Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Investigation of acoustic and visual features for pig cough classification

Nan Ji [a], Weizheng Shen [a], Yanling Yin [a,*], Jun Bao [b], Baisheng Dai [a], Handan Hou [c], Shengli Kou [d], Yize Zhao [e]

[a] School of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China
[b] School of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China
[c] School of Computer Science, Harbin Finance University, Harbin 150030, China
[d] Science and Technology Office, Northeast Agricultural University, Harbin 150030, China
[e] Department of Computer Science, Donald Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, CA, USA

ARTICLE INFO

Article history:
Received 3 December 2021
Received in revised form
10 March 2022
Accepted 16 May 2022
Published online 1 June 2022

Keywords:
Pig cough
Acoustic features
Visual features
Support vector machine

The precise detection of pig cough is a crucial step for establishing an early warning system for pig respiratory diseases. With regard to high precision pig cough recognition, feature extraction and selection are of importance. However, few studies have investigated both acoustic and visual features of pig vocalisations as input features. In this paper, we proposed a novel feature fusion method which fusing acoustic and visual features to achieve an enhanced pig cough recognition rate. We firstly extracted acoustic features from audio signals, including root-mean-square energy (RMS), mel-frequency cepstral coefficients (MFCCs), zero-crossing rates (ZCRs), spectral centroid, spectral roll-off, spectral flatness, spectral bandwidth and chroma. Then, constant-Q transform (CQT) spectrograms were employed to extract visual features involving local binary pattern (LBP) and histogram of gradient (HOG). Subsequently, a hybrid feature set was created by combining acoustic and visual features. In this stage, Pearson correlation coefficient (PCC), recursive feature elimination based on random forest (RF-RFE) and principal component analysis (PCA) were exploited for dimensionality reduction. Finally, support vector machine (SVM), random forest (RF) and k-nearest neighbours (KNN) classifiers were used to conduct a performance evaluation. It is shown that the fused acoustic features (Acoustic) combined with LBP and HOG (A-LH) achieved 96.45% pig cough accuracy. The results reveal that the fusion feature set outperforms acoustic and visual features alone.

* *Corresponding author.*
E-mail address: yinyanling@neau.edu.cn (Y. Yin).
https://doi.org/10.1016/j.biosystemseng.2022.05.010

Nomenclature

Acoustic	Combination of MFCCs and CF
A-LH	Acoustic combined with LBP and HOG
AUC	area under the curve
B	The number of bins per octave
Bandwidth	Spectral bandwidth
Centroid	Spectral centroid
CF	Combination of Bandwidth, Centroid, Chroma, Flatness, RMS, Rolloff, and ZCR
CQT	Constant-Q transform
f_1	Center frequency of the lowest-frequency bin
Flatness	Spectral flatness
FN	False negative
FP	False positive
FPR	False positive rate
HOG	Histogram of oriented gradients
KNN	K-nearest neighbours
L	Window size
LBP	Local binary pattern
MFCCs	Mel-frequency cepstral coefficients
Nb. Orient	The number of orientation bins
P	Neighbouring pixel
PCA	Principal component analysis
PCC	Pearson correlation coefficient
PFI	Permutation feature importance
R	Distance between the central pixel and a neighbouring pixel
RF	Random forest
RF-RFE	Recursive feature elimination based on random forest
RGB	Red, green, blue values of RGB colour space
RMS	Root mean square energy
ROC	Receiver operating characteristic
Rolloff	Spectral rolloff
STFT	Short time Fourier transform
TFR	Time frequency representation
SVM	Support vector machine
TP	True positive
TPR	True positive rate
TN	True negative
TTR	Training and testing split rate
ZCR	Zero crossing rate

1. Introduction

Pig cough recognition is essential for the early detection of respiratory infections, as cough-induced illnesses cause substantial pig production losses (Benjamin & Yik, 2019; Racewicz et al., 2021). Conventional detection method is time-consuming, as veterinarians or experienced breeders must enter the barn frequently to check the health of pigs. Therefore, an automatic pig cough sound recognition method needs to be developed to achieve contactless and rapid respiratory disease detection (Wang et al., 2019).

Recently, many pig vocalisation recognition and classification methods have been developed. According to previous studies, pig cough recognition involves four steps: bioacoustics signal preprocessing, voice activity detection, feature extraction and classification. In this paper, we mainly concentrate on the last two steps, feature extraction and classification. Obtaining discriminative and independent features is commonly regarded as a key step toward achieving a high classification rate (Zebari et al., 2020). Studies have shown that feature extraction follows two major trends. One involves extracting acoustic features from a one-dimensional sound signal. Various features are calculated and selected during this process. Among them, mel-frequency cepstral coefficients (MFCCs) have been extensively applied in pig sound classification tasks (Chung et al., 2013; Zhao et al., 2020). However, the performance of MFCCs degrades rapidly in noisy environments, which has been demonstrated in Zhao et al. (2019). Several other acoustic features have also been extensively explored in previous studies. Power spectral density was applied as a representative frequency characteristic for the real-time recognition of pig coughs in Exadaktylos et al. (2008). Additionally, Xie et al. (2020) proposed an ensemble of time-frequency features obtained from the decomposition of the audio signals in various frog calls. Similarly, a combined feature set composed of MFCCs, onset strengths, and tempograms was applied to distinguish horse gaits in Alves et al. (2021). Furthermore, spectrum shape-based features and average signal energy were described in terms of anuran vocalisation (Huang et al., 2014). The mentioned methods could perform well under certain conditions and provide new insights into the characteristics of various sounds. However, considering that pig sounds are different from those of other animals, we should also explore the acoustic characteristics of pigs individually and in-depth to select the most representative set of features.

Another trend aims at transforming acoustic signals into time-frequency representations (TFRs) from which visual features can be extracted. In this context, short-time Fourier transform (STFT) spectrograms are widely used in classification tasks (Stowell et al., 2015). Demir et al. (2018) investigated low-level texture features based on STFT for snore sound recognition. Recently, it has been proven that CQT is better than STFT under the low signal-to-noise ratio (SNR) condition in Xu et al. (2021). This motivates us to investigate the effectiveness of the CQT with respect to pig cough recognition. Typically, visual features, such as texture-based features, are directly computed from TFRs. Local binary pattern (LBP) (Huang et al., 2018; Li et al., 2019) and histogram of oriented gradients (HOG) (Xiang et al., 2018) are two representative texture features. Currently, although these methods have been proven to be feasible, there are still many challenges to overcome. Primarily, the effects of different pooling strategies on TFRs vary widely for different sounds from Rakotomamonjy and Gasso (2015) to Xie and Zhu (2019); thus, it makes sense to dig deeper into representative pig sound descriptors. Additionally, the performance of such methods is limited to dimensionality issues due to the number of required audio segments. Therefore, to ameliorate the accuracy of learning features and to decrease the training time, dimensionality reduction is utilised to eliminate redundant features. Jiang et al. (2020) developed a feature selection algorithm based on the least absolute

shrinkage and selection operator technique to select representative characteristics from a high-dimensional original feature space for acoustic traffic scene recognition. Principal component analysis (PCA) was selectively conducted in Xie et al. (2020) to achieve improved frog recognition performance. In general, since acoustic and visual features reflect various characteristics of sound signals at different scales, it provides an opportunity to increase the discrimination of feature sets by aggregating acoustic features with visual features. The final step is to compare the obtained feature values for classification. Machine learning techniques are commonly employed for sound comparisons due to their great improvements, including support vector machine (SVM), k-nearest neighbours classifiers (Xie & Zhu, 2019), and decision trees (Acevedo et al., 2009).

