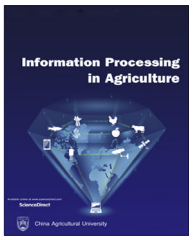
Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE xxx (xxxx) xxx

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



A new fusion feature based on convolutional neural network for pig cough recognition in field situations

Weizheng Shen [a,c], *Ding Tu* [a,c], *Yanling Yin* [a,c,*], *Jun Bao* [b,c]

[a] *College of Electrical and Information, Northeast Agricultural University, Harbin, PR China*
[b] *College of Animal Science and Technology, Northeast Agricultural University, Harbin, PR China*
[c] *Key Laboratory of Swine Facilities Engineering, Ministry of Agriculture, Northeast Agricultural University, Harbin, PR China*

ARTICLE INFO

Article history:
Received 16 February 2020
Received in revised form
12 November 2020
Accepted 17 November 2020
Available online xxxx

Keywords:
Pig cough recognition
MFCC
SVM
CNN
Sound classification

ABSTRACT

Pig cough is considered the most common clinical symptom of respiratory diseases. Thus, establishing an early warning system for respiratory diseases in pigs by monitoring and identifying their cough sounds is important. In this paper, we propose a new fusion feature, namely Mel-frequency cepstral coefficient-convolutional neural network (MFCC-CNN), to improve the recognition accuracy of pig coughs. We obtained the MFCC-CNN feature by fusing multiple frames of MFCC with multiple one-layer CNNs. We used softmax and linear support vector machine (SVM) classifiers for classification. We tested the algorithm through field experiments. The results reveal that the performance of classifiers using the MFCC-CNN feature was significantly better than those using the MFCC feature. The F1-score increased by 10.37% and 5.21%, and the cough accuracy increased by 7.21% and 3.86% for the softmax and SVM classifiers, respectively. We also analyzed the impact of different numbers of fusion frames on the classification performance. The results reveal that fusing 55 and 45 adjacent frames resulted in the best performance for the softmax and SVM classifiers, respectively. From this research, we can conclude that a system constructed by simple one-layer CNNs and SVM classifiers can demonstrate excellent performance in pig sound recognition.

1. Introduction

Porcine respiratory disease is a major reason for the reduction in breeding efficiency in the pig breeding industry [1,2,3,4,5]. Pig coughing can be used to screen and diagnose early respiratory diseases in pigs [6,7]. Hence, pig cough sound recognition is a critical issue in health monitoring.

In early studies, cough sounds were obtained through chemical induction in healthy pigs in a laboratory chamber [8,9,10]. Moshou *et al.* proposed a method of mixing two-level probabilistic neural networks and four-level multilayer perceptron networks. Compared with using a multilayer perceptron for classification, this method has a better effect, and the correct recognition rates of grunts, noises, and coughs reached 91.3%, 91.3%, and 94.8%, respectively [8,9]. Hirtum *et al.* performed a classification by applying a distance function to different frequency ranges and combining the achieved distance values in fuzzy rules [10].

* Corresponding author at: College of Electrical and Information, Northeast Agricultural University, Harbin, PR China.
E-mail address: yinyanling@neau.edu.cn (Y. Yin).

Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2020.11.003

Please cite this article as: W. Shen, D. Tu, Y. Yin et al., A new fusion feature based on convolutional neural network for pig cough recognition in field situations, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2020.11.003

Later, researchers collected spontaneous pig coughing sounds under field conditions and carried out experiments. Hirtum *et al.* used a dynamic time warping algorithm to classify extracted feature vectors (such as containing energy). The correct recognition rate of cough sounds reached 90% [11]. Guarino *et al.* extracted the feature vectors using a filter bank approach combined with amplitude demodulation, and the correct classification was 85.5% for cough sounds [12]. Exadaktylos *et al.* proposed a method for identifying the frequency content of pigs and coughing sounds using the real-time fuzzy c-means (FCM) algorithm. The total classification rate and the classification rate of sick cough sounds were 85% and 82%, respectively [13]. Ferrari *et al.* analyzed the acoustic features of cough sounds originating from a lung infection and the kind of cough sounds provoked by the inhalation of citric acid [14,15]. Chung *et al.* used the Mel-frequency cepstral coefficients (MFCC) as a characteristic parameter to identify the coughing of pigs and realized the automatic detection and recognition of cough [16]. Gong *et al.* used a vector quantization algorithm to recognize the cough sounds of pigs, with the best recognition rate reaching 91% [17]. However, the chaotic nature of the noises in pigsties resulted in a slight drop in accuracy when compared to the laboratory results.

