

UNIVERSITY OF GHANA

COLLEGE OF BASIC AND APPLIED SCIENCE



A DEEP LEARNING APPROACH FOR THE AUTOMATIC CLASSIFICATION OF ACOUSTIC
EVENTS: A CASE OF NATURAL DISASTERS

BY

EKPEZU, AKON OBU

(10704369)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF MPhil IN
COMPUTER SCIENCE DEGREE

INTEGRI PROCEDAMUS

DEPARTMENT OF COMPUTER SCIENCE

October 2020

DECLARATION

I hereby declare that I am the sole author of this thesis and all materials used from other sources or in collaboration with other researchers have been properly and fully acknowledged.

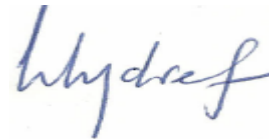


EKPEZU, AKON OBU

(Candidate)



DR. FERDINAND KATRIKU
(Supervisor)



DR. WINFRED YAOKUMAH
(Co-supervisor)

ABSTRACT

Automatic classification of acoustic events is a signal processing activity that has recently gained research interest, especially in the machine learning community. This is due to its cost-effectiveness in the long-term monitoring of larger areas and the collection of large amounts of data in real-time. A plethora of techniques have been proposed and adopted for the classification of acoustic events such as respiratory sound, animal calls/vocalizations, baby cry, speech disorders, and environmental sound. This study was aimed at developing a natural disaster sound classification model that will enable automatic classification of natural disasters. Accordingly, deep learning techniques including Convolutional Neural Network (CNN) and a Long short-term memory based-Recurrent Neural Network (RNN-LSTM) were used to develop classification models. The adopted algorithms and sound features used in this study were motivated by methodologies used in the area of speech/voice recognition. To ensure a relevant and rigorous research, this study adopted the design science research methodology which consisted of a five-phase cycle; awareness of the problem, suggestion, development, evaluation, and conclusion. Furthermore, to also ensure the real-time classification of natural disaster sounds, the detection-by-classification approach was adopted instead of detection-and-classification. The dataset used for this study consisted of five classes of natural disasters sound that was extracted from the Freesound database. The sound files were preprocessed at 16000Hz to extract 13 Mel Frequency Cepstral Coefficient (MFCC). An arbitrary time frame of 0.1s was adopted. In the end, the performance of both models was validated using the classification metrics and cross-validation. Results indicated that although CNN performed slightly better than RNN-LSTM, both models were effective at automatically discerning one disaster sound from the other in real-time. Best results of 99.95% in classification accuracy, and 0.999 in the area under the curve (AUC) score were obtained from CNN.

ACKNOWLEDGEMENT

But by the grace of God, I am what I am, and his grace toward me was not in vain. For seeing me through this phase of studies; all thanks and adoration to the Lord God almighty for his unfailing love, grace, and mercies.

I particularly thank my supervisors; Dr. Ferdinand Katsriku and Dr. Winfred Yaokumah for their wholehearted encouragement, diligent guidance, and for dutifully and meticulously guiding this thesis. Also, for the many research opportunities and experiences they have exposed me to throughout my MPhil program, may God bless you abundantly.

My appreciation also extends to Dr. Jamal Abdullahi, Dr. Isaac Wiafe, Dr. Solomon Mensah, Dr. Justice Appatti, and all the faculty members of the Department of Computer Science, University of Ghana; you have given my postgraduate journey a new and unforgettable experience. Remain blessed.

I also acknowledge with special thanks, Dr. Enoimah Umoh, and all the faculty members of the Department of Computer Science, Cross River University of Technology; their unwavering support has made this academic pursuit a success.

I am profoundly grateful to my mother Mrs. Atim Mbah, my siblings Eke and Obu Ekpezu, and the families of Dr. Emil Inyang and Dr. Isaac Wiafe whose encouragement, prayers, love, and sacrifice kept me pushing forward. May God in his infinite mercies bless you all.

To my colleagues William Apprey, Samuel Abedu, Fredrick Boafo, Abigail Wiafe, Jacqueline Kumi, and Melody Kakrabah; your immeasurable contributions and support have been amazing.

May the grace and blessings of God rest and abide with you all. Amen.

RELATED PUBLICATION

Ekpezu, A. O., Katsriku, F., Yaokumah W., and Wiafe I., **[under review]** Classification of Acoustic Signals using machine learning: A Systematic Review. Submitted to Journal of Artificial Intelligence and Soft Computing Research.

TABLE OF CONTENTS

<i>DECLARATION</i>	<i>ii</i>
<i>ABSTRACT</i>	<i>iii</i>
<i>ACKNOWLEDGEMENT</i>	<i>iv</i>
<i>RELATED PUBLICATION</i>	<i>v</i>
<i>TABLE OF CONTENTS</i>	<i>vi</i>
<i>LIST OF FIGURES</i>	<i>ix</i>
<i>LIST OF TABLES</i>	<i>x</i>
<i>LIST OF ABBREVIATIONS</i>	<i>xi</i>
<i>Chapter One INTRODUCTION</i>	<i>1</i>
1.1 BACKGROUND AND MOTIVATION	1
1.2 RESEARCH PROBLEM.....	3
1.3 RESEARCH AIM AND OBJECTIVES.....	6
1.4 EXPECTED CONTRIBUTIONS	6
1.4.1 THEORETICAL CONTRIBUTION	7
1.4.2 PRACTICAL CONTRIBUTION.....	7
1.5 THESIS OUTLINE.....	7
<i>Chapter Two RELATED STUDIES</i>	<i>9</i>
2.1 CHAPTER OVERVIEW	9
2.2 NATURAL DISASTERS	9
2.3 REVIEW SUMMARY ON RELATED WORK.....	10
2.3.1 TASK CATEGORIES	10
2.3.2 MODELLING TECHNIQUES USED	20
2.4 REVIEW SUMMARY	21
2.5 CHAPTER SUMMARY	23
<i>Chapter Three STATE OF THE ART IN SOUND CLASSIFICATION</i>	<i>25</i>
3.1 CHAPTER OVERVIEW	25
3.2 RELATED REVIEWS.....	25
3.3 REVIEW QUESTIONS	26
3.4 REVIEW APPROACH.....	27
3.4.1 LITERATURE SEARCH	27
3.4.2 INCLUSION AND EXCLUSION CRITERIA	28
3.4.3 STUDY SELECTION AND DATA EXTRACTION	29
3.5 OVERVIEW OF PUBLICATION TRENDS.....	30
3.5.1 PUBLICATION FREQUENCY	30
3.5.2 DISTRIBUTION OF JOURNALS	31

