Laporan Penelitian

Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm

Silvester Dian Handy Permana, Gusti Saputra, Budi Arifitama, Yaddarabullah, Wahyu Caesarendra, Robbi Rahim

2021

Abstract

Indonesia is a tropical country that experiences forest fires every year. Forest fires occur due to a pro- longed summer season. A major effect of forest fires that commonly occurred in Indonesia is the respira- tory disorder and the visual impairment caused by the thick smoke that can experienced by the people who lives around the forest. The effects of forest fire fumes when the forest fire spread across the bigger land or area are also experienced by neighboring countries such as Malaysia, Singapore and Brunei Darussalam. Forest is a habitat for many animals, one of which is a bird. Birds are able to communicate with their colonies through sound. The bird sound during communication to the group can be in the form of calls, marriage invitations, and warnings of danger or threat of forest fires. This paper presents bird sounds classification study using one of Deep Learning (DL) algorithms i.e. Convolutional Neural Network (CNN) method. The CNN method is used to classify bird sounds in the two conditions: (1) under normal circumstances or conditions, and (2) under threated or panic condition. The bird sound data used in this study were collected from the local birds in Indonesia. The classification result of two bird sounds based on CNN method achieved up to 96.45%. This paper is a preliminary study for bio inspired early warning of forest fires based on the sound of birds.

Keywords: Bird sound classification, CNN, Deep learning, Forest fires early warning system

1. Introduction

Indonesia is a tropical country that experiences forest fires every year. Forest fires in Indonesia continue to increase, in 2019 fires increased by 200.94% from 2018 (KLHK, 2019). As a result of forest fires one of which is respiratory disorders. Smoke from wind-borne forest fires also harms countries around Indonesia, such as Malaysia, Singapore and Brunei Darussalam. Forest fires are caused by natural and human factors. The length of summer is one of the causes that causes forest fires (Sheldon and Sankaran, 2019). Forest fires that spread too wide make it difficult for forest firefighters to extinguish fires. The extent of forests in Indonesia also makes forest guards often unaware of the signs of forest fires.

The forest is a habitat for many animals, one of which is a bird. Birds are animals that are able to communicate with their colonies using sound. Sounds issued by birds have several purposes, namely to call the colony, attract the opposite sex during the breeding sea- son, and warnings of danger when under threat (Zhang et al., 2019). The sound needs to be studied more deeply so that it can be distinguished based on the conditions. The sound of birds when they feel threatened, especially threatened when they see hotspots and smoke can be used as an early warning of a fire in the forest. This study used 4 species of birds by conducting two voting schemes. The first scheme is record the sound when normal condi- tions and the second scheme is record the sound when birds feel threatened. The four types of birds that will be recorded are Aegit- hina tiphia (Cipoh), Prinia familiaris (Prenjak), pcinonotus goiavier (Merbah cerucuk) and zosterops palpebrosus (Pleci) (Ayat, 2011). The sounds of birds used in the study are birds that are mostly found in the agroforests of Sumatra and Kalimantan islands. Both types of sounds are based the scheme of each bird species will be trained and tested so that its characteristics can be distin- guished and then classified according to the conditions of the scheme.

Deep learning is a way to automate work to predict and classify data (Stowell et al., 2018). The process of entering data, processing and presenting it will be done automatically by the machine. The positive value of this research is reducing the professional human resources who must operate it and efficiency at work. The simplest and the most commonly used deep learning method is Convolu- tional Neural Network (CNN) (Florentin et al., 2020). CNN is a method that can classify data in the form images. Meanwhile, to classify sound in CNN is a program that will change the sound which is a digital signal to be converted into a spectrogram image so that it can be classified by CNN. CNN cannot process sound directly because CNN's algorithm does not have the facilities for voice learning. However, we can take an approach so that CNN can learn sound by converting digital signals into image spectro- grams. The resulting spectrogram images can be studied by CNN and classified by the features of the CNN itself. The CNN method has been widely used in previous studies in the type of image or sound classification. A previous study that used Back-Propagation Artificial Neural Network (BP-ANN) for forest fire prediction using image data are was presented in (LeBien et al., 2020). BP-ANN is a computational algorithm that mimics the workings of human nerve cells consisting of neurons that are connected in a network. The essence of this learning algorithm is to change the weight val- ues in response to errors. The change in weights is intended to minimize the number of network errors, so that the desired output is achieved.