In this study, both acoustic and visual features are used to classify pig coughs and non-coughs. A hybrid feature vector is created by extracting visual features from TFRs and acoustic features from sound records. The fundamental contributions of this study are as follows. Firstly, to our knowledge, it is the first time to investigate both acoustic and visual features for pig cough classification. In particular, we employ feature selection to choose the optimal acoustic features set to enrich the visual features for better discriminating pig coughs. Furthermore, a novel hybrid feature set is proposed based on feature fusion, which can capture more information of pig sound characteristics. Lastly, increased classification accuracy is achieved due to the advantages of acoustic and visual feature fusion.

The remainder of this paper is organised as follows. In Section 2, the involved methods are described, including both acoustic and visual features extraction, dimensionality reduction techniques and machine learning schemes. Apart from these methods, data collection and evaluation criteria are also provided. Section 3 presents the results of our experiments. A discussion of the related results is given in Section 4. Finally, Section 5 summarises the conclusion and future work.

2. Materials and methods

2.1. Materials

2.1.1. Animals and housing

In this study, data were collected in a large commercial pig house in Harbin, Heilongjiang Province, China. One hundred and twenty-eight pigs from crossbred fattening-stage (120 d, ~60 kg) of the Northeast Folk and the Great White were reared in the barn. The schematic diagram of sound collection in pig house is shown in Fig. 1. The size (length × width × height) of the barn was 27.5 m × 13.7 m × 3.2 m. The barn was subdivided into 21 pens, 12 of which were 4.15 m × 3.6 m (length × width) in two adjacent columns and 9 of which were 3.6 m × 2.75 m (length × width) in one column, with a half-slatted floor. The number in each pen represented the number of pigs in Fig. 1. Two sick pigs with heavy coughs were separated in Pen 13, and there were also other coughing sounds occurred in other pens.

2.1.2. Data collection and preprocessing

The sound data were recorded using a microphone (LIQI LM320E, Cardioid electret microphone) connected to the sound card (Conexant Smart Audio HD) of a laptop. The microphone was fixed in a larger pen (Pen 7) adjacent to the door at a height of 1.4 m (approximately 0.8 m from the backup of pigs). The recordings were made at a sampling rate of 44.1 kHz with a resolution of 16 bits. The recordings including coughs, pig sounds and other sounds, were extracted and labelled with the assistance of a veterinary expert in a manual-labelled way. The expert labelled the sound segments by observing the sound waveforms and conducting the auditive confirmation. Then, the sounds were passed through a 10th-order Butterworth filter with a cutoff frequency of 100–16,000 Hz. In total, 3157 individual sounds were extracted from the recordings, including 1884 coughs and 1273 non-cough segments. Non-cough sounds consisted of 957 pig sounds without coughs, 188 clearing sounds produced by shovels, 100 water flowing sounds and 28 human sounds. The whole dataset was split into training set and testing set with a training testing split ratio (TTR) of 6:4. Then, we employed ten-fold cross validation to evaluate the proposed approach on the training set to identify the best parameters.

The testing set was evaluated by using Accuracy, Recall, Precision, the F1-score, and the area under the curve (AUC) (Knight et al., 2017). For the AUC, both precision-recall (PR) and receiver operating characteristic (ROC) curves were plotted for the classifier (Davis & Goadrich, 2006). ROC curves typically feature the true positive rate (TPR) on the y-axis and the false positive rate (FPR) on the x-axis. Unlike the ROC curve, the PR curve is not monotonic. A PR curve shows the tradeoff between precision and recall. A high AUC represents both high recall and high precision. PR-AUC and ROC-AUC are intuitive and effective measures for evaluating the quality of classifier outputs. The closer the value of AUC is to 1, the better the classifier is (Saito & Rehmsmeier, 2015). These metrics were calculated by using formulas (1)–(5). Here, we defined a cough as a positive sample, and a noncough was defined as a negative sample.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Recall\ (TPR) = TP/(TP + FN) \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

$$F1 - score = 2 \times (Precision \times Recall)/(Precision + Recall) \quad (4)$$

$$FPR = FP/(FP + TN) \quad (5)$$

2.2. Method

In this study, we aimed to aggregate acoustic and visual features to complete pig cough recognition tasks effectively. The conceptual framework of the research is illustrated in Fig. 2. Firstly, the audio recordings acquired in the pig house were preprocessed and segmented into sound segments. Subsequently, acoustic features were calculated directly from the preprocessed sound signals. Visual features including LBP and

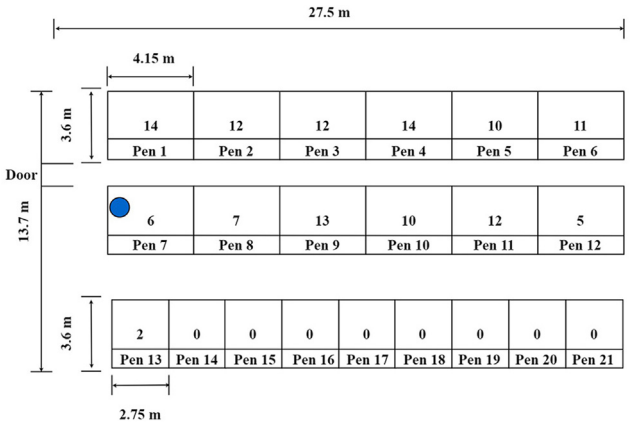

Fig. 1 – The schematic diagram of sound collection in pig house. The blue dot in Pen 7 represents the location of microphone. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

HOG were extracted from the CQT spectrograms, which had been converted by constant-Q transform from sound segments. To minimise dimensionality, Pearson correlation coefficient (PCC), recursive feature elimination based on random forest (RF-RFE) and PCA were applied for feature selection and extraction. A hybrid feature vector was obtained by integrating the obtained visual features and acoustic features through early fusion. For pig cough recognition, SVM, KNN and RF classifiers were trained for selecting the optimal classifier by employing grid search method in this paper. The details of the proposed method are described in the subsections below.

2.2.1. *Acoustic features*

Acoustic features are physical characteristics in terms of frequencies, loudness levels and amplitudes (Sharma et al., 2020). All the acoustic features were extracted from preprocessed audio segments by utilising LibROSA toolbox. Specifically, RMS (Er, 2020) and ZCR (Kedem, 1986) are extracted from time domains. MFCCs (Fu et al., 2011) is the most frequently used features in cepstral domain. In addition, the other five features are captured from frequency domains, including Centroid (Flores-Fuentes et al., 2014), Flatness (Ma & Nishihara, 2013), Bandwidth (Xie et al., 2020), Rolloff (Luz et al., 2021), and Chroma (Er, 2020). The description of acoustic features is shown in Table 1.

2.2.2. *Visual features*

LBP (Yang & Chen, 2013) and HOG (van der Walt et al., 2014) were employed as visual features for pig cough classification. This section is composed of two main components: the construction of CQT spectrograms from pig sounds and feature extraction based on LBP and HOG descriptors.

