JASA ARTICLE

Using deep learning for acoustic event classification: The case of natural disasters[a)]

Akon O. Ekpezu,[b)] Isaac Wiafe,[c)] Ferdinand Katsriku,[d)] and Winfred Yaokumah[e)]
Department of Computer Science, University of Ghana, Post Office Box 163, Legon, Accra, Ghana

ABSTRACT:
This study proposes a sound classification model for natural disasters. Deep learning techniques, a convolutional neural network (CNN) and long short-term memory (LSTM), were used to train two individual classifiers. The study was conducted using a dataset acquired online[1] and truncated at 0.1 s to obtain a total of 12 937 sound segments. The result indicated that acoustic signals are effective for classifying natural disasters using machine learning techniques. The classifiers serve as an alternative effective approach to disaster classification. The CNN model obtained a classification accuracy of 99.96%, whereas the LSTM obtained an accuracy of 99.90%. The misclassification rates obtained in this study for the CNN and LSTM classifiers (i.e., 0.4% and 0.1%, respectively) suggest less classification errors when compared to existing studies. Future studies may investigate how to implement such classifiers for the early detection of natural disasters in real time.
https://doi.org/10.1121/10.0004771

(Received 30 November 2020; revised 11 March 2021; accepted 4 April 2021; published online 30 April 2021)

[Editor: Peter Gerstoft] Pages: 2926–2935

I. INTRODUCTION

Natural disasters pose a major risk to physical and environmental securities and often result in the loss of properties and lives. Disasters such as tsunami, volcanoes, hurricanes, and earthquakes are events that exhibit infrasonic signatures with low frequencies which are inaudible to the human ear. The early detection of such low-frequency infrasonic signatures from the movement of the Earth's floor poses a significant technological challenge. However, being able to detect their occurrence early provides meteorological scientists and researchers with the ability to develop early warning systems and help limit its devastating impact (Goswami *et al.*, 2018; Wallemacq and Guha-Sapir, 2015). Accordingly, some researchers have developed techniques to predict and classify natural disaster occurrences and their severities (Doxani *et al.*, 2019; Furquim *et al.*, 2018; Khalaf *et al.*, 2015, 2018; Sayad *et al.*, 2019). Indeed, the impact and damage caused by natural disasters can be reduced if advances made in instrumentation are integrated with information and communication technology tools and emerging artificial intelligence (AI) algorithms to develop effective warning systems. For instance, AI techniques are capable of using images of disaster scenes from surveillance cameras to classify and automatically assign disaster names and send relevant information to disaster prevention networks (Okamoto *et al.*, 2018). Nonetheless, the use of images for disaster detection is challenged because meteorologists need to analyse large amounts of satellite images within a short period to make decisions. This results in unobserved occurrences and inaccurate detection reports (Antoniou and Potsiou, 2020). Besides, the use of remote sensing satellite methods are expensive and characterized by temporal delays between 48 and 72 hours for the production of relevant information (Resch *et al.*, 2018). These delays in information dissemination increase the vulnerability levels (Wisner and Adams, 2002), resulting in ineffective protection of life and property (Boustan *et al.*, 2017). For example, the effects of an unreliable information dissemination system were experienced during the 2010 earthquake in Maule, Chile: no warnings or evacuation plans were made three hours after the tsunami hit the Chilean coasts (Soulé, 2014).

The use of acoustic signals allows long-term monitoring of large areas and facilitate the collection of larger amounts of data in real time (Chen *et al.*, 2019; Malfante *et al.*, 2018). Therefore, it is more reliable when compared to image and video datasets. Acoustic signals are not affected by variations in light intensity. The microphones used for collecting acoustic data are omnidirectional, whereas existing cameras used for capturing images for such analysis are not. Also, it is easier to quantify the characteristics of acoustic data into duration, frequency, intensity, and progression over the time of the eruptions. While sonobuoys, acoustic sensors, and underwater acoustic sensing techniques have been used to obtain and classify sound/acoustic signals that precede natural disasters such as hurricanes, tornadoes, and earthquakes, these methods are expensive. Detecting an acoustic signal that precedes a disaster depends on the sensing ability (i.e., using sensor technologies). This is because the preceding sound, also known as infrasound, has lower frequencies below 20 Hz: this is below the threshold of

[a)]This paper is part of a special issue on Machine Learning in Acoustics.
[b)]ORCID: 0000-0002-9502-1052.
[c)]Electronic mail: iwiafe@ug.edu.gh, ORCID: 0000-0003-1149-3309.
[d)]ORCID: 0000-0003-2596-9478.
[e)]ORCID: 0000-0001-7756-1832.

human hearing. Again, they cannot be detected by traditional microphones. Hence, there is a need to improve and also diversify existing methods and approaches in disaster predictions and detections.

Integrating AI into disaster management systems is promising. It provides an efficient method that is capable of predicting, detecting, and classifying disasters. Nonetheless, current AI techniques, including machine learning (ML) algorithms, mostly adopt approaches that use historical numerical data, images, and videos of disaster scenes. These approaches have been proven to be either insufficient or inefficient (Boustan *et al.*, 2017) because there are discrepancies in the numerical data (Joshi, 2019), variations in light intensity on images and videos (Aziz *et al.*, 2019), and limited coverage or monitoring area (Mitilineos *et al.*, 2018).

This study, therefore, seeks to expand the knowledge base in disaster classification using sound/acoustics signals. It proposes deep learning techniques for classifying natural disaster sounds. Specifically, it adopted two ML techniques [i.e., a convolutional neural network (CNN) and a recurrent neural network (RNN) with a long short-term memory (LSTM) architecture] for classifying natural disasters. It is expected that findings from this study will provide an effective alternative approach for classifying natural disasters. It shall also provoke further investigations into the construction of early detection and classification systems that employ acoustic signals. In Sec. II, a discussion on related studies is presented with highlights on deep learning architectures adopted in this study, followed by a description of how the study was conducted. The results and findings are presented in Sec. IV and the discussion and conclusions are drawn in Secs. V and VI, respectively.

II. RELATED LITERATURE

A. Natural disaster classification

Due to the devastating effects of disasters, it is imperative to put effective methods in place for early detection and classification. Recent advancements in AI have made notable contributions to classification problems, including disasters (Antoniou and Potsiou, 2020). More importantly, the use of ML techniques to classify natural disasters has been explored by researchers using various approaches. Amit *et al.* (2016) proposed an automatic disaster detection system in which a CNN is used to analyse satellite images. The process takes advantage of the robustness of the CNN to extract event features from images and classify them into various categories with a classification accuracy between 80% and 90%. In flood severity classification, random forest (RF) has been used to classify the level of flood severity using images of disaster scenes (Khalaf *et al.*, 2018). This approach provides optimal results in the accuracy, precision, and recall when compared to other ML techniques such as the support vector machine (SVM), linear neural network (LNN), and random oracle models (ROM). In classifying disaster-prone areas, Supriyadi *et al.* (2018) used K-means clustering to obtain provincial disaster-prone areas based on historical numerical data of areas that experienced natural disasters.

