

The Use of Machine Learning Algorithms in the Classification of Sound: A Systematic Review

Akon O. Ekpezu, University of Ghana, Ghana

 <https://orcid.org/0000-0002-9502-1052>

Ferdinand Katsriku, University of Ghana, Ghana

Winfred Yaokumah, University of Ghana, Ghana*

 <https://orcid.org/0000-0001-7756-1832>

Isaac Wiafe, University of Ghana, Ghana

 <https://orcid.org/0000-0003-1149-3309>

ABSTRACT

This study is a systematic review of literature on the classification of sounds in three domains: bioacoustics, biomedical acoustics, and ecoacoustics. Specifically, 68 conferences and journal articles published between 2010 and 2019 were reviewed. The findings indicated that support vector machines, convolutional neural networks, artificial neural networks, and statistical models were predominantly used in sound classification across the three domains. Also, the majority of studies that investigated medical acoustics focused on respiratory sounds analysis. Thus, it is suggested that studies in biomedical acoustics should pay attention to the classification of other internal body organs to enhance diagnosis of a variety of medical conditions. With regard to ecoacoustics, studies on extreme events such as tornadoes and earthquakes for early detection and warning systems were lacking. The review also revealed that marine and animal sound classification was dominant in bioacoustics studies.

KEYWORDS

Acoustic Signals, Artificial Intelligence, Classification, Deep Learning, Environmental Monitoring, Machine Learning, Medical Diagnosis, Security Surveillance, Sound

INTRODUCTION

Sound or acoustic signals are gradually gaining research popularity as a tool for environmental monitoring, security surveillance, diagnoses of diseases, critical information infrastructure protection, and data transmission (Bourouhou et al., 2019; Ibrahim et al., 2018; Loey et al., 2020; Luque et al., 2018). Sound is considered as the second most important sense after sight that is capable of carrying information about the environment (Perr, 2005). Although sound varies depending on seasons, time, geographic location as well as propagation medium, it is considered as one of the most significant signals used to monitor and detect changes in the environment. Accordingly, the ability to differentiate (classify) one sound or acoustic signal type from another is pertinent that, if accomplished, would

DOI: 10.4018/IJSSMET.298667

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

result in significant progress in application areas such as early warning disaster management, medical diagnosis (Loey, Naman, & Zayed, 2020), and action or event detection. Recent studies have shown that machine learning (ML) algorithms are efficient in the domains of image and speech recognition, natural language processing, medical imaging, data extraction (Dwivedi et al., 2019; Malfante et al., 2018; Tatoian & Hamel, 2018) and text classification (Elfergany & Adl, 2020; Sangwan & Bhatnagar, 2020).

Classification aims at predicting accurately the target object and differentiating one object class from the other given a set of data. It is predominantly performed using selected features that feed classifier tools such as machine learning and neural networks (Mitilineos et al., 2018). In particular, sound classification is aimed at classifying audio segments into specific classes which requires the understanding of the fundamental structure of frequencies in acoustic signals (Dwivedi et al., 2019). This is commonly addressed with features used in speech and music processing such as MFCC (Mel frequency cepstral coefficient), linear prediction coefficients (LPC), linear prediction cepstral coefficients (LPCC) and fast Fourier transforms (Briggs et al., 2012; Chu et al., 2009; Davis & Suresh, 2019; Karbasi et al., 2011; Mitilineos et al., 2018; Oletic et al., 2012; Pramono et al., 2017; Sengupta et al., 2016). A variety of machine learning techniques have also been adopted to obtain robust sound classification models.

Considering the plethora of acoustic features and machine learning (ML) algorithms coupled with the nature of sound, it is imperative to offer researchers an indication of the major research trends and methodologies that can assist in designing and developing automatic sound classification systems. Accordingly, this study provides summaries of the existing literature on algorithms for the classification of sound and analyzes the use of ML in the various sound classification tasks. The specific objective of the review is to identify: (a) publication patterns in acoustic signal classification, (b) trends in the use of ML in acoustic signal/sound classification, (c) open questions and challenges in the use of ML algorithms in acoustic signal classification, and (d) research gaps in the subject area.

KNOWLEDGE GAP AND REVIEW QUESTIONS

Previous studies that conducted reviews on sound or acoustic signal classifications employed artificial intelligence (AI) techniques. These studies were evaluated using Greenhalgh's (1997) evaluation criteria. The criterion evaluated systematic reviews by accessing the relevance of the review question, the search strategy, the methodological quality, and the sensitivity and presentation of results and findings. Although findings from the evaluation of existing reviews in the subject area indicated that studies available provided summaries and reproducible review methodologies, they focused predominantly on the classification of biomedical acoustic signals, particularly, on heart sounds (Dwivedi et al., 2019), lung sound (Palaniappan et al., 2013), respiratory sound (Pramono et al., 2017), and speech sound disorder in children (Wren et al., 2018). Thus, systematic review on the classification of sounds is lacking. Considering the various applications of sound in various activities, the lack of sufficient summaries justifies the need for a systematic review of sound classification.

Recently, ML algorithms have been used for various classification tasks (Hao, Weiss, & Brown, 2018). However, due to the plethora of ML algorithms (Hlioui, Aloui, & Gargouri, 2020; Salama & Hassanien, 2014), choosing a suitable algorithm for a specific classification task becomes difficult. Hence, there is the need to identify open questions, publication trends, and current approaches in algorithm usage that will assist researchers to position appropriately new research activities in sound classification and detection. To address this, this review examines two broad issues. The research questions stated in Table 1 are divided into two categories. The Category A consists of questions that seek to provide an overview of publication trends whereas the Category B seeks to provide a good methodological background for a broader work by identifying research gaps and current methodologies in the domain. Specifically, it is expected that the study will provide pertinent information regarding patterns in publications since 2010, academic outlets that are dominant and attracting more studies, and countries that have focused on acoustic signal classification most within the specified period.

As mentioned earlier, questions in Category B will provide information on techniques and domains of application. Accordingly, the review questions will seek to summarize information on application domains that have dominated artificial intelligence (AI) for acoustics signal studies, the most used datasets, the machine learning techniques adopted, and the various evaluation methods that are mostly adopted. Table 1 provides a summary of the proposed review questions and the corresponding rationale for posing them.

Table 1. Research Questions and Objectives

Research Questions	Objectives
Category A	
• What are the yearly publication trends? • What journal has the highest number of publications? • What is the frequency of authors? • What is the country's origin of authors' affiliated institutions?	• To identify the frequency of primary studies per year. • To identify the frequency of publications per journal. • To identify authors who are consistent in writing on the subject area. • To identify countries with the highest number of publications.
Category B	
• What kind of sound is classified? • What is the format of the sound? • What are the sample rates of the audio recordings? • What datasets were used for the classification and (or) evaluation?	• To identify the dominant and less dominant types of classified sounds. • To identify predominantly used audio formats for classification. • To determine the maximum audio frequency that can be reproduced. • To identify datasets that are available for public use.
• What are the various application domains?	• To identify domains in which sound classification is predominantly performed.
• What ML techniques have been used for sound classification? • What measures are used to evaluate model performance?	• To identify predominantly used ML techniques and performance metrics in sound classification.

REVIEW APPROACH

A systematic search of the literature was carried out in two databases: Scopus and Acoustical Society of America (ASA). Scopus was selected because it is arguably the most extensive abstract and citation database for academic publications, whereas the ASA publications database was selected purposefully since it is the leading source of theoretical and experimental studies in acoustics-related studies. Publications were extracted from the selected databases using key search terms and their possible combination using the logical 'and' operator. The key search terms included *classification*, *sound*, *acoustic signals*, *machine learning*, *deep learning*, and *artificial intelligence*. The combination of the search terms produced the following search phrases (SP):

- SP1 Classification of sound and machine learning
- SP2 Classification of sound and deep learning
- SP3 Classification of sound and artificial intelligence
- SP4 Classification of acoustic signals and machine learning
- SP5 Classification of acoustic signals and deep learning
- SP6 Classification of acoustic signals and artificial intelligence

Inclusion and Exclusion Criteria

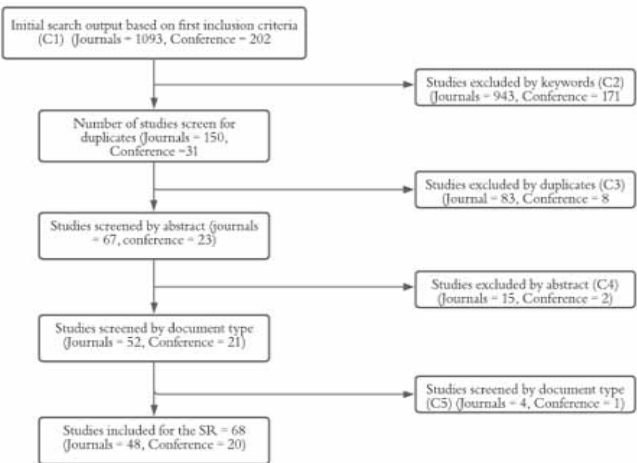
A set of specific eligibility criteria were defined and followed to limit the collection of articles to only those that fit with the research objectives. A suitability check of returned articles was performed after examining the title and removing duplicate papers. Only articles in which the aim, classification techniques, and/or results were explicitly stated in the abstract were considered. The inclusion and exclusion criteria are as follows:

- C1 Include only open-access journal articles and peer-reviewed conference papers written in English and published between the years 2010 and 2019.
- C2 Include articles whose titles contain keywords like classification and acoustic signals or sound and machine learning or deep learning or whose title suggests sound classification using artificial intelligence.
- C3 Exclude duplicate papers from the search results.
- C4 Exclude papers whose abstracts do not explicitly state the classification techniques and/or results of the evaluation metrics used.
- C5 Exclude by document type i.e., exclude secondary studies, books, thesis, reports, and letters.

Study Selection and Data Extraction

The six search phrases earlier mentioned were used to search the Scopus and ASA databases. The protocol for this systematic review has three main steps. In the first step, the retrieved articles were analyzed with an initial exclusion criterion (C1 to C2). In the second step, eligible articles were then exported to a spreadsheet (MS Excel) for further exclusion by duplicate, abstract, and type of study (C3, C4, and C5). ASA database does not have the export feature, hence this phase of exclusion was done directly from the browser and manually documented. The third step entailed downloading and reading eligible articles to extract relevant data concerning the review questions. The extracted data was collated in a spreadsheet for ease of use and analysis. Figure 1 is a flow diagram showing the results of the screening after each stage of exclusion or inclusion.

Figure 1. Flow diagram of study screening/selection



As shown in Figure 1, the initial search output contained 1,295 journal and conference articles published from 2010 to 2019. Out of these, 181 studies were included after an initial screening by title and keywords and a total of 90 articles were obtained after the removal of duplicates. Furthermore, 22 studies were excluded based on abstract and document type. Finally, 48 journal papers and 20 conference articles were selected and used in the study.

