



Detection of Respiratory Diseases Based on Poultry Vocalizations Using Deep Learning[†]

Farook Sattar 

Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC V8P 5C2, Canada; fsattar@ieee.org

[†] Presented at the 3rd International Online Conference on Agriculture (IOCAG 2025), 22–24 October 2025.

Available online: <https://sciforum.net/event/IOCAG2025>.

Abstract

In this study, we design a deep learning-based intelligent recognition method capable of accurately distinguishing abnormal chicken vocalizations among complex sound signals. Our proposed framework is based on the wavelet scattering transform (WST) and a Long Short-Term Memory (LSTM) network, and uses preprocessed chicken vocalizations processed through a denoising scheme, adopting an audio image generation model (AIGM) based on rectified STFT (Short-Term Fourier Transform). We have used a public chicken language dataset that consists of a total of segments for each of the two categories (Healthy or Sick), totaling 4000 five-second audio clips from actual farming environments, which are labeled by veterinary experts. The proposed method achieves promising performance, outperforming state-of-the-art methods for detecting poultry respiratory diseases and enabling poultry personnel to accurately assess the health and well-being of the chickens.

Keywords: chicken; health; poultry; audio recognition; respiratory sound; deep neural network; wavelet scattering transform; denoising

1. Introduction

In large-scale poultry farming, respiratory diseases affect the health of chickens, declining the quality and yield of both meat and eggs. Effective monitoring of the respiratory diseases is crucial to reduce their impact in order to enhance the quality and yield of the poultry products. At present, monitoring methods mostly rely on manual monitoring of the chicken vocalizations as it is time-consuming, labor-intensive, and requires specialized personnel. Existing intelligent methods are often limited to laboratory environments where individual chickens are monitored separately. These approaches do not meet the industrial and commercial requirements of poultry farms, which require capturing a diverse set of complex auditory signals. These audio signals include not only chicken vocalizations but also complex noises from cages, human activities, mechanical ventilation systems, and other background noises.

Most of the existing acoustical poultry health monitoring methods are based on some traditional machine learning techniques. In [1], an acoustical method is developed to detect Bronchitis and Newcastle diseases using the recorded calls from broilers. Five features, including MFCC, energy of mel sub bands, wavelet energy, wavelet entropy, and spectral flatness, are extracted, followed by using an SVM (Support Vector Machine) classifier to identify the two types of diseases. As shown in [1], the best performance is achieved using the wavelet entropy. A study in [2], evaluates a voice activity detection (VAD) algorithm



Academic Editor: Rui Costa Martins

Published: 9 February 2026

Citation: Sattar, F. Detection of Respiratory Diseases Based on Poultry Vocalizations Using Deep Learning.

Biol. Life Sci. Forum **2025**, *54*, 18.

<https://doi.org/10.3390/blsf2025054018>

blsf2025054018

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

licenses/by/4.0/).

for chicken calls. First, the algorithm applies short-term energy (STE) to separate silence from the sound segments and then classifies the sound activity into vocal and non-vocal, based on the values of Wiener Entropy (WE) and Cepstral Peak Prominence (CPP). The results shown in [2] show that the classification accuracy decreases with the increase of boiler age due to higher false rejection (FR) for the detection of voice calls as unvoiced calls. A deep learning-based Newcastle disease (ND) detection method is proposed in [3], which uses an improved multi-window spectral estimation algorithm for denoising, followed by chicken vocalization endpoint detection using MFCC features. The final step of the method is the ND poultry vocalization classification using a BiLSTM (Bidirectional Long Short-Term Memory) recurrent neural network (RNN) [3]. In [4], the study proposes a novel end-to-end model (ResNet18-TF) based on ResNet18 and a time-frequency attention mechanism (TFBlock) to automatically recognize chicken's cough and snore. As in [4], the results reveal that LogFbank features outperform the MFCC features in the task of chicken sound recognition. The incorporation of the first-order and the second-order delta features into LogFbank features further improves the recognition accuracy of the model, and better performance is achieved by the proposed ResNet18-TF model in terms of accuracy, precision, recall, and F1-score than that of the MobileViTv3-TF, EfficientNetV2-TF models [4]. The work in [5] introduces SmartEars to automatically recognize sick chicken vocalizations, allowing users to easily assess the chicken's health status. Furthermore, a recognition algorithm is developed in [5] based on RegNet to integrate into the SmartEars device, which not only monitors sound but is also computationally efficient, allowing simultaneous monitoring and decision-making. Once connected to the network, the device transmits its results to the cloud to display on the poultry health monitoring website.

