to 70°C.

4. Discussion and conclusion

This study developed an integrated audio and image-based non-intrusive fault detection and diagnosis algorithm and per-

Table 3
LB statistic and p-value for the residuals of the model.

	Low frequency		Medium frequency		High frequency	
	LB-value	p-value	LB-value	p-value	LB-value	p-value
1	0.133	0.715	0.005	0.941	0.039	0.844
2	0.700	0.704	0.161	0.922	0.040	0.980
3	0.707	0.872	0.374	0.946	1.120	0.772
4	1.791	0.774	0.423	0.980	1.788	0.774
5	5.231	0.388	0.423	0.995	2.397	0.792
6	5.618	0.467	1891	0.929	2.409	0.878
7	8.397	0.299	7.204	0.408	2.502	0.927
8	8.625	0.375	7.208	0.514	4.180	0.840
9	8.880	0.448	9.331	0.407	4.364	0.886
10	11.454	0.323	12.000	0.285	4.786	0.905
11	12.264	0.344	20.952	0.034	4.793	0.940
12	12.732	0.389	21.040	0.050	5.633	0.933

Table 4
Statistical results of the ARMA model for three different frequency domains.

Low-frequency ARMA (4,3)		Medium-frequency ARMA (5,0)		High-frequency ARMA (3,2)	
Parameter	P > |t|	parameter	P > |t|	parameter	P > |t|
ar.L1.y	0.0072	ar.L1.y	0.1535	ar.L1.y	0.0000
ar.L2.y	0.0000	ar.L2.y	0.0000	ar.L2.y	0.0445
ar.L3.y	0.0000	ar.L3.y	0.1032	ar.L3.y	0.0000
ar.L4.y	0.0000	ar.L4.y	0.0007	ma.L1.y	0.0000
ma.L1.y	0.0671	ar.L5.y	0.0000	ma.L2.y	0.0002
ma.L2.y	0.0000				
ma.L3.y	0.0000				

formed experiments in a LANDSEA chiller room to verify the validity of the proposed algorithm. Based on the proposed non-intrusive method, it is possible to verify the real-time pressure and temperature of pipelines to prevent accidents, and verify whether the working temperatures of chillers, pumps, and other equipment are normal and whether the sound emitted from pumps is abnormal. This makes it possible to schedule reasonably the working hours of operation and maintenance personnel.

In this study, based on audio, the ARIMA model and threshold method are used to determine whether the pump is in normal operation. When the abnormal audio is characterized by particularly small amplitudes in certain frequency bands, the proposed approach is difficult to detect the fault. Audio signals acquired by non-intrusive methods had complex frequency domain characteristics and indicated that they were smooth signals with strong uncertainty when pump is operating normally. In the future, we plan to collect more audio signals of pumps and using NN or other methods to analyze the audio signals in order to improve the current *audio-based diagnostic algorithm*.

Based on image, AlexNet and migration learning ideas are used to make a classifier of equipment images in the chiller room. The dial indication is read with the help of morphological methods and ideas, and OCR helps read indications in infrared images. The experiments show that the results are ideal. However, the mounting angle of some dials might have some deviation, which may cause the result to be inaccurate. There are certain requirements for the input dial image: 1) the images need to be clear to the naked eyes. 2) the dial image need to have an almost completely dark background. Moreover, the parameters in *Dial Indicator Reading* need to be adjusted for dials with different ranges beforehand,

and the current reading algorithm is not flexible enough. Continue to improve this algorithm in the future so that it can be self-adaptive to dials of different ranges and pictures of dials with different backgrounds.

The proposed FDD approach mainly detects the mechanical faults in equipment and thermodynamic faults in pipelines. If the sensors (camera modules or microphone) do not work well, the proposed approach will not applicable.

5. Prospects

The methods proposed in this study were mainly designed for pumps and dials in chiller rooms, and audio and image-based FDD approaches for chiller compressors are to be further developed. Currently the approach is completely independent of the existing BASs, we will later consider combining the unstructured data(images and audio signals) with the structured data collected by the existing BASs.