Present study uses the CNN method with a supervised approach to train the collected bird sounds. Bird sound data that have been obtained will be converted into spectrogram images or digital sig-nals. The CNN method is basically only able to process data in the form of images or visuals. By converting the sound into a spectro- gram, the sound has been changed into visual form or image form (Song and Li, 2019). The trained data will then be tested to be clas- sified based on the scheme being carried out, namely the normal state scheme and the threatened state scheme. The difference in sound waves obtained will be the main parameter in distinguish- ing the two types of bird sounds based on the scheme. The results of the classification will be tested and evaluated to measure the accuracy of the classification in the form of a percentage. The high value of accuracy will provide conclusions for the methods applied in the classification of the data (Zottesso et al., 2018).

In research conducted by (Koh, et al., 2019) with the title ‘‘Bird Sound Classification using Convolutional Neural Networks” dis- cusses the Convolutional Neural Networks (CNN) method used to classify bird species based on sound using sound spectrogram fea- ture. Research focused on conservation is included in this BirdCLEF competition. In it there are 659 types of bird species and 50 thou- sand bird recordings of various species. This research adopts two architectures that have recently been popularly used, namely the ResNet and the inception model. The steps taken include prepro- cessing in which the sound obtained is converted into a spectro- gram using the MEL-scale log-amplitude spectrogram technique, after which the application of ResNet for bird sound identification is done, after which the implementation of the inception architec- ture is continued, the results of the study are obtained after 4 or more times indicating that the evaluation in the validation set it will be better before adding Gaussian Noise. The lack of explana- tion of the best research in the past made less visible the effective- ness of the latest architectural methods (the ResNet and the inception model) used in the same study. This research shows that

CNN is still the best method for bird speech recognition to be clas- sified. However, the lack of quality sound recordings owned by researchers can reduce the level of accuracy of the results of the study. Taking sound recordings independently with an adequate device might make the sound

quality of bird recordings more awake. A good sound recording will make the research process car- ried out by CNN easier and produce higher accuracy.
In research conducted by (Zhang, 2021) with the title ‘‘Diagno- sis of COVID-19 pneumonia via a novel deep learning architecture” develop a novel deep learning architecture to cope with COVID-19 diagnosis and prevent the deep neural network from overfitting. This study using the method Alex Net be a basis of the architecture and adding batch normalization. It’s used for internal covariance change is reduced, and preparation is accelerated. The fully con- nected layer in AlexNet was replaced with three randomized neu- ral networks: SNN, ELM, and RVFL in this study. The DL system used for detection from the CTIs of the chest of COVID-19 patients was invented and tested due to this research. Their invited model has high sensitivity (0.8568), high specificity (0.9613), precision (0.9570), and accuracy (0.9091).
In a study conducted by (Sulyok, 2019) with the research title ‘‘Bird Sound Recognition Using a Convolutional Neural Network” discusses the classification of species and classes of birds based on sound using the Convolutional Neural Network (CNN) method. The use of the Short Time Fourier Transform (STFT) method in this study can easily change input cues by trimming the frequency time so that it can calculate the components or content that is in fre- quency. The pruning is useful for pre-processing in CNN to uniform input data before it is processed. However, the use of STFT direct recording data is not good because it is not able to read when the snippet is used. CNN training is carried out in two different steps. Training on the colored spectrogram dataset and training on the black and white spectrogram dataset. Both data that has been classified with CNN will be saved as frozen inference graph as test data. The results showed a little better when using a colored spectrogram compared to gray. However, this requires CNN archi- tecture with deeper layers and longer training. In cases that require faster classification results it is better to use a black and white or gray spectrogram to summarize the work of the system when con- ducting training on just one color data with not three or more colors.
This research will classify bird sounds based on the scheme by studying the dataset that is already owned. Learning outcomes will be tested with test data that has been prepared. It is hoped that this research can be useful for forest guards in knowing the poten- tial for forest fires based on bird sounds.
In the remaining sections of this paper is Section 2 about the explanation of the materials and methods to be used, Section 3 is the results of implementing the method in the case, Section 4 is a brief discussion of the cases and methods discussed in this paper and Section 5 is a conclusion taken from the entire research conducted.

