Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

An investigation of fusion strategies for boosting pig cough sound recognition

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

[a] *School of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China*
[b] *Heilongjiang Agricultural Technology Extension Station, China*

ARTICLE INFO

Keywords:
Pig cough recognition
Feature fusion
Classifier fusion
Fine-tuned CNN

ABSTRACT

The recognition of pig cough sounds is an effective way to monitor pig respiratory diseases which seriously affect healthy pig breeding. Due to the complexity of the pig-housing environment, achieving high precision cough recognition by relying only on a single feature or classifier is challenging. Therefore, in this study we investigated two fusion strategies, namely feature fusion and classifier fusion, to boost classification accuracy. For feature fusion, we improved the previously proposed feature fusion algorithm and selected better acoustic and image features for fusion. We also proposed a novel classifier fusion algorithm. In the algorithm, the support vector machine (SVM) classifiers trained by the acoustic features and deep features were fused by soft voting for pig cough prediction. The sound data collected in the pig barn were used to validate the proposed methods. Our methods achieved a substantial classification rate of 97.47% and 99.20% for feature fusion and classifier fusion, respectively. The results demonstrate that our proposed fusion strategies can significantly improve the recognition accuracy of pig cough sounds.

1. Introduction

Pig cough sound recognition is considered to be an important means of effectively warning of pig respiratory diseases (Racewicz et al., 2021). In recent years, pig cough sound classification methods have been extensively studied, and some significant results have been achieved. The accuracies of pig cough recognition in early researches in pig barns reached 85.0 %-93.0 % (Exadaktylos et al., 2008a; Guarino et al., 2008), and most of these studies used traditional sound signal recognition methods, such as dynamic time warping (DTW) (Hirtum et al., 2003; Guarino et al., 2008) and fuzzy c-means (FCM) (Hirtum and Berckmans, 2003; Exadaktylos et al., 2008b). With the development of signal processing technology, machine learning and deep learning algorithms have been gradually applied to the field of pig cough sound recognition. There has been an obvious increase in accuracy using some well-designed algorithms in feature selection and classifier optimization. Pig cough classification rates increased to 94.0 %-95.4 % (Chung et al., 2013; Yin et al., 2021). However, it is difficult to further improve classification accuracy by only relying on a single feature or classifier. Fusion strategies provide a new direction for boosting the accuracy of pig cough sound recognition.

Feature fusion is a common fusion method and is widely used in many tasks such as image recognition (Cheng et al., 2021; Shen et al., 2020; Wang et al., 2022), speech recognition (Atmajaet al., 2022; Zhao et al., 2021), and acoustic scene classification (Yang et al., 2020; Chi., 2021. Feature representations of acoustic signals can be summarized into two categories: acoustic features and visual features or image features. More effort has been spent on the fusion of acoustic features in the early research stage (Shen et al., 2021, Sharma et al., 2020). Currently, far more attention is being focused on the fusion of image features or the combination of acoustic and image features (Ji et al., 2022; Shen et al., 2022; Er., 2020). In our previous research, we tried to fuse mel-frequency cepstral coefficient (MFCC) features using multiple convolutional neural network (CNN) models (Shen et al., 2021) and we created a fusion of acoustic features and visual or deep features (Ji et al., 2022; Shen et al., 2022). The best classification accuracy reached 97.35 % (Shen et al., 2022). In recent years, classifier fusion has been widely used in many fields and has achieved good performance (Hassan et al., 2021; Yao et al., 2020). In consideration of its advantages, classifier fusion algorithm has been applied in this study. Such an application has not

* Corresponding authors at: School of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China.
E-mail addresses: jn740740@163.com (N. Ji), wzshen@neau.edu.cn (W. Shen).

https://doi.org/10.1016/j.compag.2023.107645
Received 30 July 2022; Received in revised form 10 October 2022; Accepted 8 January 2023
Available online 17 January 2023

been reported in the field of pig cough recognition as far as we know.

In this present research, we investigated both the feature fusion and the classifier fusion algorithm to boost classification accuracy, and we analyzed and discussed the advantages of the two fusion strategies in pig cough classification tasks. We propose an improved feature fusion algorithm based on previous researches, and we propose a classifier fusion algorithm. Overall, the contributions of this work can be summarized as follows:

(1) Two fusion strategies, including feature fusion and classifier fusion, are proposed to promote the accuracy of pig cough recognition.
(2) Classifier fusion is proven to outperform feature fusion.
(3) The experimental results demonstrate that our proposed classifier fusion method achieves a substantial accuracy of 99.20 %.

The remainder of this paper is organized as follows. Section 2 describes the dataset used in our work. Section 3 presents the details of our proposed fusion strategies for pig cough recognition. The experimental results are given in Section 4. A discussion of the results is provided in Section 5. Conclusions are drawn in Section 6.