2. Dataset

We have used an open-access dataset [5] consists of a total of 4000 distinct chicken calls of 5-s duration with labels of 'healthy' or 'sick' having a sampling rate of 44.1 kHz.

For illustration, the chicken calls referring to two types of 'healthy' and 'sick', are presented in Figure 1a,b, respectively. It looks like there are some outliers and an energy drop for the 'sick' calls compared to the 'healthy' calls, as shown in Figure 1a,b.

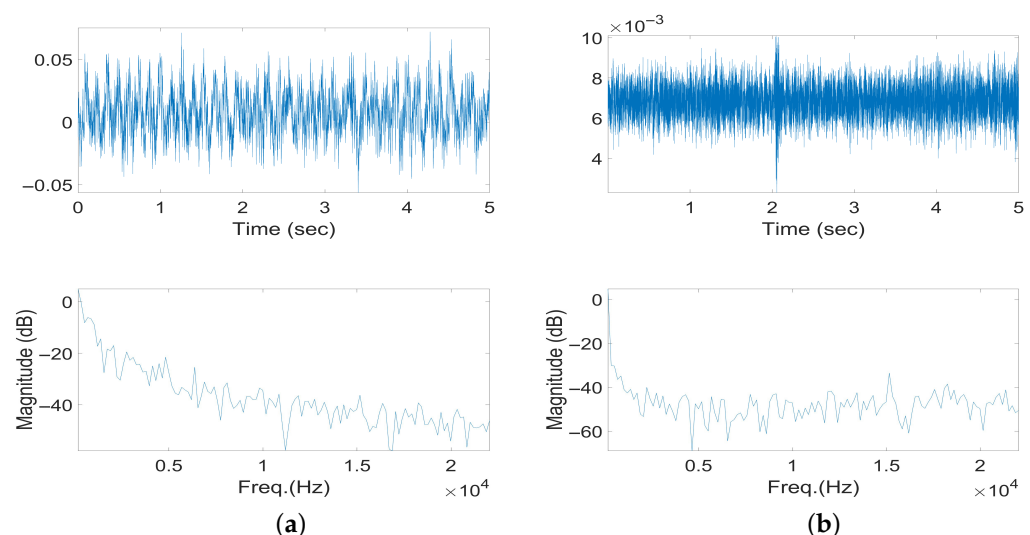


Figure 1. The sample chicken calls (**above**) and their spectrums (**below**); (a) healthy, (b) sick.

3. Proposed Method

Figure 2 shows an overall block diagram of the proposed method. It consists of preprocessing for denoising, the wavelet scattering transform for feature set generation, and deep learning-based detection. The input chicken calls are first passed through an image generation model based on rectified STFT for denoising. Then the enhanced calls are processed with the wavelet scattering transform (WST) and used as input to an LSTM (Long Short-Term Memory), a recurrent neural network (RNN), for deep learning, yielding detection of chicken calls as output.

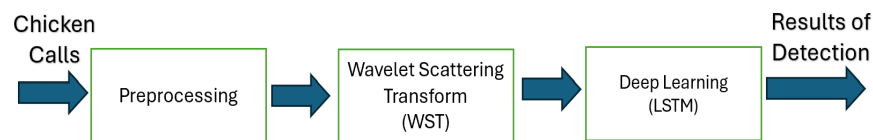


Figure 2. The overall block diagram of the proposed method.

3.1. Preprocessing

The preprocessing of the chicken calls is performed through an image generation model (IGM) based on rectified STFT (Short-Term Fourier Transform) [6].