Although ML techniques have demonstrated potentials in disaster classification studies, the majority of these studies tend to focus on the disaster severity, disaster scenes, number of casualties or affected population, and the validity of the disaster reports. Predominantly, image datasets are used. Fewer studies have sought to predict and/or detect the occurrence of disasters by classifying acoustic signals or seismic data that precede the disaster. However, Wilson and Makris (2006) have demonstrated that acoustic data can be used to predict natural disasters by using wind-generated noise to estimate the relationship between the local wind speed and underwater acoustic intensity. In their study, they collected data from ocean acoustic sensors and used it to classify hurricanes while they are still far from land. Again, the use of acoustic data for disaster detection has been demonstrated by Rouet-Leduc *et al.* (2017): RF was used to predict the laboratory earthquake based on the instantaneous analysis of acoustic signals. Although the laboratory setup closely mimicked earthquakes and the results provide evidence to encourage the use of ML techniques to analyse acoustic data, studies that adopt acoustic signals and ML techniques for disaster classification have not received adequate attention.

B. ML techniques for acoustic signal classification

Existing studies have argued that the neural network achieves the best classification results in most cases when compared to others (Aziz *et al.*, 2019; Ibrahim *et al.*, 2018; Mousa *et al.*, 2016; Zhang *et al.*, 2018). This is because neural networks are self-adaptive universal approximators. Undoubtedly, these algorithms are efficient for acoustic related investigations. For instance, Salamon and Bello (2017) demonstrated that by using a three-convolutional layer CNN, an acoustic classification model with an accuracy value of about 73% can be obtained. In classifying the environmental sound with a two-layered CNN, Khamparia *et al.* (2019) also achieved similar results. In medical acoustics, Aykanat *et al.* (2017) developed a noninvasive method of classifying respiratory sounds and improved the model performance by combining the SVM and CNN to advance pre-diagnosed respiratory sound classification. Others (Zhang *et al.*, 2018) have shown that deep neural networks (DNNs) and a RNN with a LSTM-based architecture outperform the SVM with classification accuracies of 95.60% and 91.67%, respectively. The DNN and LSTM are capable of learning dynamic characteristics of acoustic features (i.e., energy, peak ratio of autocorrelation, and log Mel-filter bank spectrum). In this study, the CNN and LSTM are used.

CNNs are mostly used in computer vision for image recognition and classification; however, in recent times, they have been demonstrated to be effective in acoustic signal classification. For example, the CNN is effective and capable of learning and identifying spectro-temporal patterns in the presence of noise (Salamon and Bello, 2017). It can capture energy modulation patterns across the time and frequency domains for a given acoustic feature (Maccagno *et al.*, 2021) by considering the spectrum of audio signals as a picture of the signal during model training processes. In effect, the spectrogram image features are used for classification. Although the CNN has been adopted in classifications of environmental sound, bioacoustics, and medical acoustics, Thakur *et al.* (2019) posits that studies on the CNN for acoustic classification are limited due to data scarcity. Thus, some studies use augmentation techniques to increase the data size for their experiments.

A RNN is a DNN with a recurrent hidden state that serves as a "memory" for the RNN, learning significant features from the upcoming data sequence, and predicting the subsequent actions based on the learned features. RNN performs well on time series data (Raza *et al.*, 2019; Verma *et al.*, 2019). They are used for modelling features that change over time and are efficient in tasks that involve sequential inputs such as speech and language. Training a RNN using the stochastic gradient descent method to capture long-term dependencies can be a challenge, however, because the gradient of the loss function decays exponentially: the "vanishing" gradient problem. A number of methods have been proposed to mitigate this problem. One such approach is the use of LSTM, consisting of a cell, an input gate, an output gate, and a forget gate. The gates control the flow of information into the cell, thus, helping the cell to maintain information for longer periods. Many studies have shown that LSTM performs better when compared to RF, decision tree, and multilayer perceptron (MLP; Raza *et al.*, 2019).

C. Acoustic signal classification

In acoustic classification, audio segments are classified into specific groups through either detection-and-classification [i.e., acoustic event detection (AED)] or detection-by-classification [i.e., acoustic event classification (AEC)]. Whereas the former involves the detection of sound before classification, the latter involves sound detection by classifying the audio segments. In detection-and-classification, no classification decision is made, rather segmentation is done when a segment boundary is detected based on a chosen threshold, followed by localization. Conversely, in detection-by-classification, the task of detection automatically translates to classification as its strategy is based on using classifiers [such as the hidden Markov model (HMM), logistic regression (LR), and artificial neural networks (ANNs) with inbuilt segmentation algorithms].

AEC has been used in bioacoustics (Allen *et al.*, 2017; Pandeya and Lee, 2018), medical acoustics (Vrbancic and Podgorelec, 2018; Yaseen *et al.*, 2018), action detection (Pandeya and Lee, 2018), environmental monitoring (Aziz *et al.*, 2019), and surveillance (Maccagno *et al.*, 2021). Yet, the challenge is that raw sound data are mostly not suitable for sound classification. They contain irrelevant information or features that impede the effectiveness of the classification models. Accordingly, relevant acoustic features are extracted and used for training classifiers. The feature extraction process facilitates classification and makes them more efficient and accurate because it reduces the size of the data and presents complex datasets as feature vectors. Most sound classification studies adopt features from speech or voice recognition systems (Karbasi *et al.*, 2011; Mitilineos *et al.*, 2018). These features predominantly include the Mel frequency cepstral coefficient (MFCC) and spectrograms (Nanni *et al.*, 2020). There are, however, other features, such as the entropy peak (Robakis *et al.*, 2018), spectral centroid (Lebien and Ioup, 2018), high frequency content (Allen *et al.*, 2017), zero-crossing rate (Zhang *et al.*, 2018), and peak ratio of autocorrelation (Oikarinen *et al.*, 2019; Zhang *et al.*, 2018), that are also suitable for extracting sound features during classification.

In this study, the MFCC is adopted because it is capable of capturing the most relevant sound features even in the presence of ambient noise (Aykanat *et al.*, 2017; Aziz *et al.*, 2019; Yaseen *et al.*, 2018) and it leverages on the robustness of the CNN-based classifiers to improve the model performance (Verma *et al.*, 2019).

