Dempster Shafer distance-based multi-classifier fusion method for pig cough recognition

Weizheng Shen[1], Xipeng Wang[1], Yanling Yin[1*], Nan Ji[1], Baisheng Dai[1], Shengli Kou[1], Chen Liang[2]

(1. *School of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China*;
2. *Heilongjiang Agricultural Technology Extension Station, Harbin 150036, China*)

Abstract: High precision pig cough recognition and low computational cost is of great importance for the realization of early warning of pig respiratory diseases. Numerous researchers have improved the recognition rate of pig cough sounds to a certain extent from feature selection and feature fusion perspectives. However, there is still a margin for the improvement in the accuracy and complexity of existing methods. Meanwhile, it is challenging to further enhance the precision of a single classifier. Therefore, this study proposed a multi-classifier fusion strategy based on Dempster Shafer distance (DS-distance) algorithm to increase the classification accuracy. Considering the engineering implementation, the machine learning with low computational complexity for fusion was chosen. First, three metrics of accuracy and diversity between classifiers were defined, including overall accuracy (OA), double fault (DF), and overall accuracy and double fault (OADF), for selecting the base classifiers. Subsequently, a two-step base classifier selection approach based on these metrics was proposed to make an optimized selection of features and classifiers. Finally, the proposed DS-distance algorithm was used to fuse the selected base classifiers to create a classification. The sound data collected in the pig barn verified the proposed algorithm. The experimental results revealed that the overall recognition accuracy of the proposed method could reach 98.76%, which was better than the existing methods. This study has achieved a high recognition accuracy through ensembled machine learning with low computational complexity. The proposed method provided an efficient way for the quick establishment of high precision pig cough recognition model in practice.

Keywords: pig cough recognition, classifier fusion, classifier selection, Dempster Shafer fusion, distance fusion
DOI: 10.25165/j.ijabe.20241704.8027

Citation: Shen W Z, Wang X P, Yin Y L, Ji N, Dai B S, Kou S L, et al. Dempster Shafer distance-based multi-classifier fusion method for pig cough recognition. Int J Agric & Biol Eng, 2024; 17(4): 245–254.

1 Introduction

Pig respiratory diseases have become one of the leading factors restricting the advancement of the pig breeding industry due to their high mortality rate and strong infectivity[1,2]. Therefore, a fast and an accurate early warning of respiratory diseases are urgently required. Recent studies have demonstrated that this can be achieved by monitoring the cough sounds of pigs, and a key step is to realize the high-precision pig cough sounds recognition[3,4]. Therefore, several researchers have paid a lot of effort into enhancing the recognition accuracy of pig cough sounds.

In early studies, some commonly used acoustic features and template matching or machine learning method were used for cough recognition[5-9]. Van Hirtum et al. extracted a feature vector, containing energy, time-derivate energy and mean power spectral density, and it was compared to the reference set using dynamic time warping (DTW), resulting in a correct classification of 90%[5]. Guarino et al. extracted feature vectors by using a filter bank approach combined with an amplitude demodulation and classified coughs by DTW. The cough classification rate was 85.5%[7]. Exadaktylos et al. used power spectral density (PSD) feature and fuzzy c-means (FCM) clustering method to make a classification, and the cough identification accuracy was 82%[8]. Chung et al. extracted mel-frequency cepstral coefficient (MFCC) and used the support vector data description (SVDD) and the sparse representation classifier (SRC) as the classifier. An accuracy of 91% was achieved[9]. Although the complexity of the machine learning model is low, the recognition accuracy of these studies needs to be further improved.

With the advancement of deep learning, numerous researchers have explored the performance of deep learning in pig cough sound recognition. Li et al.[10] designed a deep belief networks with an input feature vector of short time energy and MFCC. The cough accuracy achieved 95.08%. Yin et al. transferred learning the pretrained AlexNet[11] model and finetuned the model for pig cough recognition. The cough accuracy reached 96.8%[12]. Shen et al. explored the features of MFCC and logarithmic filter bank and the models of convolutional neural networks (CNNs) and deep feed forward sequential memory networks to make a classification. The best accuracy was 97% for cough recognition[13]. Shen et al. fused multi-frame MFCC with multiple single-layer CNNs, and a final classification accuracy of 97.7% was achieved[14]. Ji et al.[15] explored acoustic features and visual features for pig cough classification. In their research, the acoustic features including MFCC, root-mean-

Received date: 2022-11-08 **Accepted date:** 2024-03-02
Biographies: Weizheng Shen, PhD, Professor, research interest: intelligent animal husbandry, Email: wzshen@neau.edu.cn; **Xipeng Wang**, Master, research interest: intelligent monitoring of livestock diseases, Email: 1326870195@qq.com; **Nan Ji**, PhD candidate, research interest: intelligent monitoring livestock health, Email: jn740740@163.com; **Baisheng Dai**, PhD, Associate Professor, research interest: intelligent equipment production for animal, Email: bsdai@neau.edu.cn; **Shengli Kou**, PhD candidate, research interest: livestock health monitoring and equipment development, Email: koushengli@neau.edu.cn; **Chen Liang**, Bachelor, Deputy Senior Agronomist, research interest: artificial intelligent in agriculture, Email: 395225233@qq.com.

***Corresponding author: Yanling Yin**, PhD, Associate Professor, research interest: intelligent animal husbandry, College of Electrical and Information, Northeast Agricultural University, Harbin 150030, Heilongjiang, China. Tel: +86-18946022406, Email: yinyanling@neau.edu.cn.

square energy (RMS), zero-crossing rate (ZCR), spectral centroid, spectral rolloff, spectral flatness, spectral bandwidth and chroma, and the visual features including local binary pattern (LBP) and histogram of gradient (HOG) were fused for classification. The support vector machine (SVM) classifier was adopted and the best accuracy reached 96.45%. Shen et al.[16] fused the acoustic and deep features for pig cough sound recognition. The acoustic features included MFCC, RMS, ZCR, spectral centroid, spectral bandwidth, spectral rolloff, spectral contrast, and spectral flux. The deep features were extracted from the image features of short-time Fourier transform (STFT) and constant-Q transform (CQT). The best accuracy achieved 97.35% classified by the SVM classifier. The above researches enhanced the accuracy from the perspective of feature selection, feature fusion and classifier chosen. It can be seen that compared with traditional machine learning, the accuracy of pig cough recognition is significantly improved by deep learning. However, the complexity of the model is also significantly increased.

Motivated by the application of the algorithm in practical work, the study hopes to explore a high precision model with low complexity. At the same time, due to the difficulty in collecting and labeling pig cough samples, the obtained dataset is relatively small (compared with large public image and audio datasets). This study tried to realize pig cough recognition through a simple machine learning method. Classifier fusion is an effective means to improve the accuracy of the model[17,18]. Therefore, in this work, the classifier fusion model based on machine learning was studied to reduce the complexity of the model while enhancing the accuracy.

Classifier fusion collects prominent classification performance and generalization ability by combining the output outcomes of various base classifiers[19]. Classifier fusion primarily includes the following three types: single feature multi-classifier fusion, multi-feature single classifier fusion, and multi-feature multi-classifier fusion. The first two fusion methods are widely used in many studies[20-23]. Multi-feature multi-classifier fusion has significant potential in the classification. For multi-feature multi-classifier fusion, the common approach was to first select and fuse multiple features, and then input the fused feature vectors into different classifiers for fusion[24,25]. The same feature vector inputting into different classifiers may lead to a smaller output difference, thereby losing the fusion advantage. Some researchers integrated all the combinations of different features and classifiers[26], or selected a certain number of combinations by experience, such as taking the first few combinations with the highest accuracy[23,25]. Fusing all combinations is only applicable when the number of features and classifiers is small, and there may be redundancy of base classifiers. The empirical selection of partial combinations may result in the selection of a sub-optimal base classifier. Therefore, designing an efficient and reliable algorithm to select the optimal integration is essential for multi-classifier fusion.