The rectified STFT is defined as

$$X_w(n, k)' = X_w^r(n, k)' + iX_w^i(n, k)' \quad (1)$$

such that

$$\begin{aligned} X_w^r(n, k)' &= X_w^r(n, k), \\ X_w^i(n, k)' &= X_w^i(n, k) \end{aligned} ; \text{ if } A = 0 \text{ or } A = 1. \quad (2)$$

$$\begin{aligned} X_w^r(n, k)' &= X_w^r(n, k) / \sqrt{2A(n, k)}, \\ X_w^i(n, k)' &= X_w^i(n, k) / \sqrt{2(1 - A(n, k))} \end{aligned} ; \text{ otherwise.} \quad (3)$$

where

$$\begin{aligned} A(n, k) &= 0.5 + 0.5\Re\{B(n, k)\}, \\ B(n, k) &= w(o)^2 + 2 \sum_{m=1}^{(L-1)/2} w(m)^2 \cos(4\pi km / (L + Z)) \end{aligned} \quad (4)$$

In Equations (1)–(4), $X_w^r(n, k)'$ and $X_w^i(n, k)'$ are the real and imaginary parts of $X_w(n, k)'$, respectively, $w(m)$ is an energy normalized window of length L , Z is the zero padding, which gives $K = (L + Z)$ frequency bins.

We then mask the rectified STFT, $X_w(n, k)'$ along frequency by considering a masking parameter m_f :

$$X_w(n, 1 : m_f)' = \text{zeros}(n, 1 : m_f) \quad (5)$$

Lastly, we apply the inverse STFT to the masked $X_w(n, k)'$ to get the denoised chicken calls.

For illustrative purposes, the plots of the input data samples in Figure 1 together with the corresponding preprocessed signals and the rectified STFTs are displayed in Figure 3a,b.

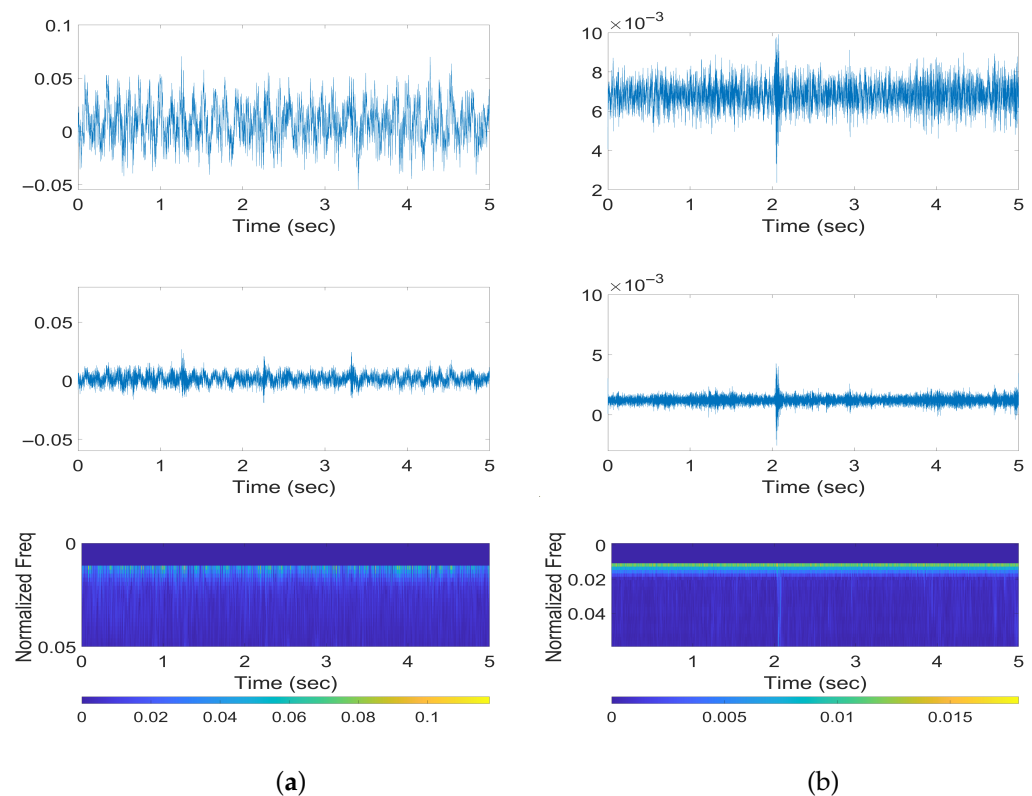


Figure 3. Illustrative plots of the input samples (**top**), preprocessed samples (**middle**) and rectified STFTs (**bottom**); (a) healthy, (b) sick.