III. STUDY DESIGN

Acoustic signal classifiers are broadly considered to be either discriminative or non-discriminative: the former entails modelling a decision boundary in the training data and matching its test input to a specific data class, whereas the latter attempts to explicitly model the actual distribution of each class. This study investigated the use of the CNN and LSTM for developing discriminative models to classify seismic acoustic events. Primarily, it seeks to assess the efficacy of these classifiers in identifying various types of natural disasters using their related acoustic signals. The study comprises four broad steps: signal preprocessing, feature extraction, training of the classification model, and its evaluation. This section presents discussions on the first two steps (i.e., signal preprocessing, feature extraction) and the architecture of the neural networks used.

A. Datasets

Data were extracted from the Freesound website.[2] The Freesound website is a repository of different sound effects of varying time frames and formats (WAV and MP3 formats). In this study, four types of natural disaster sounds in WAV file format were downloaded. Each category was comprised of 50 recordings with varying time frames and a sample rate of 4100 Hz. The time frames ranged between 2 and 120 s with a total duration of 12 937 825 s. Figure 1 shows the distribution of the four categories of natural disaster sounds, namely, earthquake, tsunami, tornado, and volcano, and Fig. 2 is the corresponding time series representation.

B. Signal preprocessing

The sounds (i.e., acoustic signals) were preprocessed in two phases to facilitate feature extraction: denoising and

https://doi.org/10.1121/10.0004771

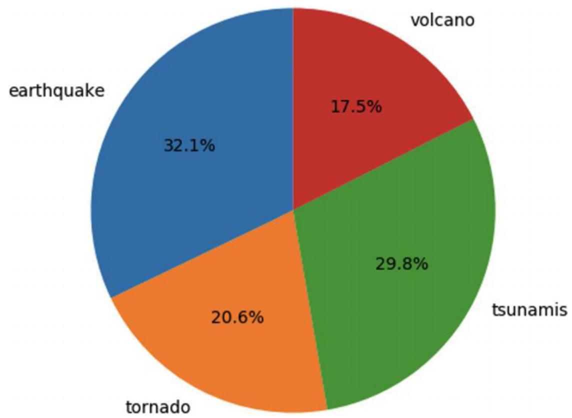

FIG. 1. (Color online) The class distribution of natural disaster sounds.

down-sampling the signals. The denoising process removed dead spaces/irrelevant data from the signals. In the down-sampling process, signal reduction techniques were employed to reduce the sound features by extracting more discriminative features. These features were used for training the classifier. With a window size of 25 ms, audio segments were down-sampled from 44 100 to 16 000 Hz. A directory was created to store the output files.

C. Feature extraction

Feature extraction reduces or eliminates the amount of reductant features to improve the model accuracy, avoid overfitting, and reduce the model training time. In acoustic datasets, signals are mostly nonstationary with superimposed background noises. These noises originate from natural ambience. Mostly, acoustic features are either related to time-frequency features (i.e., time-domain representation), spectral features (frequency-domain representation), or statistical features. Because seismic activities have a wide range of spectral content, feature extraction is performed on spectral features of audio recordings based on a filter bank-based MFCC approach. Although there are various approaches for filter bank feature extraction methods, including those used for sound feature extraction, MFCC is appropriate for audio-specific feature extraction. It is designed to describe sounds that are audible to humans (Malfante *et al.*, 2018) and has a better recognition rate when compared to the linear predictive cepstral coefficient (Prithvi and Kumar, 2016). It represents signal information efficiently (Yaseen *et al.*, 2018). More importantly, the low dimensionality of the MFCC makes it more efficient for feature extraction, considering that ML classifiers have the tendency to overfit if the dimensionality of the input representation is high. It is also commonly used for its ability to leverage the robustness of CNN-based classifiers (Verma *et al.*, 2019).

Due to the down-sampling performed earlier, the number of MFCC features was reduced from 26 to 13, whereas the fast Fourier transform (FFT) was reduced from 1103 to 512. The 13 MFCCs were calculated using a down-sampled 512-point FFT with 0.1 s window frame length. Specifically, the FFT was applied to the waveforms of each window frame before the log Mel-filter bank spectrum was extracted to obtain spectral reports of each frame. The resultant spectrum was tied to a Mel-frequency scale by passing it through a Mel-filter bank created by triangle filters. The logs of the outputs were used to develop the log Mel-filter bank spectrum for each frame, and the MFCC was obtained by applying a discrete cosine transform (DCT) to the filter banks.

D. Network architecture

Considering that class imbalances may result in low performance of some classes and also impair the process of training a classification model, an arbitrary length of time of a tenth of a second (0.1 s) was preferred. This time frame was also used to enable a faster discernment of different classifications in real time. Thus, random sampling was used to extract 0.1 s portions of the 200 sound recordings. This resulted in 12 937 sound samples. The datasets were split into two sets: 80% (10 349 samples) for training and 20% (2588) for testing. The CNN used in this experiment consists of 4 fully connected layers with 16 filters built with a 3×3 convolution (see Fig. 3 for CNN architecture used in this study). All layers had a rectified linear unit (ReLU) activation function of 1×1 stride and padding as "same," a 2×2 kernel for the max-pooling, and a dropout layer to reduce overfitting on the training data.

The LSTM had an input data shape of (*n, time, feat*); *n* is the number of sound samples (12 937), time is the time distribution (9), and feat is the number of features (13), 1 recurrent layer, a *0.5* dropout and a time distributed fully connected layer with 64 neurons (see Fig. 4 for LSTM architecture used in this study). It also used a ReLU activation, flatten, and a SoftMax activation. The time distribution was carried down from layer to layer. This allowed more parameters to be created while enabling deeper modelling.

Anaconda Python and various TensorFlow libraries, including Numpy, Matplotlib, and Scikit-learn, were used to

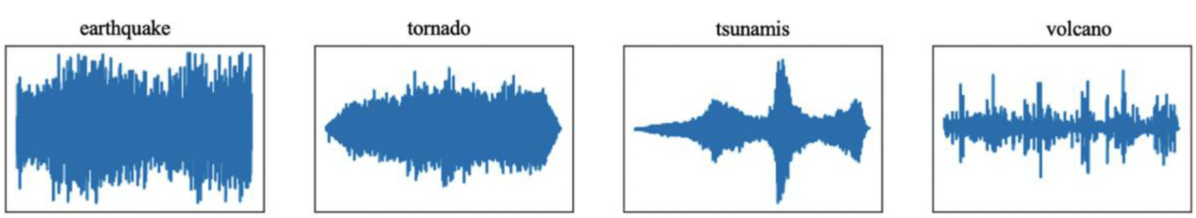

FIG. 2. (Color online) The time series representation for the types of sounds used in the study.