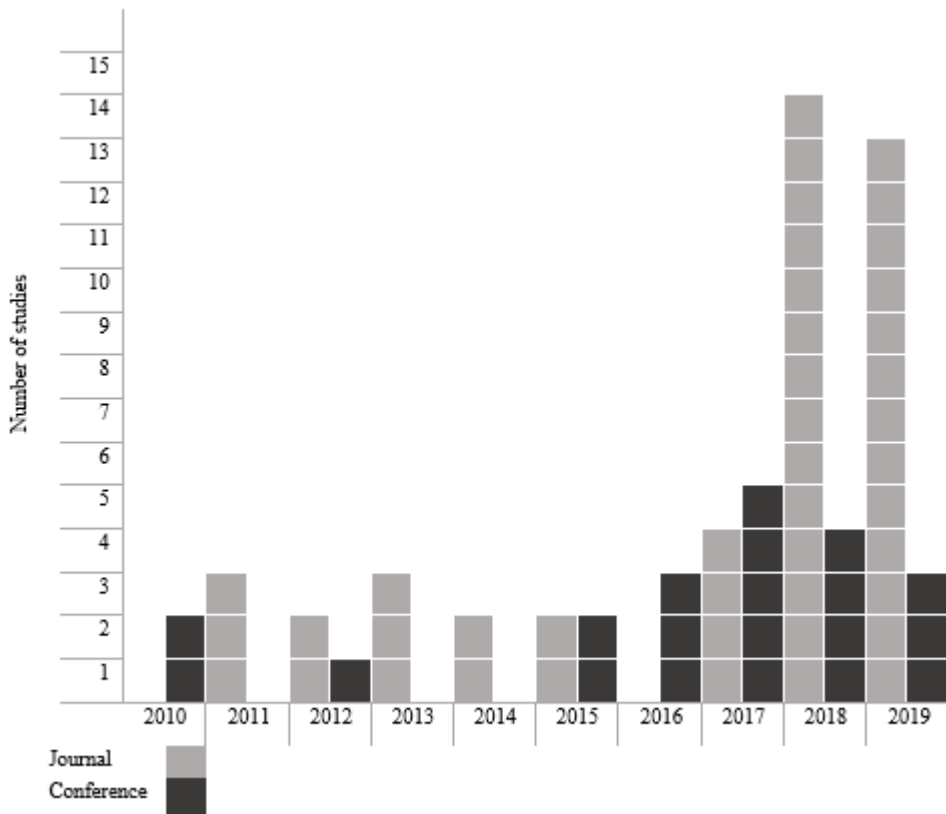
REVIEW RESULTS AND FINDINGS

Publication Trends (Category A)

This section presents the findings on the publication frequency, distribution of journals, authors, and their country of origin. As mentioned earlier, a total of sixty-eight (68) conference proceedings and journal articles were identified at the end of the selection process. This comprised 20 (29%) conference publications and 48 (71%) journal articles.

An analysis of the publication frequency (Figure 2) from 2011 to 2016 recorded between 2 and 4 publications per year. However, from 2017 there was a change in trend such that both conference and journal articles were recorded each year. Also, there was an upsurge in the publications from 2016 with a double leap from 9 publications in 2017 to 18 publications in 2018. Considering the upsurge in publication, the popularity of artificial intelligence as well as the emergence of sound as an alternative means of environmental monitoring, it is envisaged that the area of sound classification will draw more research attention.

Figure 2. Publications by year



Further, an examination of the sources of the included studies showed that the 68 studies were distributed among 35 Scopus indexed conference proceedings and journals, and 69% of the studies were published in the Journal of America Society of Acoustics (JASA). The journals and their corresponding number of included studies are JASA (24), Applied Sciences (3), Sensors (3), IEEE Access (2), APSIPA Transaction on Signal and Information Processing (1), Biomedical Journal, ELSEVIER - Computers & Electronics in Agriculture, Electronics (1), Elektronika ir Elektrotechnika (1), Eurasip Journal on Image & Video processing (1), Expert Systems with Applications (1), Frontiers in Neuroscience (1), IEEE Signal Processing Letters (1), IEICE Transactions on Information and Systems (1), International Journal of Fuzzy logic & Intelligent Systems (1), International Journal of online & biomedical engineering (1), International Journal of online engineering (1), Noise mapping (1), PeerJ (1), PLoS ONE (1). While the conference proceedings include ACM International Conference Proceeding Series (2), ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing (2), Computing in Cardiology (2), Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Bioinformatics) (2), Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH (2), 8th International Conference on Health Informatics (1), International Conference on Machine Learning and Applications, ICMLA 2012 (1), International Conference on Pattern Recognition (1), Proceedings of the International Conference on Neural Networks (1), MATEC Web of Conferences (1), IEEE International Workshop on ML for signal processing, MLSP (1), Procedia Computer Science (1), 2019 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT (1), Journal of Physics: Conference Series (1), 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC (1).

Authors and country origin. An analysis of the authors and their country origin (the country in which their affiliated institution is located) was performed to identify the author or group of authors who are consistent in writing on the subject area, as well as countries with the leading number of publications. With the number of authors per article ranging from 2 to 9, a headcount of the various authors showed that 229 authors wrote the 68 selected papers. Furthermore, 6 groups of leading authors in the subject area (i.e., authors with more than one publication) were identified (see Table 2). And it was observed that out of the 6 groups, 4 groups of authors were all interested in classifying sounds from animals (bioacoustics) and their publications were all journal articles. The 5th group was interested in classifying environmental sound and they had both conference and journal articles, while the 6th group focused on heart sounds with conference articles only.

Furthermore, the authors' country of origin (i.e., address of the authors) and the frequency of publications per year were identified. According to Figure 3, it is observed that the authors were from 31 different countries with the UK and USA leading the trend by 16% and 12% respectively. China, France, India, and Korea made up 8%, 7%, and 6% respectively. Portugal and Spain 5%, Germany 4%, while the other 22 countries make up 37% of the publication trend. Further, the highest number of publications in 2019 was from India (5), the highest in 2018 was from China (4) and Korea (4), and the highest in 2017 was from UK (5). Again, China (2) and the USA (3) had the highest publications in 2016 and 2012 respectively. Other years had one study per country. The countries with only one study include Ireland, Pakistan, Saudi Arabia, Morocco, Italy, Hong Kong, Jordan, Taiwan, Brazil, Estonia, Switzerland, Singapore, Sweden, and Austria. It is worth noting that publications from the UK are most consistent since 2012, each year there is at least one publication coming from the country.

Table 2. Leading authors

Groups	Paper Title	Year	Journal	Study
1	An approach for automatic classification of grouper vocalizations with passive acoustic monitoring.	2018	The Journal of the Acoustical Society of America	A9
	Classification of red hind grouper call types using a random ensemble of stacked autoencoders.	2019	The Journal of the Acoustical Society of America	A12
2	Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training data.	2015	The Journal of the Acoustical Society of America	A18
	A robust automatic birdsong phrase classification: A template-based approach.	2016	The Journal of the Acoustical Society of America	A14
3	Domestic cat sound classification using learned features from deep neural nets.	2018	Applied Sciences	A30
	Domestic cat sound classification using deep learning.	2018	International Journal of Fuzzy Logic and Intelligent Systems	A44
4	Non-sequential automatic classification of anuran sounds for the estimation of climate change indicators.	2018	Expert systems with Applications	A31
	Temporally aware algorithms for the classification of anuran sounds	2018	PeerJ	A33
5	Unsupervised Feature Learning for Urban Sound Classification	2015	ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing	A55
	Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound	2017	IEEE Signal Processing Letter	A46
6	Heart murmur classification with feature selection	2010	2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society	A67
	Heart murmur classification using complexity signatures	2010	Proceedings - International Conference on Pattern Recognition	A58

Sound Categories, Datasets, Classification Techniques, and Performance Metrics (Category B)

The discussions in this section are results obtained in line with Category B of the research questions. For ease of reference, the selected articles have been numbered in the order in which they were selected - A1 to A68 and will be used accordingly in further analysis (see Appendix for a list of included studies).

Figure 3. Distribution of publications by authors country of origin

2019	2	2	1	1	5	1				1	1	2		
2018	2	1	4	2		4		2					1	1
2017	2	5		1			1		3		1			
2016	1	1	2				1	1						
2015	1												1	
2014	1							1						
2013	1	1		1						1				
2012	3						1							
2011													1	
2010	1			1			1							
	UK	USA	China	France	India	Korea	Portugal	Spain	Germany	Netherlands	Australia	India	Slovenia	Japan

Classification of sound and data sources. Sound produced by plants, animals, and humans are numerous and it varies on land, air, and water depending on the medium of propagation, seasons, activities, and geographic location. There are three main sources of sound - Anthrophony (sounds made or caused by humans) such as shipping and drilling noise; Geophony (sound from the environment) such as sea surface noise like the breaking of waves, icebreaking, raindrops; and Biophony (sounds from animals) such as vocalizations of mammals, anurans, groupers. This section highlights the different kinds of sounds that were classified, data sources, sample rates, and availability of datasets as found in the selected articles. As shown in Table 3, 31 studies focused on classifying sounds caused by animals (Biophony), 19 classified sounds caused/made by human beings (Anthrophony), and the other 18 classified sounds from a combination of the three sounds categories (anthrophony, geophony, and Biophony).

In the biophony category, researchers were predominantly interested in classifying sounds from different species of Odontocetes and Mysticetes (marine mammals). While some of the researchers were interested in automatically detecting, classifying, and localizing call types from different species (Guilment et al., 2018; Halkias et al., 2013; Roch et al., 2011; Shamir et al., 2014), others were only interested in classifying vocalizations of humpback whales, whistles & pulse of dolphins, song cycles of whales and echolocation clicks of beaked whales (Allen et al., 2017; LeBien & Ioup, 2018; Ou et al., 2013; Parada & Cardenal-Lopez, 2014).

Classified sounds caused by humans (anthrophony) included respiratory sounds, human voice disorder, blast sound, snore sound, and baby cry. The baby cry was classified to identify the health state of a baby (i.e., need, pain, discomfort, or a medical condition) (Aucouturier et al., 2011) while snoring as a medical condition was classified as a means to automatically differentiate types of snore sounds (Amiriparian et al., 2017). Similarly, to automatically detect medical conditions such as cardiovascular diseases and respiratory tract diseases, EEG signals and heart sounds were classified to identify wheezes, crackles, murmur, extra-systole, normal and abnormal heartbeats. It would appear that apart from organs involved in respiratory and cardiovascular activity no other sound from internal organs of the human body was of interest. It might be instructive therefore to consider extending work to cover sounds from other internal human organs such as for example the intestines. Such work might be useful in understanding ailments that affect the digestive system. Also, Blast

sound was classified to differentiate between blast noise and non-blast noise (Cvengros et al., 2017). Sounds from the environment were predominantly classified to differentiate indoor, outdoor, natural, vocal, and non-vocal human sounds.