This study proposed a two-step basis classifier selection algorithm based on the metrics of overall accuracy (OA), double fault (DF)[27], and overall accuracy and double fault (OADF). Meanwhile, a Dempster Shafer distance (DS-distance) multi-classifier fusion algorithm was proposed to enhance the DS fusion. In this study, four acoustic features and three classifiers are used for classifier fusion, where the acoustic features include MFCC, linear prediction cepstral coefficient (LPCC), gammatone cepstral coefficient (GTCC) and PSD, and the classifiers includes SVM, *k*-nearest neighbor (KNN) and random forest (RF). The primary contributions of this study can be summarized as follows:

1) A two-step base classifier selection method based on the indicators of OA, DF and OADF was proposed to select the optimized base classifiers.

2) A DS–distance classifier fusion algorithm was proposed to reduce the classification error near the decision boundary in the DS fusion method.

2 Materials and methods

2.1 Materials

2.1.1 Animals and housing

The data were obtained in a big fattening pig barn in Harbin, Heilongjiang Province, China, in April 2018. The pig barn was 27.5 m×12.8 m×3.2 m (length×width×height), which comprised 21 pens, as depicted in Figure 1. Among them, the size of pens 1–12 was 4.15 m×3.6 m (length×width), and the size of pens 13–21 was 3.6 m×2.75 m (length × width). There were two electric fans with a diameter of 1.35 m in the pig barn, and they were opened every two hours. Each opening lasted for 15 min, and there was a considerable noise when they were opened. The employees used shovels and water to scrub the half-slatted concrete floor of the pig house twice a day. Furthermore, there were 128 pigs scattered in 1–13 pens in the pig barn. Two pigs with heavy coughs were split into pen 13 near the door. The coughs had appeared several days prior, and several coughing pigs were in each pen when the data was obtained.

Figure 2 shows the layout of the piggery environment and data

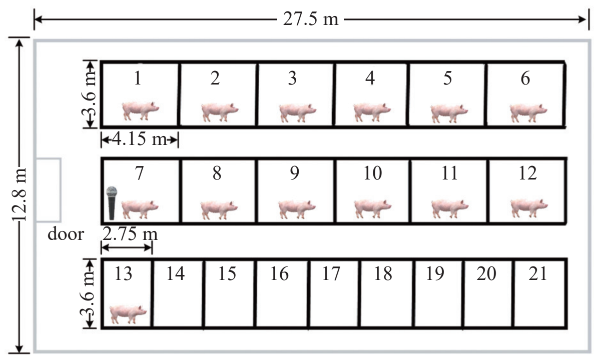
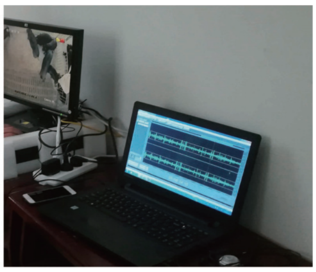


Figure 1 The layout of the pig house in this study

a. Location of the microphone and the piggery environment



b. Display and preservation of the sound signal by Cool Edit software

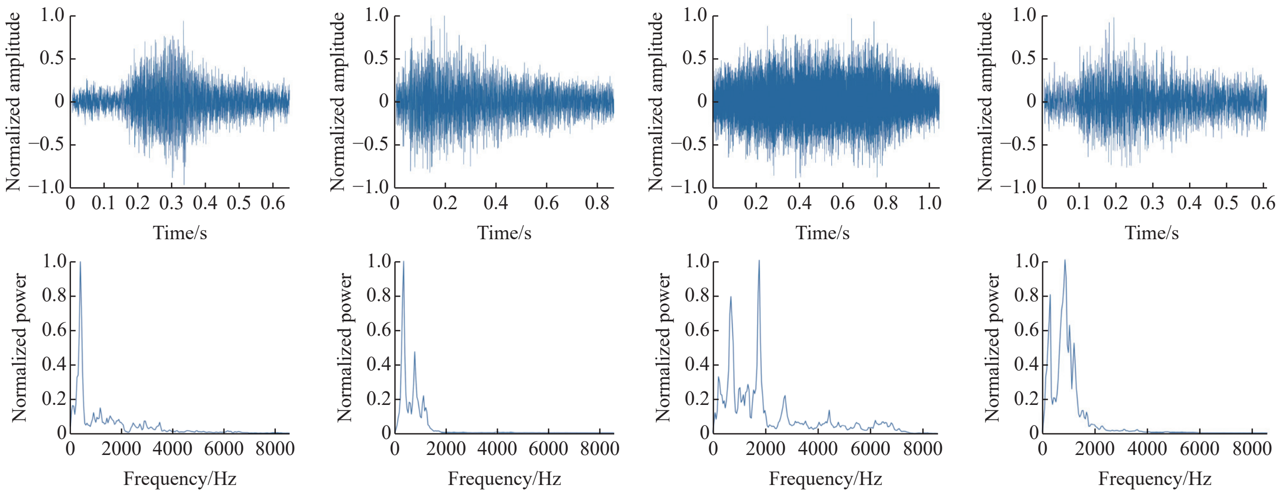
Figure 2 Experimental device layout and piggery environment collection equipment. Figure 2a shows the location of the microphone and the piggery environment. The photo was taken in front of pen 7. Figure 2b shows the display and preservation of the sound signal by Cool Edit software (Syntrillium, CA, USA).

2.1.2 Data collection

The audio acquisition device was a cardioid electret microphone (Enping Yilianda Electronic Factory, Guangdong, China) (LIQI LM320E, frequency range 50 Hz–16 kHz) and a laptop with the Conexant Smart Audio HD sound card (Conexant Systems Inc, CA, USA). Limited by the experimental conditions, the microphone was placed over pen 7, approximately 1.4 m above the ground. The audio sampling rate was 44.1 kHz and the sampling accuracy was 16 bits. The sounds were continuously recorded, and the collected audio data was saved in wave format. The continuous sounds were segmented and labelled manually with the assistance of veterinarians. The labelled individual sounds included pig coughs, groan, screams, cleaning sounds, and people's talks, among others. The screams were usually caused by the biting and fighting behaviors of pigs. The cleaning sounds were produced by workers using shovels to remove manure, including knocking sound, friction sound, etc. There were also other noises, such as fan noise, feeding noise and flushing noise. This study randomly selected 1250 cough and 1250 noncough sounds as this work's training and testing dataset.

2.1.3 Data description

The waveforms and the power spectrums of a cough and some typical noncoughs (groan, knocking, screaming) are shown in Figure 3. To better characterize the sounds, a descriptive statistic was done over the coughs and noncoughs. In this analysis the duration of single sounds and the peak frequency have been considered[28,29]. The durations and peak frequencies of cough and noncough sounds are shown in Table 1 and Figure 4. The maximum value (Max), minimum value (Min), mean value (Mean), and standard deviation (S.D.) were analyzed. The "noncough" in Table 1 and Figure 4 includes groan, scream, knocking and other noncough sounds. The sounds were filtered by a bandpass filter in a range of 100-16 000 Hz.