3.2. Wavelet Scattering Transform

The wavelet scattering transform (WST) coefficients are utilized here for feature extraction [7,8]. The 1-D WST is computed by cascading wavelet transforms along with nonlinear complex modulus operations, followed by average filtering. The WST of a 1-D signal $z(t)$ can be represented as

$$S_J z = [S_J^{(0)} z, S_J^{(1)} z, S_J^{(2)} z], \quad (6)$$

where

$$\begin{aligned} S_J^{(0)} z(t) &= z * P_J, \\ S_J^{(1)} z(t, L) &= |z * P_L^{(1)}| * P_J, \\ S_J^{(2)} z(t, L, M) &= ||z * P_L^{(1)}| * P_M^{(2)}| * P_J. \end{aligned}$$

In Equation (6), $*$ denotes the convolution operator, $P_L^{(1)}$ and $P_M^{(2)}$ are the filters representing complex wavelets having center frequencies L and M , whereas $P_J(t)$ is a real lowpass filter with zero-mean frequency.

The 1D scattering transform is implemented using a given set of wavelet filters, with their parameter values specified initially. Hence, the wavelets are fixed; however, there might be changes for other parameters, for instance, whether all of $S_J^{(0)} z$, $S_J^{(1)} z$, and $S_J^{(2)} z$, or just $S_J^{(0)} z$ and $S_J^{(1)} z$ would be computed.

For a given input signal of length N , the largest scale of the WST is set to 2^J . The other issues are the time-frequency resolutions of the wavelets, which are set as $Q = 8$ wavelets per octave for the first-order wavelets, $P_L^{(1)}$ and $Q = 1$, i.e., one wavelet per octave for the second-order wavelets $P_M^{(2)}$. Note that postprocessing is performed for the scattering

coefficients by taking the logarithm values of the scattering sequences for the first-order and second-order wavelets.

Figure 4a,b demonstrates the WST feature coefficients for the input data samples, as depicted in Figure 1a,b.

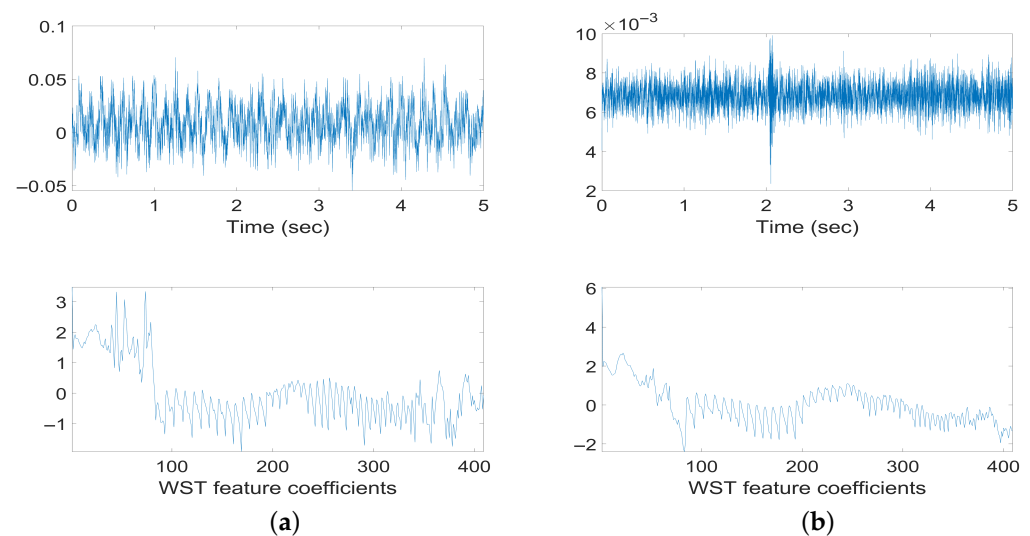


Figure 4. Illustrative plots of the input samples (**top**) and the WST feature coefficients (**bottom**); (a) healthy, (b) sick.