Table 3. Summary of classified sounds and datasets

	Sound source/ type	Link to dataset/name of the dataset	Article code
Biophony (sounds from animals)			
1	Marine mammals – Whales and Dolphins	DEFLOHYDRO, OHAS-ISBIO, DCLDE 2015, Auau Channel 2002, French Frigate Shoals (FFS), CEMMA datasets (http://www.cemma.org), https://data.gulfresearchinitiative.org	A1, A4, A13, A15, A16, A17, A19, A20, A22
2	Birds	http://www.animalsoundarchive.org/Refsys/Statistics.Php , Birdcalls71, Flight calls, Anuran, CAVI, and CUB-20002011 standard dataset	A2, A5, A14, A18, A21, A41, A7
3	Fish and Groupers	http://www.fishbase.org and http://www.dosits.org/ , SEACOUSTIC2014	A3, A9, A12, A37
4	Primates- Marmosets and Monkeys	http://home.ustc.edu.cn/~zyj008/background_noise.wav , http://marmosetbehavior.mit.edu	A8, A11, A35
5	Amphibians – Frogs and Anuran	Recordings from commercial compact discs (CD), recordings from natural habitat, http://www.fonozoo.com/	A24, A31, A33, A52
6	Domestic/farm animals – dog, cat, sheep, cattle, Maremma sheepdogs	Online video sources including YouTube, Kaggle challenge database and Flickr, and https://github.com/kyb2629/pdse .	A25, A32, A30, A44
Anthrophony (sounds made/caused by humans)			
7	Military blast sound	LRPE, East South Central, APG, SERDP-PITT, MCBC-PITT, New York (Fort Drum)	A6
8	Baby cry, human voice disorders	N/M	A24, A48

Table 3 continued on next page

Table 3 continued

	Sound source/ type	Link to dataset/name of the dataset	Article code
9	Respiratory/heart/ lung sound, EEG (electroencephalogram) signals	https://github.com/yaseen21khan/Classification-of-heart-sound-signal-using-multiple-features-/blob/master/README.md , https://physionet.org/challenge/2016/ , https://www.cs.colostate.edu/eeg_(the_Physionet_database).htm , Int. Conf. on Biomedical Health Informatics (ICBHI) scientific challenge database, Dataset B- PASCAL classifying heart sounds challenge, live recordings from patients using Bluetooth stethoscope	A27, A28, A29, A34, A38, A45, A47, A51, A53, A54, A56, A58, A61, A66,
10	Snore sound	Munich-Passau snore sound corpus	A63
Geophony (sound from the environment) and combination of various sound sources			
11	Cinematic sound	44-film dataset	A57
12	Oil, water, and gas	Life recordings	A68
13	Environmental sound	Real-world computing partnership (RWCP) sound scene dataset, DCASE challenge dataset, FindSounds database, Urban-sound 8k dataset, TIDIGITS dataset, ESC-10, ESC-50 dataset, freeseound.org, TUT database for acoustic scene classification & sound event detection, YouTube videos	A26, A35, A36, A39, A40, A42, A43, A46, A49, A50, A55, A59, A60, A62, A64, A65,

Environmental sounds classified are both indoor and outdoor sounds including air-conditioner, car horns, children playing, dog bark, drilling, engine idling, gunshot, jackhammers, siren, street music, running water, applause, footsteps, crowd, musical instruments, thunder, sea waves, etc.

Sample rate, audio format, and signal representation. The sample rate which is the number of samples of audio carried per second ranged from 0.1kHz to 192kHz. The dominantly used sample rates lied between 22 and 44.1kHz. Out of the 13 classified sounds sound categories identified from the included primary studies, the dominant audio format used was the .wav format. Others included mp3 (Parada & Cardenal-Lopez, 2014; Shamir et al., 2014), ARFF (Zhang et al., 2016), and HDF5 format (Bold et al., 2019). Furthermore, signals and audio files were predominantly visually represented as spectrograms. Spectrograms are graphical or visual representations of sound with frequency on the vertical axis, time on the horizontal axis, and a dimension of color that represents the intensity of the sound at each time-frequency location. According to (Amiriparian et al., 2017; Halkias et al., 2013; Malfante et al., 2018; Oikarinen et al., 2019; Ou et al., 2013), the classification of spectrograms as natural images allows it to be processed with available image processing tools. Additionally, it helps in removing the effect of background disturbances on the classification process (Thakur et al., 2019). Features extracted from spectrograms usually outperform hand-crafted features since spectrograms do not discriminate phrase classes with similar dominant frequency trajectories (Tan et al., 2015). However, disparate images in which the axes carry the same meaning irrespective of their location (i.e., the axes are shared weights across the vertical and horizontal dimensions), the axes of a spectrogram do not carry the same meaning (it has time and frequency as the vertical and horizontal dimensions).

Sources of data. To identify publicly available datasets, the datasets used in the reviewed articles were divided into two categories: pre-existing sound datasets and live recordings.

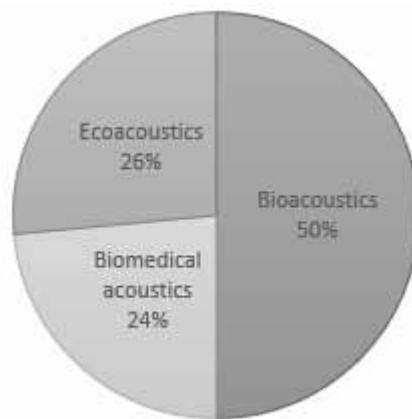
- i. Pre-existing sound datasets: This was made up of sound collected from past experiments, past projects, or existing sound databases, 28 datasets were identified from this category. Out of the 28, only 18 were stated to be publicly available, while the availability of others was either not mentioned or stated as not available due to licensing or privacy issues.
- ii. Life recordings: This category of datasets was generated by the researchers specifically for their research. It is made up of recordings of the subject of interest either in their natural habitat (Allen et al., 2017; Briggs et al., 2012; Ibrahim et al., 2019; LeBien & Ioup, 2018; Roch et al., 2011; Shamir et al., 2014), or in a controlled environment such as recording rooms and laboratories (Giret et al., 2011; Oikarinen et al., 2019; Zhang et al., 2018). In some cases, a recording device was attached to the animals while for humans a Bluetooth stethoscope was used to obtain recordings of heart sounds. Other life recordings were collected with any of the following recording units, hydrophones, passive acoustic monitoring (PAM) systems, short-gun microphones, etc. attached to divers, seafloor moving boats, or sinks. In all, 24 datasets were privately generated and only 5 are available to the public.

An important consideration in research into sound classification is the availability of datasets. Easy access to high quality dataset is critical to research success in the field. With a total of 52 mentioned data sources from both categories, only 24 are reported to be publicly available, this is a confirmation of the challenges of limited datasets stated by researchers in sound classification. Whilst researchers may be able to readily generate or record some forms of sounds that can be used in research including outdoor sounds like barking of dogs, some other forms of sound may not be so easily generated or recorded, for example volcanic activity or sound from an impending tsunami.

Distribution of classified sounds according to the application domain. Considering the different types of classified sounds, the specific sound environment, and the researcher's objective for classifying the chosen sound, the classified sounds were categorized into three broad domains. They include Bioacoustics, Biomedical acoustics, and Ecoacoustics (see Figure 4). The application domain of bioacoustics was the most explored making up 50% of the study population. This domain consists of studies that classified sounds made from animals and human beings with the predominant aim of differentiating sounds and call types between and within animal species. Variations in animal sounds were also classified based on geographical locations.

On the other hand, the biomedical domain made up 24% of the study population and consists of studies that classified snore, heart, and lungs related diseases using sound. The goal of this domain was to provide an automated and efficient sound/acoustic signal classification system that will assist medical practitioners in smart diagnosis. Studies in this category also sought to eliminate the invasive traditional vision methodologies such as the use of medical imaging (Chen et al., 2019; Oweis et al., 2015; Vrbancic & Podgorelec, 2018). Equally, 26% of the studies explored sounds from the environment (ecoacoustics) to automatically recognize environmental acoustics scenes as well as to precisely classify the detected sound. This classification will enable the identification of sound events, environmental monitoring, and surveillance. The ecoacoustics domain consisted of sounds from sub-domains such as human activities, urban environment, surveillance, machinery, weather, and musical instruments.

Figure 4. The distribution of application domains



Sound Classification Algorithms and Performance Metrics

An automatic classifier does not only identify or differentiate one sound from another, but it also reduces false detection of sounds (Binder & Paul, 2019). Thus, this section will provide a summary of the distribution of ML techniques and performance metrics used in the included studies for sound classification over the study years (i.e., between 2010 and 2019). Several ML techniques were identified from the studies and categorized as follows:

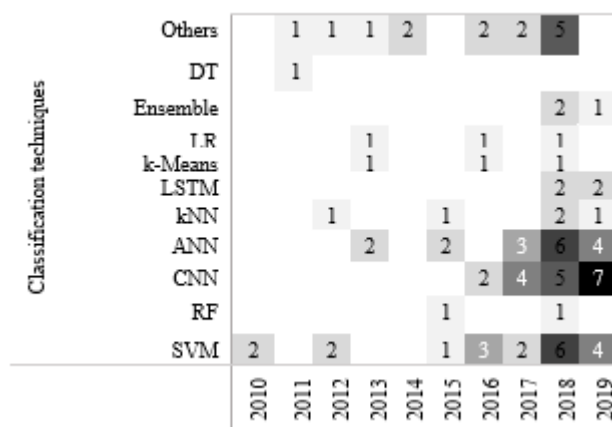
- i. Support Vector Machine (SVM) – SVM, Linear SVM, Radial Basis Function (RBF) SVM, MIML (multi-instance multi-label) SVM
- ii. Convolutional Neural Network (CNN) - CNN, Feedforward deep convolutional neural network, two-stream CNN (TSCNN-DS), CaffeNet pre-trained CNN, LeNet based CNN, SoundNet, EnvNet, multi-scale CNN (WaveMsNet), AlexNet, GoogleNet, and VGG16
- iii. Artificial Neural Network (ANN) - Deep Neural Network (DNN), Multilayer perceptron (MLP), Self-organizing map, Deep residual networks (ResNets), Convolutional deep belief network (CDBN), Sparse Auto-Encoder (SAE), Self-organizing map-Spike Neural Network (SOM-SNN)
- iv. Long Short-Term Memory - Recurrent Neural Network (RNN), LSTM-RNN, and Long short-term memory-fully convolutional network (LSTM-FCN)
- v. Random forest (RF)

- vi. K-Nearest neighbor (kNN)
- vii. Logistic Regression (LR)
- viii. Decision Tree (DT)
- ix. K-Means
- x. Ensemble Learners (EL)
- xi. Others - Sparse Representation-based Classifiers (SRC), Dynamic Time warping (DTW), Hidden Markov Model (HMM), Gaussian Mixture Models (GMM), aural classifiers, Non-Temporally Aware (NTA), Kernel-based extreme machine (KELM), Multi-view simple disagreement sampling (MV-SDS)

Overall, SVM, CNN, and ANN were the three predominantly used ML techniques in sound classification. These 3 techniques put together were adopted by 62% of the included studies (see Figure 5). Figure 5 presents a summary of the amount of research interest that each ML technique has received during the past decade. Further, it highlights the distribution of research interest in ML techniques in each publication year. It is important to note that more than one ML technique was used in some studies. Compared to other identified ML techniques, SVM, CNN, and ANN have received dominant research interest over the years with at least one of these techniques used between 2010 and 2019, except in 2014 where Gaussian mixture models (GMM) was used.