In recent years, certain deep learning algorithms have been used in pig cough sound recognition. Li *et al.* constructed a model based on deep belief networks for pig cough recognition and used principal component analysis to reduce the MFCC feature dimension, with a cough recognition rate of 95.80% [18]. They also proposed a continuous cough recognition method based on the bidirectional long short-term memory-connectionist temporal classification model. An error rate of 9.09% was obtained in the 1-h corpus outside the data set [19]. The above results indicate that recognition performance based on deep learning algorithms has significantly improved.

In this study, our goal was to improve the accuracy of cough recognition. We used multiple one-layer CNNs to fuse the MFCC feature and obtained a new feature, namely MFCC-CNN, for classification. This idea originates from the sentence classification task in Kim's work [20]. In his work, a word vector was inputted into a one-layer CNN to produce a new feature, and a softmax classifier was used to make a classification of the sentence. In this work, we input the sound feature MFCC to multiple one-layer CNNs and used a support vector machine (SVM) classifier to enhance the performance. We used the same classifier in the baseline algorithm, and the MFCC feature was used to classify different sounds. A performance comparison of different methods revealed that the proposed method has higher recognition accuracy than previous state-of-the-art algorithm.

2. Materials and methods

2.1. Overview of the proposed algorithm

The proposed algorithm is shown in Fig. 1. The system comprises five major components: data segmentation, feature extraction, data separation, classification, and performance evaluation. First, we segmented the continuous recorded sounds into individual sounds using a double-threshold endpoint detection method [21] and we had an expert labeled each individual sound. Then, each labeled sound was buffered into frames, and the MFCC and MFCC-CNN features were extracted. We trained and tested the model using data collected from two different experiments. Then, we divided the data in the first experiment into training and validation sets to train and evaluate the model and used the data in the second experiment for testing. In the classification, we used softmax [22] and linear SVM [23] classifiers. Finally, we evaluated the performance of the classifiers in terms of accuracy, precision, and F1-score.

2.2. Data

The experimental data were collected from a large pig farm in Harbin, China, in April and October 2018. We conducted two experiments in the same pig house. In the first experiment in April, there were on average 10 pigs in each big pen. We used a directional cardioid microphone to record the sounds and suspended the microphone at a distance of 1.4 m from the ground (approximately 0.8 m from the pigs' backs) near the door. The microphone was connected to a laptop, and the data were sampled at a sampling frequency of 44.1 kHz. All pigs were in the fattening stage. Fig. 2 shows the experimental pig house and the layout of the equipment. Fig. 2 (a) and (b) show the microphone and data acquisition software, respectively. Fig. 2 (c) shows a bird's-eye view of the large fence where the microphone is located, and Fig. 2 (d) shows a pig with heave cough. The experimental setup of the second experiment was almost identical to that in the first experiment. The microphone was positioned in the middle of the piggery.

In the experiment, we manually labeled all the recorded individual sounds as cough and non-cough, and the classification model was based on these different types of data. Non-coughs primarily include pig grunts, clear noises, and flowing noise, etc. We collected a total of 4551 and 383 sounds in the first and second experiments, respectively. We used the data collected in the first experiment (2744 coughs and 1807 non-coughs) for the 10-fold cross-validation to train the model and used the data collected in the second experiment (193 coughs and 190 non-coughs) for testing.

2.3. Feature extraction

2.3.1. MFCC

MFCC is a widely used feature in automatic speaker and speech recognition [24,25,26]. It is the result of the short-term real log-cosine transform of the energy spectrum, expressed as the Mel-frequency scale [27]. Fig. 3 shows the MFCC extraction process. First, the original sound was pre-processed by a pre-emphasis filter [28] and a pass-band filter. Then, the pre-processed signal was segmented into frames, and a window was added. The fast Fourier transform (FFT) operation was performed on each frame, and the spectrum was squared. Afterward, the result was filtered by a bank of Mel-filters to obtain the Mel-energy. We used the logarithm

Please cite this article as: W. Shen, D. Tu, Y. Yin et al., A new fusion feature based on convolutional neural network for pig cough recognition in field situations, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2020.11.003

