Enhancing Pig Vocalization and Non-Vocalization Classification with Deep Convolutional Neural Networks and Data Augmentation

Vandet Pann [1], Jong-Bok Kim [2*], Kyeong-Seok Kwon[3], Byeonghyeon Kim[4], Dong-Hwa Jang[5]

[1,2,3,4,5] Animal Environment Division, National Institute of Animal Science, Rural Development Administration, Wanju, 55365, South Korea

* Corresponding author. Email: jbkimj@korea.kr

Abstract

Pig vocalization classification using deep learning is a valuable tool in modern pig farming. However, collecting pig sound data for model training takes time and effort. This research study introduces a deep convolutional neural network (CNN) architecture and audio data augmentation techniques for pig vocalization and non-vocalization classification with a real pig farm. Various data augmentation techniques are utilized individually in a training set to improve the model performance and generalization, including pitch-shifting, time-shifting, time-stretching, and background-noising. The Mel-Spectrogram method is used to extract and transform the audio signal into the input representation for training the network model, and each data augmentation is applied to each audio signal before the transformation. This study explored a novel data augmentation approach called combined-all by concatenating pitch-shifting, time-shifting, time-stretching, and background-nosing augmentation techniques into a training dataset. This innovative technique improves model performance compared to individual techniques, particularly when classifying pig vocalization and non-vocalization. For the experiment dataset, a collection of 4,000 WAV files (2,000 pig vocalization and 2,000 pig non-vocalization) was gathered from an actual pig farm. Each audio file has a duration of three seconds. In this study, the performance of the predictive model was assessed using the k-fold cross-validation (k=5) technique. By conducting rigorous experiments, the average accuracy of 5-fold cross-validation on the original, pitch-shifting, time-shifting, time-stretching, background-noising, and combined-all reached the rates of 97.72%, 98.00%, 98.03%, 98.09%, 97.93%, and 98.19%, respectively. The findings validated that the combined-all technique outperforms alternative techniques, leading to an enhanced classification of pig vocalizations.

Keywords: smart livestock farming, animal vocalizations classification, audio feature extraction, deep learning network.

(This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01681003)" Rural Development Administration, Republic of Korea. This research was supported by the "RDA Research Associate Fellowship Program" of the National Institute of Animal Science, Rural Development Administration, Republic of Korea.)