Support vector machine (SVM) has been identified as a robust technique in both classification and regression tasks. It is a supervised machine learning algorithm and it seeks to find the hyperplane which optimally separates the labeled data into their various classes (Bourouhou et al., 2019; Cvengros et al., 2012; Noda et al., 2016; Qian et al., 2017; Yaseen et al., 2018). Most of the articles that used SVM were focused on improving the classification performance either by modifying existing approaches of SVM-based classification or by adding new features to it. Modifications to existing approaches included Recursive feature elimination (SVM-RFE) and linear SVM (Cvengros et al., 2012), and SVM with linear kernels (Han et al., 2016), while added features included cost parameter CSVM (Malfante et al., 2018). Generally, SVMs have been reported to be cumbersome for multi-class tasks but robust for binary sound classification.

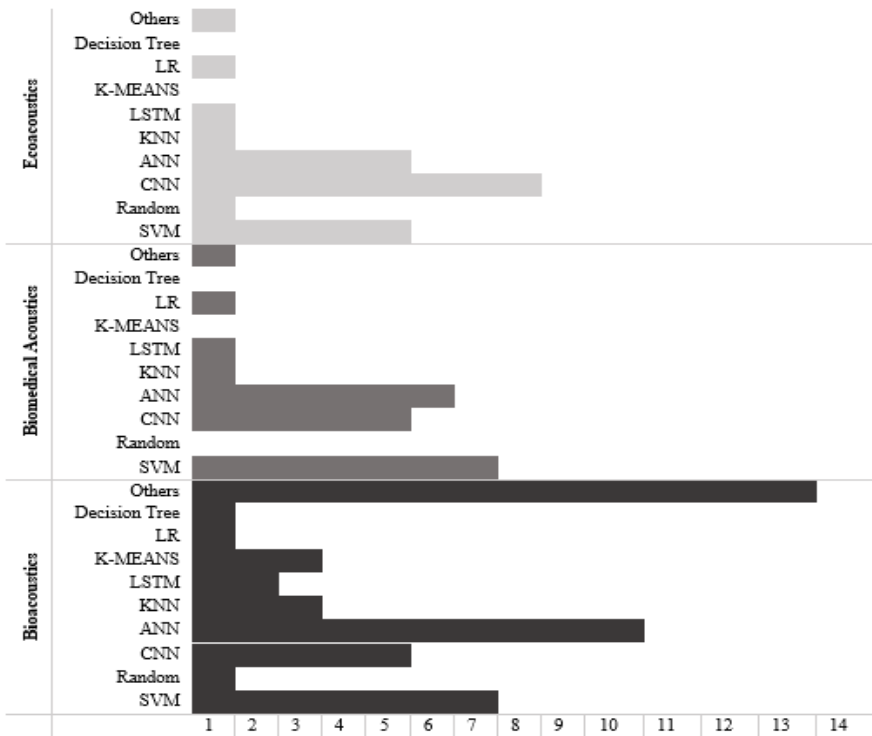
Figure 5. Distribution of ML techniques over publication year



Neural networks are algorithms that imitate the operations of a human brain to identify patterns and trends in data. Although its effectiveness is limited by the unavailability of labeled data, it is argued that they have self-organizing and adaptive learning properties with an outstanding ability to detect trends based on the sample data (Dwivedi et al., 2019). Accordingly, different types of Neural Networks in deep learning including CNN, ANN, LSTM were adopted by researchers in the included studies, and they made up 44% of the identified ML techniques.

Further, the identified classification techniques and their distribution of use were categorized as follows: supervised ML technique (76%), unsupervised ML technique (5%), semi-supervised (1%), ensemble learning (3%), and the others (sequential classifiers and statistical modeling techniques) made up 15%. Furthermore, advanced learning techniques such as transfer learning and ensemble learners were adopted by some researchers to obtain a more robust sound classification model as well as overcome the challenges of limited data, overfitting, and lack of labeled data. CNN pre-trained models such as VGG16, VGG19 LeNet based CNN, SoundNet, EnvNet, multi-scale CNN (WaveMsNet), AlexNet, GoogleNet and CaffeNet were adopted for Transfer learning (Amiriparian et al., 2017; Boddapati et al., 2017; Bold et al., 2019; Pandeya & Lee, 2018; Zhao et al., 2018; Zhu et al., 2018). While an ensemble of stacked autoencoders (Ibrahim et al., 2019), and an ensemble of supervised, unsupervised, and semi-supervised learning techniques such as random forest, kNN, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and SVM-RBF (Humayun et al., 2018; Pandeya & Lee, 2018) using majority voting and unweighted average were adopted for ensemble learning. Additionally, a semi-supervised learning technique called active learning was used to minimize the demand for human descriptions on sound classification training models (Han et al., 2016).

Figure 6. Distribution of classification techniques over application domain



Furthermore, the ML techniques were analyzed with respect to the three application domains identified in this review. More specifically, the distribution of ML techniques as earlier categorized, was mapped to the domains of bioacoustics, biomedical acoustics, and ecoacoustics (see Figure 6). As shown in Figure 6, SVM and CNN were mostly used to classify medical and environmental sounds respectively. While ANN and other sequential classifiers and statistical models were mostly used to classify sounds from animals (bioacoustics). Generally, SVM, CNN, ANN, and other statistical models were predominantly used in the three domains. Figure 6 also shows that while all the identified classification techniques were used in bioacoustics, certain ML techniques were not used in the domains of biomedical acoustics and ecoacoustics. Specifically, random forest, K-means, and decision trees were not used in classifying medical sounds. Similarly, K-means and decision trees were also not used in the classification of sounds in the environment.

Performance Metrics. An examination of the performance measures adopted by researchers to validate the reliability of their proposed ML techniques for sound classification is presented in this section. This includes evaluation measures such as cross-validation methods and classification metrics. Seven cross-validation methods were identified in the included studies for the primary purpose of evaluating model performance and computing classification accuracies. They include 10-fold cross-validation (Aucouturier et al., 2011; Han et al., 2016; Lebien & Ioup, 2018; Medhat et al., 2020; Pandeya et al., 2018; Salamon & Bello, 2015, 2017; Su et al., 2019), Leave-One-Out Cross-Validation (LOOCV) (Bourrouhou et al., 2019; Colonna et al., 2016; Oweis et al., 2015; Parada & Cardenal-Lopez, 2014; Vahabi & Selviah, 2019), 2-fold cross-validation (Ibrahim et al., 2018; Noda et al., 2016), 1 for 4-fold (Zhang et al., 2019), 5-fold cross-validation (Boddapati et al., 2017; Briggs et al., 2012; Fang et al., 2019; Mun et al., 2017; Tschannen et al., 2016; Yaseen et al., 2018), 20-fold cross-validation (Kumar et al., 2010), and 10-fold stratified cross-validation (Gingras & Fitch, 2013; Nogueira et al., 2019). Other specific reasons for the adoption of the cross-validation techniques were to determine validation error rate and estimates of algorithm performance (Han et al., 2016; Lebien & Ioup, 2018; Vahabi & Selviah, 2019).

Furthermore, classification metrics used to compare and evaluate the performance of the various ML and statistical techniques were identified. It is important to note that, more than one metric was used to evaluate the performance of a classification technique in most of the studies. Figure 7 provides an overview of the number and the proportion of reviewed studies using each performance metric. As shown in Figure 7, it is observed that accuracy is the predominantly used performance metric and it was adopted by 36% of the included studies. This is followed by Confusion Matrix (16%), Recall/sensitivity (14%), and Specificity (10%). Precision, F1-score, AUC score, ROC curve, and UAR are other adopted metrics in the included studies with an equal distribution of 4%. True Positive Rate (TPR), False Positive Rate (FPR), G-mean, and mean error rate are the least used. Generally, it was observed that the classification techniques used in the included studies predominantly had good classification accuracies.

Figure 7. Distribution of the studies over performance metrics



DISCUSSIONS

The primary objective of the systematic review was to identify publication trends, methodological approaches, and current algorithms used in the automatic classification of sounds using ML techniques. This review was restricted to open-access conferences and journal articles published between 2010 and 2019. Based on a set of inclusion and exclusion criteria, the included 68 studies were selected from Scopus and ASA databases with conference and journal articles making up 29% and 71% of the study population respectively.

This systematic review was guided by two categories of review questions which were answered accordingly. In the first category of this review, the publication trends between the years 2010 and 2019 were highlighted. It was observed that 60% of leading authors (that is authors with more than one publication) in sound classification predominantly focused on classifying sounds from animals, while the other 40% had an equal interest in classifying sounds in the environment and the biomedical domain. In addition, most of the studies originated from European and Asian countries (including the UK, USA, China, France, India, Korea, Portugal, Spain, and Germany) with a minimum of 3 publications and a maximum of 14 publications within the selected study years.

In the second category of review results, 13 groups of classified sounds that cut across the three major sound sources (anthrophony, biophony, and geophony) were identified. Further, these sound groups were divided into three application domains namely Bioacoustics, Biomedical acoustics, and Ecoacoustics. It was observed that the bioacoustics domain attracted more research interest and researchers were mostly interested in classifying sounds from marine mammals. Yet, little attention was given to classifying sounds from the underwater environment even in studies that classified environmental sound. This is a research gap considering that 70% of the earth is covered with water and the temperature of the ocean determines climate and wind patterns which in turn affects life on land and the ecosystem (Domingo, 2012). On the other hand, studies in the biomedical domain primarily focused on diagnosing respiratory diseases using sound. Although the classified sounds cut across three major application domains, the list of unclassified sounds is inexhaustive. For instance, studies in the biomedical domain should be extended to classify sounds from other internal body organs (as an alternative to radiography) to diagnose a variety of medical conditions. Studies should also investigate the classification of extreme events such as tornadoes, hurricanes, drought, earthquakes using sound. This will enable early detection and warning systems for natural disasters.

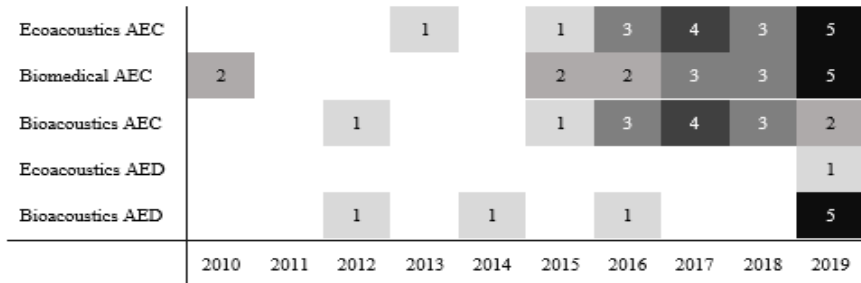
Review results also showed that a major research challenge reported by researchers was the unavailability of standardized labeled public datasets. This was particularly challenging for the biomedical domain, thus, researchers collected data (life recordings) from patients using Bluetooth stethoscope. Yet, the problem abounds because the collected data cannot be publicly available for future research. Perhaps, this could be a delimiting factor that dissuades researchers from delving into certain areas of sound classification.