In the future, we plan to fuse the proposed algorithms with the inspection robot to form an automated detection and diagnosis framework. As shown in Fig. 21, we intend to embed the non-intrusive algorithm into the inspection robot, with the sound and image sensors installed on the robot. Thus, when the robot inspects a chiller room, it can cooperate with the image sensor and the equipment image classification to identify the type of equipment in the pre-set location. Then, it can call different sensors according to the type of equipment to collect data, detect the operation of the equipment, and report to the inspection personnel if there are any abnormalities.
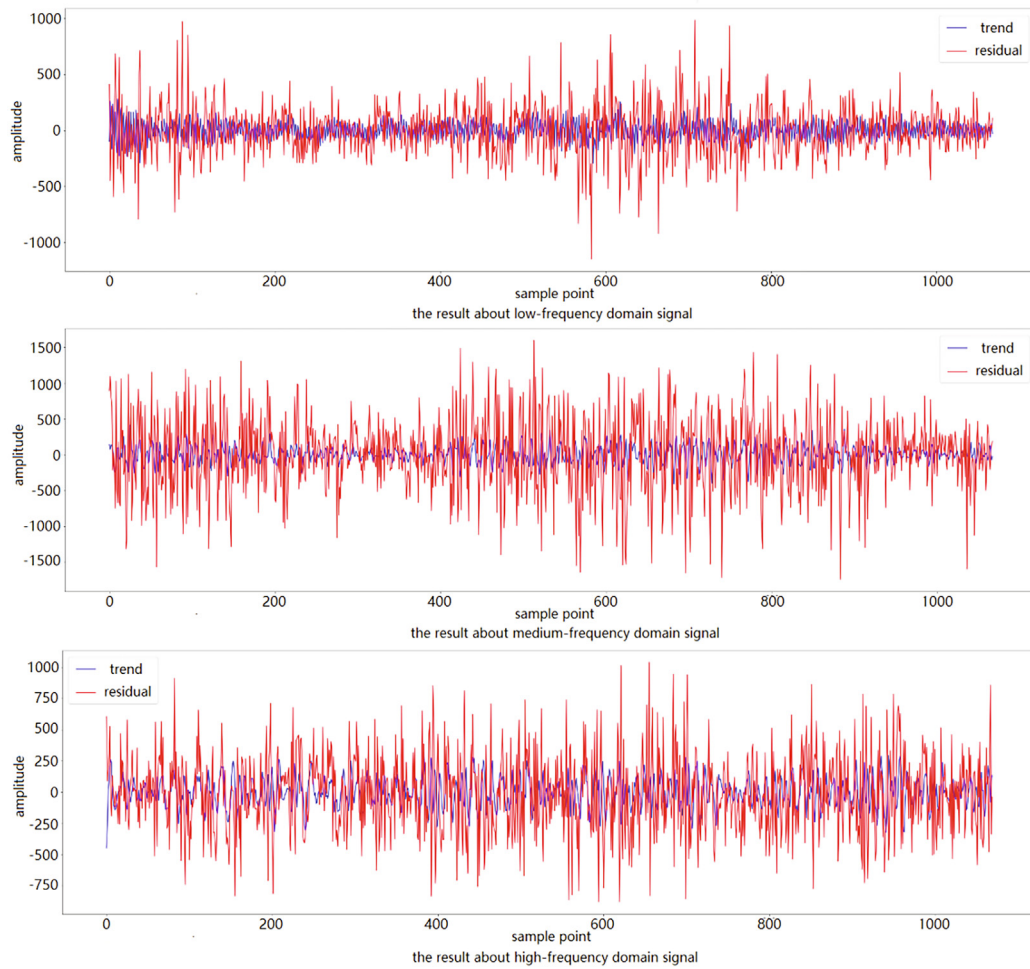