2.2.2.1. *CQT spectrograms*. Due to the nonstationary nature of sound, a large number of studies on sound recognition problems have used TFRs, especially STFT representations (Knight et al., 2020; Lim et al., 2019). In this study, the CQT was chosen for recognising pig cough sounds. In contrast to the STFT, this transform provides frequency analysis on a logarithmic scale, which makes it more suitable for sound representation. The process of this transformation can be seen in the following equations. The center frequency to bandwidth ratio is a constant value Q, which is calculated by Eq. (6).

$$Q = \frac{f_k}{\delta f_k} \tag{6}$$

$$f_k = f_1 2^{\frac{k-1}{B}} \tag{7}$$

where δf_k and f_k are the bandwidth and the central frequency of the kth filter, respectively. f_k is calculated by Eq. (7). f_1 is the center frequency of the lowest-frequency bin, and B determines the number of bins per octave.

$$Q = \frac{1}{2^{1/B} - 1} \tag{8}$$

In the CQT, the number of bins per octave (B) is related to the fidelity factor (Q) from Eq. (8). Next, we assume that N_k is the window length that changes with frequency, and f_s represents the sampling frequency. N_k is calculated in Eq. (9).

$$N_k = \frac{f_s}{\delta_{f_k}} \tag{9}$$

$$X^{CQT}(k) = \frac{1}{N_k} \sum_{n=0}^{N_k - 1} x(n) w(n,k) e^{-j\frac{2\pi Q n}{N_k}} \tag{10}$$

where $X^{CQT}(k)$ is the k component of the constant Q transforms, $x(n)$ is the input signal, and $w(n,k)$ is the window

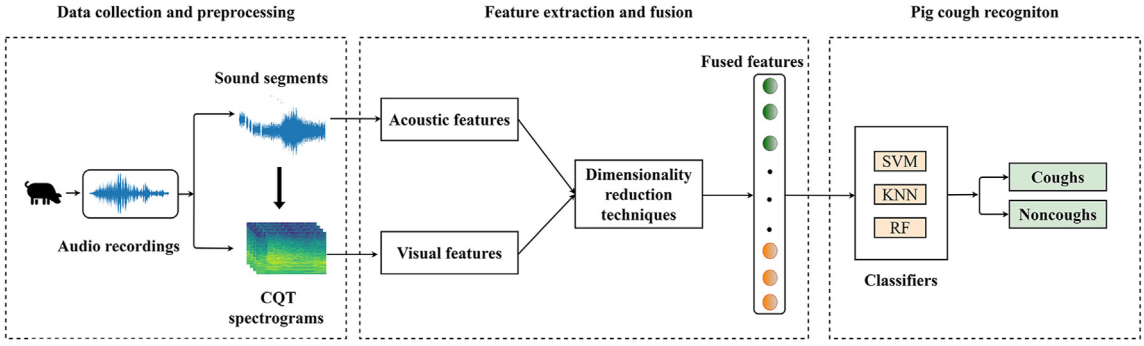

Fig. 2 – Flowchart of the pig cough recognition with the proposed method.

Table 1 – The description of acoustic features.

Feature	Dimension	Domain
MFCC	20	Cepstral
RMS	1	Time
ZCR	1	Time
Spectral bandwidth	1	Frequency
Spectral centroid	1	Frequency
Spectral flatness	1	Frequency
Spectral rolloff	1	Frequency
Chroma	1	Frequency

function with a length of N_k. As shown in Eq. (10), the transform shows that the CQT possesses its own characteristics, achieving both high-frequency resolution at low frequencies and high temporal resolution at high frequencies. Figure 3 depicts the spectrograms of the conventional STFT and CQT for pig cough segments. It is evident that the CQT spectrogram is able to detect even small variations in the spectrum. It can also be observed that the CQT spectrogram tends to be clearer and sharper than that of the STFT that in high-frequency regions. These discriminating characteristics motivate us to use CQT spectrogram image representation. The CQT spectrograms were saved at 100 × 100 pixels to obtain the TFRs with a uniform size for calculating the visual features below.

2.2.2.2. *LBP descriptors*. LBP is an effective feature extraction method for analysing image textures due to its robustness to grayscale invariance (Abidin et al., 2017). The calculation of LBP is given in Eq. (11):

$$LBP_{P,R} = \sum_{p=0}^{P-1} s\left(g_{p,R} - g_c\right) 2^p \tag{11}$$

LBP considers each pixel of an image, and it is calculated by comparing the grayscale value of each central pixel g_c with those of its neighbouring pixels g_p. Here, the distance between the central pixel and a neighbouring pixel P is denoted by R. The result of $LBP_{P,R}$ is shown in Eq. (12) and the result is converted to the binary of 1 or 0.

$$s(x) = \begin{cases} 1, \ if \ x \geq 0 \\ 0, \ otherwise \end{cases} \tag{12}$$

By varying the values of P and R, a fine-grained analysis of TFR texture patterns is introduced. We choose four (P, R) pair values for a uniform LBP: (8,1), (12,2), (16,1) and (16,2). Among them, (8,1) and (12,2) are the two best-performing common pair values. Generally, the LBP values of each pixel within an image are consolidated into a histogram as a texture descriptor (Sengupta et al., 2017). Additionally, it is meaningful to know that the histogram contains information about the distribution of the LBP features over the TFRs. Thus, analysing the LBP histogram is an intuitive approach for investigating and forecasting the classification and recognition task performance of two TFRs. Parts of the LBP histograms for the STFT and CQT spectrograms obtained with $LBP_{12,2}$ are presented in Fig. 4. Here, three common sounds, a pig cough, a pig scream and a water sound, were chosen for comparison. It can be observed that both TFRs have distinctive advantages among the three different sound segments with the LBP descriptor. Therefore, we can forecast that both TFRs are feasible for pig cough identification.

2.2.2.3. *HOG descriptors*. As another texture descriptor, HOG is calculated based on the adjacent frequency bin gradients in the time and frequency directions, which makes it possible to capture audio signals along the time axis (Rakotomamonjy & Gasso, 2015). A schematic diagram of the HOG descriptor is shown in Fig. 5 and the specific flowchart is introduced as follows. Firstly, the original colour CQT spectrograms of sound segments are converted into grayscale images. Secondly, the gradient (including its size and direction) of each pixel in the image is calculated to extract the effective information contained in the edges and outlines of the images. Then, the image is divided into small nonoverlapping cells, and the descriptor of each cell can be formed by counting the gradient histograms of each cell. Here, local groups of cells form blocks. Subsequently, the histogram vectors in the block are normalised by L2-Hys. L2-Hys is a method using L2-norm whose maximum values are limited to 0.2 and be renormalised by the L2-norm. Finally, HOG descriptors are obtained from all blocks in a dense grid of blocks. Because interference factors such as illumination were not involved in this study, normalisation was not performed on the input image. In addition, the number of orientation bins (Nb. Orient), the size of a cell and the number of cells in each block are adjustable parameters, which may influence the resulting recognition performance. Therefore, we correspondingly constructed a series of HOG feature descriptors and discussed them in detail below.

The gray-level CQT spectrogram of pig coughs and the corresponding HOG descriptors are shown in Fig. 6. As depicted in Fig. 6, the HOG descriptors are capable of sharply capturing the directions of local variations in the power spectrum. When the cell size is (16 × 16) with the same Nb. Orient (9) and block size (2 × 2) settings, the HOG representation of the signal displays a stronger capability of producing fine-grained discriminatory features. Furthermore, the features are approximately invariant to small time and frequency translations. As shown in Fig. 6(b), only subtle and discrete changes can be acquired. Therefore, it is crucial to choose appropriate cell sizes, block sizes and numbers of orientations for feature extraction. Different HOG descriptors have direct impacts on recognition performance. In this study, we chose Nb. Orient values of 8 and 9. (3 × 3) and (2 × 2) were chosen as the block sizes. The cell sizes were set as (4 × 4), (8 × 8), (16 × 16), and (32 × 32) for comparison in subsequent experiments.

2.2.3. *Dimensionality reduction techniques*

Generally, dimensionality reduction techniques can be classified into two main groups: feature selection and feature extraction. In this paper, we introduced a feature selection method combining PCC (Trzcinska et al., 2020) and RF-RFE (Demarchi et al., 2020) to select the optimal subset of acoustic features based on training set. The overall process of feature selection is illustrated in Fig. 7.