2. Datasets

2.1. Animals and housing

The data were collected in a pig fattening house in Harbin, Heilongjiang Province, China, in April 2018. The size of pig house was 27.5 m × 12.8 m × 3.2 m (length × width × height), as shown in Fig. 1. There were 12 large pens and 9 small pens in the barn. The sizes of the large pen and the small pen were4.15 m × 3.6 m (length × width) and 3.6 m × 2.75 m (length × width), respectively. The mechanical ventilation was used and the pigs were fed with mechanized equipment in the pig house. The floor was composed of a half-slatted concrete floor. The workers used shovels and water to clean up manure and flush the floor twice a day in the morning and the afternoon. There were 128 pigs largely evenly distributed in 12 large pens. Two pigs with a heavy cough were separated into a smaller pen (Pen 13) near the door. The cough had appeared for several days prior and there were several coughing pigs in each pen when we collected the data.

2.2. Data collection

The sound data were recorded using a microphone (LIQI LM320E, Cardioid electret microphone) connected to a laptop with a Conexant Smart Audio HD sound card. The frequency range of the microphone was 100 Hz–16 kHz. Limited by the experimental conditions, we affixed the microphone over a large pen near the door. The height of the microphone was 1.4 m above the floor and approximately 0.8 m above the backs of the pigs. We sampled the sound data at 44.1 kHz with a resolution of 16 bits. The sounds were continuously recorded and saved in waveform. The individual sounds including pig coughs, screams, cleaning sounds, people talking, and others were extracted and labeled with the assistance of a veterinarian. We randomly selected 1250 cough sounds and 1250 non-cough sounds as the training and testing datasets in this research.

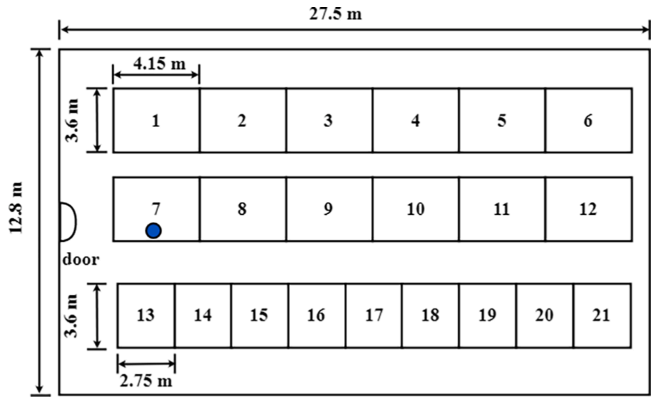

Fig. 1. Layout of the pig barn in this study. The blue circle represents the location of the microphone.

3. Proposed methods

In this section, we present the details of the proposed two fusion strategies of pig cough sound recognition. We first introduce the framework of the improved feature fusion and classifier fusion methods. The features, including acoustic features and image features and the feature extractor, are then described. Lastly, the classifier and evaluation metrics are illustrated.

3.1. Framework of the proposed methods

Improved feature fusion framework. The improved feature fusion method is based on our previous work in Shen et al. (2022). In previous work, the acoustic features and deep features were fused to make a classification of pig coughs and non-coughs. The acoustic features were formed by MFCC, root mean square, zero crossing rate, spectral centroid, spectral flatness, spectral bandwidth, spectral rolloff, spectral contrast, and spectral flux. The deep features were extracted from a CNN architecture based on the short–time Fourier transform (STFT) and constant-Q transforms (CQT).

The framework of the proposed improved feature fusion method is shown in Fig. 2. The acoustic features and image features were extracted from the pre-processed sound segments in the dataset. The pre-processing included filtering, pre-emphasis, framing and windowing. The acoustic features were composed of a stack of MFCC and linear predictive cepstral coefficient (LPCC) vectors. The image features were two-dimensional data, such as spectrograms, mel-spectrograms and CQTs. Deep features were further acquired through a fine-tuned CNN by the image features. The acoustic features and deep features were concatenated and trained with the SVM classifier.

Classifier fusion framework. The framework of the proposed classifier fusion method is shown in Fig. 3. Similarly to the feature fusion, the acoustic features and image features were first extracted from pre-processed sound segments. Then, the acoustic features were used to train an SVM classifier, and the image features were input into a fine-tuned CNN model to extract deep features. The obtained deep features were then used to train the other SVM classifier. Finally, soft voting was used for the two SVM classifiers' output to come to a final decision.

3.2. Features

Acoustic features. Acoustic features refer to a concatenated vector of MFCCs and LPCCs. MFCC is a feature based on human auditory perception. It is the result of the short-term real log-cosine transform of the energy spectrum, expressed in the mel-frequency scale. MFCC is widely used in speech recognition and speaker recognition. It has also been proven to have good performance in pig cough sound recognition (Zhao et al., 2020; Shen et al., 2022). The principles underlying MFCC are adequately described in the respective literature. We will not elaborate on them further in this paper.