Fig. 16. Results of ARMA fitting of signals in three frequency domains. The final fit showed that the fitted trend is weaker than the residuals, and in combination with the fact that the residuals have been confirmed to be white noise sequences, this indicates a strong uncertainty in the three signal components.
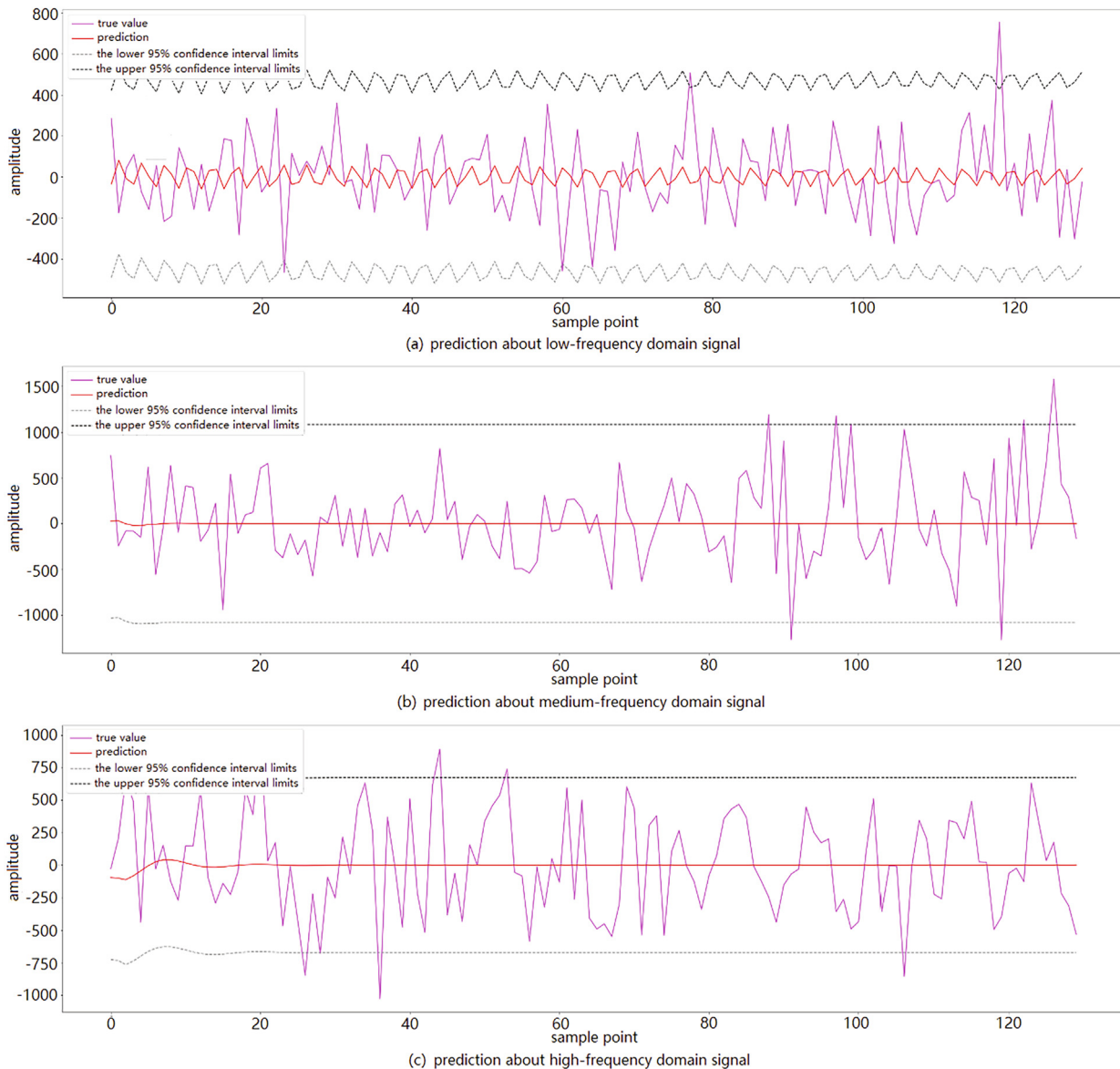
Fig. 17. Prediction for three frequency domain signals. We set the upper and lower thresholds to limit the range of normal signal fluctuations. If the sound made by the pump falls in the confidence interval, that is the sound made by the pump during the normal operation. If the sound emitted by the pump is out of the threshold, the pump is whistling and not operating properly.