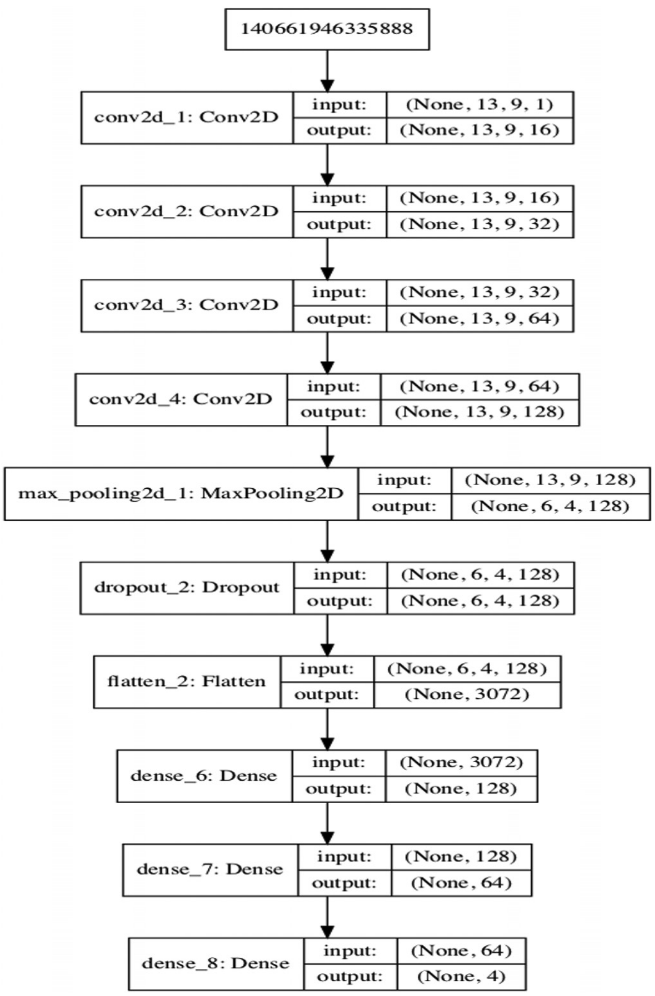

FIG. 3. The CNN architecture.

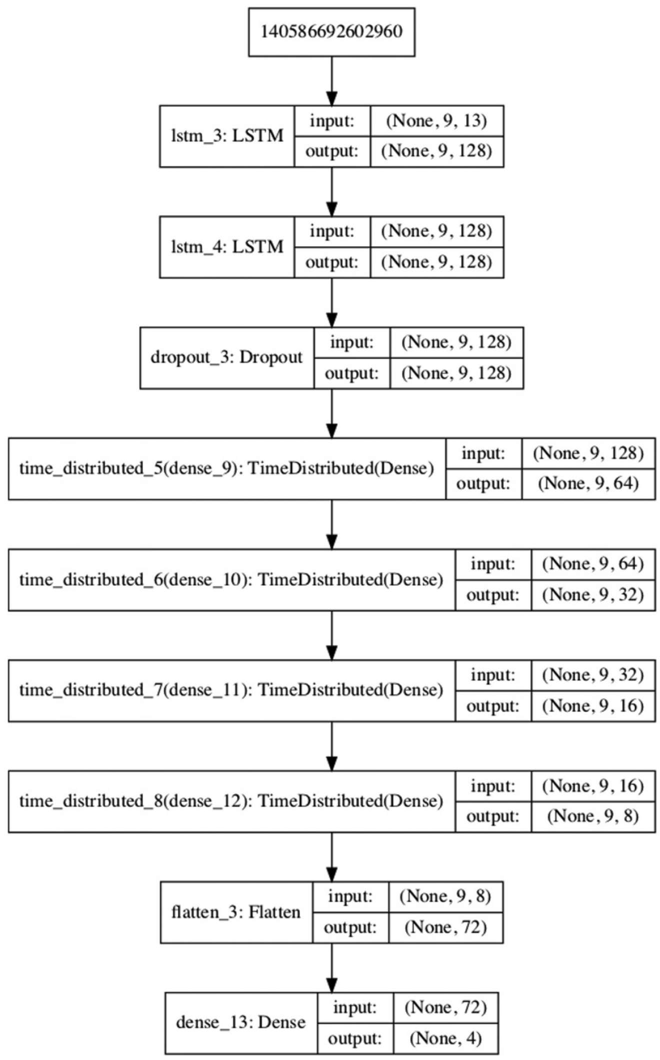

FIG. 4. The LSTM architecture.

extract features and also train the networks. Specifically, the models were prepared using a Python-based Keras library with TensorFlow graphics processing unit backend. *LibROSA* was used to read and resample the sound files and *Python_speech_features* were used for providing sound features such as the MFCC and filter banks.

IV. FINDINGS AND RESULTS

In ML, the model performance is evaluated by assessing how it performs on unseen data. This is called model validation. In this study, the models were validated using classification metrics and cross-validation. During evaluation, the trained model is evaluated with a separate portion of the same dataset (i.e., testing data). The performance of the model was evaluated using the accuracy, precision, recall, and $F1$-score [see Eqs. (1)–(4)]. The confusion matrix plays a major role in almost all of the performance scores. It provides the necessary parameters for assessing the various matrices. It is a breakdown of correct (true positives) and incorrect (errors) classifications for each class and allows the visualization of the model performance by tabulating the values of actual and predicted classes as columns and rows, respectively. Confusion matrices provide information for the calculation of *precision* [also known as the *positive predictive value* (PPV)]. This is the measure of positive instances that are actually positive. Whereas *rec*all or sensitivity or true positive rate (TPR) is a measure of the ability of a classification model to correctly select instances. *Accuracy* (Acc) is the measure of the correctly classified instance to the total instances, whereas the *F1-score* is the harmonic mean of the model's precision and recall. Equations (1)–(4) are used to calculate these values,

$$\mathrm{PPV} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})}, \tag{1}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})}, \tag{2}$$

$$\mathrm{Acc} = \frac{(\mathrm{TP}+\mathrm{TN})}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}}, \tag{3}$$

$$F1-\mathrm{score} = \frac{2(\mathrm{PPV}*\mathrm{TPR})}{(\mathrm{PPV}+\mathrm{TPR})}, \tag{4}$$

	Earthquake	Tornado	Tsunamis	Volcano
Earthquake	825	0	0	0
Tornado	0	554	0	0
Tsunamis	0	0	779	0
Volcano	2	2	0	426

FIG. 5. The CNN confusion matrix.

where TP is the true positive (i.e., the correct number of positives that are correctly identified as positives) and TN is the true negative (i.e., the correct number of negatives correctly identified as negatives). FN is the false negative (i.e., the correct number of positives that are incorrectly identified as negatives) and FP is the false positive (i.e., the correct number of negatives that are incorrectly identified as positive). In addition, the area under curve–precision-recall (AUC-PR) curve and area under curve–receiver operating characteristics (AUC-ROC) were used because they both provide relevant information for assessing the model performance, especially in cases of uneven class distribution.