2. Materials and methods

The methodology in present study includes data collection, data processing, training, testing and evaluation as presented in Fig. 1. The main method used in this research is the Convolutional Neural Network (CNN) method. CNN is one method that is often used in Deep Learning to classify image shapes automatically.
Research entitled ‘‘Handcrafted features and late fusion with deep learning for bird sound classification” Authored by (LeBien et al., 2020) discusses the importance of bird monitoring to opti- mize protection policies and increase bird population. Automatic classification helps monitoring in large-scale temporal and spatial environments. In this study a combination of bird species classifi- cation using an acoustic or sound approach, an approach with visual or imagery and Deep Learning. This study aims to maximize the performance of bird classification in order to achieve the best results. The method used in this study is the CNN to study general features and

the reduction of dimensions and fully connected lay- ers will classify bird species. Experiments carried out with 14 spe- cies of birds showed results for the acoustic approach alone 88.97%, while for the visual approach the results were 88.87% and when they were combined coupled with Deep Learning produced a value of 94.36%. This indicates that the merging of two approaches cou- pled with Deep Learning has the highest results. CNN as a method used plays an excellent role in aligning sound classifications by making them in the form of images. However, this study does not explain the greatest advantage of using each approach and also whether the methods used are in the same position in the class. By explaining these things can provide important information that traditional methods will be much better when done automatically, in addition to making it easier for users (See Figs. 2 to 4).

Research presented in (Berger, et al., 2018) discussed about new sound experiments from the wild to see whether the votes get the desired assessment results or not. This study collected data on bird sound recordings from DCASE 2018 which is one of the portals for labeled bird sound recordings. The sound used has 3 different con- ditions, namely different times, different types of microphones and different recording time durations. The study was conducted using the CNN method and Base Line method. The results of the study show that only CNN is able to adjust the complexity of the given problem. The CNN in this case has enough layers to solve complex problems. Each of the layers is related so that the process can run with better learning. The results of the study conducted tests of sounds other than the DCASE 2018 record had an average accuracy of 0.81 from a range of values from 0 to 1. The study used less birds, so the data entered could have been only bird sounds that were similar to bird sounds processed in training and testing. Testing with sound from nature will be more convincing the success of research. Because, basically the research has indirectly been in the testing phase directly with real data.

2.1. Data collection

The data used in this study is the recorded birds sound from 4 species of birds. Four types of birds used in this study were Cipoh, Prenjak, Merbah cerucuk and Pleci. These birds are typically found in agro forests on the islands of Sumatra and Kalimantan (Berger, et al., 2018). Chirping will be done directly on 4 species of birds with two voting schemes. Chirping and syllables in the sound of birds are two different things. Chirping is a series of syllables

issued by birds, while syllables are sounds issued by birds in one breath.

This study uses 580 bird sound samples consisting of 480 sounds taken using smartphone voice recording media with sam- pling rate of 48 kHz and 100 evaluation sounds downloaded from the Xeno-Canto website. The 480 sounds are the sounds of 4 spe- cies of birds namely, Cipoh, Prenjak, Merbah cerucuk and Pleci. The sound composition is 60 votes per type and per voting scheme, voting scheme when normal and voting when threatened. Threat- ened conditions in the scheme will be carried out by simulating fires with fire torches and smoke from simulated fires which will make birds feel threatened. This was done to simulate the occur- rence of fires in the forest. It is presumed that birds emit a different tweet when they are under threatened condition, which is differ- ent with the one under normal condition. This different will be studied further using CNN method.

The sound of birds singing collected is the sound of birds in one song. One song issued by a bird is usually called a syllable (Fanioudakis and Potamitis, 2017). Syllable is the reason for the brevity of bird sound recordings that are used as a dataset. A sylla- ble is a sound made by a bird in one blow. The recorded bird sound data will be processed so that it can be processed using the super- vised learning's CNN method. The sound obtained will be con- verted into an image with the name spectrogram image.

This study used four types of bird species: Cipoh, Prenjak, Mer- bah Cerucuk, and Pleci. This bird is native to the tropical forest endemic in Sumatra and Kalimantan. In this research, two sound schemes will be used, namely the sound of birds in normal condi- tions and birds' sounds in danger. Normal sounds produced by birds are birds' sounds when they are given food and drink and are properly cared. A bird's sound in danger is the sound of a bird being scared by the fire around it. In taking the voice used addi- tional microphone media for smartphone devices. The mic used is a type of unidirectional mic which is a mic with the ability to capture one-way sound or focused only from one direction only, so that the sound recording process can be more focused on bird sounds and reduce other sound disturbances.