3.3. Deep Learning—LSTM

The detection is performed using Long Short-Term Memory (LSTM) [8,9], which is a recurrent neural network (RNN) containing input gate, forget gate, output gate, temporal forward pass, and backpropagation. The input gate, forget gate, and output gate responses at time t denoted by i^t , o^t , and f^t , respectively, are associated with the forward pass in an LSTM architecture and can be expressed as follows:

$$i^t = \text{Sigmoid}(W_{ih}h^{(t-1)} + W_{ix}x^t + b_i) \quad (7)$$

$$o^t = \text{Sigmoid}(W_{oh}h^{(t-1)} + W_{ox}x^t + b_o) \quad (8)$$

$$f^t = \text{Sigmoid}(W_{fh}h^{(t-1)} + W_{fx}x^t + b_f) \quad (9)$$

The following formulations are also associated with the forward pass:

$$d^t = \text{Tanh}(W_{dh}h^{(t-1)} + W_{dx}x^t + b_d) \quad (10)$$

$$c^t = f^t \odot c^{(t-1)} + i^t \odot d^t \quad (11)$$

$$h^t = o^t \odot \text{Tanh}(c^t) \quad (12)$$

$$L^t = P(h^t) \quad (13)$$

$$L = \sum_{t=1}^T L^t \quad (14)$$

where d^t stands for the distorted input to the memory cell at time t , W_{dh} is the weight associated with $h^{(t-1)}$ and b_d is the corresponding bias vector, $\text{Tanh}(\cdot)$ is the activation function, c^t refers to the state of the memory cell at time t , h^t denotes the hidden state at

time t , ' $\odot$ ' stands for point-wise multiplication. Also, in Equation (13), P maps the hidden state to the network loss L^t at time t . Then the total network loss L is found by adding each individual network loss L^t over time, as depicted in Equation (14).

A graphical representation of Equations (7)–(13) is presented in Figure 5 [10].

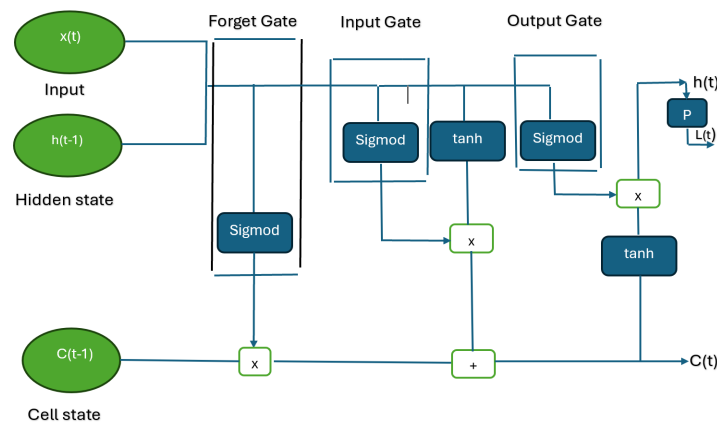


Figure 5. The structure of the LSTM cell illustrates the data flow through the forget, input, and output gates controlling the cell and hidden states.

4. Results

The experimental setup and the detection results of the proposed method are presented here.

4.1. Experimental Setup

The dataset consists of 2000 recordings of ‘healthy’ chicken calls and 2000 recordings of ‘sick’ chicken calls, sampled at 44.1 kHz. For training and testing, the input WST sequences are partitioned into two subsets, namely the training WST set and the test WST set; 70% of the dataset is used for training, and 30% of the dataset is used for testing. The WST sequences are normalized to zero mean and unit variance before being fed into the LSTM network for detection. The results are reported as binary classification accuracies (%) over five different trials. For each trial, we have used different training and test datasets with randomly changing configurations.

We used a three-layer WST and Morlet (Gabor) wavelets [11], which are commonly used complex wavelets due to their simple mathematical representation. The framework has two filter banks when the number of layers is three. The quality factor (i.e., the number of wavelet filters per octave) for the first and the second filter banks is set to $Q = 8$ and 1, respectively. It can be noted that we resampled the input signal from 44.1 kHz to 8 kHz. For an input signal of length $N = 40,000$ samples (5 s duration) and the Q values as above, the framework provides a feature matrix with size $(205 \times 8 \times 2)$ as output. The feature matrix is then generated by 205 scattering paths and 8 scattering windows for both the real and imaginary parts of the signal. Hence, the feature set contains 409 feature vectors of dimension 8, excluding those for path 1. In our training process, we have chosen the following parameters for the LSTM classifier: number of hidden layers = 512, learning rate = 0.0001, batch size = 32, ‘adm’ optimization, and number of epochs = 100. The whole process is implemented in MATLAB 2023b.