In the identification of feature extraction techniques, it was observed that, although a variety of feature extraction techniques were used, specific patterns in the use of these techniques to a particular application domain could not be established. However, it was observed that MFCCs were predominantly used in feature extraction due to their ability to imitate the hearing properties of the human ear.

Furthermore, reported approaches for sound classification involved the use of both machine learning and non-machine learning techniques. Amongst the various identified classification techniques, support vector machines (SVM), convolutional neural networks (CNN), artificial neural networks (ANN), and other probabilistic statistical models were predominantly used in the domains of bioacoustics, biomedical acoustics, and ecoacoustics. The findings on the prevalence of ANN, CNN, and SVM in the classification of medical acoustics is similar to findings from the systematic review on ML in lung sound classification (Dwivedi et al., 2019; Palaniappan et al., 2013). Indeed, the predominant use of CNN for sound classification is no surprise considering that, most of the studies adopted an image-based approach for sound classification using spectrograms. Mitilneos et al., (2018) posit that neural network are adopted for sound classification due to their ability to identify specific patterns exhibited by sound sources using the distribution of energy over frequency and time. Also, machine learning techniques are outstandingly able to differentiate target acoustic signals/sound from an acoustic background (Shamir et al., 2014). Although neural networks reportedly require high computational power and large datasets, no study reported this as a limitation or a challenge. Overall, satisfactory results were reported for the various classification techniques as observed in the results of the performance metrics. Performance metrics such as cross-validation, classification accuracy, confusion matrix, recall, and precision were used to evaluate the performance of the classifiers. In cases of an unbalanced distribution of datasets, other performance metrics such as UAR, AUC curve, ROC curve were adopted.

Finally, two types of acoustic signal classification schemes were identified, they included detection-and-classification otherwise known as acoustic event detection (AED), and detection-by-classification otherwise known as acoustic event classification (AEC). While the former involves detection of the sound and then its 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 (Temko & Nadeu, 2009). Conversely, in detection-by-classification, the task of detection automatically translates to classification as its strategy is based on using classifiers (such as HMM, logistic regression) with inbuilt segmentation algorithms (Temko & Nadeu, 2009). As shown in Figure 8, 71% of the studies focused on AEC, while 29% adopted the AED approach. Also, detection and classification were performed in the domains of bioacoustics and environment only, while detection-by-classification cut across the three identified domains with bioacoustics as the most explored.

Figure 8. Distribution of classification categories per application domain by year



LIMITATION OF THE STUDY AND CONCLUSION

This paper presented the findings of a systematic review of primary studies in the area of sound classification between the years 2010 and 2019. A major strength of this systematic review is that it was not specific to a particular sound, but it considered every kind of sound that cut across the domains of bioacoustics, ecoacoustics, and biomedical acoustics. It also identified two broad categories of sound classification schemes: acoustic event detection (AED) and acoustic event classification (AEC). Findings from the review indicated that automatic detection and classification systems were useful tools that could differentiate one acoustic event from the other, especially when deep learning techniques were used for the task.

Although the reviews provided methodologies and algorithms used in various domains of sound classification, findings indicated that the methodologies and domains (in terms of scope) were not exhaustive. For instance, there was no study on the acoustic classification or detection of extreme events such as seismic and volcanic activities or the classification of medical conditions other than respiratory tract-related diseases. Also, the unavailability of publicly benchmarked datasets for sound classification in certain domains posed a challenge to the reproducibility of research approaches. Another hindrance to reproducibility is that model architectures and methods used for training datasets were not disclosed, especially in conference articles. Considering the relevance of reproducibility in scientific research, this research gap should be addressed in future studies. Generally, future studies should seek to address research challenges such as limited bandwidth, threshold problems, lack of general applicability of classifiers and publicly available datasets. Furthermore, this study acknowledged that the search strategy is not exhaustive: limiting the search to only open-accessed Scopus and ASA creates the possibility of omitting other relevant related studies.

FUNDING AGENCY

The publisher has waived the Open Access Processing fee for this article.

REFERENCES

- Allen, J. A., Murray, A., Noad, M. J., Dunlop, R. A., & Garland, E. C. (2017). Using self-organizing maps to classify humpback whale song units and quantify their similarity. *The Journal of the Acoustical Society of America*, 142(4), 1943–1952. doi:10.1121/1.4982040 PMID:29092588
- Amiriparian, S., Gerczuk, M., Ottl, S., Cummins, N., Freitag, M., Pugachevskiy, S., Baird, A., & Schuller, B. (2017). Snore sound classification using image-based deep spectrum features. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2017*-3512–3516. doi:10.21437/Interspeech.2017-434
- Aucouturier, J.-J., Nonaka, Y., Katahira, K., & Okanoya, K. (2011). Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. *The Journal of the Acoustical Society of America*, 130(5), 2969–2977. doi:10.1121/1.3641377 PMID:22087925
- Binder, C., & Paul, H. (2019). Range-dependent impacts of ocean acoustic propagation on automated classification of transmitted bowhead and humpback whale vocalizations. *The Journal of the Acoustical Society of America*, 2480(4), 2480–2497. Advance online publication. doi:10.1121/1.5097593 PMID:31046335
- Boddapati, V., Petef, A., Rasmusson, J., & Lundberg, L. (2017). Classifying environmental sounds using image recognition networks. *Procedia Computer Science*, 112, 2048–2056. doi:10.1016/j.procs.2017.08.250
- Bold, N., Zhang, C., & Akashi, T. (2019). Cross-domain deep feature combination for bird species classification with audio-visual data. *IEICE Transactions on Information and Systems*, E102D(10), 2033–2042. 10.1587/transinf.2018EDP7383
- Bourouhou, A., Jilbab, A., Nacir, C., & Hammouch, A. (2019). Heart sounds classification for medical diagnostic assistance. *International Journal of Online and Biomedical Engineering*, 15(11), 88–103. doi:10.3991/ijoe.v15i11.10804
- Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S. J. K., Hadley, A. S., & Betts, M. G. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131(6), 4640–4650. doi:10.1121/1.4707424 PMID:22712937
- Chen, H., Yuan, X., Pei, Z., Li, M., & Li, J. (2019). Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks. *IEEE Access: Practical Innovations, Open Solutions*, 7(April), 32845–32852. doi:10.1109/ACCESS.2019.2903859
- Chu, S., Narayanan, S., & Kuo, C. C. J. (2009). Environmental sound recognition with time-frequency audio features. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6), 1142–1158. doi:10.1109/TASL.2009.2017438
- Colonna, J., Peet, T., Ferreira, C. A., Jorge, A. M., Gomes, E. F., & Gama, J. (2016). Automatic classification of anuran sounds using convolutional neural networks. *ACM International Conference Proceeding Series*, 73–78. doi:10.1145/2948992.2949016
- Cvengros, R. M., Valente, D., Nykaza, E. T., & Vipperman, J. S. (2012). Blast noise classification with common sound level meter metrics. *The Journal of the Acoustical Society of America*, 132(2), 822–831. doi:10.1121/1.4730921 PMID:22894205
- Davis, N., & Suresh, K. (2019). Environmental sound classification using deep convolutional neural networks and data augmentation. *2018 IEEE Recent Advances in Intelligent Computational Systems. RAICS, 2018*, 41–45. doi:10.1109/RAICS.2018.8635051
- Domingo, M. C. (2012). An overview of the internet of underwater things. *Journal of Network and Computer Applications*, 35(6), 1879–1890. doi:10.1016/j.jnca.2012.07.012
- Dwivedi, A. K., Imtiaz, S. A., & Rodriguez-Villegas, E. (2019). Algorithms for automatic analysis and classification of heart sounds-A systematic review. *IEEE Access: Practical Innovations, Open Solutions*, 7(c), 8316–8345. doi:10.1109/ACCESS.2018.2889437
- Elfergany, A. K., & Adl, A. (2020). Identification of Telecom Volatile Customers Using a Particle Swarm Optimized K-Means Clustering on Their Personality Traits Analysis. *International Journal of Service Science, Management, Engineering, and Technology*, 11(2), 1–15. doi:10.4018/IJSSMET.2020040101

Fang, S. H., Te Wang, C., Chen, J. Y., Tsao, Y., & Lin, F. C. (2019). Combining acoustic signals and medical records to improve pathological voice classification. *APSIPA Transactions on Signal and Information Processing*, 8(1), 1–11. doi:10.1017/ATSIP.2019.7

Gingras, B., & Fitch, W. T. (2013). A three-parameter model for classifying anurans into four genera based on advertisement calls. *The Journal of the Acoustical Society of America*, 133(October), 547–559.

Giret, N., Roy, P., Albert, A., Pachet, F., Kreutzer, M., & Bovet, D. (2011). Finding good acoustic features for parrot vocalizations: The feature generation approach. *The Journal of the Acoustical Society of America*, 129(2), 1089–1099. doi:10.1121/1.3531953 PMID:21361465

Greenhalgh, T. (1997). How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ (Clinical Research Ed.)*, 315(7109), 672–675. doi:10.1136/bmj.315.7109.672 PMID:9310574

Guilment, T., Socheleau, F.-X., Pastor, D., & Vallez, S. (2018). Sparse representation-based classification of mysticete calls. *The Journal of the Acoustical Society of America*, 144(3), 1550–1563. doi:10.1121/1.5055209 PMID:30424647

Halkias, X., Paris, S., & Glotin, H. (2013). Classification of mysticete sounds using machine learning techniques. *The Journal of the Acoustical Society of America*, 134(5), 3496–3505. doi:10.1121/1.4821203 PMID:24180760

Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., & Zhu, X. (2016). Semi-supervised active learning for sound classification in hybrid learning environments. *PLoS One*, 11(9), 1–19. doi:10.1371/journal.pone.0162075 PMID:27627768

Hao, Y., Weiss, G. M., & Brown, S. M. (2018). Identification of Candidate Genes Responsible for Age-related Macular Degeneration using Microarray Data. *International Journal of Service Science, Management, Engineering, and Technology*, 9(2), 33–60. doi:10.4018/IJSSMET.2018040102

Hlioui, F., Aloui, N., & Gargouri, F. (2020). Withdrawal Prediction Framework in Virtual Learning Environment. *International Journal of Service Science, Management, Engineering, and Technology*, 11(3), 47–64. doi:10.4018/IJSSMET.2020070104

Humayun, A. I., Tauhiduzzaman Khan, M., Ghaffarzadegan, S., Feng, Z., & Hasan, T. (2018). An ensemble of transfer, semi-supervised and supervised learning methods for pathological heart sound classification. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 127–131. doi:10.21437/Interspeech.2018-2413

Ibrahim, A. K., Chérubin, L. M., Zhuang, H., Schärer Umpierre, M. T., Dalglish, F., Erdol, N., Ouyang, B., & Dalglish, A. (2018). An approach for automatic classification of grouper vocalizations with passive acoustic monitoring. *The Journal of the Acoustical Society of America*, 143(2), 666–676. doi:10.1121/1.5022281 PMID:29495690