3.5.3	AUTHORS AND COUNTRY ORIGIN.....	32
3.6	SUMMARY OF METHODOLOGIES FROM REVIEWED ARTICLES.....	35
3.6.1	SOUND/ACOUSTIC SIGNALS CLASSIFIED AND DATA SOURCES.....	35
3.6.2	DISTRIBUTION OF CLASSIFIED SOUNDS ACCORDING TO APPLICATION DOMAIN.....	47
3.6.3	FEATURE EXTRACTION METHODS.....	49
3.6.4	SOUND CLASSIFICATION ALGORITHMS AND PERFORMANCE METRICS.....	53
3.7	REVIEW SUMMARY	59
3.8	CHAPTER SUMMARY	62
<i>Chapter Four</i>	<i>RESEARCH METHODOLOGY</i>	<i>64</i>
4.1	CHAPTER OVERVIEW	64
4.2	DESIGN SCIENCE RESEARCH METHODOLOGY (DSRM).....	64
4.3	RESEARCH APPROACH FOR THIS STUDY	65
4.3.1	AWARENESS OF THE PROBLEM.....	66
4.3.2	SUGGESTION	66
4.3.3	DEVELOPMENT.....	66
4.3.4	EVALUATION	71
4.3.5	CONCLUSION.....	72
4.4	CHAPTER SUMMARY	72
<i>Chapter Five</i>	<i>USING DEEP LEARNING FOR ACOUSTIC EVENT CLASSIFICATION</i>	<i>73</i>
5.1	CHAPTER OVERVIEW	73
5.2	SOFTWARE USED FOR THE EXPERIMENT	74
5.3	NATURAL DISASTER SOUND DATASET	74
5.4	SOUND PREPROCESSING AND FEATURE EXTRACTION	75
5.4.1	DE-NOISING THE SIGNAL.....	77
5.4.2	ACOUSTIC DOWN-SAMPLING	78
5.4.3	FILTER BANK-BASED FEATURE EXTRACTION METHOD.....	78
5.5	CLASSIFICATION TECHNIQUES.....	79
5.5.1	CONVOLUTIONAL NEURAL NETWORK.....	80
5.5.2	RECURRENT NEURAL NETWORK.....	82
5.6	CHAPTER SUMMARY	84
<i>Chapter Six</i>	<i>EVALUATION OF DEEP LEARNING TECHNIQUES.....</i>	<i>85</i>
6.1	CHAPTER OVERVIEW	85
6.2	MODEL VALIDATION.....	85
6.2.1	CROSS-VALIDATION.....	85
6.2.2	CLASSIFICATION METRICS.....	88
6.3	TESTING THE VALIDITY OF THE MODELS IN REAL-TIME CLASSIFICATION OF DISASTER SOUNDS.....	93
6.4	CHAPTER SUMMARY	95
<i>Chapter Seven</i>	<i>CONCLUSION.....</i>	<i>97</i>
7.1	CHAPTER OVERVIEW	97
7.2	THESIS SUMMARY	97
7.3	DISCUSSIONS	100

7.3.1	CLASSIFICATION CATEGORY	100
7.3.2	INPUT ACOUSTIC FEATURES.....	101
7.3.3	CLASSIFICATION PERFORMANCE	102
7.4	LIMITATION OF THE STUDY	104
7.4.1	DATASETS	104
7.4.2	DENOISING THE SIGNAL	104
7.4.3	FEATURE EXTRACTION	104
7.5	RECOMMENDATION.....	105
7.6	FUTURE WORK.....	105
<i>References:.....</i>		<i>106</i>
<i>APPENDIX A: PRIMARY STUDIES USED FOR THE SYSTEMATIC REVIEW.....</i>		<i>126</i>
<i>APPENDIX B: PYTHON CODES FOR LOADING THE DATA.....</i>		<i>130</i>
<i>APPENDIX C: PYTHON CODES FOR MODEL PREPARATION/PREDICTION</i>		<i>132</i>
<i>APPENDIX D: PYTHON CODES FOR 10-FOLD MODEL VALIDATION.....</i>		<i>134</i>
<i>APPENDIX E: PYTHON CODES FOR AUC-ROC</i>		<i>134</i>

LIST OF FIGURES

FIGURE 2.1: NATURAL DISASTER EVENTS GLOBALLY FROM 2010 TO 2019 (DUGGAR ET AL. 2016)	10
FIGURE 2.2: DISTRIBUTION OF MODELLING TECHNIQUES	20
FIGURE 3.1: PUBLICATIONS BY YEAR	31
FIGURE 3.2: DISTRIBUTION OF AUTHORS BY CONTINENT	34
FIGURE 3.3: SOUND CLASSIFICATION PUBLICATION TREND BY STUDY COUNTRY	35
FIGURE 3.4: PIE CHART SHOWING THE DISTRIBUTION OF APPLICATION DOMAINS	48
FIGURE 3.5: DISTRIBUTION OF MODELLING TECHNIQUES	58
FIGURE 4.1: LAYERS OF CNN (BORGNE & BONTEMPI, 2017)	67
FIGURE 5.1: SOUND CLASSIFICATION ARCHITECTURE.....	73
FIGURE 5.2: CLASS DISTRIBUTION OF DISASTER SOUND DATASET	75
FIGURE 5.3: TIME SERIES REPRESENTATION OF FIVE RANDOM SAMPLES BELONGING TO THE FIVE DIFFERENT CLASSES OF THE DATASET	76
FIGURE 5.4: MFCC REPRESENTATION OF FIVE RANDOM SAMPLES BELONGING TO THE FIVE DIFFERENT CLASSES OF THE DATASET.....	76
FIGURE 5.5: FILTER BANK COEFFICIENT REPRESENTATION OF FIVE RANDOM SAMPLES BELONGING TO THE FIVE DIFFERENT CLASSES OF THE DATASET	77
FIGURE 5.6: PIE-CHART SHOWING THE CLASS DISTRIBUTION OF DENOISED DISASTER SOUND DATASET	78
FIGURE 5.7: CNN MODEL DIMENSIONS.....	81
FIGURE 5.8: RNN-LSTM ARCHITECTURE (RAZA ET AL. 2019).....	82
FIGURE 5.9: RNN-LSTM MODEL DIMENSIONS	83
FIGURE 6.1: CLASSIFICATION ACCURACY AND AVERAGE ACCURACY OF THE 10-FOLDS	87
FIGURE 6.2: ACCURACY OBTAINED FROM 10-FOLD CROSS-VALIDATION FOR CNN AND RNN-LSTM	87
FIGURE 6.3: CLASSIFICATION ACCURACY FOR CNN AND RNN-LSTM	91
FIGURE 6.4: AUC-ROC FOR CNN MODEL.....	92
FIGURE 6.5: AUC-ROC FOR RNN-LSTM MODEL	93
FIGURE 6.6: CHART SHOWING ACCURACY SCORE COMPARISON FOR INITIAL AND INCREASED TIME FRAMES.....	93
FIGURE 6.7: CHART SHOWING THE CLASSIFICATION ACCURACY FOR THE REAL TIME MODEL (2x) AND AUGMENTED DATASET (4x, 6x).	95

LIST OF TABLES

TABLE 2.1: DISASTER TYPE, MODELLING TECHNIQUES, AND TASK SUMMARY.....	13
TABLE 2.2: CLASSIFICATION SCHEMES FOR NATURAL DISASTERS	22
TABLE 3.1: RESEARCH QUESTIONS AND OBJECTIVES	26
TABLE 3.2: SEARCH RESULTS PER EXCLUSION CRITERIA	30
TABLE 3.3: FREQUENCY DISTRIBUTION OF PRIMARY SOURCES	32
TABLE 3.4: LEADING AUTHORS	33
TABLE 3.5: SUMMARY OF CLASSIFIED SOUNDS AND DATASETS	38
TABLE 3.6: FEATURE EXTRACTION METHODS	49
TABLE 3.7: CLASSIFICATION TECHNIQUES USED	53
TABLE 3.8: CLASSIFICATION CATEGORIES.....	61
TABLE 4.1: DESIGN SCIENCE RESEARCH (DSR) GUIDELINES	65
TABLE 4.2: FEATURES OF THE 3-LAYERS IN A CNN	69
TABLE 4.3: CALCULATING THE CURRENT STATE, ACTIVATION FUNCTIONS AND OUTPUT IN RNN	70
TABLE 6.1: 10-FOLD CROSS-VALIDATION.....	86
TABLE 6.2: CONFUSION MATRIX SHOWING CNN PREDICTIONS	90
TABLE 6.3: CONFUSION MATRIX SHOWING RNN-LSTM PREDICTIONS.....	90
TABLE 6.4: RESULT SUMMARY OF CLASSIFICATION METRICS	96
TABLE 7.1: COMPARISON OF STUDY APPROACHES WITH OTHER STUDIES.....	103