The votes collected were 480 votes out of 60 votes per voting scheme per bird species. Total votes per bird are 120 per species and 480 votes from 4 bird species and 2 voting schemes. The sounds obtained are given a name in the format (sequence) (bird type) .wav. Example. 1Prenjak.wav. The sounds are collected in one folder with the name ‘‘clean” in the folder location in C: \Users\(username)\clean.

2.2. Preprocess

Data from bird recordings is estimated to be around sixty per sound collection scheme for each bird species, so that there will be a total of approximately 480 pieces of bird song chirping with a duration of 1 s or less. The number of bird sounds was obtained from 4 bird species with 60 normal sound data and 60 threatened sound data per species. Bird sound recordings that have been col- lected from 4 species of birds with 2 schemes will be cut to 240 sounds per scheme. So, there are 480 bird sounds. The 480 sounds will be divided into training dataset and testing dataset with a ratio of 80% of the 480 bird sounds to the training dataset and 20% of the 480 bird sounds into the testing dataset. Training dataset is used to conduct training or learning in the system to obtain high accuracy results in the final result. Dataset testing is used to test the accu- racy of systems that have ‘‘learned” from existing training data (Cakir et al., 2017). Both data that has been shared will be synchro- nized with the requirements and system requirements to be processed.

Both types of data, training dataset and testing dataset are trun- cated in order to resize the signal dimension into the same dura- tion (Qian et al., 2017). For example, the duration of recorded birdsong sound recordings in the training dataset is 0.33 s to 0.4 s. The data is truncated to 0.33 s to produce the similar data size. It is because, the CNN method is only able to process images of the same size. Alignment of the frequency image in the spectro- gram is performed to meet the requirements of the CNN method (Küçüktopcu et al., 2019). The resize data is conducted by two pro- cedures: (1) remove the part of signal that is identified not a bird sound, and (2) cutting off the top or the end of the signals to obtain a similar or uniform dataset. The accuracy of the training and the testing process will be improved if the signal spectrograph does not have ambiguous signals or signals from noise interference as well as the uniform data size. In addition, the external sound that is not from is expected to not have a large sound wave contribu- tion, because it can affect the value of accuracy when tested (Xie and Zhu, 2019).

2.3. Training

The data training stage is the stage for the system of studying the image that has been provided. The data used in the training stage is spectrogram data that has been converted from the digital sound that the birds have produced when recorded. The spectro- gram data will be used for the training process in CNN. The training dataset that has been prepared will be trained through the convo- lution and ReLu stages over and over (Niemi and Tanttu, 2018).

ReLu or Rectified Linear Unit is an activation layer available in the CNN method with the formula function f (x) = max (0, x) (Xie and Zhu, 2019). ReLu is useful for thresholding or making values less than 0 become 0. Relu Activation is illustrated as shown in Fig. 5.

ReLu is done after the input data has been convoluted or filter calculated. The convolution equation applied is as in the following function.

3. Results

In this stage there will be a discussion about the application of the Convolutional Neural Network method to Deep Learning to classify bird sounds for early warning of forest fires (See Table 1).

3.1. Similarity data

Sounds collected from 4 species of birds and 2 sound collection schemes initially are amalgamated into one sound recording with a duration of 1–2 min per recording. In this research the birds sound taking using additional microphone media with the ability to fre- quency response 100 Hz 16 kHz. An example of one bird sound signal which is presented in Fig. 9 is then converted into a spectro- gram as presented in Fig. 10.

Data acquisition is done to cut the duration of bird sounds per syllable. Cutting using the additional application Audacity. Reduc- tion of recording background noise is also done before cutting the syllable. The cut of the sound has been converted into a spectro- gram as shown in Fig. 11.

3.2. Labeling data

Prior to the CNN computation, the recording bird sound signals were labelled with the following filename: (data collection sequence)_(bird species).wav e.g. 1Prenjak.wav. All data names will be registered as one list in the number processing application (Ms. Excel). List of names and labeling is done in a number process- ing application, where normal bird sound data will have normal labels and for threatened bird sound data will have threatened labels. Examples of lists of names and labels are presented in Table 2.

The list of names is stored as ‘‘instruments.csv” so that it can be used in Deep Learning for processing more than one data. The list of names will be used to train the system in recognizing data that will be used in the process of distributing datasets, training data and testing data in Deep Learning. The list of names is stored in the folder location C: \ Users \ (username) \. Following are the data compositions used in this study (See Fig. 12).