LPCC uses linear predictive coding technology to obtain cepstral coefficients from the perspective of a human vocalization model. In speech recognition, LPCC has better description ability for vowels and poorer description ability for consonants (Tin et al., 2003, Ahmed et al., 2019). However, LPCC performance in pig cough recognition has not been clear. Thus, in this study, we investigated the potential in the task of pig cough classification based on the characteristics of LPCC.

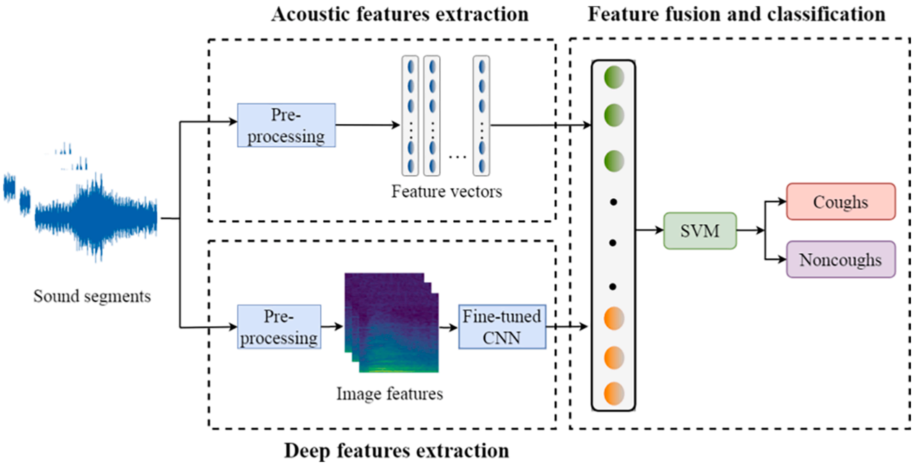

Fig. 2. Framework of the improved feature fusion method.

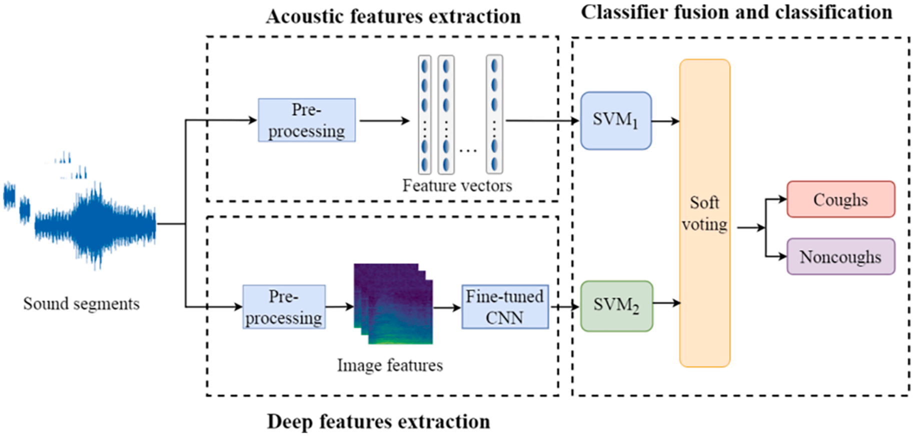

Fig. 3. Framework of the proposed classifier fusion method.

Additionally, we combined MFCC and LPCC as acoustic features to carry out classification.

LPCC can be calculated by the linear prediction coefficient (LPC). The basic premise of LPC is that the current/future sample $x[n]$ can be predicted with past p samples due to the correlation between speech samples, as shown in Eq. (1),

$$x[n] = \sum_{i=1}^{p} a_i x[n-i] \tag{1}$$

where a_i is the LPC. LPCC can be obtained by a_i from the following formula,

$$\widehat{h}(n) = \begin{cases} a_n & n = 1 \\ a_n + \sum_{i=1}^{n-1}(1 - i/n)a_i\widehat{h}(n-i) & 1 < n \leqslant p \\ \sum_{i=1}^{n-1}(1 - i/n)a_i\widehat{h}(n-i) & n > p \end{cases} \tag{2}$$

Image features. In this paper, we mainly focused on three image features, namely the spectrogram, the mel-spectrogram, and CQT. In speech recognition, the spectrogram is a commonly used image feature. A spectrogram is a visual mode of representing the signal strength, or “loudness,” of a signal over time at various frequencies presented in a particular waveform (Knight et al., 2020). It can be expressed as the short-time Fourier transform of the sound signal, as shown in Eq. (3),

$$X(m,f) = \sum_{n=-\infty}^{\infty} x(n)\omega(n-m)e^{-j2\pi fn} \tag{3}$$

where $x(n)$ is the time-domain signal, and $\omega(n)$ is the window function.