4.2. Detection Performance

The performances of the proposed method are evaluated in terms of 2-class classification accuracy (%) as well as sensitivity (%) and specificity (%), as shown below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (15a)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (15b)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (15c)$$

where TP, TN, FP, and FN denote the true positive, true negative, false positive, and false negative, respectively.

The confusion matrix for a trial with accuracy (%) is shown in Table 1.

Table 1. The confusion matrix of the proposed method (the accuracy (%) is indicated in bold face and calculated as a ratio of the sum of diagonal values to the sum of all values $\times 100$).

		Predicted Class		
		Healthy	Sick	Sensitivity (%)
True class	Healthy	462	134	85
	Sick	82	522	80
	Specificity (%)	80	85	82

The accuracies (%) of five trials are depicted in Table 2 with mean accuracy of $82.67\% \pm 0.70$.

Table 2. The accuracies (%) for different trials.

Trial #	Accuracy (%)
1	83.42
2	83.08
3	82.16
4	82.75
5	81.67

Moreover, the results are compared with those of the state-of-the-art methods [12,13]. The method in [12] considers feature patterns in spectrogram images and uses a CNN for two-class classification, whereas the study in [13] uses the cepstrum as a feature vector, followed by a deep neural network (LSTM network) for identification. The confusion matrices for the comparative methods are presented in Tables 3 and 4, showing the classification accuracies of the 53% and 78%, respectively. These comparative results also quantitatively demonstrate the impacts of the rectified STFT over the conventional spectrogram being used in [12] and the WST coefficients over the popular cepstrum coefficients as applied in [13].

Table 3. The confusion matrix for the comparative method in [12] (the accuracy (%) is indicated in bold face and calculated as a ratio of the sum of diagonal values to the sum of all values $\times 100$).

		Predicted Class		
		Healthy	Sick	Sensitivity (%)
True class	Healthy	546	14	47
	Sick	613	27	66
	Specificity (%)	66	47	53

Table 4. The confusion matrix for the comparative method in [13] (the accuracy (%) is indicated in bold face and calculated as a ratio of the sum of diagonal values to the sum of all values $\times 100$).

		Predicted Class		
		Healthy	Sick	Sensitivity (%)
True class	Healthy	436	163	82
	Sick	96	503	75
	Specificity (%)	75	82	78

5. Conclusions

In this paper, the preliminary results are presented, which look promising for detecting the ‘healthy’ chicken and ‘sick’ chickens from their acoustical calls. This would be helpful for the veterinarian in early detection of poultry diseases, managing the large-scale poultry farming efficiently to improve the quality and yield of both meat and eggs. Leveraging the detection of the spectral outliers through WST and the anomalies inferring the contextual changes in temporal data using LSTM, this method could be extended to identify the severe poultry diseases like Newcastle Disease (ND) and Bronchitis [14]. Further, we would like to expand this work by considering data on ‘no’ calls in complex noise environments to make the proposed method more robust, by introducing a learned wavelet scattering transform and optimizing the LSTM network’s model parameters. In this study, the proposed denoising scheme can effectively handle low-frequency background noise. However, to account for the complex background noise in the future, we would like to improve the proposed denoising method by incorporating an adaptive masking scheme based on clustering approaches [15,16] to handle multicomponent, highly nonstationary background noise.

Funding: This study received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in Mendeley at <https://data.mendeley.com/datasets/dy6gtvt4mk/2>, (accessed on 29 May 2025), reference number [5]. These data were derived from the following resources available in the public domain: <https://data.mendeley.com/datasets/dy6gtvt4mk/2>, (accessed on 29 May 2025).