Category result: horizontal pump   category result: compressor   category result: dial

Fig. 18. Trained model classification results.
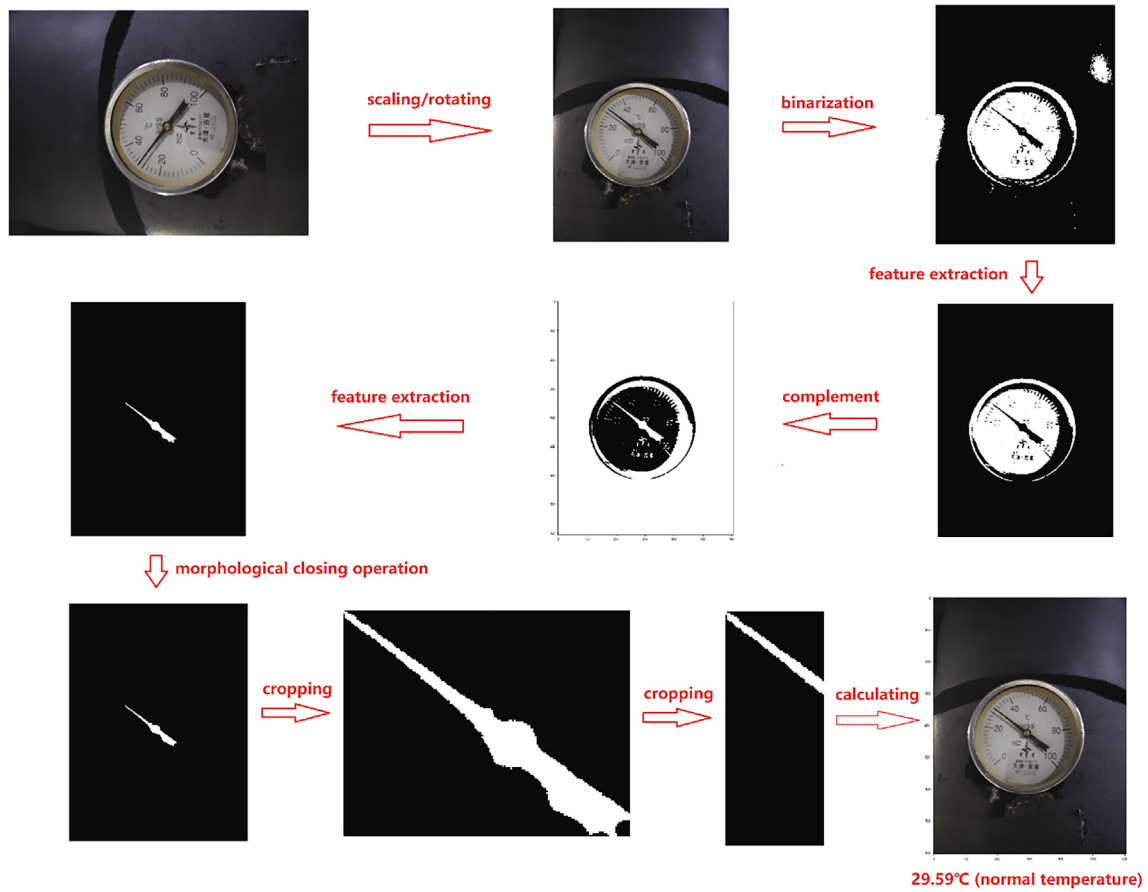
Fig. 19. Process of reading the dial indication.

Fig. 20. Actual image processing flow and results.