LIST OF ABBREVIATIONS

ABBREVIATION	FULL MEANING
AI	Artificial Intelligence
AEC	Acoustic event classification
AED	Acoustic event detection
ANN	Artificial Neural Network
ARMAX	Auto Regressive Moving Average with Exogenous Inputs
ASA	Acoustical society of America
ASR	Automatic speech recognition
CNN	Convolutional Neural Network
DCT	Discrete cosine transform
DL	Deep learning
DML	Deep metric learning
DSRM	Design science research methodology
DWT	Discrete wavelet transform
ELM	Extreme learning machine
FBC	Filter bank coefficient
FCN	Fully connected network
FFT	Fast Fourier transform
FT	Fourier transform
HMM	Hidden markov model
IoT	Internet of things
JASA	Journal of Acoustical society of America
kHz	Kilo hertz
KNN	k-Nearest Neighbor
LSTM	Long short-term memory
MFCC	Mel Frequency Cepstral Coefficient
ML	Machine learning
MLP	Multilayer Perceptron
NLP	Natural language processing
RBF	Radial Basis Function
RF	Random forest
RNN	Recurrent Neural Network
STFT	Short time Fourier transform
SVM	Support Vector Machine
.WAV	Waveform Audio

Chapter One

INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Every object either animate or inanimate produces sound in its vibrating state (Giordano, 2005; Rubinstein, 2008). Although it varies depending on seasons, time, geographic location as well as propagation medium, sound is considered as one of the most significant signals used to monitor and detect changes in the environment. However, due to varying parameters, detecting a sound of interest in a particular environment is most times challenging because the sound of interest is usually immersed in different forms of background noises such as anthrophony (man-made e.g. traffic, shipping, and aircraft noise), geophony (environmental e.g. windstorm, raindrops, thunderstorms), and biophony (animal-made e.g. dog barking, vocalizations from marine mammals) (Muir & Bradley, 2016). Additionally, in distinguishing one sound type from the other, human operations are trained to use expert knowledge; this is an overwhelming and inadequate process as false alarms are raised in most cases (Cao et al. 2017).

Consequently, the need for automatic active sound classification began to gain research interest. Generally, automatic sound classification (ASC) entails the automatic identification of ambient sound in an environment. It has so far been applicable in the domains such as disease diagnosis (Aykanat et al. 2017; Chen et al. 2019), voice classification (Fang et al. 2019), speech recognition (Turner & Joseph, 2015), bioacoustics (Kim et al. 2018; Luque et al. 2018), and action detection (Aziz et al. 2019; Maccagno et al. 2019; Salamon & Bello, 2017).

This study is focused on the automatic acoustic classification of natural disasters as a complement for vision-based classification of natural disasters. Acoustic classification instead of vision-based classification because; natural disasters are a result of seismic activities in the

ocean, these seismic activities produce layers of wave-like pulses that are invisible and inaudible to human eyes and ears yet observable as sound waves (Mone, 2007; Perlman, 2013). Accordingly, these sound waves can be recorded, monitored, and used to provide early warning of an upcoming seismic event (Monroe-Kane, 2019). It is important to note that this study is concerned with only those natural disasters which are a result of seismic activities.

Developing an automatic acoustic classification technique (model) requires the selection of appropriate acoustic features (Aziz et al. 2019), as well as a robust classification algorithm (Luque, Romero-Lemos, Carrasco, & Gonzalez-Abril, 2018). A robust classification algorithm in this context is a classifier that can distinguish sounds that belong to distinct classes of the feature space. Sound classifiers are broadly classified as discriminative and non-discriminative (Mitilineos et al. 2018). While the former entails modelling the decision boundary in the training data and matching its test input to a specific data class, the latter attempts to explicitly model the actual distribution of each class (Chu et al. 2009; Mitilineos et al. 2018). Examples of discriminative classifiers include logistic regression, nearest neighbor, k-means, support vector machines, and traditional neural networks such as the multilayer perceptron. Non-discriminative classifiers include Naïve Bayes, Markov random fields, Hidden Markov Models, (HMM) and, Bayesian networks.

In this study, a discriminative classifier such as a neural network will be adopted for the classification task. Neural networks such as convolutional neural networks (CNNs) and recurrent neural networks (RNN) are well-known classifiers that have been efficient in a wide variety of practical applications (Binkhonain & Zhao, 2019). While CNN is known for its high-performance accuracy in image classification and recognition (Arel et al. 2010; Epelbaum, 2017), RNN is known for its efficiency in processing sequential and time-series information (Arel et al. 2010; Pouyanfar et al. 2018). Additionally, these two classifiers have been shown

to perform excellently in signal and speech/sound classification (Lecun et al. 2015; Mitilneos et al. 2018). This study will use these two neural networks in the automatic acoustic classification of natural disasters.

1.2 RESEARCH PROBLEM

Natural disasters such as tsunamis, volcanoes, hurricanes, and earthquakes are powerful events with an infrasonic signature and low frequencies that are inaudible to the human ear (Gopaldaswami, 2018; Perlman, 2013). With the low frequency of sound from the movement of the earth floor, it is no surprise why there are challenges in the early detection of natural disasters (Mone, 2007; Perlman, 2013). Consequently, the yearly cost of natural disasters is expensive in terms of financial loss and loss of human lives. It is estimated that annually ninety thousand (90000) people lose their lives and nearly 160 million are also affected as a result of natural disasters (Tobergte & Curtis, 2013). For instance, in the year 2018, the financial cost of natural disasters was estimated to be in the region of \$91 billion in the US alone (Chappell, 2019).

While meteorological scientists and researchers are looking for better techniques to combat these disasters, studies have shown that natural disasters cannot be stopped (Goswami et al. 2018; Khalaf et al. 2018; Wallemacq, 2015). However, being able to predict well in advance its occurrence and to differentiate between various kinds of disasters and their severity can mitigate their impact (Khalaf et al. 2018). According to Chen et al. (2013), the damage caused by natural disasters can be reduced significantly if information technology tools such as remote sensing, and satellite data are employed. Okamoto et al. (2018) also add that images obtained from surveillance cameras can be used to automatically assign disaster names to the disaster prevention networks. Conversely, Panagiota et al. (2011) argue that the use of images in pre-disaster imagery most times prevents accurate change detection. On the other hand, remote

sensing methods are limited by cost-effective availability and temporal delays of about 48 to 72 hours before information can be produced (Resch et al. 2018). According to Wisner & Adams (2002), these temporal delays in information dissemination increases the level of vulnerability regarding the physical and environmental security. Studies have shown that natural disasters disrupt measures taken to protect buildings, systems, and business operations (Tierney, 2019).

With the resurgence of interest in the use of Artificial Intelligence (AI), researchers have found that it can be used to manage, predict, or detect disasters. However, since the AI techniques like machine learning algorithms are based on data from records, it is difficult for artificial intelligence to predict long-term trends of various natural disasters (Joshi, 2019). More particularly, predictions are most times inaccurate, underestimated or overestimated due to discrepancies in the data used for the prediction (Joshi, 2019); typical cases include Indonesia's Tsunami early warning system (Singhvi et al. 2018), false earthquake warning in Japan (BBC News, 2018), and the case of the California earthquake warning which was sent 92 years late (<https://www.bbc.com/news/technology-40366816>, 2017). Additionally, existing natural disaster detection systems are either unable to identify an event (natural disaster) in real-time or unable to send early warning signals due to the unavailability of a reliable detection and alert system that runs 24 hours a day (UNDP, 2012). A typical case of an ineffective alert system is the 2010 earthquake in Maule, Chile; no warnings or evacuation plans were made even after three hours of the tsunamic hitting the Chilean coasts (Soule, 2014).

Machine learning algorithms used in several studies to classify natural disasters have used images, videos, text, and numerical data for the analysis. In all these, the damaging effects of natural disasters still abound. Current approaches for managing these disasters are either insufficient or inefficient (Boustan et al. 2017; Duggar et al. 2016; Evans, 2011).

Given that:

- the temperature of the ocean determines climate and wind patterns which in turn affects life on land and the ecosystem (Domingo, 2012),
- disasters like hurricanes, tsunamis, earthquake, tornadoes are as a result of seismic activities in the ocean,
- most of these disasters produce sound during formation, even though some of the produced sounds are below the range of human hearing, with a sampling rate of 2.5 to 5kHz (Mone, 2007),
- sound travels faster in the water at a speed of 1500m/s compared to the speed of 353m/s in air, this implies that water is a favorable environment for sound propagation (Aziz et al. 2019; Stojanovic & Beaujean, 2016).
- the use of satellite or airborne images is limited due to the inability of light to travel beyond shallow water depth as well as access underwater information (Domingo, 2012; Hassiotis, 2018).