3.3. Preprocessing data

The preprocessing data stage is necessary for preparing useful (i.e. uniform) data to be processed. The voice that has been

acquired will be synchronized to the same duration as the smallest tempo of the syllable available. In the votes collected the smallest syllable is 0.9 s long and the syllable with the longest duration is 1.2 s. All sounds will be uniformly resized using Audacity to a dura- tion of 0.9 s to be used in the training and testing process to improve the accuracy in bird sound classification. The uniformed sounds will be divided into two, namely the training dataset and the testing dataset. After importing the data, the next step is to divide the data into training dataset and testing dataset (See Table 3).

Dividing the dataset into 80% as a training dataset and 20% test- ing the dataset. The data is stored in each created variable. X is a simile of rows and y is a column in a matrix. The X_train variable stores the training row dataset from the total dataset that has been shared and X_test variable stores the testing dataset from the total dataset that has been shared. Likewise, with the data in variable y which is a column simile in the matrix of the dataset. An example of data presentation including the variables and the classification labels is presented in Fig. 13.

3.4. Training

The training process utilized the train dataset contained in X_- train and y_train. The output of the training is a learning model obtained from learning classification of the two different datasets (normal condition and threated condition). In the training process, the convolution or filtering process using 2-dimensional convolu- tion with filters 16, 32, 64 and 128, and kernel size 3x3. The convo- lution uses ReLu activation with 1x1 stride and padding. The dimensions of the input dimensions are summarized in the vari- able ‘‘input_shape” of 9 ———————————— 13. The CNN processes the hidden layer so that many calculations are not displayed, such as filtering, pool- ing, flatten and fully connected. The hidden layer has several layers so that the output displays the calculation results and dimensions used as information about the size of the filter used, what stages are carried out, and what are the weights or parameters of each stage. The output of the training is presented in Fig. 14.

4. Discussion

As explained previously that this study is a preliminary study of bio inspired forest fire early warning by utilizing bird sounds. In this study, researchers will make machines to be able to distin- guish normal and threatened sounds from the same bird species. The characteristic sound of a bird when threatened and normal can be recognized when converted into an image or spectrogram. The sound is processed in deep learning using the CNN method where the method is the method most often used in the case of sound classification including bird sounds.

Each method has advantages and disadvantages, as with the CNN method. CNN method is able to classify sound in the form of images very well, the results in this study alone have a success value of 96.45%. This success also resulted from the lack of a system in determining prediction or classification results. Some classification data that has the same value between choices will prefer the label that was first given. The system does not give a special mark on the predicted value parallel to the available label choices. These deficiencies are expected to be given a solution so that the system can explain the results of true and false also with results that cannot be determined or parallel.

CNN has the advantage of classifying images quickly and in real-time. CNN has an accurate method of recognizing images. However, even though it has benefits, CNN does not escape its shortcomings. Both other deep learning methods are building and conducting training on these costly methods, both from here the time and computational burden. Even with a Graphics Process- ing Unit (GPU), which can speed up parallel processing, training modern CNNs on large-scale datasets, such as ImageNet, is still rel- atively expensive.

5. Conclusions

A proposed bird sound classification method based on the CNN algorithm has been presented. Some important points are drawn as follow: (1) the greater the epoch in training will have a positive effect on the accuracy value curve, while the greater epoch when doing training will have a negative impact on the loss value curve; (2) classification of bird sounds under normal conditions and threatened using CNN method has 96.45% accuracy; (3) according to the training process using 50 epoch obtained an average value of accuracy of 0.9276 and an average value of accuracy validation of 0.9121. The method will be extended in future study for prediction of the early warning forest fire using the actual bird sound and other animal sound by initiating a joint research between Researchers and Ministry of Forestry in Indonesia.

References

Ayat, A. (2011) Burung-burung Agroforest di Sumatera.

Berger, F. et al. (2018) ‘Bird Audio Detection - DCASE 2018’, pp. 1–4. Yellow-vented Bulbul (Pycnonotus goiavier) :: xeno-canto (no date). Available at:

https://www.xeno-canto.org/species/Pycnonotus-goiavier (Accessed: 20 July 2020).

Cakir, E. et al. (2017) ‘Convolutional recurrent neural networks for bird audio detection’, 25th European Signal Processing Conference, EUSIPCO 2017, 2017- Janua, pp. 1744–1748. doi: 10.23919/EUSIPCO.2017.8081508.