The mel-spectrogram applies a frequency-domain mel-filter bank to audio signals that are windowed in time (Zhang et al., 2020). The audio input is first buffered into frames with an overlap. The specified window is applied to each frame, and then the frame is converted to a frequency-domain representation. Each frame of the frequency-domain representation passes through a mel-filter bank. Finally, the output of each mel-filter bank is summed and concatenated together, which contributes to the mel-spectrogram.

CQT is generally used for the analysis of music signals. CQT is a filter bank in which the center frequency is distributed according to the relevant exponential law, and the filter bandwidth is different, but the ratio of the center frequency to bandwidth is the constant Q (Xu et al., 2021).The expression of Q is shown in Eq. (4),

$$Q = \frac{f_k}{\delta f_k} = \frac{f_k}{f_{k+1} - f_k} = \frac{1}{2^{1/B} - 1}, \quad k = 0, 1, \dots K - 1 \tag{4}$$

where$f_k = f_1 2^{\frac{k-1}{B}}$, and f_1 and f_k are the center frequency of the lowest frequency bin and the kth filter, respectively. B associates with the number of bins per octave. CQT can be expressed byEq.(5),

$$X^{CQT}(k) = \frac{1}{N_k}\sum_{n=0}^{N_k-1} x(n)\omega_{N_k}(n)e^{-\frac{j2\pi Q}{N_k}n} \tag{5}$$

where$N_k = Q\frac{f_s}{f_k}$, $k = 0, 1, ..., K-1$, is the window length, F_s is the sampling frequency, and $\omega_{N_k}(n)$ represents the window function with a length of N_k.

3.3. Feature extractor

CNN has shown excellent performance in numerous recognition tasks, such as image, video, and speech recognition. CNN models can not only be used for classification but can also be treated as feature extractors for object recognition and image classification (Shen et al., 2020). Two methods are mainly used to extract features. One is to use the pre-trained models directly as feature extractors, the other is to fine-tune and retrain the pre-trained models with a private dataset for a specific image classification task. Feature extraction using the pre-trained model is relatively simple and quick, with a small computation load. However, it is more suitable for image classification tasks if the images are similar to the pre-trained network, such as the nature images in the ImageNet database (Deng et al., 2009). If the data is very different from the ImageNet data (for example, small images, spectrograms, or non-image data), fine-tuning the pre-trained network may work better. In our work, both a pre-trained CNN model and a fine-tuned CNN model were considered as deep feature extractors for investigating the performance of fusion strategies for pig cough recognition. The architecture of the proposed deep feature extractor is shown in Fig. 4.

The fine-tuned CNN model was based on the framework of the AlexNet model (Krizhevsky et al., 2012). The AlexNet model consists of five convolutional layers and three fully connected layers. Three max pooling layers are followed by Convolutional layers one, two, and five, respectively. We removed the last two fully connected layers of the model. The output of the fully connected layer is the deep feature representation. Further, we retrained the model using the dataset as described in Section 2.2 in the form of spectrogram, mel-spectrogram, and CQT images.

3.4. Experimental parameters and evaluation metrics

The experiments were conducted using a configuration of an Intel(R) Core (TM) i7-1165G7 CPU running at 2.80 GHz with 16 GB of memory, and an NVIDIA GeForce MX450Ti GPU with 2 GB of memory. The software used in this work was *MATLAB 2021a*.

A widely used SVM was adopted as the basic classifier. It has shown good performance in multiple classification tasks. In the experiment, the *fitcsvm* function in the *Statistics and Machine Learning Toolbox* of *MATLAB 2021a* was used. The ‘kernelFunction’ was set to ‘RBF’ and the ‘kernelScale’ was set to ‘auto’. In the pre-processing, a FIR bandpass filter with the frequency range of 100–14,000 Hz was used. The coefficient of the pre-emphasis filter was 0.9375. The continuous sound files were framed into 20 ms with 10 ms overlap and the hamming window was used. The coefficient number of MFCC and LPCC was 13 and 24, respectively.

The parameters used for fine-tuning the pre-tained CNN are listed as follows: batch size: 50, learning rate: 0.0001, epoch: 30, optimizer: Adam, loss function: cross entropy.