This study proposes the use of acoustic signals/sound instead of images, video, text, or numerical data to differentiate one disaster type from another for the following reasons. Using a sound-based approach is less invasive, inexpensive, and it allows long-term monitoring of large areas and the collection of large amounts of data in real-time (Chen et al. 2019; Malfante et al. 2018). Sound classification is more reliable compared to image and video classification because it is not affected by variations in light intensity (Aziz et al. 2019). Additionally, the wide-angle camera lenses used in computer vision are not as omnidirectional as the microphones used for sound (Aziz et al. 2019; Mitilineos et al. 2018).

1.3 RESEARCH AIM AND OBJECTIVES

Putting the above-mentioned factors (see section 1.2) into consideration, this thesis aims to develop an automatic sound classification model for natural disasters. The disaster sound classification model will enable the automatic classification of these disasters amid ambient noise as well as sounds below the human hearing range. Particularly, this study will use deep learning techniques to build a sound classification model for the automatic classification of natural disasters. The specific objectives of this study include:

- i. To explore existing literature and identify the approaches and methods used for managing and detecting natural disasters.
- ii. To explore literature and identify state-of-the-arts in the classification of acoustic signals/sounds (events) in various application domains that use AI techniques.
- iii. To develop a natural disaster sound classification model using deep learning techniques.
- iv. To analyze diversified sound features as well as evaluate the models using existing model validation techniques.

1.4 EXPECTED CONTRIBUTIONS

The outcome of this study is expected to make key contributions to the fields of Internet of Things (IoT), intelligent systems, environmental monitoring, natural disaster detection, and acoustic signal processing. Since approximately 71% of the earth's surface is covered by the ocean, the expected contribution of this study to IoT also extends to Internet of Underwater Things (IoUT) and its applications such as underwater exploration, disaster prevention, and military surveillance.

1.4.1 THEORETICAL CONTRIBUTION

This thesis explores existing literature in natural disasters, sound/acoustic signals, and artificial intelligence. It provides a theoretical background on the relationship between acoustic signals and natural sounds in the environment. It also highlights how machine learning and deep learning techniques are used to distinguish one sound type from the other.

It therefore, proposes an acoustic model that will be useful in the early detection of natural disasters. More particularly, the proposed model is expected to simplify natural disaster detection processes and the detection of natural sounds in general for both researchers and practitioners.

1.4.2 PRACTICAL CONTRIBUTION

The developed model is expected to facilitate the early and automatic classification of natural disasters. The automatic classification of natural disasters using sound/acoustic signals may also be useful for environmental monitoring, disaster detection, and consequently mitigate the massive loss of life and properties caused by natural disasters.

Furthermore, the developed model is expected to facilitate acoustic sensing in IoT devices. This is because Internet of things (IoT), as an enabling technology for remote monitoring requires a clear sound detection and identification system that is capable of effectively sensing and analyzing environmental and natural sounds.

1.5 THESIS OUTLINE

The thesis is organized as follows. Chapter one presents the background of the research by briefly stating the research problem, aim, and objectives. The chapter also highlights the expected practical and theoretical contributions of the research to knowledge. An outline of the thesis structure is also presented in this chapter.

Chapter two provides review summaries on studies related to artificial intelligence methods of mitigating the effects of natural disasters. It provides an extensive literature study on current techniques for disaster detection, prevention, and management concerning various types of disasters.

Chapter three presents state of the art in the area of general sound classification, and also reports feature extraction and classification techniques that exist in literature.

Chapter four presents the research methodology. More particularly, it presents the design science research methodology (DSRM) and how this study fits into it.

Chapter five will report the steps used in developing the model and conducting the classification tasks.

Chapter six will evaluate the performance of the developed models as well as provide a comparison of the deep learning techniques used in classifying natural disasters sound. While Chapter seven will provide a summary of the thesis.

Chapter Two

RELATED STUDIES

2.1 CHAPTER OVERVIEW

This chapter provides a review summary of studies on natural disasters mitigation schemes using artificial intelligence (AI). More particularly, it identifies and summarizes the different measures proposed by researchers to either detect, prevent, or manage natural disasters.

2.2 NATURAL DISASTERS

Every year around the world, there has been significant damage to properties and animal life, as well as the loss of thousands of human lives, this is due to various forms of natural disasters. Natural disasters can be classified into three main categories; those caused by movements of the earth otherwise known as geophysical events (earthquake, tsunamis, volcanic eruptions), weather-related disasters (hurricanes, tornadoes, extreme heat or cold) and others (floods, landslides, famine) (Evans, 2011).

Although meteorologists, environmental scientists, computer scientists, and researchers have put in a lot of work to predict, detect, and manage these disasters, their effects still abound. For instance, between the years 2010 to 2019, there has been an estimated total of 7342 natural disaster events globally, (see figure 2.1). The most disaster-prone areas have been coastal regions and the most affected people are the low-income earners (Boustan et al. 2017).

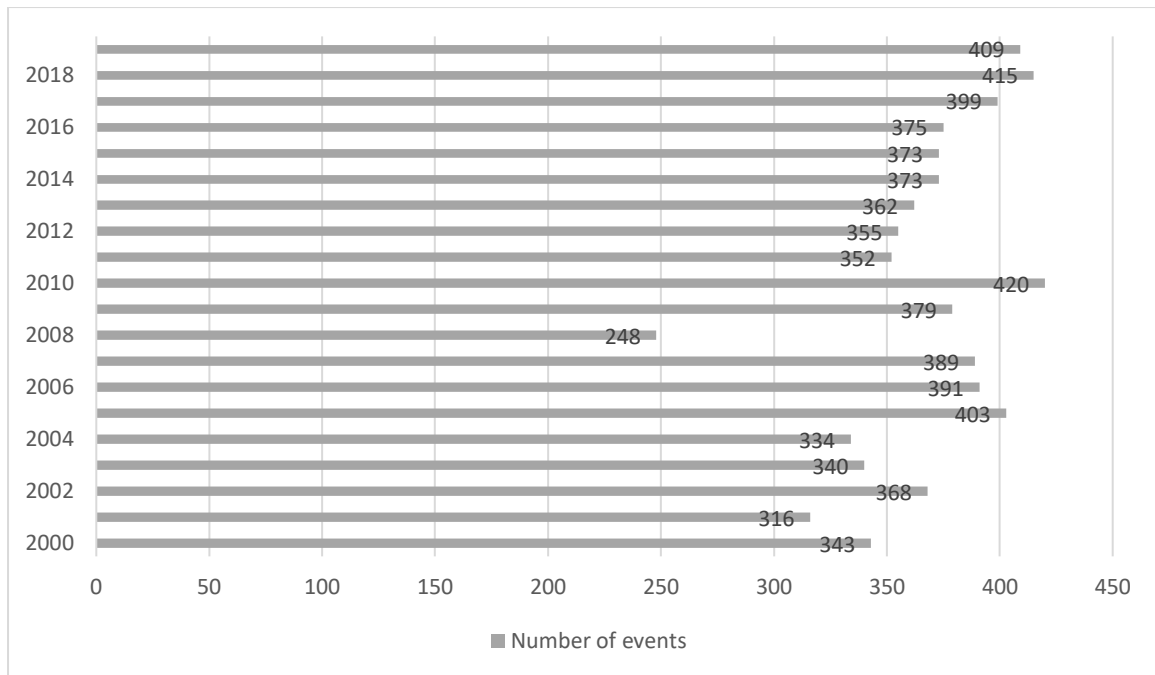


Figure 0.1: Natural disaster events globally from 2010 to 2019 (Duggar et al. 2016)

2.3 REVIEW SUMMARY ON RELATED WORK

In this section, a summary of existing literature on the current methodologies for managing natural disasters using artificial intelligence will be provided. The findings are summarized in Table 2.1.