Fanioudakis, L. and Potamitis, I. (2017) ‘Deep Networks tag the location of bird vocalisations on audio spectrograms’. Available at: http://arxiv.org/abs/ 1711.04347.

Florentin, J., Dutoit, T., Verlinden, O., 2020. Detection and identification of European woodpeckers with deep convolutional neural networks. Ecol. Inf. 55, 101023. https://doi.org/10.1016/j.ecoinf.2019.101023.

Common Iora (Aegithina tiphia): xeno-canto (2020). Available at: https://www. xeno-canto.org/species/Aegithina-tiphia (Accessed: 20 July 2020).

KLHK (2019) ‘Rekapitulasi Luas Kebakaran Hutan dan Lahan (Ha) Per Provinsi Di Indonesia Tahun 2014-2019’, SiPongi - Karhutla Monitoring Sistem, (1), pp. 26– 27. doi: 10.1103/PhysRevB.65.020504.

Koh, C. et al. (2019) ‘Bird Sound Classification using Convolutional Neural Networks’, pp. 9–12.

Küçüktopcu, O., Masazade, E., Ünsalan, C., Varshney, P.K., 2019. A real-time bird sound recognition system using a low-cost microcontroller. Appl. Acoust. 148, 194–201. https://doi.org/10.1016/j.apacoust.2018.12.028.

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J.P., Dodhia, R., Ferres, J.L., Aide, T. M., 2020. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecol. Inf. 59, 101113. https://doi.org/10.1016/j.ecoinf.2020.101113.

Niemi, J., Tanttu, J.T., 2018. Deep learning case study for automatic bird identification. Applied Sciences (Switzerland) 8 (11), 1–15. https://doi.org/ 10.3390/app8112089.

Bar-winged Prinia (Prinia familiaris) :: xeno-canto (2020). Available at: https:// www.xeno-canto.org/species/Prinia-familiaris (Accessed: 20 July 2020).

Qian, K., Zhang, Z., Baird, A., Schuller, B., 2017. Active learning for bird sound classification via a kernel-based extreme learning machine. J. Acoust. Soc. Am. 142 (4), 1796–1804. https://doi.org/10.1121/1.5004570.

Sheldon, T.L., Sankaran, C., 2019. Averting behavior among singaporeans during indonesian forest fires. Environ. Resour. Econ. 74 (1), 159–180. https://doi.org/ 10.1007/s10640-018-00313-8.

Song, J., Li, S., 2019. Bird sound detection based on binarized convolutional neural networks. Lecture Notes Electr. Eng. 568, 63–71. https://doi.org/10.1007/978- 981-13-8707-4_6.

Stowell, D., Wood, M. and Pamuła, H. (2018) 'arXiv : 1807 . 05812v1 [cs . SD] 16 Jul 2018 Automatic acoustic detection of birds through deep learning : the fi rst Bird Audio Detection challenge', (15), pp. 1–21.

Sulyok, C. (2019) 'Bird Sound Recognition Using a Convolutional Neural Network', (September 2018). doi: 10.1109/SISY.2018.8524677.

Ambon White-eye (Zosterops kuehni) : xeno-canto (202). Available at: https:// www.xeno-canto.org/species/Zosterops-kuehni (Accessed: 20 July 2020).

xeno-canto: Sharing bird sounds from around the world (no date). Available at: https://www.xeno-canto.org/ (Accessed: 19 July 2020).

Xie, J., Zhu, M., 2019. Ecological informatics handcrafted features and late fusion with deep learning for bird sound classi fi cation. Ecol. Inf. 52, 74–81. https://doi. org/10.1016/j.ecoinf.2019.05.007.

Zhang, X. et al., 2021. Diagnosis of COVID-19 pneumonia via a novel deep learning architecture. J. Comput. Sci. Technol. https://doi.org/10.1007/s11390-020- 0679-8.

Zhang, X., Chen, A., Zhou, G., Zhang, Z., Huang, X., Qiang, X., 2019. Spectrogram- frame linear network and continuous frame sequence for bird sound classification. Ecol. Inf. 54, 101009. https://doi.org/10.1016/j.ecoinf.2019. 101009.

Zottesso, R.H.D., Costa, Y.M.G., Bertolini, D., Oliveira, L.E.S., 2018. Bird species identification using spectrogram and dissimilarity approach. Ecol. Inf. 48, 187– 197. https://doi.org/10.1016/j.ecoinf.2018.08.007.