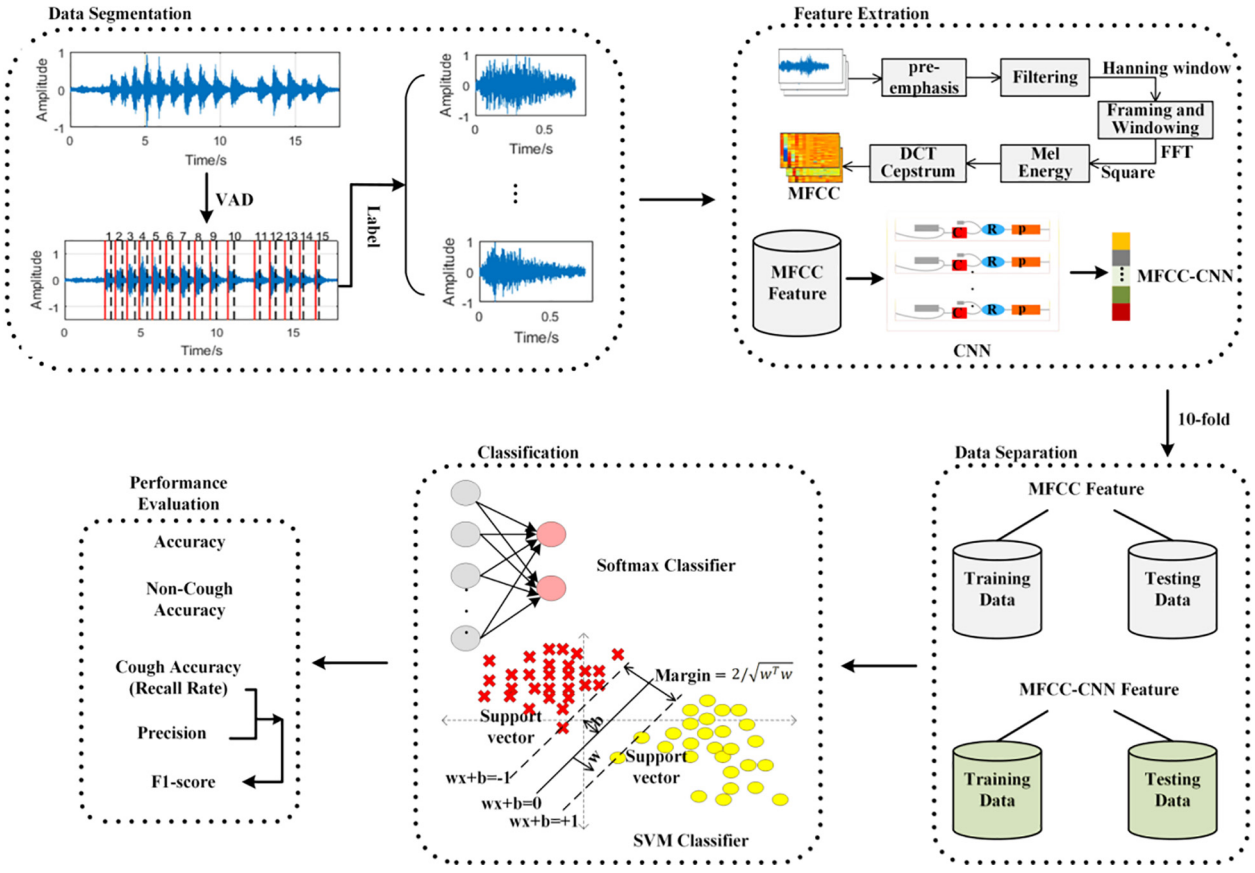


Fig. 1 – Overall process of the proposed algorithm.

Fig. 2 – Experimental pig house and the layout of the equipment.

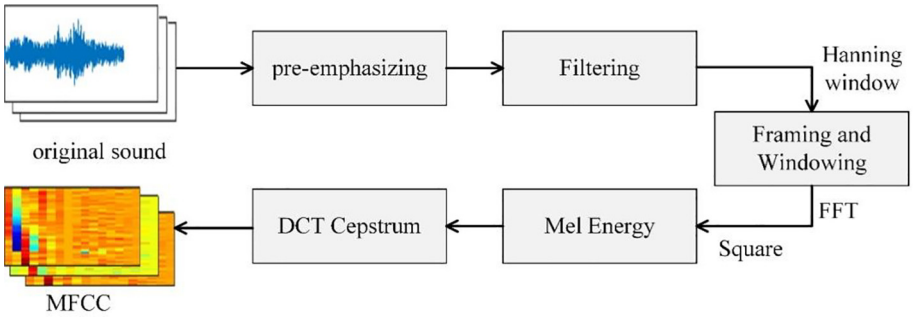


Fig. 3 – MFCC extraction process.