Figure 5 shows the classification made using the CNN. The test dataset is made up of a total of 2588 sound recordings, out of which 2584 instances were correctly classified. Misclassifications were observed in the following instances: volcano as earthquake (two) and volcano as tornado (two). Earthquake, tornado, and tsunami were correctly predicted at all instances.

Similarly, Fig. 6 shows the confusion matrix for classification made using the LSTM model. Out of the 2588 sound recordings, 22 errors were recorded across the 4 classes. Earthquake was misclassified as tsunami in four instances, tornado was misclassified as earthquake, tsunami, and volcano at two, one, and two instances, respectively. Volcano was misclassified as tornado at one instance. Tsunami was misclassified as earthquake (four) and volcano (three), whereas volcano was misclassified as earthquake (ten) and tornado (one). None of the classes were correctly predicted at all instances with the LSTM.

	Earthquake	Tornado	Tsunamis	Volcano
Earthquake	780	0	4	22
Tornado	2	546	1	2
Tsunamis	4	0	751	3
Volcano	10	1	0	462

FIG. 6. The LSTM confusion catrix.

TABLE I. Performance metrics for the CNN and LSTM.

Models	CNN	LSTM
Accuracy	99.85%	98.11%
Precision	99.85%	98.11%
Recall	99.85%	98.11%
$F1$-score	99.85%	98.11%
AUC-ROC	0.999	0.999
AP	100%	100%

A. Performance metrics

Based on the confusion matrix, classification metrics, such as accuracy, precision, recall, and $F1$-score, were computed for both the CNN and LSTM. It was observed that for each classifier, the accuracy, precision, recall, and $F1$-score produced similar results (see Table I), indicating that both models are impressive. As shown in Table I, although the CNN outperformed the LSTM, both classifiers had good performance in accuracy, precision, recall, and $F1$-scores of 99.85% and 98.11%, respectively.

Because a time frame of a tenth of a second (0.1 s) was used for the classification, other time frames were used to train the model to test the real-time classification validity of the model. These time frames were 0.2 and 0.4 s. From the results (see Table II), it can be observed that the classification accuracies at time 0.2 s was lower than the accuracies at time 0.4 s and 0.6 s for both the CNN and LSTM. At time 0.1 s, however, the classification accuracies were higher for both the CNN and LSTM. Hence, it can be concluded that both models perform better at the shorter time frame of 0.1 s. The models with the 0.1 s time frame were then adopted for further analysis (i.e., AUC-ROC and AUC-PR).

B. AUC-ROC and AUC-PR

Due to the dominating effect of majority classes in imbalanced datasets, classification metrics, such as accuracy, precision, and recall, are not sufficient for evaluating the performance of a model (Luque *et al.*, 2018; Weng and Poon, 2008; Yang *et al.*, 2015). Consequently, the AUC-ROC and AUC-PR are visualization tools for comparing classification models and were used to evaluate the dominating effect of the imbalance dataset (Ling *et al.*, 2003; Weng and Poon, 2008; Yang *et al.*, 2015). The AUC-ROC in Figs. 7 and 8 are plots of the false positive rate (FPR) on the x axis and TPR on the y axis for theCNN and LSTM, respectively. It shows the variation between the number of correctly predicted (classified) positive instances and incorrectly predicted negative instances, i.e., how much the model can distinguish between classes. It can be observed

TABLE II. Accuracy scores for various time frames.

Models	0.1 s	0.2 s	0.4 s	0.6 s
CNN	99.85%	99.61%	99.69%	99.77%
LSTM	98.11%	96.59%	96.75%	97.68%

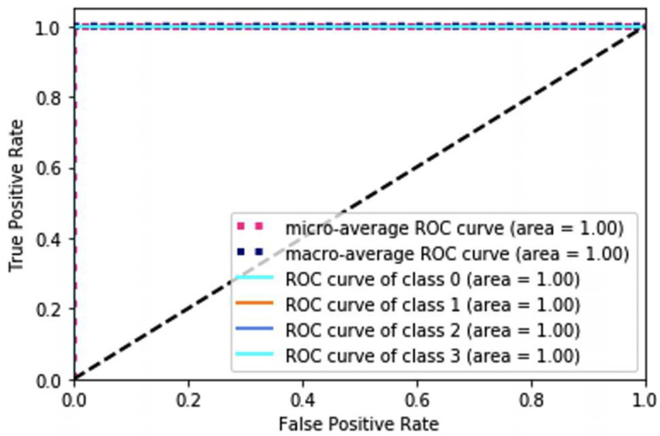

FIG. 7. (Color online) The AUC-ROC for the CNN model.

from the AUC-ROC plots in Figs. 7 and 8 that both models were not affected by class imbalance. This demonstrates that the classifiers have good measures for classifying disaster sounds: both models had an AUC-ROC score of 0.999.

The AUC-PR shows the relationship between a positive predicted value (precision) on the y axis and a TPR (recall) on the x axis. The high precision and recall scores of 99.85% and 98.11% for the CNN and LSTM, respectively (see Table I), indicate that the models are able to classify the majority of the acoustic signal correctly (high recall) with less inaccurate classification (high precision). Furthermore, the AUC-PR score, which is obtained by finding the area under the AUC-PR (see Figs. 9 and 10), was computed for the four classes in the dataset. As shown in Table I, an AUC-PR score of 1.00 (100%) was obtained for both models. The AUC-PR score indicates a good test as the models are able to discriminate between the various disaster sounds with 100% recall and 100% precision.

C. Cross-validation

A K-fold cross-validation, with $k = 10$ was adopted to further validate the performance of the models. $k = 10$ was used because it has the ability to produce a better performance by the model hyper-parameters (Chen *et al.*, 2017; Davis and Suresh, 2019; Han *et al.*, 2016; Pandeya and Lee, 2018; Thakur *et al.*, 2019). In the tenfold cross-validation, the entire training dataset is divided into ten subsets such that in each iteration, all of the subsets in the datasets are

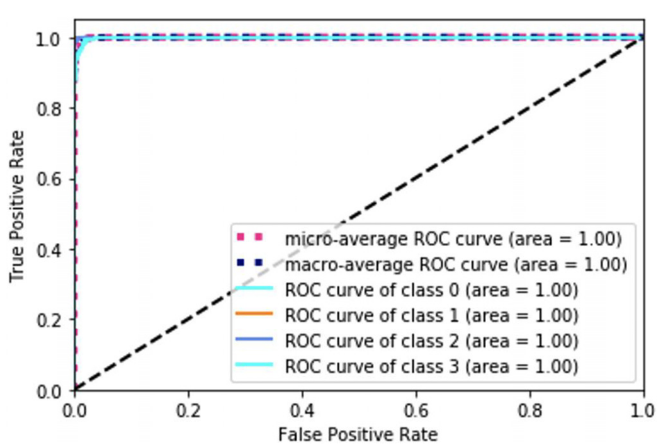

FIG. 8. (Color online) The AUC-ROC for the LSTM model.