Ibrahim, A. K., Zhuang, H., Chérubin, L. M., Schärer-Umpierre, M. T., & Erdol, N. (2018). Automatic classification of grouper species by their sounds using deep neural networks. *The Journal of the Acoustical Society of America*, 144(3), EL196–EL202. doi:10.1121/1.5054911 PMID:30424627

Ibrahim, A. K., Zhuang, H., Chérubin, L. M., Umpierre, M. T. S., Ali, A. M., Richard, S., Sch, M. T., Ali, A. M., Nemeth, R. S., & Erdol, N. (2019). Classification of red hind grouper call types using random ensemble of stacked autoencoders. *The Journal of the Acoustical Society of America*, 146(4), 2155–2162. doi:10.1121/1.5126861 PMID:31671953

Kaewtip, K., Alwan, A., O'Reilly, C., & Taylor, C. E. (2016). A robust automatic birdsong phrase classification: A template-based approach. *The Journal of the Acoustical Society of America*, 140(5), 3691–3701. doi:10.1121/1.4966592 PMID:27908084

Karbasi, M., Ahadi, S. M., & Bahmanian, M. (2011). Environmental sound classification using spectral dynamic features. *ICICS 2011 - 8th International Conference on Information, Communications and Signal Processing*, 2–7. doi:10.1109/ICICS.2011.6173513

Kumar, D., Carvalho, P., Antunes, M., Paiva, R. P., & Henriques, J. (2010). Heart murmur classification with feature selection. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10, June 2014*, 4566–4569. doi:10.1109/IEMBS.2010.5625940

- Kumar, D., Carvalho, P., Couceiro, R., Antunes, M., Paiva, R. P., & Henriques, J. (2010). Heart murmur classification using complexity signatures. *Proceedings - International Conference on Pattern Recognition*, 2564–2567. doi:10.1109/ICPR.2010.628
- Lebien, J., & Ioup, J. (2018). Species-level classification of beaked whale echolocation signals detected in the northern Gulf of Mexico. *The Journal of the Acoustical Society of America*, 144(1), 387–396. doi:10.1121/1.5047435 PMID:30075691
- Loey, M., ElSawy, A., & Afify, M. (2020). Deep Learning in Plant Diseases Detection for Agricultural Crops: A Survey. *International Journal of Service Science, Management, Engineering, and Technology*, 11(2), 41–58. doi:10.4018/IJSSMET.2020040103
- Loey, M., Naman, M. R., & Zayed, H. H. (2020). A Survey on Blood Image Diseases Detection Using Deep Learning. *International Journal of Service Science, Management, Engineering, and Technology*, 11(3), 18–32. doi:10.4018/IJSSMET.2020070102
- Luque, A., Romero-Lemos, J., Carrasco, A., & Gonzalez-Abril, L. (2018). Temporally-aware algorithms for the classification of anuran sounds. *PeerJ*, 6(e4732), 1–40. doi:10.7717/peerj.4732 PMID:29740517
- Malfante, M., Mars, J. I., Dalla Mura, M., & Gervaise, C. (2018). Automatic fish sounds classification. *The Journal of the Acoustical Society of America*, 143(5), 2834–2846. doi:10.1121/1.5036628 PMID:29857733
- Medhat, F., Chesmore, D., & Robinson, J. (2020). Masked Conditional Neural Networks for sound classification. *Applied Soft Computing*, 90(608014), 1–13. doi:10.1016/j.asoc.2020.106073
- Mitilineos, S. A., Potirakis, S. M., Tatlas, N. A., & Rangoussi, M. (2018). A two-level sound classification platform for environmental monitoring. *Journal of Sensors*, 2018(5828074), 1–13. doi:10.1155/2018/5828074
- Mun, S., Shon, S., Kim, W., Han, D. K., & Ko, H. (2017). A novel discriminative feature extraction for acoustic scene classification using RNN based source separation. *IEICE Transactions on Information and Systems*, E100D(12), 3041–3044. 10.1587/transinf.2017EDL8132
- Noda, J. J., Travieso, C. M., & Sánchez-Rodríguez, D. (2016). Automatic taxonomic classification of fish based on their acoustic signals. *Applied Sciences (Switzerland)*, 6(12), 443. Advance online publication. doi:10.3390/app6120443
- Nogueira, D. M., Ferreira, C. A., Gomes, E. F., & Jorge, A. M. (2019). Classifying Heart Sounds Using Images of Motifs, MFC, C and Temporal Features. *Journal of Medical Systems*, 43(6), 186–203. doi:10.1007/s10916-019-1286-5 PMID:31056720
- Oikarinen, T., Srinivasan, K., Meisner, O., Hyman, J. B., Parmar, S., Fanucci-Kiss, A., Desimone, R., Landman, R., & Feng, G. (2019). Deep convolutional network for animal sound classification and source attribution using dual audio recordings. *The Journal of the Acoustical Society of America*, 145(2), 654–662. doi:10.1121/1.5087827 PMID:30823820
- Oletic, D., Arsenali, B., & Bilas, V. (2012). Towards continuous wheeze detection body sensor node as a core of asthma monitoring system. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 83 LNICST, 165–172. 10.1007/978-3-642-29734-2_23
- Ou, H., Au, W., Zurk, L., & Lammers, M. (2013). Automated extraction and classification of time-frequency contours in humpback vocalizations. *The Journal of the Acoustical Society of America*, 133(1), 301–310. doi:10.1121/1.4770251 PMID:23297903
- Oweis, R. J., Abdulhay, E. W., Khayal, A., & Awad, A. (2015). An alternative respiratory sounds classification system utilizing artificial neural networks. *Biomedical Journal*, 38(2), 153–161. doi:10.4103/2319-4170.137773 PMID:25179722
- Palaniappan, R., Sundaraj, K., & Ahamed, N. U. (2013). Machine learning in lung sound analysis: A systematic review. *Biocybernetics and Biomedical Engineering*, 33(3), 129–135. doi:10.1016/j.bbe.2013.07.001
- Pandeya, Y. R., Kim, D., & Lee, J. (2018). Domestic cat sound classification using learned features from deep neural nets. *Applied Sciences (Switzerland)*, 8(10), 1–17. doi:10.3390/app8101949
- Pandeya, Y. R., & Lee, J. (2018). Domestic cat sound classification using transfer learning. *International Journal of Fuzzy Logic and Intelligent Systems*, 18(2), 154–160. doi:10.5391/IJFIS.2018.18.2.154

- Parada, P. P., & Cardenal-Lopez, A. (2014). Using Gaussian mixture models to detect and classify dolphin whistles and pulses. *The Journal of the Acoustical Society of America*, 135(6), 3371–3381. doi:10.1121/1.4876439 PMID:24907800
- Perr, J. (2005). Basic acoustics and Signal Processing. *LinuxFocus.Org*, 1(271), 1–22. <http://linuxfocus.org>
- Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. *PLoS One*, 12(5), e0177926. Advance online publication. doi:10.1371/journal.pone.0177926 PMID:28552969
- Qian, K., Zhang, Z., Baird, A., & Schuller, B. (2017). Active learning for bird sound classification via a kernel-based extreme learning machine. *The Journal of the Acoustical Society of America*, 142(4), 1796–1804. doi:10.1121/1.5004570 PMID:29092546
- Roch, M. A., Klinck, H., Baumann-Pickering, S., Mellinger, D. K., Qui, S., Soldevilla, M. S., & Hildebrand, J. A. (2011). Classification of echolocation clicks from odontocetes in the Southern California Bight. *The Journal of the Acoustical Society of America*, 129(1), 467–475. doi:10.1121/1.3514383 PMID:21303026
- Salama, M. A., & Hassanien, A. E. (2014). Fuzzification of Euclidean Space Approach in Machine Learning Techniques. *International Journal of Service Science, Management, Engineering, and Technology*, 5(4), 29–43. doi:10.4018/ijssmet.2014100103
- Salamon, J., & Bello, J. P. (2015). Unsupervised Feature Learning for Urban Sound Classification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 171–175.
- Salamon, J., & Bello, J. P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters*, 24(3), 279–283. doi:10.1109/LSP.2017.2657381
- Sangwan, N., & Bhatnagar, V. (2020). Comprehensive Contemplation of Probabilistic Aspects in Intelligent Analytics. *International Journal of Service Science, Management, Engineering, and Technology*, 11(1), 116–141. doi:10.4018/IJSSMET.2020010108
- Sengupta, N., Sahidullah, M., & Saha, G. (2016). Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine*, 75, 118–129. doi:10.1016/j.combiomed.2016.05.013 PMID:27286184
- Shamir, L., Yerby, C., Simpson, R., von Benda-Beckmann, A. M., Tyack, P., Samarra, F., Miller, P., & Wallin, J. (2014). Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *The Journal of the Acoustical Society of America*, 135(2), 953–962. doi:10.1121/1.4861348 PMID:25234903
- Su, Y., Zhang, K., Wang, J., & Madani, K. (2019). Environment sound classification using a two-stream CNN based on decision-level fusion. *Sensors (Switzerland)*, 19(7), 1–15. doi:10.3390/s19071733 PMID:30978974
- Tan, L. N., Alwan, A., Kossan, G., Cody, M. L., & Taylor, C. E. (2015). Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training data a). *The Journal of the Acoustical Society of America*, 137(3), 1069–1080. Advance online publication. doi:10.1121/1.4906168 PMID:25786922
- Tatoian, R., & Hamel, L. (2018). Self-organizing map convergence. *International Journal of Service Science, Management, Engineering, and Technology*, 9(2), 61–84. doi:10.4018/IJSSMET.2018040103
- Temko, A., Nadeu, C., Macho, D., Malkin, R., Zieger, C., & Omologo, M. (2009). Acoustic Event Detection and Classification. *Computers in the Human Interaction Loop*, (December), 61–73. doi:10.1007/978-1-84882-054-8_7
- Thakur, A., Thapar, D., Rajan, P., & Nigam, A. (2019). Deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss. *The Journal of the Acoustical Society of America*, 146(1), 534–547. doi:10.1121/1.5118245 PMID:31370640
- Tschannen, M., Kramer, T., Marti, G., Heinzmann, M., & Wiatowski, T. (2016). Heart sound classification using deep structured features. *Computers in Cardiology*, 43, 565–568. doi:10.22489/CinC.2016.162-186
- Vahabi, N., & Selviah, D. R. (2019). Convolutional Neural Networks to Classify Oil, Water and Gas Wells Fluid Using Acoustic Signals. *2019 IEEE 19th International Symposium on Signal Processing and Information Technology, ISSPIT 2019*. doi:10.1109/ISSPIT47144.2019.9001845