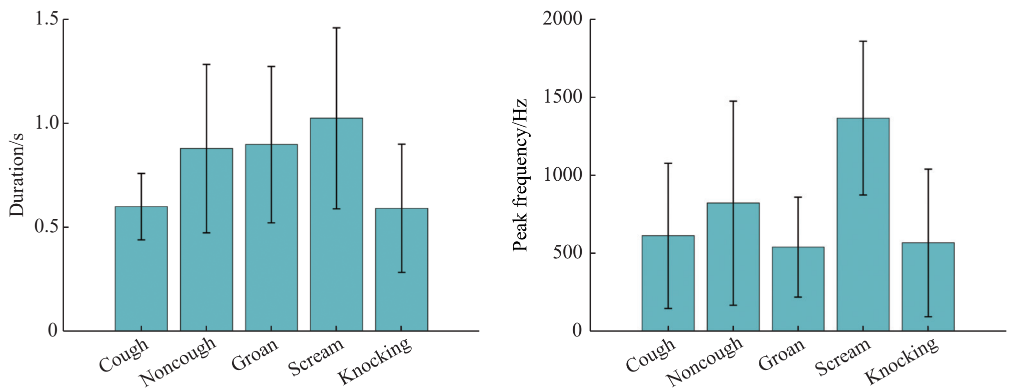
a. Cough b. Croan c. Scream d. Knocking

Figure 3 The waveforms (above) and power spectrum (below) of a cough and some typical noncoughs

Table 1 Descriptive statistic of cough and noncough sounds

Type of sounds	Duration				Peak frequency			
	Min/s	Max/s	Mean/s	S.D.	Min/Hz	Max/Hz	Mean/Hz	S.D.
Cough	0.22	1.09	0.60	0.16	139.97	4478.91	610.61	460.26
Noncough	0.19	1.79	0.88	0.40	139.97	6933.69	823.21	654.09
Groan	0.20	1.63	0.90	0.38	139.97	1453.49	541.17	319.58
Scream	0.26	1.76	1.02	0.44	215.33	3090.01	1368.35	494.11
Knocking	0.19	1.62	0.59	0.31	150.73	2971.58	569.09	473.17

According to the statistical results, the duration of cough was significantly different from that of groan and scream ($p < 0.0001$), but it was highly similar to that of knocking ($p > 0.05$). The peak frequency of cough was significantly different from that of scream ($p < 0.0001$), but it was highly similar to that of groan and knocking ($p > 0.05$). The "noncough" in Table 1 contains different types of sounds. Therefore, the statistical values show a big range



Note: Vertical lines indicate the standard deviations for each type

Figure 4 Barplot of the duration and peak frequency in cough and noncough sounds and S.D. Overall, there are some differences between coughs and some type of noncoughs, meanwhile, there is also partial overlap in duration and peak frequency for some sounds.

2.2 Proposed methods

This section first described the entire structure of the proposed method. The features used in the study were subsequently concisely introduced. Next, the two-step base classifier selection algorithm was described. Lastly, the proposed DS–distance fusion strategy was demonstrated.

2.2.1 Overall structure

The flowchart of the proposed algorithm is depicted in Figure 5. First, the sound signals were preprocessed. The acoustic features were extracted from the preprocessed sound signals. The acoustic features included MFCC, LPCC, GTCC, and PSD. Subsequently, the acoustic features were input into SVM, KNN, and RF classifiers. 12 base classifiers were constructed. Next, filtered the number of base classifiers to a small number. A two-step base classifier selection method was suggested to filter the classifiers. In the first step, OA and DF indicators were used to select the base classifier. The OADF indicator was used in the second step to choose classifier combinations. Finally, the DS–distance fusion algorithm fused the filtered base classifiers to obtain the classification results.

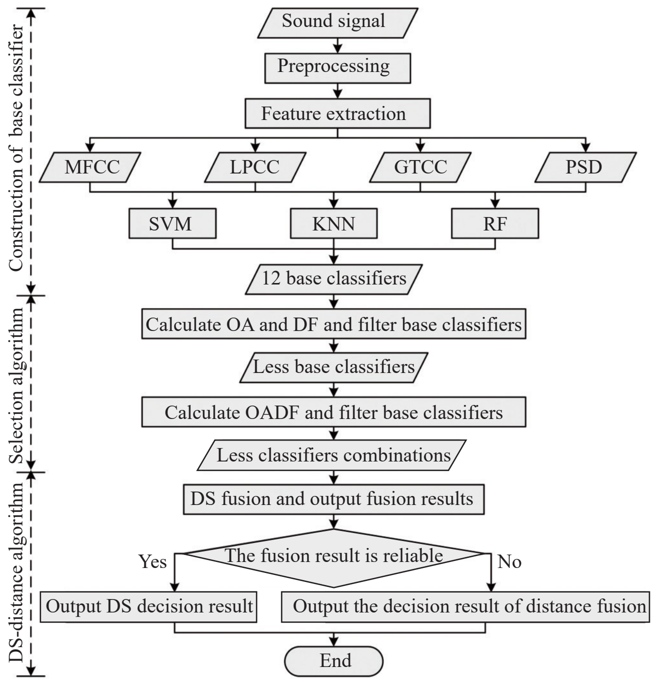


Figure 5 Algorithm flowchart proposed in this study

2.2.2 Preprocessing

Data preprocessing includes filtering, pre-emphasis, framing, and windowing. Among them, the role of filtering is to reduce the interference of out-of-band noises. The role of pre-emphasis is to increase the high frequency component in the sound signal and compensate the loss of the high frequency component in the transmission process. Framing is to convert non-stationary sounds into short-time stationary signals for analysis. The role of windowing is to reduce spectrum leakage.

2.2.3 Feature extraction

The features used in this study include MFCC, LPCC, GTCC and PSD. The sound was preprocessed first, then the features of each frame was extracted, and finally the mean value of all frames was calculated to obtain a feature vector for subsequent classification.

1) MFCC

MFCC is a frequently used acoustic feature in the field of sound signal processing. It also performs well in pig cough sound recognition[10,13,14]. MFCC is designed based on the human auditory mechanism, which is transformed to the perception frequency domain. The MFCC extraction process is shown in Figure 6. The preprocessing of the audio signal began with pre-emphasis, framing, and windowing. Subsequently, the preprocessed signal was converted to a frequency domain representation by applying the fast Fourier transform. Next, calculated the power of each frame and pass them via a set of mel-filter banks. Finally, the logarithm and discrete cosine transformation was performed to obtain the MFCC.

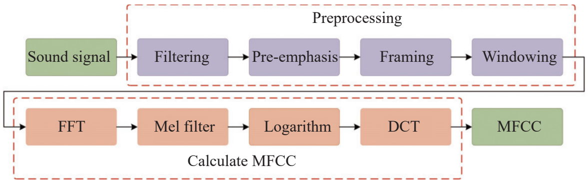


Figure 6 MFCC extraction process

2) LPCC

The fundamental idea of linear predictive analysis is that any sampling point of a sound signal can be represented by a linear combination of past sampling points. LPCC is the representation of Linear Prediction Coefficient (LPC) in spectrum domain. Let the current sample is x_n, the past p samples were used to predict x_n, then:

$$x_n = \sum_{i=1}^{p} a_i x_{n-1} \tag{1}$$

where, a_i is the LPC coefficient. LPCC can be obtained as follows:

$$\hat{h}(n) = \begin{cases} a_n, \quad n = 1 \\ a_n + \sum_{n=1}^{n-1} \left(1 - \frac{i}{n}\right) a_i \hat{h}(n-1), \quad 1 < n \le p \\ \sum_{n=1}^{n-1} \left(1 - \frac{i}{n}\right) a_i \hat{h}(n-1), \quad n > p \end{cases} \tag{2}$$

3) GTCC

The extraction process of GTCC is similar to that of MFCC, which is shown in Figure 7. The difference is that the gammatone-filter bank replaces the mel-filter bank. Gammatone-filter banks are designed based on the frequency decomposition of the cochlea. Several researchers have used gammatone-filter banks rather than traditional mel-filter banks to identify sound signals and achieved good outcomes[30-32].