Please cite this article as: W. Shen, D. Tu, Y. Yin et al., A new fusion feature based on convolutional neural network for pig cough recognition in field situations, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2020.11.003

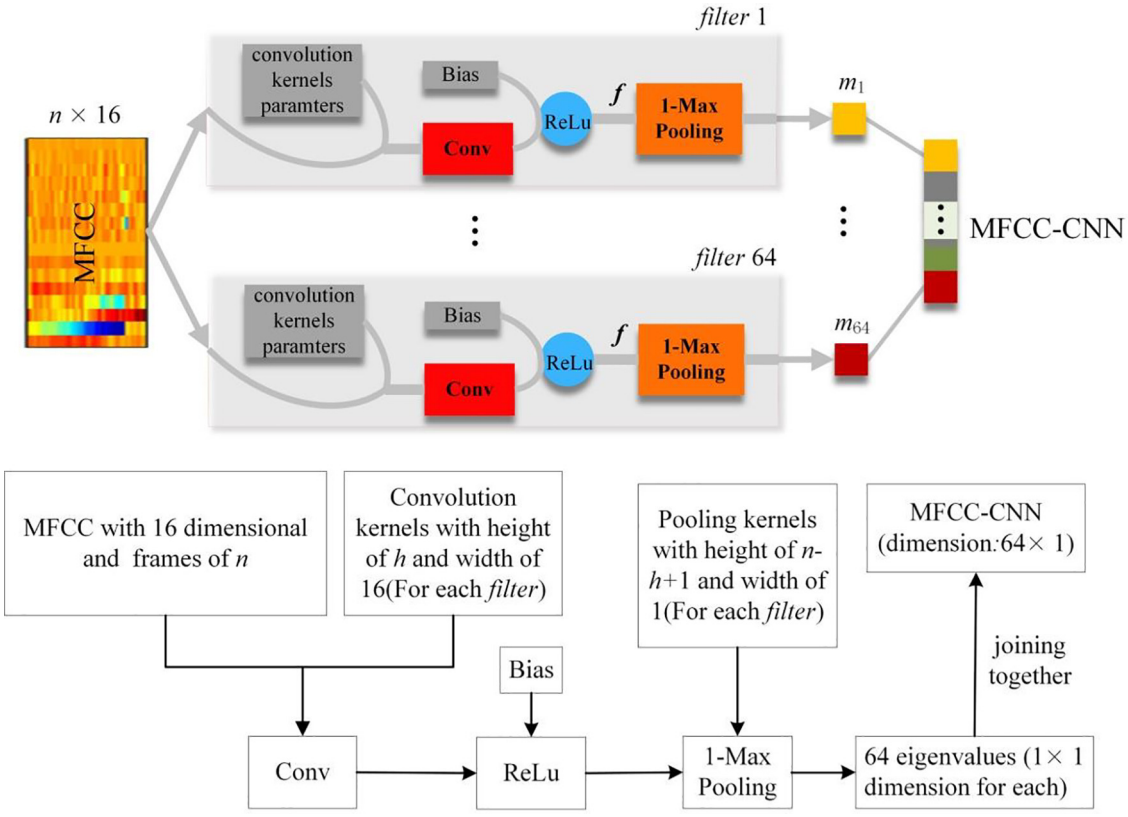


Fig. 4 – Model architecture for the MFCC-CNN feature.

of the Mel-energy and then performed discrete cosine transform (DCT). Finally, we derived the MFCC.

2.3.2. MFCC-CNN

Fig. 4 illustrates the proposed MFCC-CNN feature construction process. Because different sound signals have different numbers of frames, we padded zeros at the end of each MFCC to the same frame number according to the longest frame number. Let $\mathbf{X} \in \mathbb{R}^{n \times m}$ be the padded MFCC feature matrix, where n is the frame number, and m is the dimension of the MFCC in a frame. In this work, m is equal to 16, which comprises eight MFCCs and eight first-order differential MFCCs. An $h \times m$-dimensional convolutional *window* was applied to the MFCC feature, and the *window* slid from the beginning along the line in a single step. Here, h is the height of the convolutional kernel, which also represents the number of fusion frames. Hence, the input can be expressed in a matrix $[X_{1:h}; X_{2:h+1}; \ldots; X_{n-h+1:n}]$. We define a one-layer CNN as a *filter* extracts one feature. A simple one-layer CNN only contains one convolutional layer, one rectified linear unit (ReLU) layer, and one max-pooling layer. The convolutional kernel sizes of all CNNs are $h \times m \times 1$, and the stride is one without zero padding. The output of the convolutional layer is a $(n - h + 1) \times 1$-dimensional *feature map*, that is, $\boldsymbol{f} = [f_1, f_2, \ldots, f_{n-h+1}]$. Feature f_i is generated from a *window* of MFCC features $X_{i:i+h-1}$ by