- Vrbancic, G., & Podgorelec, V. (2018). Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks. *Elektronika ir Elektrotechnika*, 24(4), 1–7. doi:10.5755/j01.eie.24.4.21469
- Wren, Y., Harding, S., Goldbart, J., & Roulstone, S. (2018). A systematic review and classification of interventions for speech-sound disorder in preschool children. *International Journal of Language & Communication Disorders*, 53(3), 446–467. doi:10.1111/1460-6984.12371 PMID:29341346
- Yaseen, S., Son, G.-Y., & Kwon, S. (2018). Classification of heart sound signal using multiple features. *Applied Sciences (Basel, Switzerland)*, 8(12), 1–14. doi:10.3390/app8122344
- Zhang, Y., Lv, D., & Zhao, Y. (2016). Multiple-view active learning for environmental sound classification. *International Journal of Online Engineering*, 12(12), 49–54. doi:10.3991/ijoe.v12i12.6458
- Zhang, Y.-J., Huang, J.-F., Gong, N., Ling, Z.-H., & Hu, Y. (2018). Automatic detection and classification of marmoset vocalizations using deep and recurrent neural networks. *The Journal of the Acoustical Society of America*, 144(1), 478–487. doi:10.1121/1.5047743 PMID:30075670
- Zhao, H., Huang, X., Liu, W., & Yang, L. (2018). Environmental sound classification based on feature fusion. *MATEC Web of Conferences*, 173, 1–5. doi:10.1051/mateconf/201817303059
- Zhu, B., Wang, C., Liu, F., Lei, J., Lu, Z., & Peng, Y. (2018). Learning Environmental Sounds with with Multi-scale Convolutional Neural Network. *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 1–8. doi:10.1109/IJCNN.2018.8489641

APPENDIX A - PRIMARY STUDIES USED FOR THE SYSTEMATIC REVIEW

Table 4. Primary studies

Ref no.	Bibliography
A1.	Shamir, L., Yerby, C., Simpson, R., von Benda-Beckmann, A. M., Tyack, P., Samarra, F., Miller, P., & Wallin, J. (2014). Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. <i>The Journal of the Acoustical Society of America</i> , 135(2), 953–962. https://doi.org/10.1121/1.4861348
A2.	Qian, K., Zhang, Z., Baird, A., & Schuller, B. (2017). Active learning for bird sound classification via a kernel-based extreme learning machine. <i>The Journal of the Acoustical Society of America</i> , 142(4), 1796–1804. https://doi.org/10.1121/1.5004570
A3.	Malfante, M., Mars, J. I., Dalla Mura, M., & Gervaise, C. (2018). Automatic fish sounds classification. <i>The Journal of the Acoustical Society of America</i> , 143(5), 2834–2846. https://doi.org/10.1121/1.5036628
A4.	Halkias, X. C., Paris, S., & Glotin, H. (2013). Classification of mysticete sounds using machine learning techniques. <i>The Journal of the Acoustical Society of America</i> , 134(5), 3496–3505. https://doi.org/10.1121/1.4821203
A5.	Thakur, A., Thapar, D., Rajan, P., & Nigam, A. (2019). Deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss. <i>The Journal of the Acoustical Society of America</i> , 146(1), 534–547. https://doi.org/10.1121/1.5118245
A6.	Cvengros, R. M., Valente, D., Nykaza, E. T., & Vipperman, J. S. (2012a). Blast noise classification with common sound level meter metrics. <i>The Journal of the Acoustical Society of America</i> , 132(2), 822–831. https://doi.org/10.1121/1.4730921
A7.	Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S. J. K., Hadley, A. S., & Betts, M. G. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. <i>The Journal of the Acoustical Society of America</i> , 131(6), 4640–4650. https://doi.org/10.1121/1.4707424
A8.	Robakis, E., Watsa, M., & Erkenwick, G. (2018). Classification of producer characteristics in primate long calls using neural networks. <i>The Journal of the Acoustical Society of America</i> , 144(1), 344–353. https://doi.org/10.1121/1.5046526
A9.	Ibrahim, A. K., Chérubin, L. M., Zhuang, H., Schärer Umpierre, M. T., Dalglish, F., Erdol, N., Ouyang, B., & Dalglish, A. (2018). An approach for automatic classification of grouper vocalizations with passive acoustic monitoring. <i>The Journal of the Acoustical Society of America</i> , 143(2), 666–676. https://doi.org/10.1121/1.5022281
A10.	Zhang, Y.-J., Huang, J.-F., Gong, N., Ling, Z.-H., & Hu, Y. (2018). Automatic detection and classification of marmoset vocalizations using deep and recurrent neural networks. <i>The Journal of the Acoustical Society of America</i> , 144(1), 478–487. https://doi.org/10.1121/1.5047743
A11.	Oikarinen, T., Srinivasan, K., Meisner, O., Hyman, J. B., Parmar, S., Fanucci-Kiss, A., Desimone, R., Landman, R., & Feng, G. (2019). Deep convolutional network for animal sound classification and source attribution using dual audio recordings. <i>The Journal of the Acoustical Society of America</i> , 145(2), 654–662. https://doi.org/10.1121/1.5087827
A12.	Ibrahim, A. K., Zhuang, H., Chérubin, L. M., Umpierre, M. T. S., Ali, A. M., Richard, S., Sch, M. T., Ali, A. M., Nemeth, R. S., & Erdol, N. (2019). Classification of red hind grouper call types using a random ensemble of stacked autoencoders. 2155. https://doi.org/10.1121/1.5126861
A13.	Guilment, T., Socheleau, F.-X., Pastor, D., & Vallez, S. (2018). Sparse representation-based classification of mysticete calls. <i>The Journal of the Acoustical Society of America</i> , 144(3), 1550–1563. https://doi.org/10.1121/1.5055209
A14.	Kaewtip, K., Alwan, A., O'Reilly, C., & Taylor, C. E. (2016). A robust automatic birdsong phrase classification: A template-based approach. <i>The Journal of the Acoustical Society of America</i> , 140(5), 3691–3701. https://doi.org/10.1121/1.4966592

Table 4 continued on next page

Table 4 continued

Ref no.	Bibliography
A15.	Binder, C., & Paul, H. (2019). Range-dependent impacts of ocean acoustic propagation on automated classification of transmitted bowhead and humpback whale vocalizations. 2480. https://doi.org/10.1121/1.5097593
A16.	Roch, M. A., Newport, D., Baumann-pickering, S., Mellinger, D. K., Qui, S., Soldevilla, M. S., & Hildebrand, J. A. (2011). Classification of echolocation clicks from odontocetes in the Southern California Bight. <i>The Journal of the Acoustical Society of America</i> , 129(January), 467–476. https://doi.org/10.1121/1.3514383
A17.	Allen, J. A., Murray, A., Noad, M. J., Dunlop, R. A., & Garland, E. C. (2017). Using self-organizing maps to classify humpback whale song units and quantify their similarity. <i>The Journal of the Acoustical Society of America</i> , 142(4), 1943–1952. https://doi.org/10.1121/1.4982040
A18.	Tan, L. N., Alwan, A., Kossan, G., Cody, M. L., & Taylor, C. E. (2015). Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training data a). 137(3). https://doi.org/10.1121/1.4906168
A19.	Ou, H., Au, W., Zurk, L., & Lammers, M. (2013). <i>Automated extraction and classification of time-frequency contours in humpback vocalizations</i> . 133(January).
A20.	LeBien, J. G., & Ioup, J. W. (2018). Species-level classification of beaked whale echolocation signals detected in the northern Gulf of Mexico. <i>The Journal of the Acoustical Society of America</i> , 144(1), 387–396. https://doi.org/10.1121/1.5047435
A21.	Giret, N., Roy, P., Albert, A., Pachet, F., Kreutzer, M., & Bovet, D. (2011). Finding good acoustic features for parrot vocalizations: The feature generation approach. <i>The Journal of the Acoustical Society of America</i> , 129(2), 1089–1099.
A22.	Parada, P. P., & Cardenal-Lopez, A. (2014). Using Gaussian mixture models to detect and classify dolphin whistles and pulses. <i>The Journal of the Acoustical Society of America</i> , 135(June), 3371–3381. https://dx.doi.org/10.1121/1.4876439
A23.	Gingras, B., & Fitch, W. T. (2013). <i>A three-parameter model for classifying anurans into four genera based on advertisement calls</i> . 133(October 2012), 547–559.
A24.	Aucouturier, J.-J., Nonaka, Y., Katahira, K., & Okanoya, K. (2011). Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. <i>The Journal of the Acoustical Society of America</i> , 130(5), 2969–2977. https://doi.org/10.1121/1.3641377
A25.	Bishop, J. C., Falzon, G., Trotter, M., Kwan, P., & Meek, P. D. (2019). Livestock vocalization classification in farm soundscapes. <i>Computers and Electronics in Agriculture</i> , 162(April), 531–542. https://doi.org/10.1016/j.compag.2019.04.020
A26.	Aziz, S., Awais, M., Akram, T., Khan, U., Alhussein, M., & Aurangzeb, K. (2019). Automatic scene recognition through acoustic classification for behavioral robotics. <i>Electronics (Switzerland)</i> , 8(5). https://doi.org/10.3390/electronics8050483
A27.	Chen, H., Yuan, X., Pei, Z., Li, M., & Li, J. (2019). Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks. <i>IEEE Access</i> , 7(April), 32845–32852. https://doi.org/10.1109/ACCESS.2019.2903859
A28.	Bourouhou, A., Jilbab, A., Nacir, C., & Hammouch, A. (2019). Heart sounds classification for medical diagnostic assistance. <i>International Journal of Online and Biomedical Engineering</i> , 15(11), 88–103. https://doi.org/10.3991/ijoe.v15i11.10804
A29.	Yaseen, Son, G. Y., & Kwon, S. (2018). Classification of heart sound signal using multiple features. <i>Applied Sciences (Switzerland)</i> , 8(12). https://doi.org/10.3390/app8122344
A30.	Pandeya, Y. R., Kim, D., & Lee, J. (2018). Domestic cat sound classification using learned features from deep neural nets. <i>Applied Sciences (Switzerland)</i> , 8(10), 1–17. https://doi.org/10.3390/app8101949
A31.	Luque, A., Romero-Lemos, J., Carrasco, A., & Barbancho, J. (2018). Non-sequential automatic classification of anuran sounds for the estimation of climate-change indicators. <i>Expert Systems with Applications</i> , 95, 248–260. https://doi.org/10.1016/j.eswa.2017.11.016