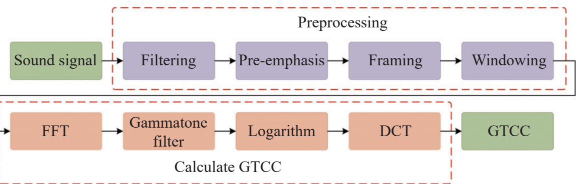


Figure 7 GTCC extraction process

4) PSD

PSD reflects the relationship between the power of the sound signal and the frequency. PSD extraction methods include an autocorrelation function, periodogram, and an average periodogram. The average periodogram method is the most commonly used, which divides the sound signal into multiple segments, calculates the PSD in each segment using the periodogram method, and finally averages the PSD over all segments. Let L is the length of each segmented signal, then the PSD of the i^{th} signal $x_i(n)$ can be obtained by the following formula:

$$P(\omega)=\frac{1}{L}|fft(window(x_i(n)))|^2 \tag{3}$$

where, window and *fft* represent windowing and fast Fourier transform on the signal, respectively.

2.2.4 Two-step base classifier selection method based on heterogeneous classifiers metrics

There are twelve various combinations of four features and three classifiers. Therefore, choosing an excellent and different base classifier is crucial for classifier fusion. That is, the base classifier should have high accuracy, and there should be some disparity between them[19,27,33]. To this end, the accuracy and diversity indicators of the base classifier was defined to examine and filter the base classifier. The indicators include OA, DF, and the combination of these two metrics: OADF. These indicators were used to filter the base classifiers.

NA is suppose to be the total number of cough and noncough samples included in the classification; NR_i is the number of samples correctly classified by the i^{th} base classifier; NF_{ij} is the number of samples incorrectly classified by both i^{th} and j^{th} base classifiers.

The OA of the i^{th} base classifier is defined as the following:

$$\mathrm{OA}_i=\frac{\mathrm{NR}_i}{\mathrm{NA}} \tag{4}$$

OA is the ratio of the number of correctly classified samples to the total number of samples, and its value range is [0,1]. The larger the OA, the better the performance of the base classifier. Meanwhile, DF_{ij} between the i^{th} base classifier and the j^{th} base classifier is defined as follows:

$$\mathrm{DF}_{ij}=\frac{\mathrm{NF}_{ij}}{\mathrm{NA}} \tag{5}$$

DF is the proportion of samples that are misclassified at the same time between the two base classifiers, and its value range is [0,1]. The smaller the DF, the more significant the difference between the two base classifiers.

The OADF_{ij} between the i^{th} base classifier and the j^{th} base classifier is defined as follows:

$$\mathrm{OADF}_{ij}=\frac{\mathrm{OA}_i+\mathrm{OA}_j}{2}-\mathrm{DF}_{ij} \tag{6}$$

For a system consisting of N base classifiers, the OADF is defined as follows:

$$\mathrm{OADF}=\frac{1}{N}\sum_{i=1}^{N}\mathrm{OA}_i-\frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{DF}_{ij} \tag{7}$$

OADF systematically considers the accuracy and disparity between the base classifiers. Theoretically, the larger the OADF, the better the performance of the base classifier after fusion.

The base classifier was defined as inputting a feature into a classifier. Four features and three classifiers resulted in twelve different base classifiers. For convenience, C_i is use to represent the i^{th} base classifier, where i is from 1 to 12, and the specific numbering rules are listed in Table 2. For instance, C_1 denotes the base classifier collected by inputting the LPCC into the SVM classifier. Let (C_i, C_j) signify the fusion of C_i and C_j.

As the number of base classifiers involved in the fusion increases, the required storage space will increase and may not necessarily bring better results. Therefore, using fewer base classifiers to obtain better fusion results is expected, and how to select the base classifier is the key to obtain good fusion results. In this study, a two-step base classifier selection method for good base classifiers selection is proposed.

Table 2 Numbering rules of 12 base classifiers

Classifier	LPCC	MFCC	GTCC	PSD
SVM	C_1	C_2	C_3	C_4
KNN	C_5	C_6	C_7	C_8
RF	C_9	C_{10}	C_{11}	C_{12}

Note: C_i represent the i^{th} base classifier.

The first-step base classifier selection approach is depicted in Figure 8. This step achieved the purpose of selection by continuously removing the base classifiers with high DF and low OA. Assuming that the number of original base classifiers is L_1 and the DF threshold is $\mathrm{DF}_{\mathrm{thr}}$, the OA of each base classifier and DF_{ij} between each base classifier were calculated, where $A=\{C_1,C_2,\ldots,C_{L_1}\}$. If the DF_{ij} between the two base classifiers is greater than the threshold $\mathrm{DF}_{\mathrm{thr}}$, it indicates that the error similarity between the two base classifiers is high. Therefore, at least one of the classifiers should be removed. Hence, the OA of the two base classifiers is judged, the base classifier with lower OA value is eliminated, and the base classifier with higher OA value is temporarily reserved. Next, all the DF are traversed and the base classifiers with high DF and low OA are continuously removed, and finally the result of the first-step selection is obtained. Assume that the number of base classifiers after the first-step selection is L_2.

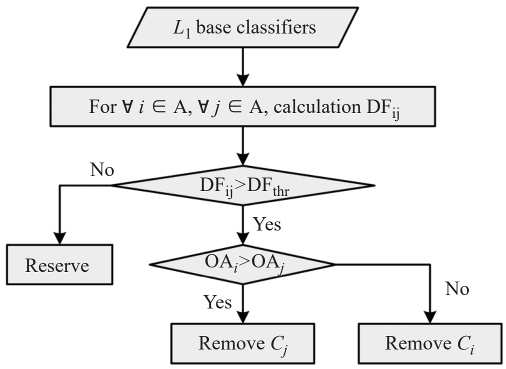


Figure 8 Flowchart of the first-step base classifier selection algorithm.

There will be several combinations for the first-step selected L_2 base classifiers (if the number of base classifiers is more than three) according to the number of fusion classifiers. Therefore, an additional selection need to be made. The second-step base classifier selection algorithm is shown in Figure 9. Firstly, various combinations of all L_2 classifiers were obtained and grouped based on the number of fusion classifiers. Subsequently, the OADF of each set was calculated, and the OADF and the corresponding combinations were sorted from big to small. Finally, the combinations with top N percent in each group of OADF are reserved for classifier fusion.