$$f_i = R(W * X_{i:i+h-1} + b), 1 \le i \le (n - h + 1), \quad (1)$$

where R is the ReLU function [29]; $*$ represents the convolution operation; W is the convolutional kernel parameter, and b is the bias. The initial W obeys a normal distribution with a standard deviation of 0.1. Then, we applied a max-pooling operation [30] over the *feature map* and used the maximum value $m = \max\{\boldsymbol{f}\}$ as the feature corresponding to this particular *filter*. The process described above is one feature extracted from one *filter*. Sixty-four *filters* were used to obtain 64 features. Vector $\mathbf{M} = [m_1, m_{12}, \ldots, m_{64}]$ is the MFCC-CNN feature.

3. Results and discussion

3.1. Experimental parameter configuration

The main hardware environment and parameter configuration of the experiment are listed in Table 1. The program was implemented based on the Python language and the TensorFlow deep learning framework. In the process of extracting the MFCC feature, the main libraries used include *scipy*, *numpy*, *math*, and *sigprocess*. First, we used the *read* function in *scipy* to read the collected audio data and used the *pre-emphasis* function in *sigprocess* to pre-emphasize data, where the pre-emphasis filter coefficient was 0.93. Then, we used the *numpy* and *math* libraries to perform band-pass filtering and frame windowing on the data. The bandwidth of the band-pass filter was 0.1–16 kHz, the frame length was 256 points (5.8 ms), overlap length was 128 points, and window function was the Hamming window. Next, we used the *spectrum_power* function in *sigprocess* to obtain the Mel-energy and used the *dct* function in *scipy* to calculate the cepstrum. The FFT length was 512, and the number of Mel-filters was

Please cite this article as: W. Shen, D. Tu, Y. Yin et al., A new fusion feature based on convolutional neural network for pig cough recognition in field situations, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2020.11.003

Table 1 – Main hardware environment and parameter configuration.

Parameter name	Specific configuration
CPU	i7-7800x with 3.50 GHz
GPU	NVIDIA GeForce GTX 1060Ti
GPU memory	3 GB
LIBSVM Version	3.23
Optimizer	Adam optimizer
batch size	64
learning_rate	3×10^{-4}
frame_num	max_len/16
mfcc_dim	16
dropout_prob	0.5
epoch_max	200

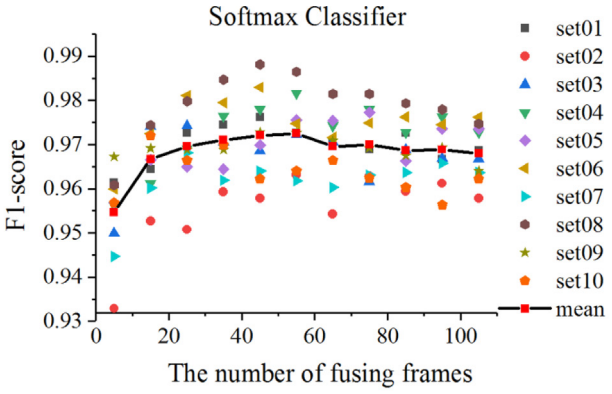
Remark: max_len is the padded MFCC length.

8. Simultaneously, the first-order differential MFCC was extracted to form an MFCC vector with a dimension of 16.

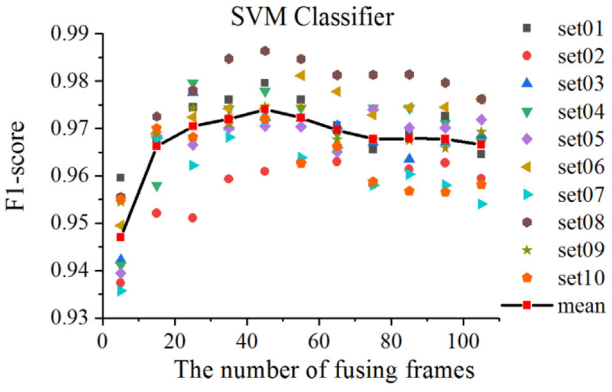
In the processing of feature fusion and cough recognition, the Adam optimizer was used to optimize the cross-entropy loss of the model. The data were processed in batches of 64. The main libraries used include *sklearn*, *tensorflow*, *numpy*, and *libsvm* [31]. The MFCC features were padded by the *pad_sequences* function in *tensorflow* to the same size. Then, we used the *KFold* function in the *sklearn* package to divide the data set. All the data and model parameters were passed to the defined CNN model in the form of a list and a dictionary, respectively. The parameters (keys) in the dictionary include the number of MFCC frames (frames_num), dimensions of each frame of MFCC (mfcc_dim), number of filters (filters_num), height of the convolutional kernel (conv_kernel_h), dropout node reservation probability (dropout_porb), batch size (batch_size), maximum number of epochs in the model training (epoch_max), and the initial learning rate of the Adam optimizer (learning_rate). We performed experiments and analyses on different values of two parameters (the numbers of *filters* were 32, 64, 96, and 128; the heights of the convolution kernel were 5, 10, 15, . . .). The values of the remaining parameters in the dictionary are listed in Table 1.