In Table 2.1, different research trends, and techniques in combating natural disasters have been summarized into task categories, objectives of the tasks, modeling techniques, type and sources of data. The ensuing sections will further elaborate on the findings summarized in Table 2.1.

2.3.1 TASK CATEGORIES

From Table 2.1, three major categories of tasks for managing natural disasters can be identified. They include (i) prediction, (ii) detection, and (iii) disaster recovery and management strategies.

PREDICTION: Predicting a natural disaster is an ideal solution to mitigating the effects of natural disasters. It involves forecasting the type, time, place, and magnitude of a disaster and is commonly based on data gathered from past occurrences, disaster-prone areas as well as the attributes of the disaster. According to Khalaf et al. (2018) using a predictive classification

approach assists in tackling the severity of a natural disaster and it depends on the features identified from the available datasets. Chen et al. (2017) proposed a hybrid model of rotation forest ensembles and naïve Bayes tree classifiers that can improve the accuracy of a disaster predictive model. Kim et al. (2017) on the other hand developed a smart-eye platform for disaster recognition and response.

However, Goswami et al. (2018) argue that although the disaster-prone areas can be predicted, combating natural disasters cannot be solved with available data and techniques. Conversely, Kim et al. (2017) posit that prediction aims to reduce the damage caused by disasters by using data from past disasters to recognize current disaster situations.

DETECTION: Detection involves promptly detecting a disaster as soon as it occurs. According to Gupta & Doshi, (2018), the primary objective of disaster detection is to reduce the level of damage and destruction on lives and properties, hence it should be fast and accurate. Traditionally, the task of natural disaster detection is done by meteorological observatories, however, the news of detection with exact spots and repositories takes a long time to get to the appropriate authorities (Goswami et al. 2018). Consequently, the demand for real-time situation reports in disaster situations has birthed the need for the development of disaster detection methods and systems (Wieland et al. 2016). Some of the flood detection systems found in literature include:

- i. an alert generating system for flood detection using sensors technology, particularly global communication and mobile system modems (Khalaf et al. 2015),
- ii. Disaster and Agriculture Sentinel Applications (DiAS) for remote sensing and a processing chain for the analysis of Sentinel data towards flood detection (Doxani et al. 2019),
- iii. a sensing device that can monitor and detect flash floods, pluviometry, water presence and water level (Mousa et al. 2016),

iv. a smart Automatic Warning System (AWS) that uses an automatic water level recorder (AWLR) sensor. The working mechanism of the AWLR sensors is based on cognitive artificial intelligence (CAI) (Asnaning & Putra, 2018),

v. SENDI (System for detecting and forecasting Natural Disaster based on IoT); a fault-tolerant system based on IoT, ML, and wireless sensor network (WSN) for the detection and forecasting of natural disasters (Furquim et al. 2018).

vi. An automatic image-based natural disaster naming system that uses AI (Okamoto et al. 2018).

DISASTER RECOVERY AND MANAGEMENT: Disaster recovery and management is made up of three phases; awareness/early warning, response, and post-disaster assessments (Tarasconi et al. 2017). It involves response and rescue activities in addition to communication measures that ensure prompt identification and support for casualties (Goswami et al. 2018). According to Hwang et al. (2018), communication plays a key role in the survival of victims during and after disasters. Data from summaries in Table 2.1 indicates that more studies tend to focus on response and post-disaster assessments (i.e. investigating the recovery and monitoring mechanisms), rather than awareness/early warning. Some of the post-disaster systems developed include;

- i. a web-based knowledge system for emergency preparedness and response (Sermet & Demir, 2018).
- ii. a flood monitoring system based on computer vision where the uploaded images are analyzed using deep learning algorithms (Vallimeena et al. 2018).
- iii. an alert based system that uses WSN to sense environmental changes in temperature (Gupta & Doshi, 2018).

Table 0.1: Disaster type, modelling techniques, and task summary

Task	Objective	Modelling Technique Used	Source of Data	Type of Data
Flood				
Prediction	To build a model for the prediction of flood severity (Khalaf et al. 2018)	<i>Random forest</i> , Artificial Neural Network (ANN), Levenberg-Marquardt learning algorithms (LEVNN), Support Vector Machine (SVM)	Meteorological data	Numerical flood data
Detection	To develop a flood alert generating system (Khalaf et al. 2015)	Random forest, bagging, decision	Historical data from the environment agency website, UK	Numerical flood data

		tree, and hyper pipes algorithms		
Disaster recovery & management strategies	To develop an intelligent system designed to improve societal preparedness for flood (Sermet & Demir, 2018)	Natural language processing (NLP) techniques	-	Ontology based data
Detection	To introduce a processing chain for the analysis of Sentinel data towards flood detection (Doxani et al. 2019)	Decision support system	Sentinel and Synthetic aperture radar (SAR) data	SAR and optical Images
Detection	To develop a sensing device for water level detection (Mousa et al. 2016)	ANN, ARMAX	8F48 and D3CB sensors	raw water level measurement data
Detection	To develop a smart warning system (AWS) (Asnaning & Putra, 2018)	Cognitive artificial intelligence	Automatic water level recorder (AWLR) sensor	raw water level measurement data recorded spreadsheets

Disaster recovery & management strategies	To develop a computer vision-based algorithm for flood depth estimation (Vallimeena et al. 2018)	CNN	Smartphones	Raw crowdsourced Images
Detection and Prediction	To develop a fault-tolerant system for forecasting and issuing warnings of natural disasters-based on IoT (Furquim et al. 2018)	Multilayer Perceptron (MLP)	IP-based (Internet Protocol) sensor networks	Raw numerical data collected from rivers
Prediction	To produce landslide susceptibility maps (LSM) for the planning and management of areas vulnerable to landslides (Chen et al. 2017)	Rotation Forest ensembles (RFEs) and naïve Bayes tree (NBT)	Historical records	Satellite images
Detection	To develop an efficient image object detection method for use in a disaster prevention network (Okamoto et al. 2018)	Convolutional Neural Network (CNN), Natural	Information Centric Networking (ICN)	Text and images

		language processing (NLP) techniques		
Disaster management	To identify spatial distributions of both risk and damage cost of the wildfire (Nasanbat et al. 2018)	Multi-criteria Evaluation Analysis (MCEA)	GIS vector, GIS Thematic maps, Climate data, Field data, and Satellite data	Satellite imagery, map of the land cover thematic, field and climatic data
Prediction	To implement a model that predicts wildfires using Remote Sensing (Sayad et al. 2019)	ANN and SVM	Data was collected from Moderate Resolution Imaging Spectroradiometer	Images
Detection	To develop an alert system using sensor network and LEACH algorithm (Gupta & Doshi, 2018)	Low Energy Adaptive Clustering Hierarchy (LEACH) algorithm	Wireless sensor network transmitting data from the cluster head to the base station and then to the radio receiver	Raw analog data from sensors converted to digital data

Prediction	To show that regression works better than classification for detection of forest fires (Kansal et al. 2016)	Root mean square error (RMSE), linear regression, SVM, decision trees, GRNN	Meteorological data from UCI repository	Forest fire data
Prediction	To reduce disaster damages by training a deep learning model for forest fire prediction (Kim et al. 2017)	CNN	Optical sensor	Aerial images
Detection and prevention	To develop an algorithm for early detection of forest fire (Li et al. 2018)	Multiple Linear regression (MLR) and Decision Tree (DT)	Forest fires dataset	Numerical and categorical data
Disaster recovery & management strategies	To develop a distributed data center that carries information from relief distribution centers to the affected areas for emergency needs (Dubey et al. 2018)	Distributed service broker policy algorithm (DSBP)	Logistic information systems (LIS)	Statistical data