2.2.5 DS evidence theory

DS evidence theory is an uncertain reasoning approach proposed by Dempster[34] and further developed by his student Shafer[35]. DS evidence theory can deal with uncertain information

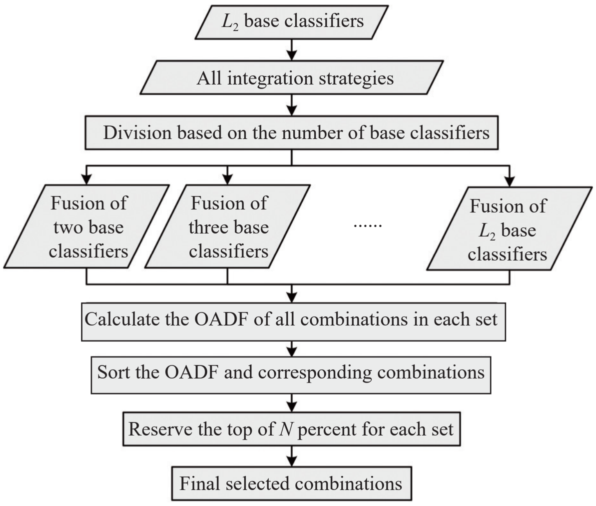


Figure 9 Flowchart of the second-step base classifier selection algorithm

and is broadly used in pattern recognition and other fields[26,36-40]. For example, let Ω be the identification framework, the function $m(\cdot)$is a mapping from 2^{Ω} to $[0,1]$, A is a subset of Ω and satisfy:

$$\begin{cases} m(\varnothing) = 0 \\ \sum_{A\subseteq\Omega} m(A) = 1 \end{cases} \tag{8}$$

Subsequently, $m(A)$ is the basic probability assignment (BPA) function of proposition A, which denotes the support degree of evidence to proposition A. For proposition A, when there are several different BPAs$(m_1, m_2, \ldots, m_n)$ on Ω, the composition rule is given as follows:

$$m(A) = \frac{1}{1-K}\sum_{A_1\cap A_2\cap\ldots\cap A_n = A} m_1(A_1)m_2(A_2)\ldots m_n(A_n) \tag{9}$$

while,

$$K = \sum_{A_1\cap A_2\cap\ldots\cap A_n = \varnothing} m_1(A_1)m_2(A_2)\ldots m_n(A_n) \tag{10}$$

where, K is the conflict coefficient, the greater the K, the greater the conflict between the evidence.

DS fusion is the fusion of the output probability of each classifier. The KNN output probability may appear to be 0 or 1. At this time, DS fusion will lead to a “one-vote veto,” that is, if there is an output probability of 0 in a single base classifier, the output probability must be 0 after DS fusion. This may result in the fusion of classifiers not getting any advantage or even result in performance degradation. Thus, the KNN output probability of 1 and 0 were transformed to α and $1-\alpha$. The range of α is $[0.5, 1]$. The optimal value of α can be obtained by the linear search.

2.2.6 Distance fusion

Distance fusion refers to the fusion of distances between the testing set and training set samples of various features[41]. x_j^i is used to represent the i^{th} feature vector of the j^{th} training sample and y_i to denote the i^{th} feature vector of the current testing sample, where $i = 1,2,3,4$ represent LPCC, MFCC, GTCC, and PSD, respectively. Let $\rho(j)$ be the class of the j^{th} training sample and d_j^i represents the distance between y_i and x_j^i The Manhattan distance is used to estimate the distance between the training and testing samples. The fusion distance D_j from the current testing sample to the j^{th} training sample is defined as follows:

$$D_j = \sum_{i=1}^{4}\frac{d_j^i}{\bar{d}^i} \tag{11}$$

where,

$$\bar{d}^i = \frac{1}{M}\sum_{j=1}^{M} d_j^i \tag{12}$$

where, M is the total number of training samples, the class R of the testing sample is presented as follows:

$$R = \rho\left(\underset{j}{\operatorname{argmin}}\, D_j\right) \tag{13}$$

2.2.7 DS–distance algorithm

DS evidence theory achieves a better classification by fusing the output probabilities of various base classifiers. However, in practical applications, when the fusion result is close to the decision boundary of 0.5, DS fusion’s classification result is unreliable due to noise interferences and other factors. Thus, this study proposed a DS–distance fusion algorithm to mitigate this problem. For the unreliable interval, the distance fusion algorithm was used rather than the DS fusion. The flowchart of the DS–distance algorithm is shown in Figure10.

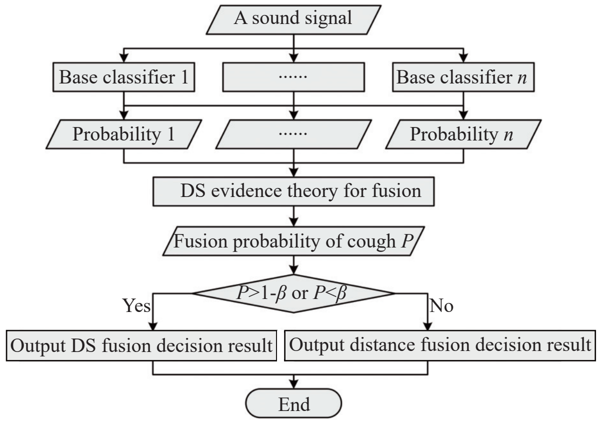


Figure 10 Flowchart of the DS–distance algorithm, where β is the conversion boundary

For testing sample x, let the probability of coughing by DS fusion is P and the classification result of distance fusion is RD, the final classification result R of the testing sample is given as follows:

$$R = \begin{cases} 1, & P > 1-\beta \\ \mathrm{RD}, & \beta \le P \le 1-\beta \\ 0, & P < \beta \end{cases} \tag{14}$$

among them, 1 represents a cough, 0 illustrates noncough, β is the conversion boundary, the value range is $[0.3, 0.5]$. In this study, a linear search is used to find the optimal β for all base classifier fusion strategies.

This approach effectively combines DS fusion and distance fusion by transferring decision-making power within an unreliable interval, thereby enhancing classification accuracy.

3 Results

The software used in this work was Matlab2018a (MathWorks, MA, USA). The machine learning functions used were in the Statistics and Machine Learning Toolbox. In the preprocessing, the bandpass filter bandwidth was 100-16 000 Hz, the pre-emphasis coefficient was 0.9375, the window was hamming window, the frame length was 20 ms, and the overlap length was 10 ms. The coefficient numbers of MFCCs, LPCCs, and GTCCs were 13, 24, and 13, respectively. The number of FFT points for PSD was 1024. The SVM kernel function was set to “RBF,” and the “KernelScale” was set to “auto.” The RF decision tree was set at 100. The hyperparameters of KNN were all set to “auto.” In classifier fusion, the KNN output probability transformation parameter α was 0.65.

This study choses the commonly used classification evaluation indicators: accuracy (equivalents to OA), recall, precision, and F1-score to evaluate the model. The dataset was shuffled, and 4/5 of them were randomly selected as the training set, and the remaining 1/5 was used as the testing set. All the experimental results were obtained by averaging 10 runs to reduce the error.

The OA of each base classifier is listed in Table 3, where the row represents the classifier, and the column represents the feature. It can be observed that the OA of each base classifier was above 91.14%. The SVM classification performed slightly better than the other two classifiers except for LPCC features. The LPCC and MFCC features were marginally better than those of GTCC and PSD.

Table 3 OA of the single base classifier

Classifier	LPCC	MFCC	GTCC	PSD
SVM	93.88%	94.84%	94.24%	91.70%
KNN	95.48%	92.68%	92.04%	91.88%
RF	93.96%	92.14%	91.14%	92.48%

The DF between each base classifier is shown in Figure 11, where 1–12 denotes C_1-C_{12}. The DF threshold was set to 2.5%. Following the first-step selection, the base classifiers include C_1, C_2, C_5, C_8, and C_9.