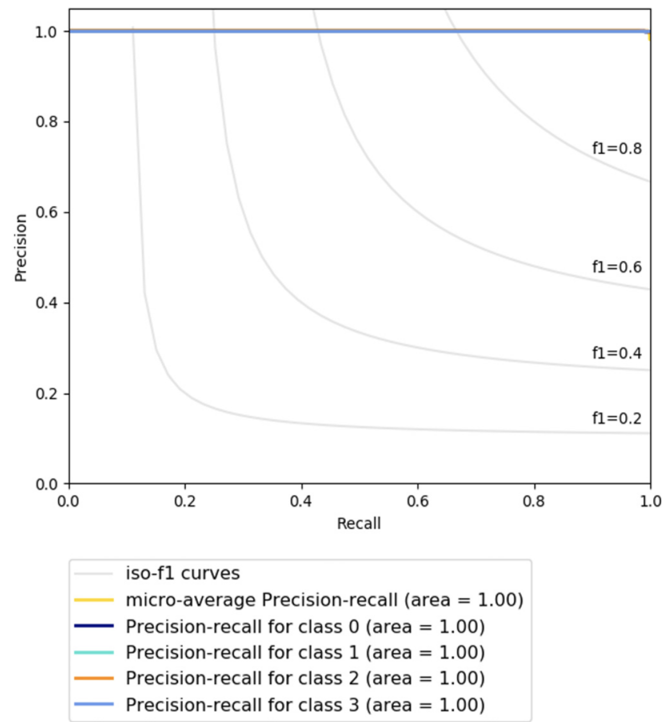

FIG. 9. (Color online) The AUC-PR for the CNN.

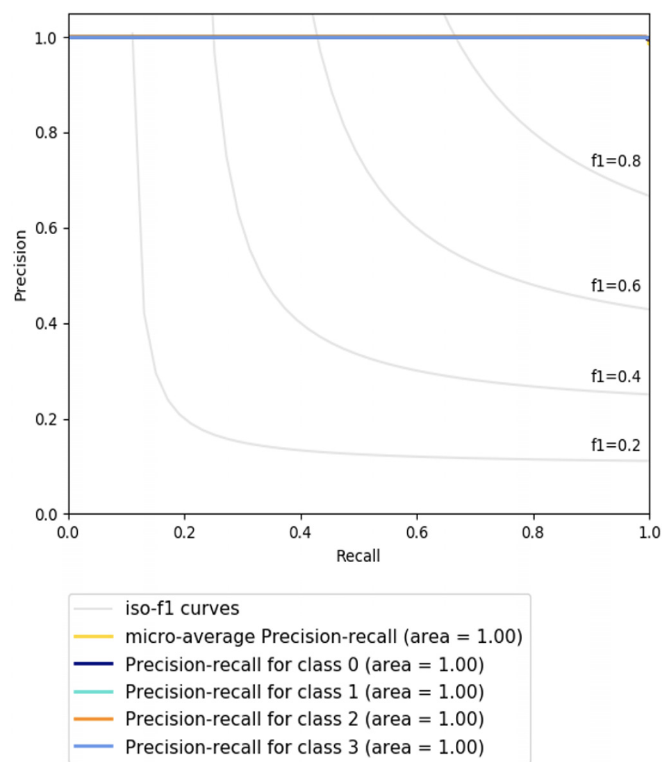

FIG. 10. (Color online) The AUC-PR for the LSTM.

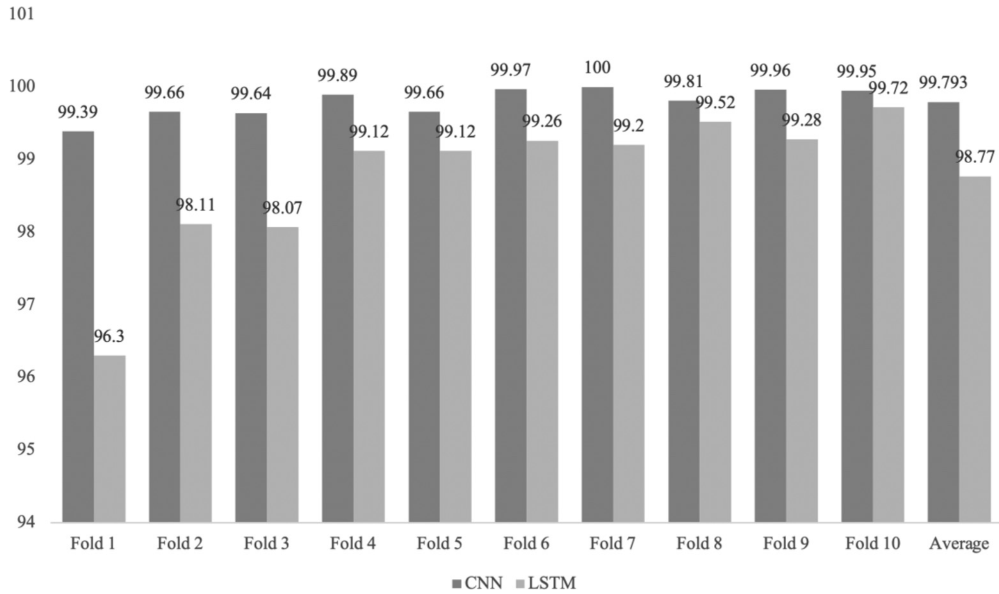

FIG. 11. The tenfold cross-validation for both models.

trained except one subset (hold out), which is reserved and used for testing. Each successive iteration outputs an accuracy, and the overall accuracy is calculated by taking an average of the accuracy results returned from each fold. The CNN and LSTM had average accuracies of 99.79% and 98.77%, respectively, at time 0.1 s. Figure 11 shows the accuracy for each fold as well as the average of the ten. CNN recorded an average accuracy of 99.79% and LSTM recorded an average accuracy of 98.77%.

V. DISCUSSION

In the model validation stage, it was observed that the CNN performed slightly better than the LSTM. CNNs are predominantly known to achieve high accuracies in image classification and recognition tasks. However, findings from this study indicate that CNNs can be successfully trained to classify natural sounds. It also shows that it can identify patterns across inputs in a time-frequency domain as well as differentiate acoustic signals. The LSTM, on the other hand, achieved a good performance in terms of accuracy, precision, and recall. Even though the results were lower than those of the CNN, it further affirms that RNNs perform well on sound and time series data.

Further, the findings of this study provoke the need for further studies on the use of acoustics signals or sound for early detection of natural disasters. As mentioned earlier, studies that sought to investigate the use of acoustic signals for prediction, detection, or classification of natural disasters using ML techniques have been less explored. However, using acoustic methods presents a number of advantages over the existing methods. Predominantly, most ML techniques use imaging and social network datasets. Yet, information or data from social networks cannot be validated and, thus, it cannot be considered as an authentic approach because it is prone to errors. Similarly, the image dataset is limited in terms of the coverage area and, thus, cannot cover a larger spectrum of disaster data relative to sound.