Commonly used evaluation metrics were adopted to evaluate the models. These are Accuracy, Recall, Precision, and F1-score. These metrics were calculated by using Equations (6)–(9). Here, true positive, true negative, false positive, and false negative are referred to as TP, TN, FP, and FN, respectively. We defined cough as a positive sample and non-cough as a negative sample.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{6}$$

$$Recall = TP/(TP + FN) \tag{7}$$

$$Precision = TP/(TP + FP) \tag{8}$$

$$F1 - score = 2 \times (Precision \times Recall)/(Precision + Recall) \tag{9}$$

4. Results

In this section, we describe the extensive experiments that were conducted to evaluate the performance of our proposed methods for pig cough sound recognition. First, the performances of the deep features from both the pre-trained and fine-tuned CNN with different image inputs are compared. Second, the results of acoustic and deep feature fusion are given and analyzed. Finally, the classifier fusion results with different input features are presented.

The validation accuracy and loss curves of the pre-trained model for three image features are shown in Fig. 5 and Fig. 6, respectively. It can be found that the models converge rapidly, and the accuracies basically

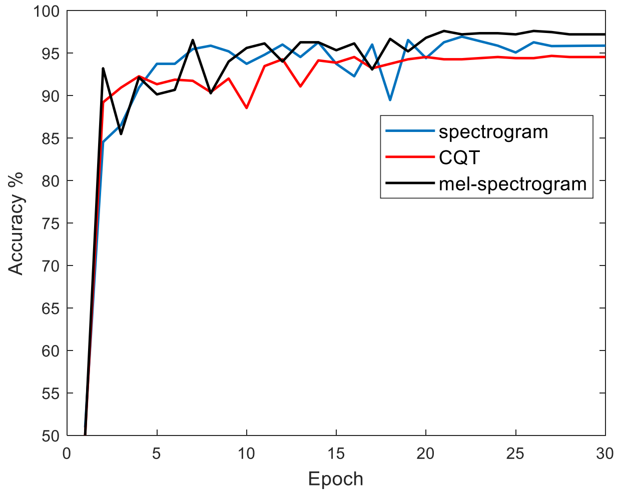

Fig. 5. The validation accuracies of the pre-trained models.

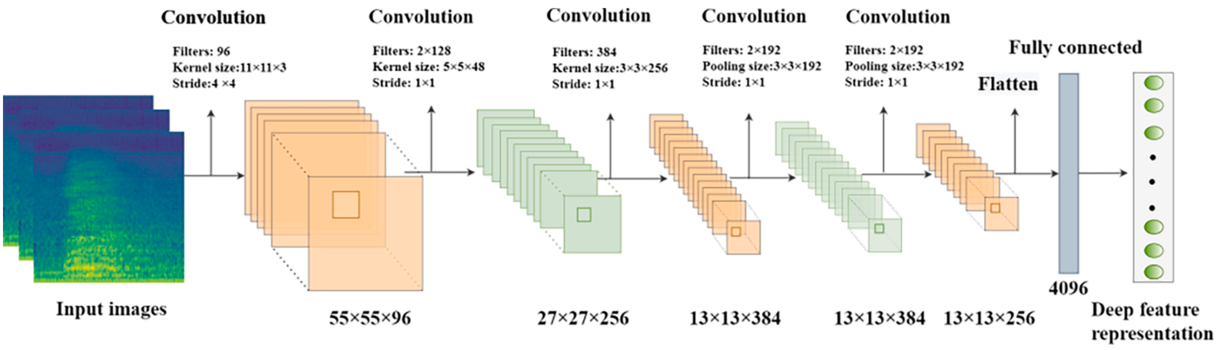

Fig. 4. Architecture of a fine-tuned CNN model.

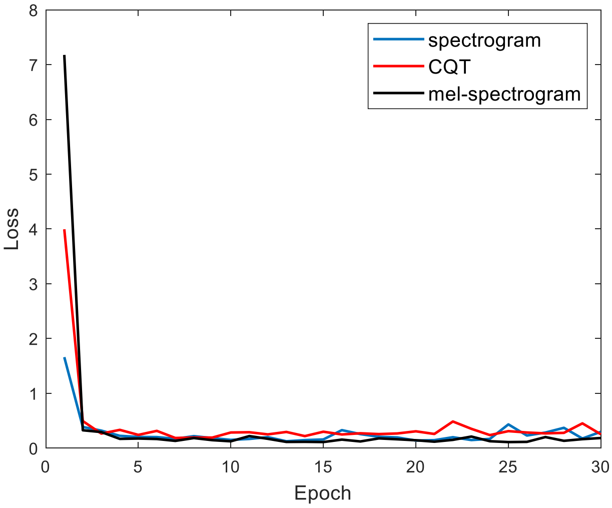

Fig. 6. The validation losses of the pre-trained models.

hold steady after 25 epoches. It apparently shows that the mel-spectrogram performed better than the other two features.