As shown in Fig. 7, PCC was firstly calculated among all the features extracted from the audio segments. The PCC is based on the relevance (predictive power) of each feature, which

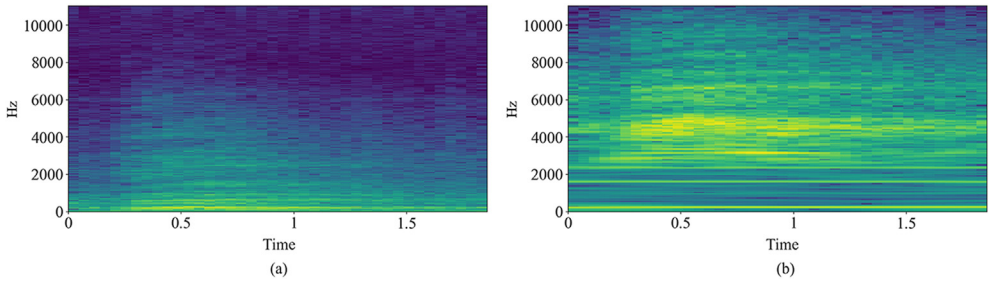

Fig. 3 – Two-dimensional spectrograms of cough. (a) Cough spectrum obtained using the STFT. (b) Cough spectrum obtained using the CQT.

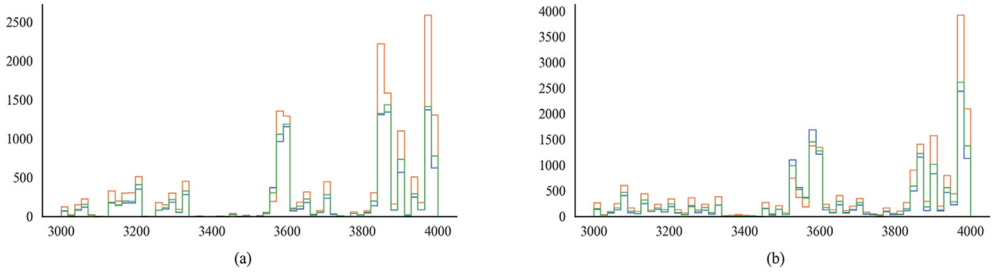

Fig. 4 – (a) LBP histogram based on CQT spectrograms and (b) LBP histogram based on STFT spectrograms. The bars of blue, orange and green represent cough, scream, and water sounds, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

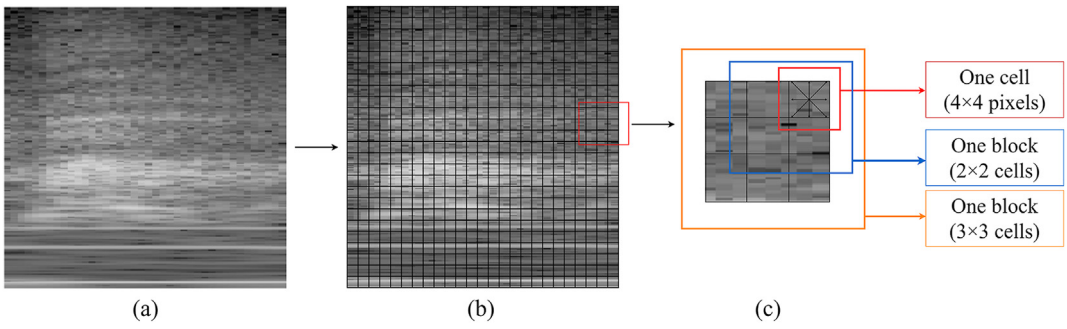

Fig. 5 – Schematic representation of the HOG descriptor. (a) Representation of the gray CQT spectrogram of a pig cough. The CQT spectrogram was divided into 25 × 25 grids in (b). One cell size with 4 × 4 pixels was obtained, while two different blocks were separately chosen with sizes of (2 × 2) and (3 × 3) in (c).

measures the level of linear correlation between two variables. The PCC is defined in Eq. (13):

$$PCC = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}} \quad (13)$$

where $x_i$and y_i are sample observations of x and y, respectively. $\overline{x}$ and $\overline{y}$ are the average values of x and y. The PCC ranges from −1 to 1, with a coefficient close to 0 indicating no linear correlation. Furthermore, 1 represents a positive correlation, while −1 represents a negative correlation. After calculating the PCC, the features that have no correlations with other features are analysed independently in the follow-up experiments. Additionally, the other remaining features are further selected by RF-RFE. Due to its unbiased and stable results in different fields, the RF has been proven to perform well in terms of predictive accuracy (Jeon & Oh, 2020). The key aspect of RFE is to select the least number of features with the best predictive accuracy by calculating feature importance. In this study, permutation feature importance (PFI) was employed to measure feature importance on training dataset since PFI favours low-cardinality features such as categorical variables with small numbers of possible categories. PFI is defined as the decrease in model accuracy when a single feature value is randomly shuffled (Huang et al., 2016). The drop in the model accuracy is related to the importance of the feature. The importance i_x for feature f_j is defined in Eq. (14):

$$i_x = a - \frac{1}{K}\sum_{k=1}^{K} a_{k,x} \quad (14)$$

where x represents each feature in the dataset, a is the accuracy of the RF, K is the number of repetitions employed to

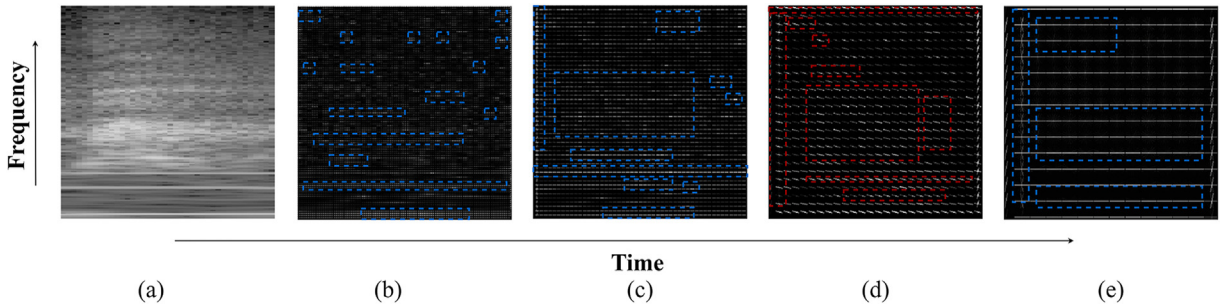

Fig. 6 – The gray-level CQT spectrogram of pig coughs, as well as the corresponding HOG descriptors. (a) represents the gray CQT spectrogram of pig coughs. Cell sizes of (4 × 4), (8 × 8), (16 × 16) and (32 × 32) are mapped to (b–e), respectively. The horizontal and vertical axes denote time and frequency, respectively.