Disaster recovery & management strategies	Using AI to identify risk areas and determine future needs (Ivić, 2019)	ANN and SVM	Pre-earthquake and post-earthquake remote sensing images Landsat and Sentinel images	Images
Disaster recovery & management strategies	To present an approach to analyzing social media posts to assess the footprint of the damage caused by natural disasters (Resch et al. 2018)	Latent Dirichlet Allocation (LDA) and Local spatial autocorrelation	Social media data - Twitter	Tweet text
Disaster recovery & management strategies	To develop a classification model to solve the trigger design challenge (Calvet et al. 2017)	SVM, Naïve Bayes, logistic regression, Neural network	Simulated CAT data models	Numerical and categorical data
Detection	To develop a model for early detection of disaster (Wieland et al. 2016)	SVM	SAR data	SAR images
Disaster recovery &	To process social media textual and imagery data to generate visual and	Stanford sentiment analysis, K-means	Twitter	Tweets text and images

management strategies	descriptive summaries of hurricanes (Alam et al. 2019)	algorithm, Random forest, and LDA,		
General disasters				
Disaster recovery & management strategies	The application of multihop ad hoc networks in disaster response scenarios (Hwang et al. 2018)	Simulation	-	-
Disaster recovery & management strategies	To develop a model for performing information extraction on generic, hazard-related social media data streams (Tarasconi et al. 2017)	Natural Language Processing (NLP) technique	Twitter	Tweet text

2.3.2 MODELLING TECHNIQUES USED

The impact of natural disasters can be reduced by developing predictive algorithms (Li et al. 2018). Thus, researchers have adopted machine learning (ML) techniques as tools for either detecting, predicting, or managing natural disasters. Table 2.1 and Figure 2.2 shows the various modelling techniques and the distribution of the techniques respectively. The modelling techniques are divided into four categories; traditional machine learning techniques such as random forest, SVM, and logistic regression, neural networks (deep learning) such as ANN and CNN, natural language processing techniques (NLP) such as standford sentiment analysis, and Latent Dirichlet Allocation (LDA), and other statistical tools and algorithms such as root mean square error (RMSE), multiple linear regression (MLR), local spatial autocorrelation, and linear regression.

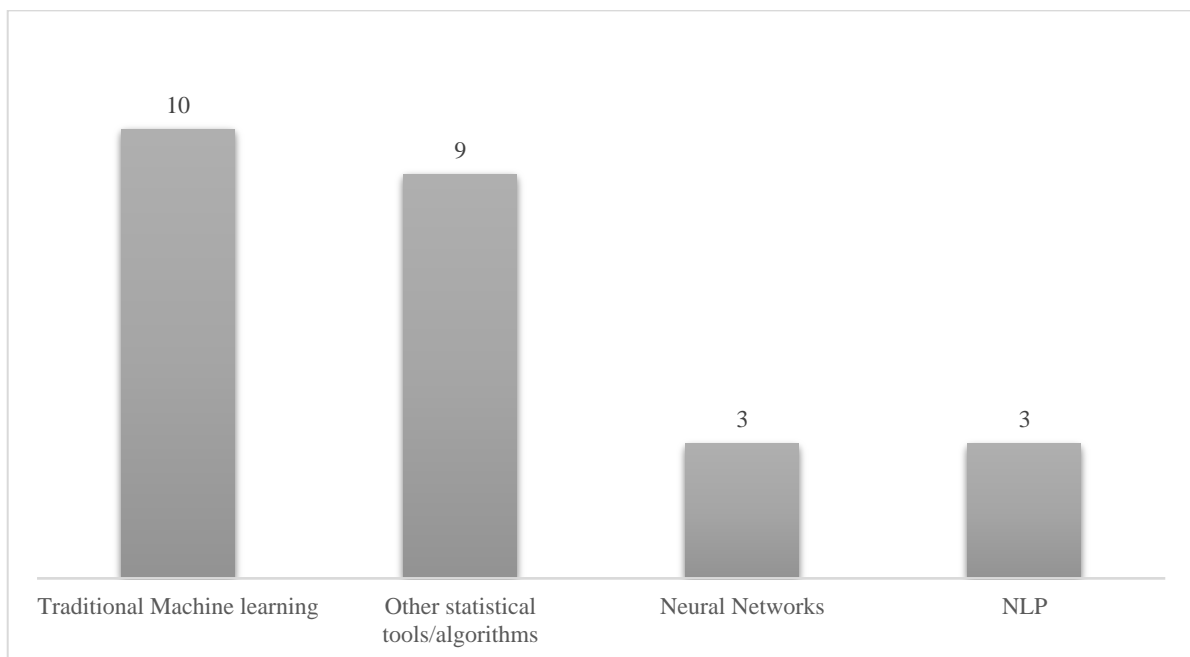


Figure 0.2: Distribution of modelling techniques

As shown in Figure 2.2, the modelling techniques used were predominantly machine learning techniques, closely followed by statistical tools/algorithms. The least used techniques were neural networks and NLP techniques. It was also observed that, although various modelling techniques were used, there are a plethora of techniques that were not explored.

2.4 REVIEW SUMMARY

Although the review summary provided in this chapter is not exhaustive as the search was narrowed to articles from the Scopus database only. Yet, 24 articles from the past 5 years (2014 - 2020) were reviewed and the following observations were made.

- i. It was observed that the subject of natural disaster control is a multidisciplinary one as it involved researchers from computer & mathematical sciences, engineering, informatics, agriculture, and social sciences.
- ii. Flood, earthquake, tsunami, forest fire, wildfire, landslides, and hurricane made up the investigated disasters.
- iii. Out of the 24 articles, research was predominantly conducted on disaster detection and disaster recovery & management strategies. Also, these tasks had varying methodologies.
- iv. Some of the researchers who proposed detection, prediction, or disaster management strategies as a means of mitigating the effects of natural disaster adopted the machine learning classification approach using either images of disaster scenes, text content or numerical measurements. Table 2.2 summarizes the various classification schemes identified in the reviewed articles.

Table 0.2: Classification schemes for natural disasters

Task	Classes	Type of data
Classification of flood severity (Khalaf et al. 2015)	Normal or dangerous	Numerical measurements
Classification of flood severity (Khalaf et al. 2018)	Normal, abnormal, or high-risk floods	Numerical measurements
Classification of flood data (Doxani et al. 2019)	Flood or no flood	Numerical measurements
Classification of disaster reports (Imran et al. 2017)	Relevant or irrelevant messages and mild or severe disasters	Textual and image contents
Classification of fire scenes (Kim et al. 2017)	Fire or non-fire scenes	Images
Classification of disaster scenes (Ivić, 2019)	affected, vegetation, sea and rivers, bare land, and clouds	Images
Classification of faces, gender and age group of flood victims (Vallimeena et al. 2018)	Deep or shallow water depth	Numerical measurements

It was observed that classification schemes were used as post-disaster management or assessment strategies to determine either the magnitude of the disaster, or the number of casualties, or the affected population.

Furthermore, the following classification metrics were used to evaluate the performance of the models; true positive (TP) – cases where the disasters are correctly predicted, false negative (FN) – cases where a disaster is incorrectly predicted, false positive (FP) – cases where a non-disaster is incorrectly predicted, true negative (TN) – cases where a non-disaster is correctly

predicted. Data sources included online tweets from social media (twitter), historical data, meteorological data, simulated data, and wireless sensors.

Although data forms such as text, images, and statistical data were used to classify and identify patterns of natural disasters, these lists of data types are not exhaustive. For instance, no study was identified on the use of sound/acoustic signals for detecting or predicting natural disasters, even though this comes with its advantages. Monroe-Kane, (2019), states that acoustic data makes it possible to quantify the characteristics of a volcano including the duration, frequency, intensity, and the progression over time of eruptions. Also, distinguishing events using acoustics enables the measurement of wind patterns as well as determining the destructive power of a natural disaster such as a hurricane (Wilson & Makris, 2006). Similarly, natural disasters produce acoustic signals, and detecting the infrasound pulses and T-waves from disasters such as earthquake, crashing waves, hurricanes, volcanoes, and tornadoes can supplement the information gathered from satellites (satellite images) and airplanes (aerial views) (Hassiotis, 2018; Mone, 2007). An effective disaster sound detection and monitoring systems can increase early detection rates (true positive) as well as reduce false alarms (Simmonds & MacLennan, 2005).