The impressive accuracy rate for both the CNN and LSTM as individual classifiers supports the argument that ML techniques are suitable for disaster classification. In particular, the ability to produce optimal results with sound truncated at 0.1 s with an accuracy of 99.96% (CNN) and 99.90% (LSTM) is promising. It suggests that the CNN and LSTM can be used to classify and, perhaps, detect natural disasters at less than a second. Further investigations need to be conducted to explore automatic systems that will harness the capabilities of these models for early detection. Although existing studies have used pre-disaster sound to classify natural disasters, such as earthquake and hurricane, for early detection, they are mostly expensive. They are also limited to their dependence on sensors and sonobouys, utility vehicles, aircrafts, or vessels to retrieve information carried by the sensors. For instance, a minimum of \$10 000 is required to deploy sonobouys from nonspecialized ships in the path of an upcoming hurricane (Wilson and Makris, 2006).

VI. CONCLUSION

Significant developments have been achieved with systems that enable physical environmental monitoring, remote sensing, extreme event detection, and natural disaster detection. With regard to natural disasters, several types of AI solutions (systems) have been proposed by researchers and meteorologists to predict, detect, and manage natural disasters. This study proposed the use of sound/acoustic signals to classify natural disasters. It developed and compared sound classification models using the CNN and LSTM. With varying time frames ranging between 0.1 and 0.4 s, higher classification results were obtained for both of the models at the least time of 0.1 s.

The results of this study show a promising future for disaster detection, environmental monitoring, remote sensing, and the security of the physical environment. Although it is not expected to replace existing methods, it provides an effective alternative for disaster classification. Also, although acoustic signals are not affected by light, they may be affected by other environmental sound and, thus, there is a need for further investigations into how to address this limitation. Further validation needs to be conducted to authenticate the models with different disaster sound datasets from different locations. Particularly, there is the need to investigate the dataset that contains other non-disaster sound because although the present study performed well, the dataset used for the investigation contained only disaster data. However, acoustic signals are usually made of different sounds from different sources and, therefore, will not only contain disaster sound. Additionly, there is the need to investigate how the models in this study can be deployed in systems to facilitate early disaster detections and warnings.

[1]See freesoundeffects.com for the dataset (Last viewed March 30, 2020).
[2]The selected sound files used in this study were downloaded from freesoundeffects.com/free-sounds/ambience-10005/. However, the specific selection used is saved at https://drive.google.com/drive/folders/1CPGJFfTv8jf7pEMFawkxIJz6qrRXoIRC (Last viewed March 5, 2020).

Allen, J. A., Murray, A., Noad, M. J., Dunlop, R. A., and Garland, E. C. (**2017**). "Using self-organizing maps to classify humpback whale song units and quantify their similarity," J. Acoust. Soc. Am. **142**(4), 1943–1952.

Amit, S. N. K. B., Shiraishi, S., Inoshita, T., and Aoki, Y. (**2016**). "Analysis of satellite images for disaster detection," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, Beijing, China, pp. 5189–5192.

Antoniou, V., and Potsiou, C. (**2020**). "A deep learning method to accelerate the disaster response process," Remote Sens. **12**(3), 544.

Aykanat, M., Kılıç, Ö., Kurt, B., and Saryal, S. (**2017**). "Classification of lung sounds using convolutional neural networks," Eurasip J. Image Video Process. **2017**(1), 65.

Aziz, S., Awais, M., Akram, T., Khan, U., Alhussein, M., and Aurangzeb, K. (**2019**). "Automatic scene recognition through acoustic classification for behavioral robotics," Electronics (Switzerland) **8**(5), 483.

Boustan, P. L., Kahn, M. E., Rhode, P. W., and Yanguas, M. L. (**2017**). "The effect of natural disasters on economic activity in us counties: A century of data," in National Bureau of Economic Research (NBER) Working Paper Series, available at http://www.nber.org/papers/w23410 (Last viewed March 30, 2020).

Chen, H., Yuan, X., Pei, Z., Li, M., and Li, J. (**2019**). "Triple-classification of respiratory sounds using optimized S-transform and deep residual networks," IEEE Access **7**(April), 32845–32852.

Chen, W., Shirzadi, A., Shahabi, H., Ahmad, B. B., Zhang, S., Hong, H., and Zhang, N. (**2017**). "A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naive Bayes tree classifiers for a landslide susceptibility assessment in Langao County," China. Geomatics, Nat. Hazards Risk **8**(2), 1955–1977.

Davis, N., and Suresh, K. (**2019**). "Environmental sound classification using deep convolutional neural networks and data augmentation," in *2018 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2018*, pp. 41–45.

Doxani, G., Siachalou, S., Mitraka, Z., and Patias, P. (**2019**). "Decision making on disaster management in agriculture with sentinel applications," Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci. **XLII**(3/W8), 121–126.

Furquim, G., Filho, G. P. R., Jalali, R., Pessin, G., Pazzi, R. W., and Ueyama, J. (**2018**). "How to improve fault tolerance in disaster predictions: A case study about flash floods using IoT, ML and real data," Sensors (Switzerland) **18**(3), 907–920.

Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A., and Chakraborty, B. (**2018**). "A review on application of data mining techniques to combat natural disasters," Ain Shams Eng. J. **9**(3), 365–378.

Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., and Zhu, X. (**2016**). "Semi-supervised active learning for sound classification in hybrid learning environments," PLoS One **11**(9), e0162075.

Ibrahim, A. K., Chérubin, L. M., Zhuang, H., Schärer Umpierre, M. T., Dalgleish, F., Erdol, N., Ouyang, B., and Dalgleish, A. (**2018**). "An approach for automatic classification of grouper vocalizations with passive acoustic monitoring," J. Acoust. Soc. Am. **143**(2), 666–676.

Joshi, N. (**2019**). "How AI can and will predict disasters," https://www.forbes.com/sites/cognitiveworld/2019/03/15/how-ai-can-and-will-predict-disasters/#4b76c3c55be2 (Last viewed April 21, 2021).

Karbasi, M., Ahadi, S. M., and Bahmanian, M. (**2011**). "Environmental sound classification using spectral dynamic features," in *ICICS 2011—8th International Conference on Information, Communications and Signal Processing*, August 2016, pp. 2–7.

Khalaf, M., Hussain, A. J., Al-Jumeily, D., Baker, T., Keight, R., Lisboa, P., Fergus, P., and Al Kafri, A. S. (**2018**). "A data science methodology based on machine learning algorithms for flood severity prediction," in *2018 IEEE Congress on Evolutionary Computation*, pp. 1–8.