Acknowledgments: The author would like to express his sincere thanks to the anonymous reviewers for their valuable comments and suggestions that helped to improve this paper significantly.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Mahdavian, A.; Minaeia, S.; Marchettob, P.M.; Almasganj, F.; Rahimi, S.; Yang, C. Acoustic Features of Vocalization Signal in Poultry Health Monitoring. *Appl. Acoust.* **2021**, *175*, 107756. [\[CrossRef\]](#)
2. Mahdavian, A.; Minaei, S.; Yang, C.; Almasganj, F.; Rahimi, S.; Marchetto, P.M. Ability Evaluation of a Voice Activity Detection Algorithm in Bioacoustics: A Case Study on Poultry Calls. *Comput. Electron. Agric.* **2020**, *168*, 105100. [\[CrossRef\]](#)
3. Cuan, K.; Zhang, T.; Li, Z.; Huang, J.; Ding, Y.; Fang, C. Automatic Newcastle Disease Detection Using Sound Technology and Deep Learning Method. *Comput. Electron. Agric.* **2022**, *194*, 106740. [\[CrossRef\]](#)
4. Ma, H.; Xin, P.; Ma, J.; Yanga, X.; Zhang, R.; Liang, C.; Liu, Y.; Qi, F.; Wang, C. End-to-End Detection of Cough and Snore Based on ResNet18-TF for Breeder Laying Hens: A Field Study. *Artif. Intell. Agric.* **2026**, *16*, 412–422. [\[CrossRef\]](#)
5. Huang, J. SmartEars: A Practical Framework for Poultry Respiratory Monitoring via Spectrogram-Based Audio Classification and AI-Assisted Labeling. *Mendeley Data* **2025**, *241*. [\[CrossRef\]](#)
6. Millioz, F.; Martin, N. Circularity of the STFT and Spectral Kurtosis for Time-Frequency Segmentation in Gaussian Environment. *IEEE Trans. Signal Process.* **2011**, *59*, 515–524. [\[CrossRef\]](#)
7. Bruni, V.; Cardinali, M.L.; Vitulano, D. An MDL-Based Wavelet Scattering Features Selection for Signal Classification. *Axioms* **2022**, *11*, 376. [\[CrossRef\]](#)
8. Sattar, F. A New Acoustical Autonomous Method for Identifying Endangered Whale Calls: A Case Study of Blue Whale and Fin Whale. *Sensors* **2023**, *23*, 3048. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Houdt, G.V.; Mosquera, C.; Napoles, G. A Review on the Long Short-Term Memory Model. *Artif. Intell. Rev.* **2020**, *53*, 5929–5955. [\[CrossRef\]](#)
10. Amir, A.; Butt, M. Improved Sap Flow Prediction: A Comparative Deep Learning Study Based on LSTM, BiLSTM, LRCN, and GRU with Stem Diameter Data. *Smart Agric. Technol.* **2025**, *12*, 101105. [\[CrossRef\]](#)
11. Chen, M.; Zuo, M.J.; Wang, X.; Hoseini, M.R. An Adaptive Morlet Wavelet Filter for Time-of-Flight Estimation in Ultrasonic Damage Assessment. *Measurement* **2010**, *43*, 570–585. [\[CrossRef\]](#)
12. Zhou, C.; Zhang, L.; Zhang, X.; Wu, Y.; Wu, D.; Tao, Z. Classification of Normal and Pathological Voices Using Convolutional Neural Network. In Proceedings of the 2020 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence (ICSMD), Xi'an, China, 15–17 October 2020; pp. 325–329. [\[CrossRef\]](#)
13. Fang, S.H.; Tsao, Y.; Hsiao, M.J.; Chen, J.Y.; Lai, Y.H.; Lin, F.C.; Wang, C.T. Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. *J. Voice* **2019**, *33*, 634–641. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Banakara, A.; Sadeghia, M.; Shushtarib, A. An Intelligent Device for Diagnosing Avian Diseases: Newcastle, Infectious Bronchitis, Avian Influenza. *Comput. Electron. Agric.* **2016**, *127*, 744–753. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Sattar, F.; Driessen, P.F. Non-Stationary Signals Separation Using STFT and Affinity Propagation Clustering Algorithm. In Proceedings of the 2013 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), Victoria, BC, Canada, 27–29 August 2013.
16. Barkat, B.; Sattar, F.; Meraim, K.A. Sources Separation of Instantaneous Mixtures Using A Linear Time-Frequency Representation and Vectors Clustering. In Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, France, 14–19 May 2006.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.