The results of the pre-trained and fine-tuned models with different image inputs are shown in Table 1. Overall, we can see that the accuracies of the fine-tuned network were significantly better than those of the pre-trained network. For the pre-trained network, mel-spectrogram performed best with an accuracy of 95.47 % and CQT slightly outperformed spectrogram. The mel-spectrogram showed a good balance in Recall and Precision. CQT and spectrogram achieved a higher Recall and a lower Precision. After fine-tuning the model, the accuracy of the mel-spectrogram was improved significantly, reaching a good classification accuracy of 97.33 %. Although the accuracy of both the spectrogram and CQT were found to be improved compared to the results of the pre-trained model, the spectrogram provided better performance than the CQT in general. Motivated by the above findings, we utilized the fine-tuned CNN model as the feature extractor in the following experiments.

The results of the acoustic features and different feature fusion strategies are summarized in Table 2. As expected, the use of both MFCC and LPCC was proven feasible in the task of pig sound classification, and the results of the two features were comparable. 'Acoustic' in the fourth line represents the fusion of MFCC and LPCC. It shows that the performance of Acoustic was superior to that of the two separate acoustic features. Moreover, the classification results were all increased by fusing Acoustic with the other three image features. Based on the above findings, it is evident that the exploitation of inter-feature variability can compensate for the limitations of a single feature. As a result, each fusion strategy further enhances recognition performance. It is worth noting that model accuracy failed to improve and remained stable as more image features were added. At the same time, it required an increase in computation.

The results of the proposed classifier fusion method are shown in Table 3. As can be seen, there is a significant increase in each evaluation indicator for the results of the proposed classifier fusion compared with the results of Table 1. An exciting classification accuracy of 99.20 % was achieved by the fusion of Acoustic and mel-spectrogram features. For spectrogram and CQT features, good performance was also shown after the classifier fusion. Comparing the results in Table 2 and Table 3, it is not difficult to see that the proposed classifier fusion algorithm outperforms the feature fusion method in the pig cough sound recognition task.

Table 1
Performance of the pre-trained and fine-tuned feature extractors for different input images without any fusion.

Feature extractor	Features	Accuracy	Recall	Precision	F1-score
Pre-trained	spectrogram	92.67 %	93.60 %	91.88 %	92.73 %
	CQT	92.93 %	96.00 %	90.45 %	93.14 %
	mel-spectrogram	95.47 %	95.47 %	95.47 %	95.47 %
Fine-tuned	spectrogram	95.87 %	96.80 %	95.03 %	95.90 %
	CQT	94.93 %	96.27 %	93.77 %	95.00 %
	mel-spectrogram	97.33 %	97.87 %	96.83 %	97.35 %

Table 2
Performance of the acoustic features and the proposed feature fusion method.

Features	Accuracy	Recall	Precision	F1-score
MFCC	94.67 %	95.47 %	93.96 %	94.71 %
LPCC	95.51 %	95.29 %	95.31 %	94.92 %
Acoustic	96.85 %	96.31 %	96.22 %	96.90 %
Acoustic + spectrogram	96.40 %	96.80 %	96.03 %	96.41 %
Acoustic + CQT	95.20 %	97.07 %	93.57 %	95.29 %
Acoustic + mel-spectrogram	97.47 %	98.40 %	96.60 %	97.49 %
Acoustic + spectrogram+ mel-spectrogram	97.60 %	98.40 %	96.85 %	97.62 %
Acoustic + CQT+ mel-spectrogram	97.60 %	97.60 %	97.60 %	97.60 %
Acoustic + CQT+ spectrogram	97.60 %	97.60 %	97.60 %	97.60 %

Table 3
Performance of the proposed classifier fusion method.

Features	Accuracy	Recall	Precision	F1-score
MFCC + melsepctrogram	98.13 %	98.67 %	97.63 %	98.14 %
LPCC + melsepctrogram	98.40 %	98.93 %	97.89 %	98.41 %
Acoustic + mel-spectrogram	99.20 %	99.47 %	98.94 %	99.20 %
Acoustic + spectrogram	98.80 %	99.47 %	98.16 %	98.81 %
Acoustic + CQT	98.13 %	99.20 %	97.13 %	98.15 %

5. Discussion

In our previous research (Shen et al., 2022), we found that fusion of the acoustic features and CQT and spectrogram image features achieved a better improvement than the single feature in accuracy. In this present research, we improved the algorithm by using the mel-spectrogram image feature instead of CQT and spectrogram features, as well as switching to a fine-tuned CNN model as the feature extractor. Our results indicate that the improved feature fusion method performed similarly compared to the previous feature fusion strategy. However, computation load decreased significantly. Since it has been proven that there was no gain from deep convolutional networks, the issue is not discussed at extent in this paper. We have also proven that more fully-connected layers would lead to an even worse performance in the pig cough sound recognition task. Hence, in this study, we removed the last two fully connected layers.