Fig. 21. Detection and diagnosis framework based on robotic inspections.
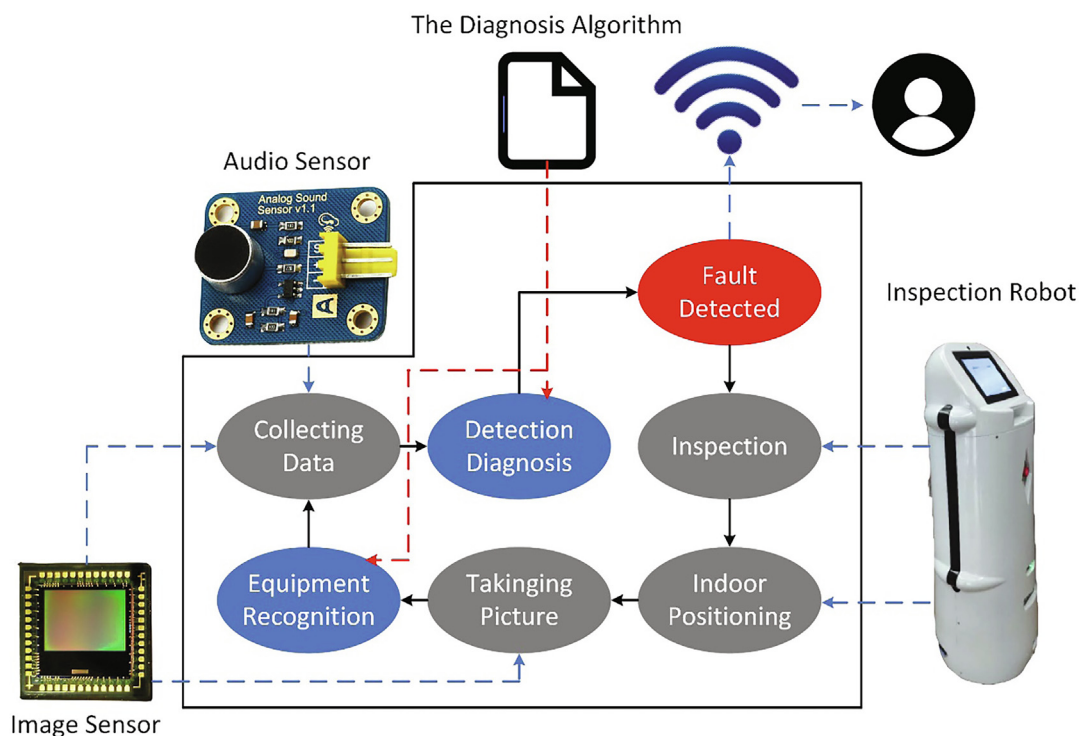
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

References

[1] Y. Yu, D. Woradechjumroen, D. Yu, A review of fault detection and diagnosis methodologies on air-handling units[J], Energy Build. 82 (2014) 550–562.

[2] L. Xing. Fault detection and diagnosis in low detla-T syndrome of chilled systems in large submetering commercial buildings[D]. Tongji University. 2017.

[3] N.Z. Zhao. Intelligent Building Automation Technology[M]. China Electric Power Press. 2009.

[4] S. Bagavathiappan, B.B. Lahiri, T. Saravanan, J. Philip, T. Jayakumar, Infrared thermography for condition monitoring – A review[J], Infrared Phys. Technol. 60 (2013) 35–55.

[5] I. Bandyopadhyay, P. Purkait, C. Koley, A combined image processing and Nearest Neighbor Algorithm tool for classification of incipient faults in induction motor drives[J], Comput. Electr. Eng. 54 (2016) 296–312.

[6] G. Wang, Y.u. Sun, J. Wang, Automatic image-based plant disease severity estimation using deep learning[J], Computat. Intell. Neurosci. 2017 (2017) 1–8.

[7] I.S.G.H. Alex Krizhevsky, Imagenet classification with deep convolutional neural networks[J], Advan. Neural Info. Process. Sys. 25 (2) (2012).

[8] A. Glowacz, Z. Glowacz, Diagnosis of the three-phase induction motor using thermal imaging[J], Infrared Phys. Technol. 81 (2017) 7–16.

[9] R. Vilela, J.C. Metrolho, J.C. Cardoso, Machine and industrial monitorization system by analysis of acoustic signatures[M], IEEE (2004:) 277–279.

[10] G. Iannace, G. Ciaburro, A. Trematerra, A. Heating, Ventilation, and air conditioning (HVAC) noise detection in open-plan offices using recursive partitioning, Buildings 8 (12) (2018) 169, https://doi.org/10.3390/buildings8120169.

[11] G. Iannace, G. Ciaburro, A. Trematerra, Fault diagnosis for UAV blades using artificial neural network, Robotics 8 (3) (2019) 59, https://doi.org/10.3390/robotics8030059.

[12] Y.M. Yang. Digital Signal Process[M]. China Machine Press. 2017.

[13] Y. Lei, M.J. Zuo, Z. He, Y. Zi, A multidimensional hybrid intelligent method for gear fault diagnosis[J], Expert Syst. Appl. 37 (2) (2010) 1419–1430.

[14] P.K. Kankar, S.C. Sharma, S.P. Harsha, Fault diagnosis of ball bearings using machine learning methods[J], Expert Syst. Appl. 38 (3) (2011) 1876–1886.

[15] N.R. Sakthivel, V. Sugumaran, B.B. Nair, Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump[J], Mech. Syst. Sig. Process. 24 (6) (2010) 1887–1906.