2.5 CHAPTER SUMMARY

In this chapter, summaries of studies related to the use of AI techniques to mitigate the effects of natural disasters on life and properties as well as the environmental and physical security were identified. The modelling techniques, data source, data types as well as classification schemes were also summarized.

The main focus of this study is in differentiating one natural disaster type from the other in real-time using sound/acoustic signals. However, no study on the subject was found from the review as most of the studies were focused on post-disaster management strategies. In the next

chapter, this study will seek to investigate various sound classification tasks, their methodologies and algorithms. This will be achieved by a systematic review of the literature.

Chapter Three

STATE OF THE ART IN SOUND CLASSIFICATION

3.1 CHAPTER OVERVIEW

This chapter presents a systematic review of the literature that analyzes the use of machine learning in the various sound classification tasks. The goal of this chapter is to;

- i. Identify publication patterns in the area of acoustic signal/sound classification.
- ii. Identify trends in the use of machine learning in acoustic signal/sound classification.
- iii. Identify open questions and challenges in the use of machine learning algorithms in sound/acoustic signal classification.
- iv. Identify research gaps in the subject area.

In this chapter, the terms sound and acoustic signals may be used interchangeably.

3.2 RELATED REVIEWS

This section identifies studies that have attempted to systematically review the literature on the classification of sounds or acoustic signals using artificial intelligence (AI) techniques. Existing systematic reviews in the subject area were identified and then evaluated using Greenhalgh, (1997) evaluation criteria. These criteria have been adopted in a number studies including (Tranfield et al. 2003; Van Dulmen et al. 2007)

Findings from existing systematic reviews in the subject area indicated that, available studies satisfied the predefined evaluation criteria as well as provided summaries and reproducible review methodologies. However, researchers 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). In other words, no secondary study (systematic review) on the classification of sounds in general was found. Considering the various applications of sound in

our day to day activities, the lack of sufficient summaries justifies the need for a systematic review.

3.3 REVIEW QUESTIONS

Acoustic signals or sound rather than imaging or computer vision is gradually gaining research popularity as a tool for environmental monitoring, diagnosing diseases, and data transmission. Recently, machine learning algorithms have been used for various classification tasks (Aucouturier et al. 2011; Bishop et al. 2019). However, due to the plethora of machine learning algorithms, selecting a suitable algorithm for a specific classification task is difficult. Hence, the need to identify open questions and state-of-the-art trends and tools that will assist researchers to appropriately position new research activity in this domain. This review examines the following research questions to identify publication trends and provide answers that will provide researchers with information about current approaches in algorithm usage.

The research questions stated in Table 3.1 are divided into two categories. Category one is made up of questions that seek to provide an overview of publication trends in the area of sound classification and machine learning. On the other hand, category two seeks to provide a good methodological background for a broader work by identifying research gaps and up-to-date methodologies.

Table 0.1: Research questions and objectives

	RESEARCH QUESTIONS	OBJECTIVES
Category one		
1.	<ul style="list-style-type: none"> - What are the yearly publication trends? - What Journal has the highest number of publications? 	<ul style="list-style-type: none"> - To identify the frequency of primary studies per year.

	<ul style="list-style-type: none"> - What is the frequency of authors? - What is the country's origin of authors affiliated institutions? 	<ul style="list-style-type: none"> - 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 two

<p>2.</p>	<ul style="list-style-type: none"> - 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? 	<ul style="list-style-type: none"> - To identify the different 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.
<p>3.</p>	<p>What are the various application domains?</p>	<p>To identify domains in which sound classification is predominantly performed.</p>
<p>4.</p>	<ul style="list-style-type: none"> - What features were extracted or what feature extraction technique was used? - What classifiers were used? 	<p>To identify predominantly used feature extraction techniques as well as classification techniques.</p>

3.4 REVIEW APPROACH

3.4.1 LITERATURE SEARCH

A systematic search of the literature was carried out in two databases, Scopus and Acoustical Society of America (ASA). We sought to review scientific articles from high ranking journals,

hence our choice of SCOPUS. Additionally, ASA was included as the second database since the primary interest of this study is on sound.

Publications were extracted from 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 these 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

3.4.2 INCLUSION AND EXCLUSION CRITERIA

A set of specific eligibility criteria were defined and followed to limit our collection of articles to only those that fit with our research objectives. A suitability check of returned articles was performed after examining the title and removing duplicate papers. Only articles in which the methodologies and results were explicitly stated in the abstract and or conclusion and which are primary studies were considered eligible for the review. The inclusion and exclusion criteria are as follows:

- C1 Include only open-access journal articles written in English and published between the years 2010 and 2019.
- C2 Include articles whose title contains 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 repeated papers from the search results.

C4 Exclude papers in which the abstract and (or) conclusions do not explicitly state the classification techniques and or results for sound classification.

C5 Exclude secondary studies.

3.4.3 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 repetition, abstract and or conclusion and by 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 from it concerning the review questions. The extracted data was collated in a spreadsheet for ease of use. Table 3.2 shows the search results gotten from the selected databases after each stage of exclusion or inclusion.

As shown in Table 3.2, the initial search output contained 1,028 research articles published from 2010 to 2019. Out of these, 150 articles were included after an initial screening by title and keywords and a total of 67 articles were obtained after the removal of duplicates. Furthermore, 19 articles were omitted based on abstract and secondary studies. Finally, a total of 48 journal articles were selected. It is important to note that this search was until the 22nd of December 2019.

Table 0.2: Search results per exclusion criteria

Databases		C1	C2	C3	C4	C5
ASA	SP1	50	10	10	8	8
	SP2	117	12	7	7	7
	SP3	112	11	0	0	0
	SP4	291	21	6	5	5
	SP5	117	16	0	0	0
	SP6	111	21	5	4	3
SCOPUS	SP1	104	19	17	13	12
	SP2	37	19	11	10	10
	SP3	27	6	3	2	1
	SP4	27	6	3	2	1
	SP5	23	7	3	0	0
	SP6	12	2	1	1	1
TOTAL		1,028	150	67	43	48

3.5 OVERVIEW OF PUBLICATION TRENDS

This section will answer category one of the research questions stated in Table 3.1. It highlights the frequency of publications, distribution of journals, leading authors in the subject area and their country origin.

3.5.1 PUBLICATION FREQUENCY

The publication trend covers articles published between the years 2010 and 2019. Figure 3.1 shows the frequency distribution of the selected articles.

During the search, it was observed that within the selected year range, Scopus had no open access publications (exclusion criteria 1) in the subject area until the year 2013. As shown in

Figure. 3.1, there has been a moderate publication trend within the years 2011 and 2015 with a minimum of two publications per year. The publication trend started increasing from 2016 with a big jump in the year 2018 and a slight drop in the year 2019; probably because the study was completed before the end of the year. It can therefore be concluded that researchers are beginning to develop considerable interests in this area of research.