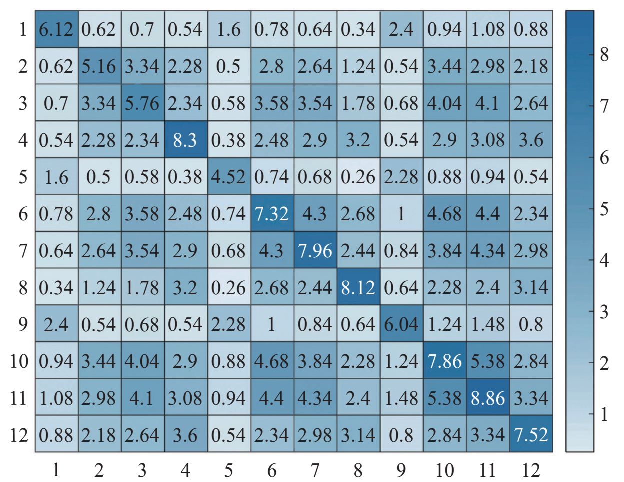


Figure 11 DF between base classifiers (%)

According to the number of fusion base classifiers, four sets of classifiers with different numbers of 2, 3, 4, and 5 could be obtained which were recorded in sets 2 to set 5. Among them, set 5 had only one combination, which was (C_1, C_2, C_5, C_8, C_9). Therefore, this set was reserved for additional comparison. The OADF of various combinations in each set are shown in Figure 12. By obtaining the top 20% OADF in each set, the resulting fusion combinations included (C_2, C_5), (C_2, C_9), (C_1, C_2, C_5), (C_2, C_5, C_9), (C_1, C_2, C_5, C_8), (C_1, C_2, C_5, C_8, C_9).

The DS fusion and DS-distance fusion results of the above-selected fusion combinations and the corresponding conversion boundary β are listed in Table 4. Compared with the results in Table 3, it is not difficult to find that DS and DS–distance fusion recognition accuracy is significantly enhanced. Generally, the accuracies of DS-distance are better than DS fusion results. The accuracy of DS fusion is associated with the number of fusion classifiers. This demonstrates an increasing trend as the number of fused classifiers increases as a whole. The best result is 98.48% for the fusion of four base classifiers. The accuracy of DS-distance fusion is largely unaffected by the number of fused classifiers. The result of the fusion of two classifiers achieved the accuracy of DS fusion of four classifiers. These primarily benefits from the effects of misclassification data nearing the decision boundary were categorized by the distance fusion strategy rather than DS fusion. The proposed method has obvious advantage, that is, the classifier result is not significantly affected by the number of fusion classifiers, and the result of fusing two classifiers is close to the result of combining five classifiers. Therefore, good classification result can be obtained by using a few classifiers.

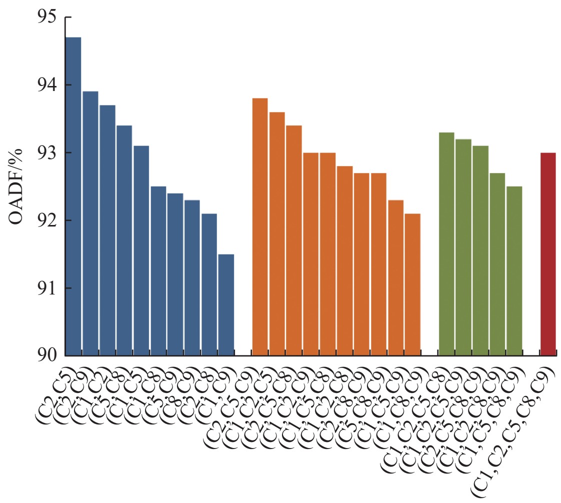


Figure 12 OADF of all base classifier combinations for each set

Table 4 OA of DS fusion, DS-distance fusion and the corresponding conversion boundary β

Classifier fusion	DS	DS–distance	β
(C_2, C_5)	97.36%	98.04%	0.40
(C_2, C_9)	97.52%	98.48%	0.32
(C_1, C_2, C_5)	98.20%	98.76%	0.35
(C_2, C_5, C_9)	97.88%	98.60%	0.30
(C_1, C_2, C_5, C_8)	98.48%	98.68%	0.46
(C_1, C_2, C_5, C_8, C_9)	98.24%	98.76%	0.31

In this section, the effectiveness of the classifier and feature fusion is compared. Feature fusion is the process of classifying data by joining several acoustic features into a single feature vector. Figure 13 depicts the OA variation of the three base classifiers with the number of fused features. The classification accuracy of SVM and RF classifiers gradually improves with the increase of the number of fused features. The classifiers have the highest accuracy when all the features are combined. For the KNN classifier, the best performance occurs when three features are fused. However, there is a significant drop when fusing the four features. Overall, SVM has the highest classification accuracy when fusing four features, which reaches 97.64%.

The optimal results of feature fusion of the three base classifiers and the best DS-distance classifier fusion results are presented in Table 5. It can be observed that the algorithm proposed in this study is obviously better than the feature fusion. Comparing the results in Table 4, even fusing two classifiers, the accuracy is higher than the feature fusion. Although four features were used in the feature fusion and only two were used in the classifier fusion, classifier fusion still gains more benefits by adding one more classifier. These outcomes show that classifier fusion has significant benefits and is promising in pig cough sound recognition. It can offer a practical way to accomplish highly accurate and stable pig cough sound recognition in complex pig houses.

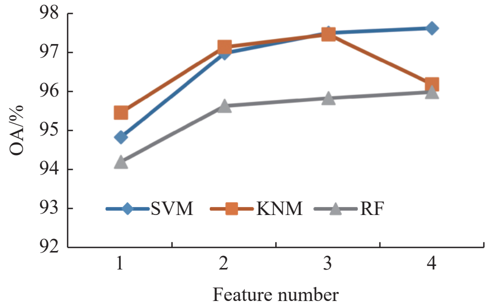


Figure 13 OA variation of the three base classifiers with the number of fused features

Table 5 Performance comparison between the proposed method and feature fusion method

Method	OA	Recall	Precision	F1-score
SVM	97.64%	96.40%	98.85%	97.61%
KNN	97.48%	96.96%	97.98%	97.46%
RF	96.00%	94.96%	96.99%	95.95%
Proposed method	98.76%	98.94%	98.61%	98.76%

In order to further illustrate the advantages of the proposed method, its performance was compared with the existing typical methods using deep learning models and feature fusion which have been introduced in the Introduction part in recent two years. Both of them are all based on the same dataset. The performance comparison is listed in Table 6. It can be seen that the highest accuracy of the existing research is 97.35% by fusion of acoustic (two time-domain features, six frequency-domain features and MFCC) and deep features (extracted by CNN model) and SVM model. The proposed model used four features and three machine learning models to construct the base classifiers and selected the best combination by two-step base classifier selection method. The accuracy of the proposed method achieved 98.76 %, which is higher than the existing state-of-the-art methods. Meanwhile, compared with the method proposed by Shen et al.[16], the complexity and the computational load of the proposed method are much lower.

Table 6 Performance comparison between the proposed method and previous methods.