[16] Y. Pan, J. Chen, L. Guo, Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description[J], Mech. Syst. Sig. Process. 23 (3) (2009) 669–681.

[17] Y. Lei, Z. He, Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery[J], Expert Syst. Appl. 35 (4) (2008) 1593–1600.

[18] Y. Lei, Z. He, Y. Zi, X. Chen, New clustering algorithm-based fault diagnosis using compensation distance evaluation technique[J], Mech. Syst. Sig. Process. 22 (2) (2008) 419–435.

[19] W. Li, C.K. Mechefske, Detection of induction motor faults: A comparison of stator current, vibration and acoustic methods[J], J. Vib. Control 12 (2) (2006) 165–188.

[20] I.S. Koo, W.W. Kim, The development of reactor coolant pump vibration monitoring and a diagnostic system in the nuclear power plant[J], ISA Trans. 39 (3) (2000) 309–316.

[21] Y. Lei, Z. He, Y. Zi, Application of the EEMD method to rotor fault diagnosis of rotating machinery[J], Mech. Syst. Sig. Process. 23 (4) (2009) 1327–1338.

[22] Z.K. Peng, P.W. Tse, F.L. Chu, A comparison study of improved Hilbert-Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing[J], Mech. Syst. Sig. Process. 19 (5) (2005) 974–988.

[23] J. Rafiee, M.A. Rafiee, P.W. Tse, Application of mother wavelet functions for automatic gear and bearing fault diagnosis[J], Expert Syst. Appl. 37 (6) (2010) 4568–4579.

[24] N. Baydar, A. Ball, Detection of gear failures via vibration and acoustic signals using wavelet transform[J], Mech. Syst. Sig. Process. 17 (4) (2003) 787–804.

[25] G. Ciaburro, G. Iannace, Improving smart cities safety using sound events detection based on deep neural network algorithms[J], Informatics 7 (3) (2020) 23, https://doi.org/10.3390/informatics7030023.

[26] Y. Zhang, M. Martínez-García. Machine Hearing for Industrial Fault Diagnosis [M]. 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Hong Kong, Hong Kong. 2020. 849–854.

[27] M. Martínez-García, R. S. Kalawsky, T. Gordon, T. Smith, Q. Meng, F. Flemisch. Communication and Interaction With Semiautonomous Ground Vehicles by Force Control Steering[M]. IEEE Transactions on Cybernetics.

[28] M. Martínez-García, Y.u. Zhang, T. Gordon, Memory pattern identification for feedback tracking control in human-machine systems, Hum. Factors. 63 (2) (2021) 210–226.

[29] A. Glowacz, W. Glowacz, Z. Glowacz, J. Kozik, Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals[J], Measurement 113 (2018) 1–9.

[30] A. Glowacz, Fault diagnosis of single-phase induction motor based on acoustic signals[J], Mech. Syst. Sig. Process. 117 (2019) 65–80.

[31] M. Liu. Research of motor bearing fault diagnosis methods based on virbation signal[D]. Harbin Institute of Technology. 2016.

[32] Y. Li, The Research on Fault Diagnosis Methods of Motor Bearing Based on Data Driven[D], Shenyang Ligong University, 2017.

[33] Seed. ReSpeaker Mic Array v2.0[Z] https://wiki.seeedstudio.com/ReSpeaker_Mic_Array_v2.0/.

[34] Taobao.RER-USB13MAF-V7S camera module[Z] https://item.taobao.com/item.htm?id=591973602466.

[35] FLIR. FLIR A315[Z] https://www.flir.com/products/a315/.

[36] W. Yan, Applied Time Series Analysis[M], China Renmin University Press, 2012.

[37] J. Minichino, Learning OpenCV 3 computer Vision with Python[M], China Machine Press (2016).

[38] Jin YJ. Tongjixue[M]: China Renmin University Press, 2010.

[39] M. Guo, LiDAR Technology and Structure Analysis Method[M], China Surveying Mapping Press. (2017).

[40] Xinyu W. Motor Fault Diagnosis Based On Neural Network[D]. Shanghai Jiao Tong University. 2013.

[41] IEC 60034-1:2017. Rotating electrical machines Part 1: Rating and performance, IDT[S]. China Quality and Standards Publishing & Media Co., Ltd. 2019.