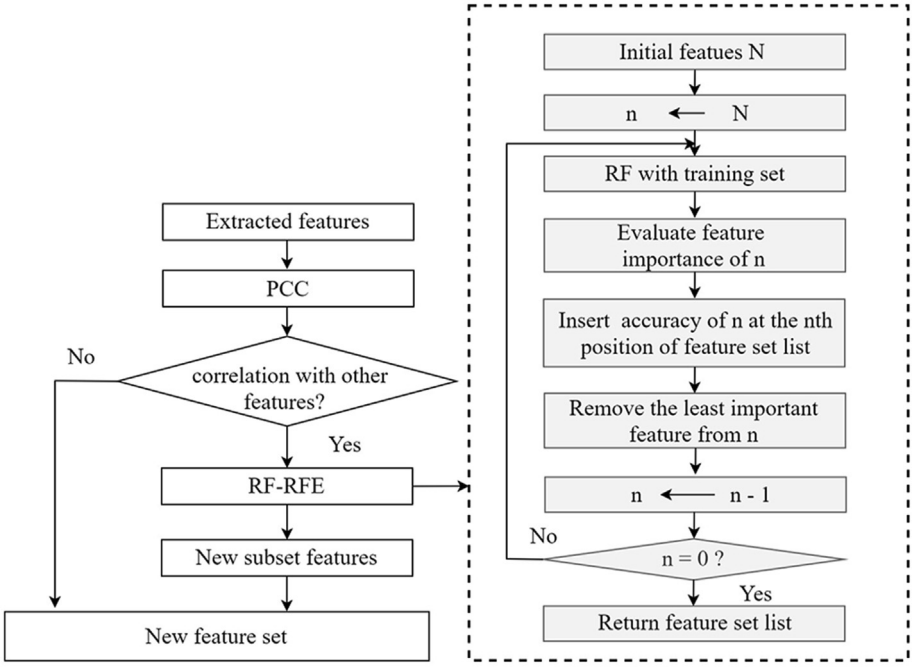

Fig. 7 – Process of the feature selection method. N denotes the number of features with correlations. n denotes the number of subset features.

permute a feature, and $a_{k,x}$ is computed by the RF on the data generated by the randomly shuffled column x of the dataset. Here, K was set to 10. From Fig. 7, the predictive accuracy of the RF was ranked according to the list of feature sets based on PFI. And then, the least important feature was removed. All the processes were executed in a loop until no features remained. Finally, the overall ranking of the feature subsets was obtained and then the optimal number of features was selected.

PCA is a popular method for decorrelating feature sets and reducing their respective dimensions. The key idea of PCA is to find orthogonal directions that represent the given data with the lowest error. PCA is calculated in Eq. (15). It is assumed that a feature set (f_1, f_2,..., f_n) can be transformed from a d-dimensional space to n vectors ($f'_1, f'_2 \ldots, f'_n$) in a new d' dimensional space.

$$f'_i = \sum_{k=1}^{d'} a_{k,i} e_k, d' \leq d \tag{15}$$

where e_k denotes the eigenvectors corresponding to the d' largest eigenvalues for the scatter matrix $S(S = E[x_i x_i^T], i = 1, \ldots n)$ and the $a_{k,i}$ are the projections of the original vectors f_i on the eigenvectors e_k. These projections are names as the principal components. In this study, PCA was applied to visual features extraction and feature fusion.

2.2.4. *Machine learning schemes*

Technological breakthroughs have recently been achieved in the domain of machine learning, yielding enhanced improvements in accuracy and utilisation (Zhang et al., 2019). To evaluate the discrimination abilities of the extracted acoustic and visual features, three frequently used models like SVM, RF and k-nearest neighbours (KNN), were applied for the pig cough recognition task.

However, numerous hyperparameters in machine learning need to be tuned and controlled properly, for preventing model from overfitting. As a consequence, we chose the grid

search method with ten-fold cross validation based on training set, to select the optimal hyperparameter settings. The main parameters for the three classifiers are shown in Table 2.

3. Results

3.1. Acoustic features

The correlation coefficients between pairs of attributes are as shown in Fig. 8. Since the calculated PCCs between MFCCs and the other seven features were quite small (all less than 0.2), MFCCs could be regarded as independent features and put into the feature set. Apart from the MFCCs, chroma showed a weak correlation with the remaining features. Additionally, strong correlations were observed between ZCR, Flatness, Rolloff, Centroid and Bandwidth. It leaded to a difficulty in performing effective and accurate feature selection. Therefore, we attempted to use RF-RFE with cross validation (RF-RFE) to find the optimal numbers of features (other than MFCCs) by eliminating possible dependencies and covariances in the model. Figure 9(a) shows boxplots comparing the mean values of seven features with the PFI algorithm. The ZCR attained the highest score (0.1768), and its span was significantly larger than those of the other diverse features. Additionally, the mean values of feature importance permuting on Rolloff (0.0848) and Flatness (0.0828) were approximated. Bandwidth was the least impactful feature, with a median value of 0.0252. According to the results in Fig. 9(a), different subsets of features were scored, and the set of features with the best accuracy was selected by RF-RFE, as shown in Fig. 9(b). The curve leaps to a better accuracy when two informative features are captured. Then, as more acoustic informative characteristics are added to the model, the accuracy gradually increases. When the number of features selected reaches seven, the highest accuracy of 0.863 is achieved. Moreover, an increasement in the number of features beyond five yields little significant improvement in the overall accuracy.

ROC and PR curves of the three classifiers with CF are shown in Fig. 10(a) and Fig. 10(b). Here, the CF includes the seven features mentioned above. The performances of the SVM and RF appeared to be comparable (0.93) in terms of their ROC-AUCs. However, it could be found that the SVM has a clear advantage over the RF, achieving a PR-AUC of 0.92. As expected, the PR curves can expose differences between algorithms that are not apparent in the ROC curves. From the results, the best pig cough performance was obtained by the SVM classifier with the parameter settings (kernel-RBF, C = 1 and Gamma = 0.1). Based on the above experiments, SVM and three feature sets in terms of the CF, MFCCs and the combination of the MFCCs and CF (named as Acoustic) were selected as baselines for comparison and subsequent analysis.

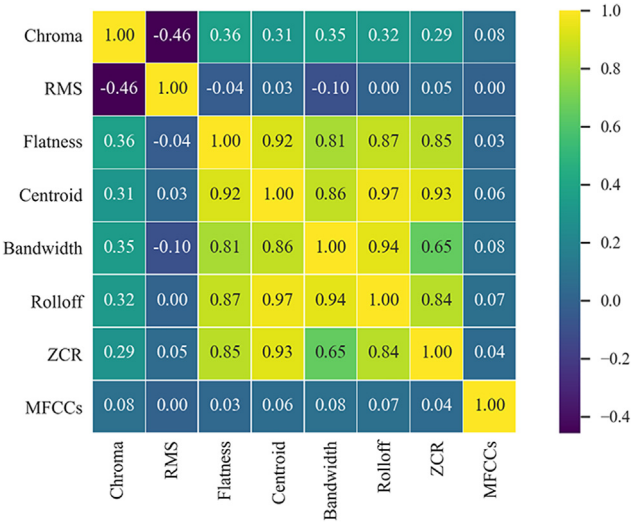

Fig. 8 – PCC diagram.

The performance of the MFCCs, CF and Acoustic with a TTR of 6:4 is shown in Table 3. The overall accuracies of the CF, MFCCs and Acoustic features were 87.25%, 94.22% and 95.49%, respectively. This indicates that all of them are capable of successfully recognising pig cough sounds. The acoustic feature set achieves the best performance compared to the other two features. It is demonstrated that feature fusion in different domains is feasible and beneficial in pig cough recognition.

3.2. Visual features

In this part, two visual features including LBP and HOG were investigated in the following experiments. On one hand, the performances of the LBP based on STFT and CQT spectrograms were firstly compared, and the multiresolution of LBP was further discussed. On the other hand, we considered the influences of different HOG parameters on pig cough recognition. Meanwhile, PCA was applied during the process to reduce the dimensions. The minimum number of principal components were computed by preserving 95% of cumulative explained variance. And then the number of visual features dimensions was reduced to 1054, for the following experiments.

To investigate the performance of LBP on the pig cough recognition task, we extracted LBP features from both STFT and CQT spectrograms. For the linear-scaled STFT spectrogram, transforms with a long window size (L = 2048 samples) and a short window size (L = 1024 samples) were both used. Here, L denoted the window size, and the hop size was fixed at L/2 in both cases. For the CQT, we set B to 32 and 64, and f_1 was set to 22.05 Hz. Here, the LBP from the gray-level TFRs

Table 2 – Grid search of the main parameters for each classifier.