Method	Feature	Model	OA	Recall	Precision	F1-score
Yin et al.[12]	Spectrogram	AlexNet	95.40%	96.80%	95.50%	96.20%
Shen et al.[14]	MFCC	CNNs+SVM	96.68%	97.72%	96.81%	97.26%
Ji et al.[15]	Acoustic+Visual	SVM	96.45%	97.33%	96.83%	97.08%
Shen et al.[16]	Acoustic+Deep	CNN+SVM	97.35%	96.51%	98.41%	97.46%
Proposed method*	MFCC+LPCC	SVM+KNN+ DS-distance	98.76%	98.94%	98.61%	98.76%

Note: *It shows the best combination (C_1, C_2, C_5) result. The base classifiers are composed by MFCC, LPCC features and SVM, KNN models.

4 Discussion

The purpose of the data collection experiment is to obtain the sounds including pig coughs and noncoughs in a piggery in order to explore and study the method of pig cough recognition. Limited by the experimental conditions, only one microphone near the door was put to collect the sounds. The pickup distance of the microphone is approximately 10 m and the piggery is 27.5 m long. There is no guarantee of collecting coughs from all the pigs in an entire barn, and that is not necessary. The purpose of this experiment is not to identify all the coughs of an entire barn. Other studies usually use a recording pen to pick up pigs' cough at close range. In this experiment, the data were collected continuously for one week. Due to the depletion of the microphone battery, some data for part of the time was lost. The data were segmented and labeled manually with the help of experts. The overlapped and hard distinguished sounds were not considered. The dataset contained the sounds with different signal-to-noise ratios and limited types of noncoughs. More data will be needed to verify the proposed algorithm and train the model to make it more robust.

The features involved in the sound signal processing can be summarized into three categories: acoustic features, image/visual features and deep features. Image features and deep features were commonly used in the deep neural networks. In this study, only four acoustic features were considered and no deep neural network was involved. Besides the frequently used MFCC and PSD features, the LPCC and GTCC features were also considered, which had performed well in some other classification tasks. However, they were seldom discussed in the pig cough recognition before. In this work, first, the performance of these two features in pig cough recognition were discussed. Meanwhile, more excellent and different base classifiers were created for classifier fusion. The results in Table 3 revealed that they were comparable with MFCC. LPCC was even better than MFCC for some classifiers. The feature fusion was also researched for comparison, as listed in Table 5. The best accuracy of feature fusion reached 97.64%, which was better than the results in Ji et al.[15] (acoustic and visual feature fusion: 96.45%) and in Shen et al.[16] (acoustic and deep feature fusion: 97.35%). The accuracy was further increased by the proposed multi-classifier fusion. This is because multi-classifier fusion essentially implied feature fusion and it increased the accuracy by adding classifiers. Meanwhile, the performance was improved by the proposed DS-distance algorithm.

Other classifier fusion methods have also been tried. Firstly, the features were selected and fused as a feature vector, and then the feature vector were input into three classifiers for fusion. When four features were fused, the classification accuracy was 97.80%. According to the results in Table 5, when each classifier selected the optimal feature for fusion, the classification accuracy reached 98.04%. They were all lower than the proposed method. The results indicate that the selection of feature and classifier had a certain impact on multi-classifier fusion. Although the fusion feature was better than the single feature, due to the similarity of the features, the difference between each classifier output was small, so the accuracy was not significantly improved for this kind of fusion. For the proposed algorithm, the base classifiers with large classification differences were selected for fusion through the error similarity indicator, so as to obtain a good fusion result.

Although deep neural networks have shown excellent performance in many classification tasks, machine learning has more potential in the rapid establishment of pig respiratory disease early warning system. In the current study, whether pig cough sounds at different growth stages or different farms can be identified by the same training model has not been clearly concluded. If a new model needs to be trained from scratch for different growth stages or different pig barns, it is hope that an accurate model can be quickly established by a small amount of training data. However, due to the rapid development of deep learning, the research in recent years has spent too much energy on deep learning, ignoring the basic acoustic features and the advantages of machine learning methods. For this reason, further research on machine learning and acoustic features has been done, and the fusion of multiple acoustic features and classifiers has been deeply studied. Table 6 shows that the proposed method has achieved a good results and the accuracy is higher than the recent deep learning models. In order to ensure the fairness of the comparison, only the latest researches using the same dataset as in this paper were compared. The accuracy of early studies was relatively low. In recent years, although the accuracy has been significantly improved, most of them are based on deep learning models. The accuracy of the proposed algorithm is 98.76%, which is higher than the existing methods. The proposed method only uses traditional acoustic features and machine learning models. Compared with the existing methods in recent years, this method has the advantages of short training time and low complexity which is more conducive to practical engineering applications.

For multi-classifier fusion, a critical procedure is to choose the excellent base classifiers which are related to the selection of features and classifiers. In this work, only four features and three classifiers were provided for selection. More representative features and classifiers are deserved to be explored to gain a more robust recognition. Although the proposed method achieved good performance, the results were only tested on a private dataset which collected in one pig barn. Meanwhile, the dataset in this study was segmented manually which did not consider the accuracy loss in the voice active detection process. The overlapped sounds and the sounds which cannot be distinguished by experts did not considered in the dataset. In a real-time pig cough sound detection scenario in a large-scale piggery, the data will be more complex than the dataset in this work. More experiment will be conducted and the new data will be used to test the algorithm in the future. This study hope that a high-quality public pig sound dataset will be opened soon for more researchers, contributing their good ideas and approaches to accelerate the application of this technique.

5 Conclusions

In this work, the multi-classifier fusion based on machine learning was proved to be an effective way to enhance the accuracy of pig cough recognition. The features of MFCC, LPCC, GTCC and PSD provide the characteristics of different sound from different perspectives. The machine learning model of SVM, KNN and RF performed well in the pig cough recognition task. Among the combinations of different base classifiers, the combinations of (MFCC+SVM, LPCC+KNN) and (MFCC+SVM, LPCC+RF) showed an outstanding performance. The combination of (LPCC+SVM, MFCC+SVM, LPCC+KNN) achieved the best performance. The accuracy of the proposed algorithm reached 98.76%, which is better than 97.64% of feature fusion and 97.35% of state-of-the-art method. This proves that classifier fusion has significant advantages and the potential in the establishment of pig respiratory disease early warning system.

Acknowledgements

This work was supported by the Outstanding Youth Program of the Natural Science Foundation of Heilongjiang Province of China (Grant No. YQ2023C012); the project of the National Natural Science Foundation of China (Grant No. 32172784, 31902210); the Academic Backbone Project of Northeast Agricultural University; the National Key Research and Development Program of China (Grant No. 2019YFE0125600); the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (Grant No. UNPYSCT-2020092); and the earmarked fund for CARS-36 and CARS-35.

[References]

[1] Ji N, Yin Y L, Shen W Z, Kou S L, Dai B S, Wang G W. Pig sound analysis: A measure of welfare. Smart Agriculture, 2022; 4(2): 19–35.

[2] Maes D, Sibila M, Pieters M, Haesebrouck F, Segalés J, de Oliveira L G. Review on the methodology to assess respiratory tract lesions in pigs and their production impact. Veterinary Research, 2023; 54(1): 8.

[3] Racewicz P, Ludwiczak A, Skrzypczak E, Składanowska-Baryza J, Biesiada H, Nowak T, et al. Welfare health and productivity in commercial pig herds. Animals, 2021; 11(4): 1176.

[4] Lagua E B, Mun H-S, Ampode K M B, Chem V, Kim Y-H, Yang C-J. Artificial intelligence for automatic monitoring of respiratory health conditions in smart swine farming. Animals, 2023; 13(11): 1860.