3.2. *Experimental results*

In the proposed algorithm, the convolutional kernel size determines the number of fusion frames that may influence the performance of the classification. Fig. 5 (a) and (b) show the F1-score variation trend of the two classifiers with an increase in fusion frames. In the figures, "sub01-sub10" represents 10 sub-experiments of the 10-fold cross-validation experiment, and "mean" represents the average F1-score of the 10 sub-experiments. We found that the F1-score increased at the beginning and tended to be stable, and then decreased slightly as the number of fusion frames increased. The softmax and SVM classifiers demonstrated the best performance when the fusion frames were 55 and 45, respectively. We used the data in the second experiment to test the model with the best performance for the two classifiers, and the results are



(a) Softmax classifier



(b) SVM classifier

Fig. 5 – Classification performance of softmax (a) and linear SVM (b) classifier for different MFCC-CNNs (the number of fusion MFCC frames increases from 5 to 105).

shown in Table 2 (the 4th column, MFCC-CNN). The parameters of sensitivity and specificity represent the accuracies of the correctly classified coughs and non-coughs. From the table, we can see that the best sensitivity results are 95.51% and 97.72% and the F1-scores are 96.41% and 97.26% for the

Please cite this article as: W. Shen, D. Tu, Y. Yin et al., A new fusion feature based on convolutional neural network for pig cough recognition in field situations, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2020.11.003

Table 2 – MFCC and MFCC-CNN performance comparison.

Classifier	Performance Indicator	MFCC	MFCC-CNN	Δ
Softmax	Accuracy	82.73%	95.82%	13.09%
	Sensitivity	88.30%	95.51%	7.21%
	Specificity	74.30%	96.28%	21.98%
	Precision	83.96%	97.33%	13.37%
	F1-score	86.04%	96.41%	10.37%
SVM	Accuracy	90.22%	96.68%	6.46%
	Sensitivity	93.86%	97.72%	3.86%
	Specificity	84.75%	95.01%	10.26%
	Precision	90.34%	96.81%	6.47%
	F1-score	92.05%	97.26%	5.21%

softmax and SVM classifiers, respectively. Thus, the SVM performance is better than the softmax performance.

3.3. *Performance comparison*

To compare the classification performance of MFCC-CNN and MFCC, we consider an algorithm with MFCC as a feature as the baseline algorithm. Because different sounds have different lengths, the numbers of MFCC frames are also different. To maintain the same feature length, the FCM [13] algorithm was used to make a cluster of the MFCC features. Ten classes were clustered to return the average MFCC features of the central mass point, and ultimately, 160-dimensional MFCC features were used in the classification [32]. The baseline model directly uses the clustered MFCC features to train and test the softmax and SVM classifiers.

Table 2 compares the performance of the proposed algorithm to that of the baseline algorithm. The last column of the table shows the increment of MFCC-CNN. Table 2 shows that, when MFCC-CNN is used as a feature to classify different sounds, the performance is improved. The F1-score increases by 10.37% and 5.21% for the softmax and SVM classifiers, respectively. This performance improvement was primarily caused by the CNN. The CNN can be treated as a feature extractor. We used multiple CNNs to capture the deeper features of the sound and the correlations between adjacent frames to improve the classification accuracy.

Table 3 compares the recognition performance of the method proposed in this work to the existing work on the same evaluation index. The algorithm was validated in the laboratory in [8] and in the pig house in [12] and [16], as stated above. Although the experimental environment in this study was considerably complex and the collected number of pigs was larger than that in the previous studies, the performance was still enhanced.