Models	Ranges of hyperparameters for model
SVM	Kernel: 'linear', 'rbf'; C: 1, 10, 100; Gamma: 0.1, 0.01, 0.001
KNN	n_neighbors: (1,50)
RF	n_estimators: (10,50); min_samples_split: (2,9)

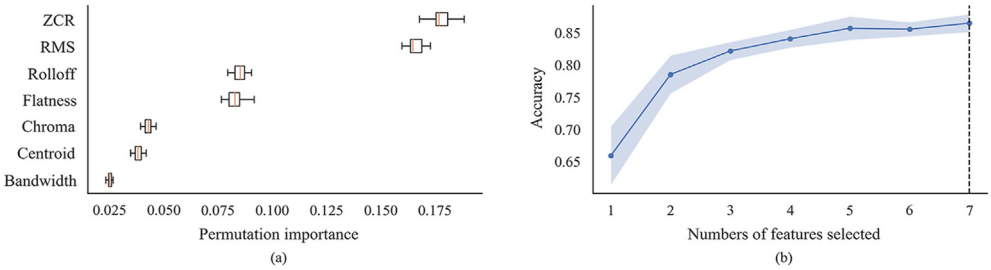

Fig. 9 – (a) Boxplots for the mean feature importance levels of seven features. (b) RF-RFECV with different numbers of features. Dashed line indicates the optimal predicted accuracy.

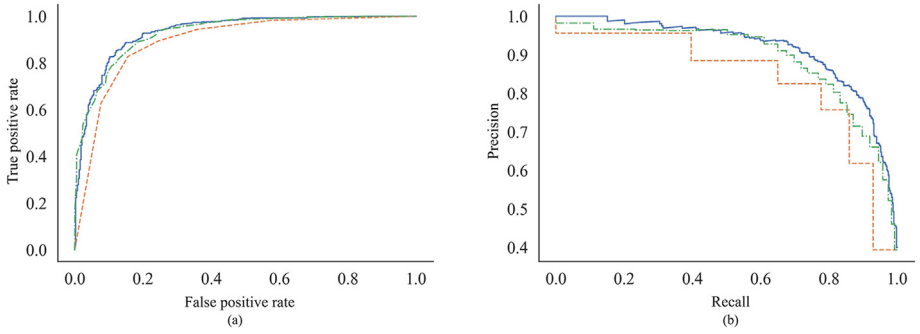

Fig. 10 – AUCs of the CF for pig cough recognition. (a) ROC-AUC. (b) PR-AUC. The blue solid line, orange dashed line, and green dash-dotted line indicate SVM, KNN and RF, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3 – Cough recognition rates of CF, MFCCs and Acoustic.

Features	Evaluation metrics			
	Accuracy	F1-score	Precision	Recall
CF	87.25%	89.82%	87.12%	92.69%
MFCCs	94.22%	95.26%	94.71%	95.82%
Acoustic	95.49%	96.31%	95.51%	97.13%

Table 4 – Pig cough classification results obtained by multiresolution analysis.

Features	Evaluation metrics			
	Accuracy	F1-score	Precision	Recall
$LBP_{12,2}$	89.07%	91.17%	89.75%	92.30%
$LBP_{12,2} + LBP_{12,2}$	89.07%	91.09%	90.15%	92.04%
$LBP_{12,2} + LBP_{8,1}$	88.68%	90.76%	89.88%	91.64%

was calculated for comparison. The accuracies and F1-scores of the STFT and CQT with different settings are shown in Fig. 11. As a whole, high-frequency resolution provides better performance. In other words, the long window size (L = 2048) for the STFT and the short window size (B = 32) for the CQT produce higher-frequency resolutions. Furthermore, different pairs of (P, R) for LBP provide different accuracies. $LBP_{12,2}$ achieves the best performance, with an F1-score of 91.17%, and $LBP_{8,1}$ ranks second. For these two parameters, the CQT exhibits more competitive performance than the STFT. Hence, $LBP_{12,2}$ and $LBP_{8,1}$ of the CQT is chosen as the two resolutions for multiresolution fusion in the subsequent experiments. LBP extensions such as multiresolution analysis may further enhance the performance of the LBP operator with regard to providing fine-grained analysis (Abidin et al., 2017). We try to extend the feature discrimination ability of LBP by conducting a multiresolution analysis for pig cough recognition.

As summarised in Table 4, $LBP_{12,2} + LBP_{12,2}$ is superior to $LBP_{12,2} + LBP_{8,1}$ and closer to $LBP_{12,2}$. However, it is noted that the multiresolution LBP performance yields a slight drop in accuracy compared that of to the single-resolution LBP. Therefore, only $LBP_{12,2}$ is utilised for feature fusion in the follow-up study.

To conduct a comparison with the LBP, a grayscale CQT was first adopted to compare different HOG descriptors. Since HOG feature parameter selections varied from those used in the recognition task, the cell size, block size and number of bins were discussed in this study, as shown in Fig. 12. In general, in contrast to the block size and the number of bins, the cell size imposes a stronger influence on the recognition task. As illustrated in Fig. 12, the maximum recognition rate occurs when the cell size is (16 × 16), followed by (32 × 32). Simultaneously, the cell size of (4 × 4) underperforms in terms of pig cough recognition. Additionally, when the block size is (2 × 2), a better performance is obtained than that achieved when the block size is (3 × 3). In comparison with 8 bins, the 9-bin setting yields a slight improvement in the recognition results. As a whole, the optimal F1-score of 95.15% is obtained with the following parameters: the cell size is (16 × 16), the block size is (2 × 2) and the number of orientation bins is nine.

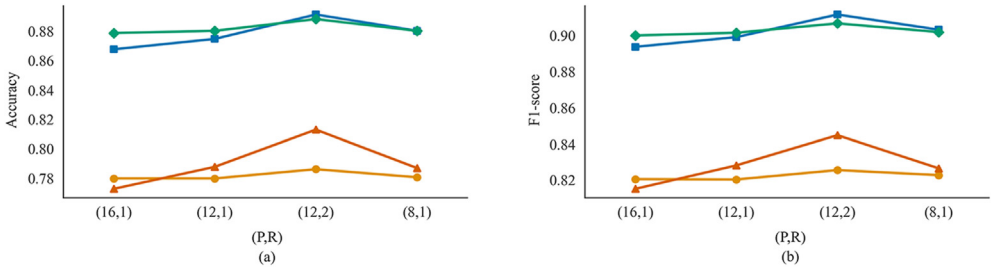

Fig. 11 – Performance of LBP with the CQT and STFT. (a) Accuracies of LBPs with four TFRs. (b) F1-scores of LBPs with four TFRs. Blue lines with squares represent CQT (32B); green lines with diamonds represent STFT (2048L); orange lines with triangles represent STFT (1024L); yellow lines with circles represent CQT (64B). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

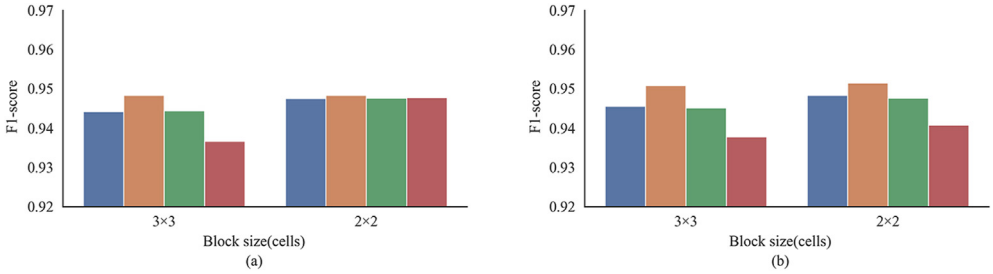

Fig. 12 – F1-score of pig cough recognition with different HOG descriptors. Blue, orange, green and red bars represent four cell sizes with (32 × 32), (16 × 16), (8 × 8) and (4 × 4). F1-score with 8 and 9 orientation bins are shown in (a) and (b), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

After choosing the parameters, the colour CQT spectrograms were also considered concerning pig cough recognition. Here, we chose a three-colour combination with red (R), green (G) and blue (B) channels for analysis; this combination was referred as HOG-Colour. Similarly, HOG-Gray referred to the HOG extracted from gray-level CQT spectrograms, as previously discussed. As compared in Table 5, the results indicate that there are no significant differences between two features. Hence, the gray-level HOG was adopted for feature fusion in the following section.