[5] Van Hirtum V, Guarino M, Costa A, Jans P, Hhesquiere K, Aerts J M, et al. Automatic detection of chronic pig coughing from continuous registration in field situations. In: Models and analysis of vocal emissions for biomedical applications: 3rd International workshop, Firenze: Firenze University Press. 2003; pp.251–254. doi: 10.1400/39888.

[6] Exadaktylos V, Silva M, Ferrari S, Guarino M, Taylor C J, Aerts J M, et al. Time-series analysis for online recognition and localization of sick pig (*Sus scrofa*) cough sounds. The Journal of the Acoustical Society of America, 2008; 124(6): 3803–3809.

[7] Guarino M, Jans P, Costa A, Aert J M, Berckmans D. Field test of algorithm for automatic cough detection in pig house. Computers and Electronics in Agriculture, 2008; 62(1): 22–28.

[8] Exadaktylos V, Silva M, Aerts J-M, Taylor C J, Berckmans D. Real-time recognition of sick pig cough sounds. Computers and Electronics in Agriculture, 2008; 63(2): 207–214.

[9] Chung Y, Oh S, Lee J, Park D, Chang H-H, Kim S. Automatic detection and recognition of pig wasting diseases using sound data in audio surveillance systems. Sensors, 2013; 13(10): 12929–12942.

[10] Li X, Zhao J, Gao Y, Lei M G, Liu W H, Gong Y J. Recognition of pig cough sound based on deep belief nets. Transactions of the CSAM, 2018; 49(3): 179–186. (in Chinese)

[11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017; 60(6): 84–90.

[12] Yin Y L, Tu D, Shen W Z, Bao J. Recognition of sick pig cough sounds based on convolutional neural network in field situations. Information Processing in Agriculture, 2021; 8(3): 369–379.

[13] Shen M X, Wang M Y, Liu L S, Chen J, Tai M, Zhang W. Recognition method of pig cough based on deep neural network. Transactions of the CSAM, 2022; 53(5): 257–266. (in Chinese)

[14] Shen W Z, Tu D, Yin Y L, Bao J. A new fusion feature based on convolutional neural network for pig cough recognition in field situations. Information Processing in Agriculture, 2021; 8(4): 573–580.

[15] Ji N, Shen W, Yin Y L, Bao J, Dai B S, Hou H D, et al. Investigation of acoustic and visual features for pig cough classification. Biosystems Engineering, 2022; 219: 281–293.

[16] Shen W Z, Ji N, Yin Y L, Dai B S, Tu D, Sun B H, et al. Fusion of acoustic and deep features for pig cough sound recognition. Computers and Electronics in Agriculture, 2022; 197: 106994.

[17] Yin Y L, Ji N, Wang X P, Shen W Z, Dai B S, Kou S L, et al. An investigation of fusion strategies for boosting pig cough sound recognition. Computers and Electronics in Agriculture, 2023; 205: 107645.

[18] Wang X P, Yin Y L, Dai X P, Shen W Z, Kou S L, Dai B S. Automatic detection of continuous pig cough in a complex piggery environment. Biosystems Engineering, 2024; 238: 78–88.

[19] Sagi O, Rokach L. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018; 8(4): e1249.

[20] Huang X, Zhang L P. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution

remotely sensed imagery. IEEE transactions on Geoscience and Remote Sensing, 2012; 51(1): 257–272.

[21] Li X Y, Fan P P. Study on characteristic spectrum and multiple classifier fusion with different particle size in marine sediments. IEEE Access, 2020; 8: 157151–157160.

[22] Mishra S, Shaw K, Mishra D, Patil S, Kotecha K, Kumar S, et al. Improving the accuracy of ensemble machine learning classification models using a novel bit-fusion algorithm for healthcare AI systems. Frontiers in Public Health, 2022; 10: 858282.

[23] Kaur T, Gandhi T K. Classifier fusion for detection of COVID-19 from CT scans. Circuits, Systems, and Signal Processing, 2022; 41: 3397–3414.

[24] Wang X, Wu Q X, Lin X J, Zhuo Z Q, Huang L P. Pedestrian identification based on fusion of multiple features and multiple classifiers. Neurocomputing, 2016; 188: 151–159.

[25] Huan Z, Chen X J, Lv S Y, Geng H Y. Gait recognition of acceleration sensor for smart phone based on multiple classifier fusion. Mathematical Problems in Engineering, 2019; 2019: 6471532.

[26] Perez A, Tabia H, Declercq D, Zanotti A. Using the conflict in Dempster–Shafer evidence theory as a rejection criterion in classifier output combination for 3D human action recognition. Image and Vision Computing, 2016; 55: 149–157.

[27] Shiue Y-R, You G-R, Su C-T, Chen H. Balancing accuracy and diversity in ensemble learning using a two-phase artificial bee colony approach. Applied Soft Computing, 2021; 105: 107212.

[28] Ferrari S, Silva M, Guarino M, Aerts J M, Berckmans D. Cough sound analysis to identify respiratory infection in pigs. Computers and Electronics in Agriculture, 2008; 64(2): 318–325.

[29] Ferrari S, Silva M, Guarino M, Berckmans D. Analysis of cough sounds for diagnosis of respiratory infections in intensive pig farming. Transactions of the ASABE, 2008; 51(3): 1051–1055.

[30] Zhao X J, Shao Y, Wang D L. CASA-based robust speaker identification. IEEE Transactions on Audio, Speech, and Language Processing, 2012; 20(5): 1608–1616.

[31] Valero X, Alias F. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. IEEE Transactions on Multimedia, 2012; 14(6): 1684–1689.

[32] Fedila M, Bengherabi M, Amrouche A. Gammatone filterbank and symbiotic combination of amplitude and phase-based spectra for robust speaker verification under noisy conditions and compression artifacts. Multimedia Tools and Applications, 2018; 77: 16721–16739.

[33] Woźniak M, Grana M, Corchado E. A survey of multiple classifier systems as hybrid systems. Information Fusion, 2014; 16: 3–17.

[34] Dempster A P. Upper and Lower Probabilities Induced by a Multivalued Mapping. The Annals of Mathematical Statistics, 1967; 38(2): 325–339.

[35] Shafer G. A mathematical theory of evidence. Princeton: Princeton University Press. 1976; 292p. doi: 10.1080/00401706.1978.10489628.

[36] Zhang S X, Lin J H, Su L, Zhou Z P. pDHS-DSET: Prediction of DNase I hypersensitive sites in plant genome using DS evidence theory. Analytical Biochemistry, 2019; 564–565: 54–63.

[37] Sandula P, Okade M. A novel video saliency estimation method in the compressed domain. Pattern Analysis and Applications, 2022; 25: 867–878.

[38] Pan Y, Zhang L M, Wu X G, Skibniewski M J. Multi-classifier information fusion in risk analysis. Information Fusion, 2020; 60: 121–136.

[39] Mi J H, Li Y-F, Peng W W, Huang H-Z. Reliability analysis of complex multi-state system with common cause failure based on evidential networks. Reliability Engineering & System Safety, 2018; 174: 71–81.

[40] Jin J, Qu T N, Xu R, Wang X Y, Cichocki A. Motor imagery EEG classification based on Riemannian sparse optimization and dempster-shafer fusion of multi-time-frequency patterns. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023; 31: 58–67.

[41] Wang B, Li H, You J W, Chen X, Yuan X H, Feng X Z. Fusing deep learning features of triplet leaf image patterns to boost soybean cultivar identification. Computers and Electronics in Agriculture, 2022; 197: 106914.