3.4. *Discussion*

When discussing the impact of different numbers of fusion frames on the classification performance, we found that the best performances occurred with the fusion of 55 and 45 frames for the softmax and SVM classifiers, respectively. However, this result may not be entirely accurate because the interval of the selected fusion frames was 5, therefore, the result is necessarily a multiple of 5. Perhaps, the maximum F1-score will appear at other values. Fig. 5 shows that the F1-score is not significantly different around the optimal number of fusion frames. Moreover, the optimal number of fusion frames is related to the feature length, as discussed in [33]. The optimal number of fusion frames may increase if the feature length is longer.

In the experiment, we used 64 *filters* to fuse the feature. We also tried other numbers of *filters*, such as 32, 96, and 128. We found that using more filters results in a higher classification accuracy. For example, the F1-scores for the softmax and SVM classifiers were 96.28%, 96.41%, 96.55%, and 96.68%, and 97.08%, 97.26%, 97.35%, and 97.46% for 32, 64, 96, and 128 *filters*, respectively. However, more *filters* will consume a longer time to train the model. In practice, one can perform a line search over the *filter* size to recognize the "best" size and choose the appropriate *filter* size to achieve a balance between performance and training time.

4. Conclusions

Table 3 – Performance comparison between the proposed method and some typical methods.

References	Number		Correctly classified	
	Cough	Other sounds	Cough	Total
Moshou *et al.* [8]	212	142	94.8%	90.2%
Guarino *et al.* [12]	159	433	85.5%	86.2%
Chung *et al.* [16]	300	200	94.0%	94.2%
Proposed method (SVM)	2937	1997	97.7%	96.6%

Please cite this article as: W. Shen, D. Tu, Y. Yin et al., A new fusion feature based on convolutional neural network for pig cough recognition in field situations, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2020.11.003

In this research, we primarily focused on classification methods for pig cough sounds and proposed a new feature, named MFCC-CNN, for identifying pig coughs. We discussed the performance of softmax and SVM classifiers and tested them in field experiments. The results reveal that our algorithm is better than the MFCC-based baseline algorithm. The proposed algorithm realized a cough recognition rate (sensitivity) of 97.72% and an overall recognition rate (accuracy) of 96.68%. For the proposed method, the recognition accuracy is related to the number of fusion frames and the number of *filters*. The optimal values of these parameters depend on the input data. In general, the recognition accuracy exhibits a growth trend as the two parameters increase within a certain range.

In future studies, we will try to identify more specific cough categories in pigs, such as dry coughs and wet coughs. Because dry coughs and wet coughs typically reflect different pathologies, their identification is extremely significant in the diagnosis of pig diseases.

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.

Acknowledgments

This work was supported by the grant from the National Key Research and Development Program of China under Grant 2016YFD0700204-02, the Earmarked Fund for China Agriculture Research System under Grant CARS-35, the "Young Talents" Project of Northeast Agricultural University under Grant 17QC20, the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province under Grant UNPYSCT-2020092 and UNPYSCT-2018142, and the Heilongjiang Post-doctoral Subsidy Project of China under Grant LBH-Z17020.