3.3. *Feature fusions*

As characteristics are complementary, fusion is a frequently used technique for increasing recognition rates. Here, early fusion was applied to concatenate the feature vectors into enhanced vectors in this study.

Table 6 displays the performance of the various feature fusions. It is extremely obvious that feature fusion as inputs exceeds typical single features in classification performance. The best pig cough recognition result is obtained by Acoustic + $LBP_{12,2}$ + HOG-Gray (A-LH), with 97.33%. In addition, the accuracy and F1-scores are 96.45% and 97.08%, respectively. In general, we conclude that relying exclusively on acoustic or visual features is insufficient for capturing audio information. Additionally, the concatenation of Acoustic, LBP, and HOG features as a hybrid feature set can boost classification performance.

4. Discussion

Undoubtedly, feature engineering in terms of feature selection and feature extraction is rather significant and essential for the pig cough recognition task. In contrast to investigating representative features, previous studies on pig coughs have almost exclusively concentrated on classification algorithms. As demonstrated in Zhao et al. (2020) and Wang et al. (2019), SVM with MFCCs features performed better with smaller sound samples, but the overall accuracy decreased with increasing sample size. Although Yin et al. (2021) achieved satisfying results by employing a finetuned Alexnet model to recognise pig coughs, high time cost and hardware resources were required by using deep convolutional neural network. Compared to the approach of Yin et al. (2021), our proposed method can achieve a better result by adopting representative features. In future work, we will consider utilising machine learning techniques with more distinct features for pig cough recognition.

The assessment of feature importance is a critical part of the classification task. Although our results indicate that the CF achieves feasible results without eliminating subfeatures, it is worthwhile to investigate the role of acoustic features in recognition performance. Figure 9(a) reveals that the ZCR and RMS contain considerable amounts of information regarding the spectral characteristics of pig sounds. This is related to higher magnitudes within frames and the swift spectral shifts

Table 5 – Classification results obtained by HOG-Gray and HOG-Colour.

Features	Evaluation metrics			
	Accuracy	F1-score	Precision	Recall
HOG-Gray	94.06%	95.15%	94.35%	95.95%
HOG-Colour	93.90%	95.00%	94.45%	95.56%

Table 6 – Classification results of feature fusion.

Features	Evaluation metrics			
	Accuracy	F1-score	Precision	Recall
Acoustic	95.49%	96.31%	95.51%	97.13%
$LBP_{12,2}$ + HOG-Gray	93.51%	94.70%	93.73%	95.69%
Acoustic + $LBP_{12,2}$	95.53%	96.30%	96.76%	95.84%
Acoustic + HOG-Gray	96.28%	96.90%	97.87%	95.95%
Acoustic + $LBP_{12,2}$ + HOG-Gray	96.45%	97.08%	96.83%	97.33%

of pig coughs over time, which have significant potential for recognition tasks. As expected, the MFCCs achieve success in pig cough classification. However, integration with the CF produces superior effectiveness. This result provides strong encouragement for merging the complementary information carried by acoustic features with other low-level features in the future, such as spectral fluxes and peaks. Additionally, feature selection should be undertaken prior to classification since it would be beneficial to select features as the number of input features expands.

Our study confirms that CQT spectrograms are promising tools for distinguishing pig cough sounds under field conditions. It should also be noted that the performance of the CQT is slightly better than that of the STFT, but the difference is not significant in our experiments. This result is consistent with the findings of many previous studies in Xu et al. (2021). In addition, it is obvious that the adjustment applied to the parameters used to construct the CQT has a noticeable effect on recognition performance in various domains. Specifically, the F1-scores of pig cough recognition vary from 82.56% to 91.17%, which is highly influenced by the utilised window length. It generally means that a larger window size further emphasises the importance of frequency resolution in pig cough classification, which is in accordance with the findings of Knight et al. (2020). In general, our findings imply that frequency is the most significant factor for pig cough classification. Moreover, the benefit of wideband transforms over narrowband transforms was not consistent across various sounds. As a result, further insight into the audio characteristics of the sounds in pig houses would help in determining which variations would be more advantageous. Overall, it is feasible to apply CQT spectrograms for pig cough recognition in our research, and further optimisation of CQT parameters will be one of our future works.

From Fig. 11, $LBP_{12,2}$ yields the best performance on the recognition task, which are aligned with the findings of Xie and Zhu (2019). While in the study of Demir et al. (2018), $LBP_{8,1}$ produced stronger snore sound discrimination results. Therefore, it is reasonable to assume that (P, R) values vary across various datasets for obtaining better recognition accuracy. As shown in our experiments, the LBPs are less effective in detecting pig coughs than the HOGs. A possible reason for this phenomenon is that global LBP representations are calculated in this work. LBP is powerful texture pattern discriminator. However, pig houses are frequently subjected to low-frequency noise. In particular, cough and scream signals are not apparent and are indistinguishable when the frequency bands are 100–2000 Hz. Therefore, it is possible to better capture local LBP features from TFRs by using segmentation, thereby achieving an accuracy improvement. On the other hand, contrary to our expectations, the LBPs extracted from multiresolution fusion sets do not provide satisfactory results. This may be because positive information is enhanced at the expense of amplifying the interference factors. A tradeoff is formed between the positives and the negatives aspects when using multiresolution LBP-based feature extraction.

When extracting features by relying on TFRs as their inputs, one of the key factors is the resolution of the data generated in the preprocessing step. A previous study showed that the recognition rate does not increase with increasing image resolution when utilising HOG features for facial recognition (Xiang et al., 2018). Therefore, this inspired us to choose a size of 100 × 100 pixels to reduce the computation time of the whole experimental process in this study. However, it is clear to us that choosing an appropriate image resolution for extracting visual features is essential. This will be further verified in future experiments.

Due to the limitation of experiment conditions, only one microphone was placed in Pen 7 near the door. Inevitably, data quality collected by the microphone was subject to the distance. Some coughs near the microphone were louder and the far-away data were sounded weaker. In the manual labelling stage, both coughs and non-coughs had been captured and labelled by an expert in the way of observing sound waveforms and performing auditory confirmation. The reason was to do our best to cover acoustic variability throughout the experiment in order to overcome the bias posed by devices. Consequently, to some extent, classification accuracy may be expected to degrade in real-world applications, but classification performance would not be not significantly affected in general. In the future, we will try to improve the audio quality collected in the pig houses by adding more microphones and upgrading collection devices.

5. Conclusion

In this study, we investigated the applicability of various feature sets based on the aggregation of acoustic and visual features for pig cough recognition. The experimental results revealed that HOG descriptors were much more robust than LBP descriptors. Moreover, it was demonstrated that $LBP_{12,2}$ + HOG-Gray + Acoustic with the SVM achieved the best accuracy rate (96.45%).