Table 4 continued on next page

Table 4 continued

Ref no.	Bibliography
A32.	Kim, Y., Sa, J., Chung, Y., Park, D., & Lee, S. (2018). Resource-efficient pet dog sound events classification using LSTM-FCN based on time-series data. <i>Sensors (Switzerland)</i> , 18(11). https://doi.org/10.3390/s18114019
A33.	Luque, A., Romero-Lemos, J., Carrasco, A., & Gonzalez-Abril, L. (2018). Temporally aware algorithms for the classification of anuran sounds. <i>PeerJ</i> , 2018(5), 1–40. https://doi.org/10.7717/peerj.4732
A34.	Aykanat, M., Kılıç, Ö., Kurt, B., & Saryal, S. (2017). Classification of lung sounds using convolutional neural networks. <i>Eurasip Journal on Image and Video Processing</i> , 2017(1). https://doi.org/10.1186/s13640-017-0213-2
A35.	Zhang, Yan, Lv, D., & Zhao, Y. (2016). Multiple-view active learning for environmental sound classification. <i>International Journal of Online Engineering</i> , 12(12), 49–54. https://doi.org/10.3991/ijoe.v12i12.6458
A36.	Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., & Zhu, X. (2016). Semi-supervised active learning for sound classification in hybrid learning environments. <i>PLoS ONE</i> , 11(9), 1–19. https://doi.org/10.1371/journal.pone.0162075
A37.	Noda, J. J., Travieso, C. M., & Sánchez-Rodríguez, D. (2016). Automatic taxonomic classification of fish based on their acoustic signals. <i>Applied Sciences (Switzerland)</i> , 6(12). https://doi.org/10.3390/app6120443
A38.	Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S., & On, B. W. (2019). Heartbeat sound signal classification using deep learning. <i>Sensors (Switzerland)</i> , 19(21), 1–15. https://doi.org/10.3390/s19214819
A39.	Su, Y., Zhang, K., Wang, J., & Madani, K. (2019). Environment sound classification using a two-stream CNN based on decision-level fusion. <i>Sensors (Switzerland)</i> , 19(7), 1–15. https://doi.org/10.3390/s19071733
A40.	Khamparia, A., Gupta, D., Nguyen, N. G., Khanna, A., Pandey, B., & Tiwari, P. (2019). Sound classification using convolutional neural network and tensor deep stacking network. <i>IEEE Access</i> , 7(January), 7717–7727. https://doi.org/10.1109/ACCESS.2018.2888882
A41.	Bold, N., Zhang, C., & Akashi, T. (2019). Cross-domain deep feature combination for bird species classification with audio-visual data. <i>IEICE Transactions on Information and Systems</i> , E102D(10), 2033–2042. https://doi.org/10.1587/transinf.2018EDP7383
A42.	Verma, D., Jana, A., & Ramamritham, K. (2019). Classification and mapping of sound sources in local urban streets through AudioSet data and Bayesian optimized Neural Networks. <i>Noise Mapping</i> , 6(1), 52–71. https://doi.org/10.1515/noise-2019-0005
A43.	Wu, J., Chua, Y., Zhang, M., Li, H., & Tan, K. C. (2018). A spiking neural network framework for robust sound classification. <i>Frontiers in Neuroscience</i> , 12(NOV), 1–17. https://doi.org/10.3389/fnins.2018.00836
A44.	Pandeya, Y. R., & Lee, J. (2018). Domestic cat sound classification using transfer learning. <i>International Journal of Fuzzy Logic and Intelligent Systems</i> , 18(2), 154–160. https://doi.org/10.5391/IJFIS.2018.18.2.154
A45.	Vrbancic, G., & Podgorelec, V. (2018). Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks. <i>Elektronika Ir Elektrotehnika</i> , 24(4), 1–7. https://doi.org/10.5755/j01.eie.24.4.21469
A46.	Salamon, J., & Bello, J. P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. <i>IEEE Signal Processing Letters</i> , 24(3), 279–283. https://doi.org/10.1109/LSP.2017.2657381

Table 4 continued on next page

Table 4 continued

Ref no.	Bibliography
A47.	Oweis, R. J., Abdulhay, E. W., Khayal, A., & Awad, A. (2015). An alternative respiratory sounds classification system utilizing artificial neural networks. <i>Biomedical Journal</i> , 38(2), 153–161. https://doi.org/10.4103/2319-4170.137773
A48.	Fang, S. H., Wang, C. Te, Chen, J. Y., Tsao, Y., & Lin, F. C. (2019). Combining acoustic signals and medical records to improve pathological voice classification. <i>APSIPA Transactions on Signal and Information Processing</i> , 8(2019), 1–11. https://doi.org/10.1017/ATSIP.2019.7
A49.	Wang, W., Meratnia, N., Seraj, F., & Havinga, P. J. M. (2019). Privacy-aware environmental sound classification for indoor human activity recognition. <i>ACM International Conference Proceeding Series</i> , 36–44. https://doi.org/10.1145/3316782.3321521
A50.	Kroos, C., Bones, O., Cao, Y., Harris, L., Jackson, P. J. B., Davies, W. J., Wang, W., Cox, T. J., & Plumbley, M. D. (2019). Generalization in Environmental Sound Classification: The “Making Sense of Sounds” Data Set and Challenge. <i>ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings</i> , 2019-May, 8082–8086. https://doi.org/10.1109/ICASSP.2019.8683292
A51.	Humayun, A. I., Tauhiduzzaman Khan, M., Ghaffarzadegan, S., Feng, Z., & Hasan, T. (2018). An ensemble of transfer, semi-supervised and supervised learning methods for pathological heart sound classification. <i>Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH</i> , 2018-September(i), 127–131. https://doi.org/10.21437/Interspeech.2018-2413
A52.	Colonna, J., Peet, T., Ferreira, C. A., Jorge, A. M., Gomes, E. F., & Gama, J. (2016). Automatic classification of anuran sounds using convolutional neural networks. <i>ACM International Conference Proceeding Series</i> , 20-22-July-2016, 73–78. https://doi.org/10.1145/2948992.2949016
A53.	Tschannen, M., Kramer, T., Marti, G., Heinzmann, M., & Wiatowski, T. (2016). Heart sound classification using deep structured features. <i>Computing in Cardiology</i> , 43, 565–568. https://doi.org/10.22489/cinc.2016.162-186
A54.	Yang, X., Yang, F., Gobeawan, L., Yeo, S. Y., Leng, S., Zhong, L., & Su, Y. (2016). A multi-modal classifier for heart sound recordings. <i>Computing in Cardiology</i> , 43, 1165–1168. https://doi.org/10.22489/cinc.2016.339-225
A55.	Salamon, J., & Bello, J. P. (2015). Unsupervised Feature Learning for Urban Sound Classification. <i>ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings</i> , 171–175.
A56.	Kocuvan, P., & Torkar, D. (2015). Classification of the heart auscultation signals. <i>HEALTHINF 2015 - 8th International Conference on Health Informatics, Proceedings; Part of 8th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2015</i> , 534–539. https://doi.org/10.5220/0005264005340539
A57.	Silva, P. (2012). Classification, segmentation, and chronological prediction of cinematic sound. <i>Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012</i> , 2, 369–374. https://doi.org/10.1109/ICMLA.2012.172
A58.	Kumar, D., Carvalho, P., Couceiro, R., Antunes, M., Paiva, R. P., & Henriques, J. (2010). Heart murmur classification using complexity signatures. <i>Proceedings - International Conference on Pattern Recognition</i> , 2564–2567. https://doi.org/10.1109/ICPR.2010.628
A59.	Zhu, B., Wang, C., Liu, F., Lei, J., Lu, Z., & Peng, Y. (2018). Learning Environmental Sounds with Multi-scale Convolutional Neural Network. <i>Proceedings of the International Joint Conference on Neural Networks (IJCNN)</i> , 1–8. https://doi.org/10.1109/IJCNN.2018.848964
A60.	Zhao, H., Huang, X., Liu, W., & Yang, L. (2018). Environmental sound classification based on feature fusion. <i>MATEC Web of Conferences</i> , 173, 1–5. https://doi.org/10.1051/mateconf/201817303059

Table 4 continued on next page

Table 4 continued

Ref no.	Bibliography
A61.	Hu, W., Lv, J., Liu, D., & Chen, Y. (2018). Unsupervised Feature Learning for Heart Sounds Classification Using Autoencoder. <i>Journal of Physics: Conference Series</i> , 1004(1). https://doi.org/10.1088/1742-6596/1004/1/012002
A62.	Bisot, V., Serizel, R., Essid, S., & Richard, G. (2017). Leveraging deep neural networks with nonnegative representations for improved environmental sound classification. <i>IEEE International Workshop on Machine Learning for Signal Processing, MLSP, 2017-September</i> , 1–6. https://doi.org/10.1109/MLSP.2017.8168139
A63.	Amiriparian, S., Gerczuk, M., Ottl, S., Cummins, N., Freitag, M., Pugachevskiy, S., Baird, A., & Schuller, B. (2017). Snore sound classification using image-based deep spectrum features. <i>Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2017-August</i> , 3512–3516. https://doi.org/10.21437/Interspeech.2017-434
A64.	Boddapati, V., Petef, A., Rasmusson, J., & Lundberg, L. (2017). Classifying environmental sounds using image recognition networks. <i>Procedia Computer Science</i> , 112, 2048–2056. https://doi.org/10.1016/j.procs.2017.08.250
A65.	Medhat, F., Chesmore, D., & Robinson, J. (2020). Masked Conditional Neural Networks for sound classification. <i>Applied Soft Computing Journal</i> , 90(608014), 1–13. https://doi.org/10.1016/j.asoc.2020.106073
A66.	Nogueira, D. M., Ferreira, C. A., Gomes, E. F., & Jorge, A. M. (2019). Classifying Heart Sounds Using Images of Motifs, MFCC, and Temporal Features. <i>Journal of Medical Systems</i> , 43(6), 186–203. https://doi.org/10.1007/s10916-019-1286-5
A67.	Kumar, D., Carvalho, P., Antunes, M., Paiva, R. P., & Henriques, J. (2010). Heart murmur classification with feature selection. <i>2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10, June 2014</i> , 4566–4569. https://doi.org/10.1109/IEMBS.2010.5625940
A68.	Vahabi, N., & Selviah, D. R. (2019). Convolutional Neural Networks to Classify Oil, Water, and Gas Wells Fluid Using Acoustic Signals. <i>2019 IEEE 19th International Symposium on Signal Processing and Information Technology, ISSPIT 2019</i> . https://doi.org/10.1109/ISSPIT47144.2019.9001845

Akon O. Ekpezu is a Lecturer in the Department of Computer Science, Cross River University of Technology (CRUTECH), Nigeria. She holds; a Bachelor of Science (B.Sc) in Mathematics and Statistics from the University of Calabar, Nigeria, a Post Graduate Diploma (PGD) in Computer Science from the same university, a Master of Science (M.Sc.) in Information Technology from the National Open University of Nigeria (NOUN) and a Master of Philosophy (MPhil) in Computer Science from the University of Ghana. She is currently pursuing a PhD in Information Processing Science, University of Oulu, Finland. She is interested in the following areas of research; Persuasive Systems, Behavior Change Support Systems, Machine Learning and Information Security.

Winfred Yaokumah is a researcher, cyber security expert and senior faculty at the Department of Computer Science of the University of Ghana. His work appears in several reputable journals including *Information and Computer Security*, *International Journal of Distributed Artificial Intelligence*, *Journal of Information Technology Research*, *Information Resources Management Journal*, *IEEE Xplore*, *International Journal of e-Business Research*, and *International Journal of Enterprise Information Systems*. He is an editor of the *Modern Theories and Practices for Cyber Ethics and Security Compliance*. His research interest includes Cyber Security, Machine Learning, Network Security, and Information Systems Security. He also serves on an International Review Board for the *International Journal of Technology Diffusion*.