In the classifier fusion algorithm, we chose the 'RBF' kernel function of SVM for classification. We also tried other kernel functions such as 'linear' and 'polynomial'. The results showed that 'RBF' performed slightly better than 'linear' and was comparable to 'polynomial' for pig sound features. For acoustic features, we also considered the fusion of the first order and the second order differential of MFCC and LPCC. There was no gain or even a loss when fusing more acoustic features. More image features were also considered and fused to the final decision in classifier fusion. This did not bring any performance improvement, except for increasing computation.

In this current research, we treated the CNN model as a feature extractor to obtain the deep features of the image input and an SVM classifier was used to make the classification (CNN + SVM). As we know, a CNN model can also make a classification by using the softmax layer (CNN + softmax). In this study, we found that the framework of CNN +

SVM performed better than CNN + softmax. There were 1.10 %~ 2.66 % accuracy increases for the fine-tuned CNN + SVM. The pre-trained CNN + SVM outperformed CNN + softmax slightly.

For the proposed classifier fusion strategy, we used different feature representations and frameworks to create a classifier fusion, which essentially implied feature fusion. Therefore, the results reveal that classifier fusion performed better than feature fusion. This provides us insights and ideas for future research. It is difficult to improve classification performance only by fusing more features. Selecting an optimized classifier algorithm for different features and fusing different classifiers may further improve classification accuracy.

Although our proposed method achieved high classification accuracy, the results were only tested on our private dataset. In the future, we will use more new data collected in the real pig barns to test the algorithm, and we hope that the algorithm can be verified by more different datasets by other researchers. Some other limitations still stand in our experiments and the limitations have been described and discussed in our previous work (Shen et al., 2022).

6. Conclusions

In this paper, we investigated the feature fusion and classifier fusion strategies for boosting pig cough recognition. We found that the acoustic features of MFCC and LPCC showed a good performance in pig cough recognition and the fusion of two features improved the accuracy significantly. The image feature of mel-spectrogram performed better than CQT and spectrogram in the pre-trained CNN model. Compared with the previous feature fusion algorithm, the proposed fusion of acoustic and mel-spectrogram features reduced the computational load while maintaining high accuracy. The proposed classifier fusion showed a better performance than feature fusion, achieving an accuracy of 99.20 %. In the future research, more representative features and classifiers are deserved to be explored to gain more robust recognition.

CRediT authorship contribution statement

Yanling Yin: Conceptualization, Software, Methodology, Writing – original draft, Investigation, Formal analysis, Funding acquisition. **Nan Ji:** Writing – review & editing, Methodology, Visualization. **Xipeng Wang:** Software, Methodology, Investigation. **Weizheng Shen:** Conceptualization, Methodology, Funding acquisition, Investigation. **Baisheng Dai:** Software, Funding acquisition. **Shengli Kou:** Supervision, Data curation. **Chen Liang:** Supervision, Data curation.

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.

Data availability

No data was used for the research described in the article.

Acknowledgements

Funding: This work was supported by the project of the National Natural Science Foundation of China [grant numbers 32172784, 31902210]; the Academic Backbone Project of Northeast Agricultural University; the National Key Research and Development Program of China [grant number 2019YFE0125600]; the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province [grant number UNPYSCT-2020092]; and the Earmarked fund [CARS-36, CARS-35].

References

Ahmed, A.I., Chiverton, J.P., Ndzi, D.L., Becerra, V.M., 2019. Speaker recognition using PCA-based feature transformation. Speech Comm. 110, 33–46.

Atmaja, B.T., Sasou, A., Akagi, M., 2022. Survey on bimodal speech emotion recognition from acoustic and linguistic information fusion. Speech Comm. 140, 11–28.

Cheng X., Tan L., Ming F., 2021. Feature fusion based on convolutional neural network for breast cancer auxiliary diagnosis. Mathematical Problems in Engineering. 2021, ID 7010438.

Chi, M., 2021. SAFFNet: Self-Attention-Based feature fusion network for remote sensing few-shot scene classification. Remote Sens. (Basel) 13, 2532.

Chung, Y., Oh, S., Lee, J., Park, D., Kim, S., 2013. Automatic detection and recognition of pig wasting diseases using sound data in audio surveillance systems. Sensors 13, 12929–12942.

Deng, J., Dong, W., Socher, R., Li, L.J., Kai, L., Li, F.F., 2009. ImageNet: A large-scale hierarchical image database. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2009, 248–255.

Er, M.B.B., 2020. A novel approach for classification of speech emotions based on deep and acoustic features. IEEE Access 8, 221640–221653.

Exadaktylos, V., Silva, M., Ferrari, S., Guarino, M., Taylor, C.J., Aerts, J.M., Berckmans, D., 2008a. Time-series analysis for online recognition and localization of sick pig (Sus scrofa) cough sounds. J. Acoust. Soc. Am. 124, 3803–3809.