Khalaf, M., Hussain, A. J., Al-Jumeily, D., Fergus, P., and Idowu, I. O. (**2015**). "Advance flood detection and notification system based on sensor technology and machine learning algorithm," in *2015 22nd International Conference on Systems, Signals and Image Processing*, pp. 105–108.

Khamparia, A., Gupta, D., Nguyen, N. G., Khanna, A., Pandey, B., and Tiwari, P. (**2019**). "Sound classification using convolutional neural network and tensor deep stacking network," IEEE Access **7**(1), 7717–7727.

LeBien, J. G., and Ioup, J. W. (**2018**). "Species-level classification of beaked whale echolocation signals detected in the northern Gulf of Mexico," J. Acoust. Soc. Am. **144**(1), 387–396.

Ling, C. X., Huang, J., and Zhang, H. (**2003**). "AUC: A statistically consistent and more discriminating measure than accuracy," in *IJCAI International Joint Conference on Artificial Intelligence*, pp. 519–524.

Luque, A., Romero-Lemos, J., Carrasco, A., and Gonzalez-Abril, L. (**2018**). "Temporally-aware algorithms for the classification of anuran sounds," PeerJ **6**(5), e4732.

Maccagno, A., Mastropietro, A., Mazziotta, U., Scarpiniti, M., Lee, Y.-C., and Uncini, A. (**2021**). "A CNN approach for audio classification in construction sites," in *Progresses in Artificial Intelligence and Neural Systems*, edited by A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero (Springer, Singapore), pp. 371–381.

Malfante, M., Mars, J. I., Dalla Mura, M., and Gervaise, C. (**2018**). "Automatic fish sounds classification," J. Acoust. Soc. Am. **143**(5), 2834–2846.

Mitilineos, S. A., Potirakis, S. M., Tatlas, N. A., and Rangoussi, M. (**2018**). "A two-level sound classification platform for environmental monitoring," J. Sens. **2018**, 5828074.

Mousa, M., Zhang, X., and Claudel, C. (**2016**). "Flash flood detection in urban cities using ultrasonic and infrared sensors," IEEE Sens. J. **16**(19), 7204–7216.

Nanni, L., Rigo, A., Lumini, A., and Brahnam, S. (**2020**). "Spectrogram classification using dissimilarity space," Appl. Sci. (Switzerland) **10**(12), 4176.

Oikarinen, T., Srinivasan, K., Meisner, O., Hyman, J. B., Parmar, S., Fanucci-Kiss, A., Desimone, R., Landman, R., and Feng, G. (**2019**). "Deep convolutional network for animal sound classification and source attribution using dual audio recordings," J. Acoust. Soc. Am. **145**(2), 654–662.

Okamoto, K., Mochida, T., Nozaki, D., Wen, Z., Qi, X., and Sato, T. (**2018**). "Content-oriented surveillance system based on ICN in disaster scenarios," in *21st International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Chiang Rai, Thailand, pp. 484–489.

Pandeya, Y. R., and Lee, J. (**2018**). "Domestic cat sound classification using transfer learning," Int. J. Fuzzy Logic Intell. Syst. **18**(2), 154–160.

Prithvi, P., and Kumar, T. K. (**2016**). "Comparative analysis of MFCC, LFCC, RASTA-PLP," Int. J. Sci. Eng. Res. **4**(5), 1–4.

Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S., and On, B. W. (**2019**). "Heartbeat sound signal classification using deep learning," Sensors (Switzerland) **19**(21), 4819.

Resch, B., Usländer, F., and Havas, C. (**2018**). "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment," Cartography Geogr. Inf. Sci. **45**(4), 362–376.

Robakis, E., Watsa, M., and Erkenswick, G. (**2018**). "Classification of producer characteristics in primate long calls using neural networks," J. Acoust. Soc. Am. **144**(1), 344–353.

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., and Johnson, P. A. (**2017**). "Machine learning predicts laboratory earthquakes," Geophys. Res. Lett. **44**(18), 9276–9282, http://dx.doi.org/10.1002/2017GL074677.

Salamon, J., and Bello, J. P. (**2017**). "Deep convolutional neural networks and data augmentation for environmental sound classification," IEEE Signal Process. Lett. **24**(3), 279–283.

Sayad, Y. O., Mousannif, H., and Al Moatassime, H. (**2019**). "Predictive modeling of wildfires: A new dataset and machine learning approach," Fire Safety J. **104**(3), 130–146.

Soulé, B. (**2014**). "Post-crisis analysis of an ineffective tsunami alert: The 2010 earthquake in Maule, Chile," Disasters **38**(2), 375–397.

Supriyadi, B., Windarto, A. P., Soemartono, T., and Mungad, (**2018**). "Classification of natural disaster prone areas in Indonesia using *K*-means," Int. J. Grid Distributed Comput. **11**(8), 87–98.

Thakur, A., Thapar, D., Rajan, P., and Nigam, A. (**2019**). "Deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss," J. Acoust. Soc. Am. **146**(1), 534–547.

Verma, D., Jana, A., and Ramamritham, K. (**2019**). "Classification and mapping of sound sources in local urban streets through AudioSet data and Bayesian optimized neural networks," Noise Mapp. **6**(1), 52–71.

Vrbancic, G., and Podgorelec, V. (**2018**). "Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks," Elektron. Elektrotech. **24**(4), 1–7.

Wallemacq, P., and Guha-Sapir, D. (**2015**). "Global natural disasters: Patterns and trends," in *The Human Cost of Natural Disasters: A Global Perspective*, 1st ed., edited by R. House (CRED, Brussels).

Weng, C. G., and Poon, J. (**2008**). "A new evaluation measure for imbalanced datasets," in *Conferences in Research and Practice in Information Technology Series*, Vol. 87, pp. 27–32.

Wilson, J. D., and Makris, N. C. (**2006**). "Ocean acoustic hurricane classification," J. Acoust. Soc. Am. **119**(1), 168–181.

Wisner, B., and Adams, J. (**2002**). *Environmental Health in Emergencies and Disasters: A Practical Guide* (World Health Organization, Geneva, Switzerland), Vol. 62, issue 5.

Yang, J., Wang, Y. X., Qiao, Y. Y., Zhao, X. X., Liu, F., and Cheng, G. (**2015**). "On evaluating multi-class network traffic classifiers based on AUC," Wireless Pers. Commun. **83**(3), 1731–1750.

Yaseen, Son, G. Y., and Kwon, S. (**2018**). "Classification of heart sound signal using multiple features," Appl. Sci. (Switzerland) **8**(12), 2344.

Zhang, Y.-J., Huang, J.-F., Gong, N., Ling, Z.-H., and Hu, Y. (**2018**). "Automatic detection and classification of marmoset vocalizations using deep and recurrent neural networks," J. Acoust. Soc. Am. **144**(1), 478–487.