In future work, we intend to conduct in-depth research and analysis on the recognition of other pig vocalisations for disease detection and stress assessment. Since the vocalisations of an animal are useful for assessing the state of the animal,

we will focus our study on the recognition of pig coughs and other pig sounds encountered in the field for abnormal state detection and warning.

Funding

The work was supported by the project of the National Natural Science Foundation of China [grant numbers 32172784, 31902210]; 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 China Agriculture Research System of MOF and MARA [grant number CARS-35, CARS-36].

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.

REFERENCES

Abidin, S., Togneri, R., & Sohel, F. (2017). Enhanced LBP texture features from time frequency representations for acoustic scene classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 626–630).

Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics, 4*(4), 206–214.

Alves, A. A. C., Andrietta, L. T., Lopes, R. Z., Bussiman, F. O., Silva, F. F. e, Carvalheiro, R., Brito, L. F., Balieiro, J. C. de C., Albuquerque, L. G., & Ventura, R. V. (2021). Integrating audio signal processing and deep learning algorithms for gait pattern classification in Brazilian gaited horses. *Frontiers in Animal Science, 2*, 681557.

Benjamin, M., & Yik, S. (2019). Precision livestock farming in swine welfare: A review for swine practitioners. *Animals, 9*(4), 133.

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

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning - ICML 06* (pp. 233–240).

Demarchi, L., Kania, A., Ciężkowski, W., Piórkowski, H., Oświecimska-Piasko, Z., & Chormański, J. (2020). Recursive feature elimination and random forest classification of natura 2000 grasslands in lowland river valleys of Poland based on airborne hyperspectral and lidar data fusion. *Remote Sensing, 12*(11), 1842.

Demir, F., Sengur, A., Cummins, N., Amiriparian, S., & Schuller, B. (2018). Low level texture features for snore sound discrimination. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 413–416).

Er, M. 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., Aerts, J.-M., Taylor, C. J., & Berckmans, D. (2008). Real-time recognition of sick pig cough sounds. *Computers and Electronics in Agriculture, 63*(2), 207–214.

Flores-Fuentes, W., Rivas-Lopez, M., Sergiyenko, O., Gonzalez-Navarro, F. F., Rivera-Castillo, J., Hernandez-Balbuena, D., & Rodríguez-Quiñonez, J. C. (2014). Combined application of power spectrum centroid and support vector machines for measurement improvement in optical scanning systems. *Signal Processing, 98*, 37–51.

Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia, 13*(2), 303–319.

Huang, C.-J., Chen, Y.-J., Chen, H.-M., Jian, J.-J., Tseng, S.-C., Yang, Y.-J., & Hsu, P.-A. (2014). Intelligent feature extraction and classification of anuran vocalizations. *Applied Soft Computing, 19*, 1–7.

Huang, N., Lu, G., & Xu, D. (2016). A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. *Energies, 9*(10), 767.

Huang, W., Zhu, W., Ma, C., Guo, Y., & Chen, C. (2018). Identification of group-housed pigs based on gabor and local binary pattern features. *Biosystems Engineering, 166*, 90–100.

Jeon, H., & Oh, S. (2020). Hybrid-recursive feature elimination for efficient feature selection. *Applied Sciences, 10*(9), 3211.

Jiang, D., Huang, D., Song, Y., Wu, K., Lu, H., Liu, Q., & Zhou, T. (2020). An audio data representation for traffic acoustic scene recognition. *IEEE Access, 8*, 177863–177873.

Kedem, B. (1986). Spectral analysis and discrimination by zero-crossings. *Proceedings of the IEEE, 74*(11), 1477–1493.

Knight, E. C., Hannah, K. C., Foley, G. J., Scott, C. D., Brigham, R. M., & Bayne, E. (2017). Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology, 12*(2), art14.

Knight, E. C., Poo Hernandez, S., Bayne, E. M., Bulitko, V., & Tucker, B. V. (2020). Pre-processing spectrogram parameters improve the accuracy of bioacoustic classification using convolutional neural networks. *Bioacoustics, 29*(3), 337–355.

Li, S., Li, D., & Yuan, W. (2019). Wood defect classification based on two-dimensional histogram constituted by LBP and local binary differential excitation pattern. *IEEE Access, 7*, 145829–145842.

Lim, S. J., Jang, S. J., Lim, J. Y., & Ko, J. H. (2019). Classification of snoring sound based on a recurrent neural network. *Expert Systems with Applications, 123*, 237–245.

Luz, J. S., Oliveira, M. C., Araújo, F. H. D., & Magalhães, D. M. V. (2021). Ensemble of handcrafted and deep features for urban sound classification. *Applied Acoustics, 175*, 107819.

Ma, Y., & Nishihara, A. (2013). Efficient voice activity detection algorithm using long-term spectral flatness measure. *EURASIP Journal on Audio Speech and Music Processing, 2013*(1), 87.

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*(4), 1176.

Rakotomamonjy, A., & Gasso, G. (2015). Histogram of gradients of time–frequency representations for audio scene classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23*(1), 12.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One, 10*(3), Article e0118432.

Sengupta, N., Sahidullah, M., & Saha, G. (2017). *Lung sound classification using local binary pattern. ArXiv:1710.01703*.

Sharma, G., Umapathy, K., & Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics, 158*, 107020.

Stowell, D., Giannoulis, D., Benetos, E., Lagrange, M., & Plumbley, M. D. (2015). Detection and classification of acoustic scenes and events. *IEEE Transactions on Multimedia, 17*(10), 1733–1746.

Trzcinska, K., Janowski, L., Nowak, J., Rucinska-Zjadacz, M., Kruss, A., von Deimling, J. S., Pocwiardowski, P., & Tegowski, J. (2020). Spectral features of dual-frequency multibeam echosounder data for benthic habitat mapping. *Marine Geology, 427*, 106239.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). scikit-image: image processing in Python. *PeerJ, 2*, e453.

Wang, X., Zhao, X., He, Y., & Wang, K. (2019). Cough sound analysis to assess air quality in commercial weaner barns. *Computers and Electronics in Agriculture, 160*, 8–13.

Xiang, Z., Tan, H., & Ye, W. (2018). The excellent properties of a dense grid-based HOG feature on face recognition compared to Gabor and LBP. *IEEE Access, 6*, 29306–29319.

Xie, J., Towsey, M., Zhang, J., & Roe, P. (2020). Investigation of acoustic and visual features for frog call classification. *Journal of Signal Processing Systems, 92*(1), 23–36.

Xie, J., & Zhu, M. (2019). Investigation of acoustic and visual features for acoustic scene classification. *Expert Systems with Applications, 126*, 20–29.

Xu, L., Wei, Z., Zaidi, S. F. A., Ren, B., & Yang, J. (2021). Speech enhancement based on nonnegative matrix factorization in constant-Q frequency domain. *Applied Acoustics, 174*, 107732.

Yang, B., & Chen, S. (2013). A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image. *Neurocomputing, 120*, 365–379.

Yin, Y., Tu, D., Shen, W., & Bao, J. (2021). Recognition of sick pig cough sounds based on convolutional neural network in field situations. *Information Processing in Agriculture, 8*(3), 369–379.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends, 1*(2), 56–70.

Zhang, J., Wang, Y., Molino, P., Li, L., & Ebert, D. S. (2019). Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models. *IEEE Transactions on Visualization and Computer Graphics, 25*(1), 364–373.

Zhao, Q., Guo, F., Zu, X., Li, B., & Yuan, X. (2019). An acoustic-based feature extraction method for the classification of moving vehicles in the wild. *IEEE Access, 7*, 73666–73674.

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. *International Journal of Agricultural and Biological Engineering, 13*(3), 186–193.