Exadaktylos, V., Silva, M., Aerts, J.M., Taylor, C.J., Berckmans, D., 2008b. Real-time recognition of sick pig cough sounds. Comput. Electron. Agric. 63, 207–214.

Guarino, M., Jans, P., Costa, A., Aert, J.M., Berckmans, D., 2008. Field test of algorithm for automatic cough detection in pig house. Comput. Electron. Agric. 62, 22–28.

Hassan, M.F., Abdel-Qader, I., Bazuin, B., 2021. A new method for ensemble combination based on adaptive decision making. Knowl.-Based Syst. 2021 (233), 107544.

Hirtum, A.V., Berckmans, D., 2003. Fuzzy approach for improved recognition of citric acid induced piglet coughing from continuous registration. J. Sound Vib. 266, 677–686.

Hirtum, A.V., Guarino, M., Costa, A., Jans, P., Hhesquiere, K., Aerts, J.M., Navarotto, P. L., Berckmans, D., 2003. Automatic detection of chronic pig coughing from continuous registration in field situations. In: 3rd International Workshop MAVEBA, pp. 251–254.

Ji, N., Shen, W., Yin, Y., Bao, J., Dai, B., Hou, H., Kou, S., Zhao, Y., 2022. Investigation of acoustic and visual features for pig cough classification. Biosyst. Eng. 219, 281–293.

Knight, E.C., Poo Hernandez, S., Bayne, E.M., Bulitko, V., Tucker, B.V., 2020. Preprocessing spectrogram parameters improve the accuracy of bioacoustic classification using convolutional neural networks. Bioacoustics 29, 337–355.

Krizhevsky A., Sutskever I., Hinton G. E., 2012. ImageNet classificationwith deep convolutional neural networks. The Proceedings of the 25th International Conference on Neural Information Processing System. 1097-1105.

Racewicz, P., Ludwiczak, A., Skrzypczak, E., Składanowska-Baryza, J., Biesiada, H., Nowak, T., Nowaczewski, S., Zaborowicz, M., Stanisz, M., Ślósarz, P., 2021. Welfare health and productivity in commercial pig herds. Animals 11, 1176.

Sharma, G., Umapathy, K., Krishnan, S., 2020. Trends in audio signal feature extraction methods. Appl. Acoust. 158, 107020.

Shen W., Tu D., Yin Y., et al.,2021.A new fusion feature based on convolutional neural network for pig cough recognition in field situations. Inf. Process. Agric. 8, 573–580.

Shen, W., Hu, H., Dai, B., Wei, X., Sun, J., Jiang, L., Sun, Y., 2020. Individual identification of dairy cows based on convolutional neural networks. Multimed. Tools Appl. 75, 14711–14724.

Shen, W., Ji, N., Yin, Y., Dai, B., Tu, D., Sun, B., Hou, H., Kou, S., Zhao, Y., 2022. Fusion of acoustic and deep features for pig cough sound recognition. Comput. Electron. Agric. 197, 106994.

Tin L. N., Say W. F. Li yanage C. D. S., 2003. Speech emotion recognition using hidden Markov models. Speech Commun. 41, 603–623.

Wang, B., Li, H., You, J., Chen, X., Yuan, X., Feng, X., 2022. Fusing deep learning features of triplet leaf image patterns to boost soybean cultivar identification. Comput. Electron. Agric. 197, 106914.

Xu, L., Wei, Z., Zaidi, S.F.A., Ren, B., Yang, J., 2021. Speech enhancement based on nonnegative matrix factorization inconstant-Q frequency domain. Appl. Acoust. 174, 107732.

Yang, L., Tao, L., Chen, X., Gu, X., 2020. Multi-scale semantic feature fusion and data augmentation for acoustic scene classification. Appl. Acoust. 163, 107238.

Yao, Z., Wang, Z., Liu, W., Liu, Y., Pan, J., 2020. Speech emotion recognition using fusion of three multi-task learning-based classifiers: HSF-DNN MS-CNN and LLD-RNN. Speech Commun. 2020 (120), 11–19.

Yin, Y., Tu, D., Shen, W., Bao, J., 2021. Recognition of sick pig cough sounds based on convolutional neural network in field situations. Inf. Process. Agric. 8, 369–379.

Zhang, S., Tao, X., Chuang, Y., Zhao, X., 2020. Learning deep multimodal affective features for spontaneous speech emotion recognition. Speech Comm. 127, 73–81.

Zhao, J., Li, X., Liu, W., Gao, Y., Lei, M., Tan, H., Yang, D., 2020. DNN-HMM based acoustic model for continuous pig cough sound recognition. Int. J. Agric. Biol. Eng. 13, 186–193.

Zhao, S., Xu, T., Wu, X.J., Zhu, X.F., 2021. Adaptive feature fusion for visual object tracking. Pattern Recogn. 111, 107679.