REFERENCES

[1] Aarnink A, Wagemans M. Ammonia volatilization and dust concentration as affected by ventilation systems in houses for fattening pigs. Trans ASAE 1997;40(4):1161–70.
[2] Wang X, Zhang Y, Zhao LY. Effect of ventilation rate on dust spatial distribution in a mechanically ventilated airspace. Trans ASAE 2000;43(6):1877–84.
[3] Wathes CM, Jones JB, Kristensen HH, Jones E. Aversion of pigs and domestic fowl to atmospheric ammonia. Trans ASAE. 2002;45(5):4289–303.
[4] Pijpers A, Schoevers EJ, Van Gogh H. The influence of disease on feed and water consumption and on pharmacokinetics of orally administered oxytetracycline in pigs. J Anim Sci 1991;69(7):2947–54.
[5] Greiner LL, Stahly TS, Stabel TJ. Quantitative relationship of systemic virus concentration on growth and immune response in pigs. J Anim Sci 2000;78(10).
[6] Baumann B, Bilkei G. Emergency-culling and mortality in growing/fattening pigs in a large Hungarian "farrow-to-finish" production unit. DTW. Deutsche tierärztlicheWochenschrift 2002;109(1):26–33.
[7] Davis JA. Pathological Cry, Stridor, and Cough in Infants. Arch Dis Child 1983;58(4):319–20.
[8] MoshouD CA, Hirtum AV, et al. An Intelligent Alarm for Early Detection of Swine Epidemics Based on Neural Networks. Trans ASABE 2001;44(1):167–74.
[9] MoshouD CA, Hirtum AV, et al. Neural recognition system for swine cough. Math Comput Simul 2001;56(4–5):475–87.
[10] Hirtum A, Berckmans D. Fuzzy approach for improved recognition of citric acid induced piglet coughing from continuous registration. J Sound Vib 2003;266(3):677–86.
[11] Hirtum A, Guarino M, Costa A, Jans P, et al. Automatic detection of chronic pig coughing from continuous registration in field situations. Comput Electron Agric 2004;1–20.
[12] Guarino M, Jans P, Costa A, et al. Field test of algorithm for automatic cough detection in pig houses. Comput Electron Agric 2008;62(1):22–8.
[13] Exadaktylos V, Silva M, Aerts J, et al. Real-time recognition of sick pig cough sounds. Comput Electron Agric 2008;63(2):207–14.
[14] Ferrari S, Silva M, Guarino M, Aerts JM, Berckmans D. Cough sound analysis to identify respiratory infection in pigs. Comput Electron Agric 2008;64(2):318–25.
[15] Ferrari S, Silva M, Guarino M, Berckmans D. Analysis of Cough Sounds for Diagnosis of Respiratory Infections in Intensive Pig Farming. Trans ASABE 2008;51(3):1051–5.
[16] Chung Y, Oh S, Lee J, Park D, Chang H, Kim S. Automatic Detection and Recognition of Pig Wasting Diseases Using Sound Data in Audio Surveillance Systems. Sensors. 2013;13(10):12929–42.
[17] Yongjie Gong, Xuan Li, Yun Gao, Minggang Lei, Wanghong Liu, Zhuan Yang. Recognition of pig cough sound based on vector quantization. J Huazhong Agric Univ 2017;36(3):119–24.
[18] Xuan Li, Jian Zhao, Yun Gao, Minggang Lei, Wanghong Liu, Yongjie Gong. Recognition of pig cough sound Based on Deep Belief Nets. Trans Chin Soc Agric Mach 2018;49(3):179–86.
[19] Xuan Li, Jian Zhao, Yun Gao, Wanghong Liu, Minggang Lei, Hequn Tan. Pig continuous cough sound recognition based on continuous speech recognition technology. Trans Chin Soc Agric Eng 2019;35(6):174–80.
[20] Yoon Kim. Convolutional neuralnetworks for sentence classification. 2014; arXiv preprintarXiv:1408.5882.
[21] Davis A, Nordholm S, Togneri R. "Statistical Voice Activity Detection Using Low-Variance Spectrum Estimation and an Adaptive Threshold. IEEE Trans Audio, Speech, Lang Process 2006;14(2):412–24.
[22] Jiang M, Liang Y, Feng X, et al. Text classification based on deep belief network and soft max regression. Neural Comput Appl 2018;29:61–70.
[23] Gold C, Sollich P. Model Selection for Support Vector Machine Classification. Neurocomputing 2002;55(1):221–49.
[24] Darabkh KA, Haddad L, Sweidan SZ, et al. An efficient speech recognition system for arm-disabled students based on isolated words. Comp Appl Eng Educ 2018;26(2):285–301.
[25] ParthasaradiV Kailasapathi P. A novel MFCC-NN learning model for voice communication through Li-Fi for motion control of a robotic vehicle. Soft Comput 2019;23(18):8651–60.
[26] Ali H, Tran SN, Benetos E, et al. Speaker recognition with hybrid features from a deep belief network. Neural Comput Appl 2018;29(6):13–9.
[27] Zheng TF, Zhang G, Song Z. Comparison of Different Implementations of MFCC. J Comput Sci Technol 2001;16(6):582–9.
[28] LoweimiE Ahadi S M, Drugman T, et al. On the Importance of Pre-emphasis and Window Shape in Phase-Based Speech

Recognition. 6th International Conference on Non-Linear Speech Processing, 2013.

[29] Krizhevsky A, Sutskever I, Hinton Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. Commun ACM 2017;6(60):84–90.

[30] Collobert R, Weston J, Bottou L, Karlen M, Kavukcuglu K, Kuksa P. Natural LanguageProcessing (Almost) from Scratch. J Mach Learn Res 2011;12:2493–537.

[31] Chang C, Lin C. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3):1–27.

[32] Min Wu, Shanshan Zhu. Language Recognition Method of Convolutional Neural Network Based on Spectrogram. J Educ, Teach Soc Stud 2019;1(2):113–29.

[33] Zhang Y, Wallace B. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. Computer. Science 2015. arXiv preprint arXiv:1510.03820.

Please cite this article as: W. Shen, D. Tu, Y. Yin et al., A new fusion feature based on convolutional neural network for pig cough recognition in field situations, Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2020.11.003