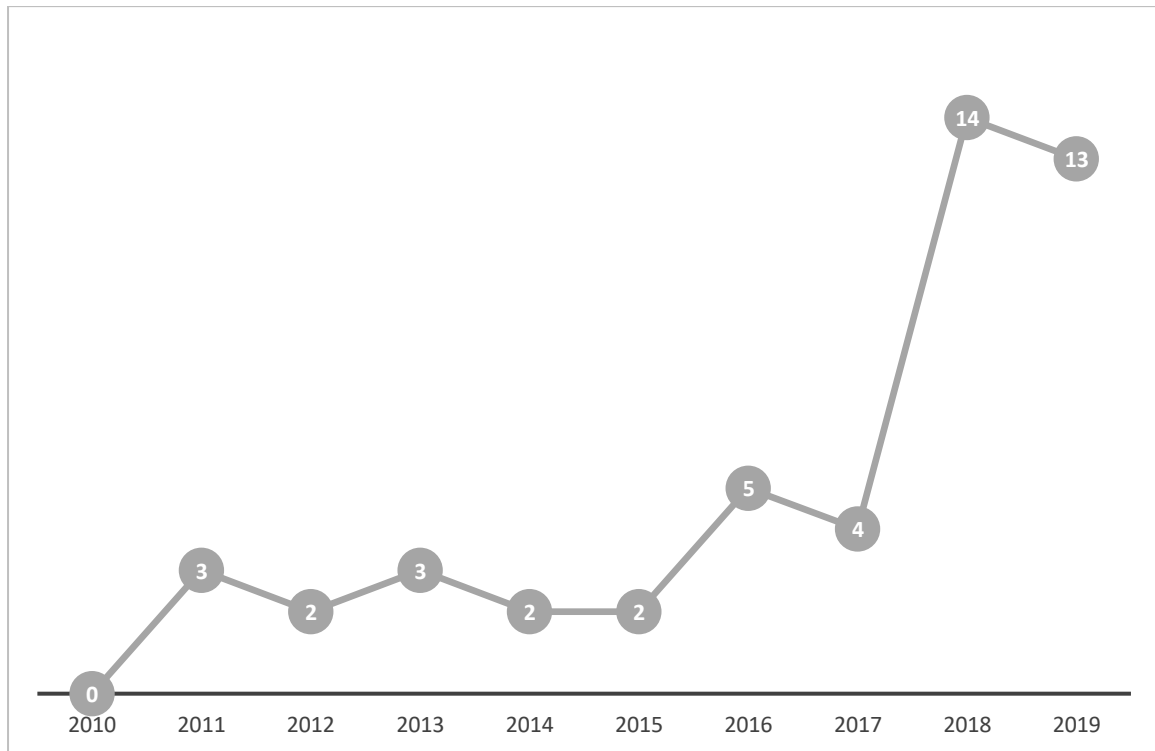


Figure 0.1: Publications by year

A further increase in subsequent years is envisaged, considering the popularity of artificial intelligence as well as the emergence of sound as an alternative form of data transmission.

3.5.2 DISTRIBUTION OF JOURNALS

The search results from SCOPUS had publications from several journals including the Journal of Acoustical Society of America (JASA). JASA is a journal in the ASA database with numerous publications in the area of sound/acoustic signals. It had more publications from an independent search has shown in Table 3.3.

It can be observed that a total of 20 Scopus journals published articles in the subject area within the years 2010 and 2019. Out of the 20 journals, 35% of the publications cut across 17 journals with a maximum of one publication. Applied Sciences and Sensors journal each made up 6% of the publications respectively, followed by IEEE Access with 4%. As earlier mentioned, JASA with 49% had the highest number of publications.

Table 0.3: Frequency distribution of primary sources

	Journals	Freq
1.	APSIPA Transaction on Signal and Information Processing	1
2.	Biomedical Journal	1
3.	Electronics	1
4.	Elektronika ir Elektrotechnika	1
5.	Eurasip Journal on Image & Video processing	1
6.	Expert Systems with Applications	1
7.	Computers & Electronics in Agriculture	1
8.	Frontiers in Neuroscience	1
9.	IEEE Signal Processing Letters	1
10.	IEICE Transactions on Information and Systems	1
11.	International Journal of Fuzzy logic & Intelligent Systems	1
12.	International Journal of online & biomedical engineering	1
13.	International journal of online engineering	1
14.	Noise mapping	1
15.	PeerJ	1
16.	PLoS ONE	1
17.	IEEE Access	2
18.	Applied Sciences	3
19.	Sensors (Switzerland)	3
20.	JASA	24

3.5.3 AUTHORS AND COUNTRY ORIGIN

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; an analysis of the authors and their country origin (the country in which their affiliated institution is located) was done.

With the number of authors per article ranging from 2 to 9, a headcount of the various authors showed that 209 authors wrote the 48 selected articles. Since one of the objectives of research question one is to identify authors who are consistent in writing on the subject area, Table 3.4 provides details of authors who wrote more than one article.

Table 3.4 highlights four groups of authors who wrote more than one article as either corresponding authors or co-authors. It was observed that the four groups of authors were all interested in bioacoustics.

Table 0.4: Leading authors

AUTHORS NAMES	YEAR	JOURNAL	REFERENCE
Ali K. Ibrahim, Laurent M. Chérubin, Hanqi Zhuang, Michelle Umpierre, Fraser Dagleish, Nurgun Erdol, B. Ouyang, and A. Dagleish	2018	JASA	(Ibrahim et al. 2018)
	2019	JASA	(Ibrahim et al. 2019)
Abeer Alwann and Charles E. Taylor	2015	JASA	(Tan et al. 2015)
	2016	JASA	(Kaewtip et al. 2016)
Yagya Pandeya, Joonwhoan Lee	2018	Applied Sciences	(Pandeya et al. 2018)
	2018	International Journal of Fuzzy Logic and Intelligent Systems	(Pandeya & Lee, 2018)
Amalia Luque, Javier Romero-Lemos, Alejandro Carrasco	2018	Expert systems with Applications	(Luque et al. 2018)
	2018	PeerJ	(Luque et al. 2018)

Furthermore, authors' country origin (countries of authors institutions) and the frequency of publications per year was also identified. Figure 3.2 shows the distribution of authors' country origin grouped into continents.

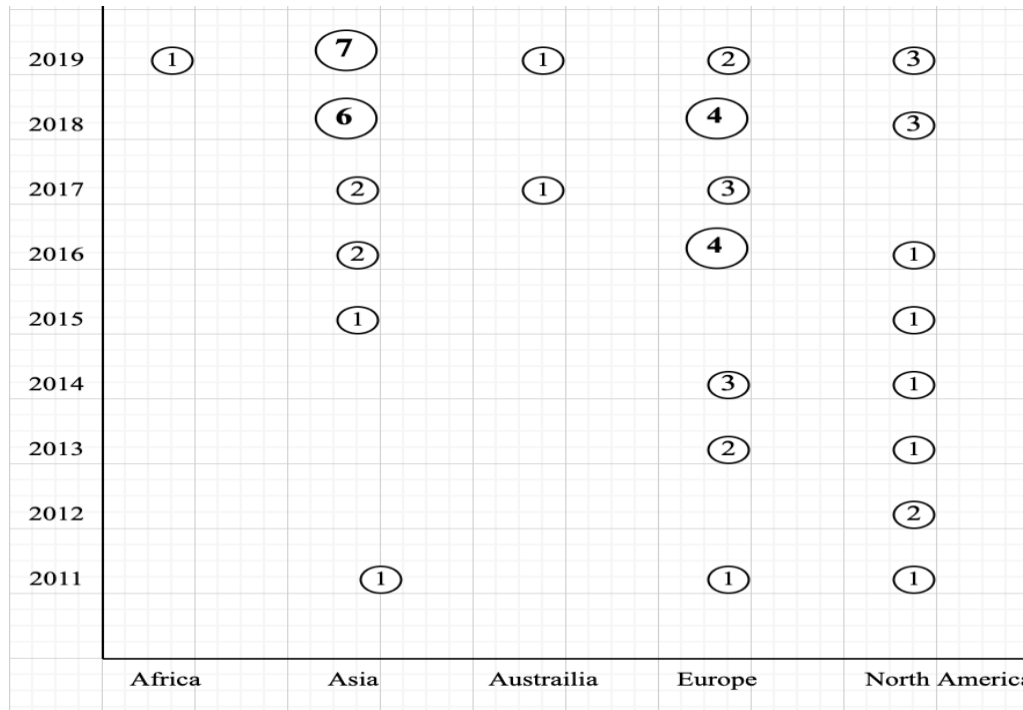


Figure 0.2: Distribution of authors by continent

As shown in Figure 3.2, the authors were distributed across five continents. Africa had the least number of publications (1), closely followed by Australia (2), while Europe had the highest (19). South America recorded no publications; hence it is not represented in Figure 3.2. Figure 3.3 provides a distribution of the publication trend by study country.

As shown in Figure 3.3, studies have been conducted in 22 different countries with the USA leading the trend with a total of 23% of all the studies. USA is followed by China, Korea and France with 10% each. Spain is in third place with a total of 8%, while India and UK are the fourth place, having 6% each of all the studies.

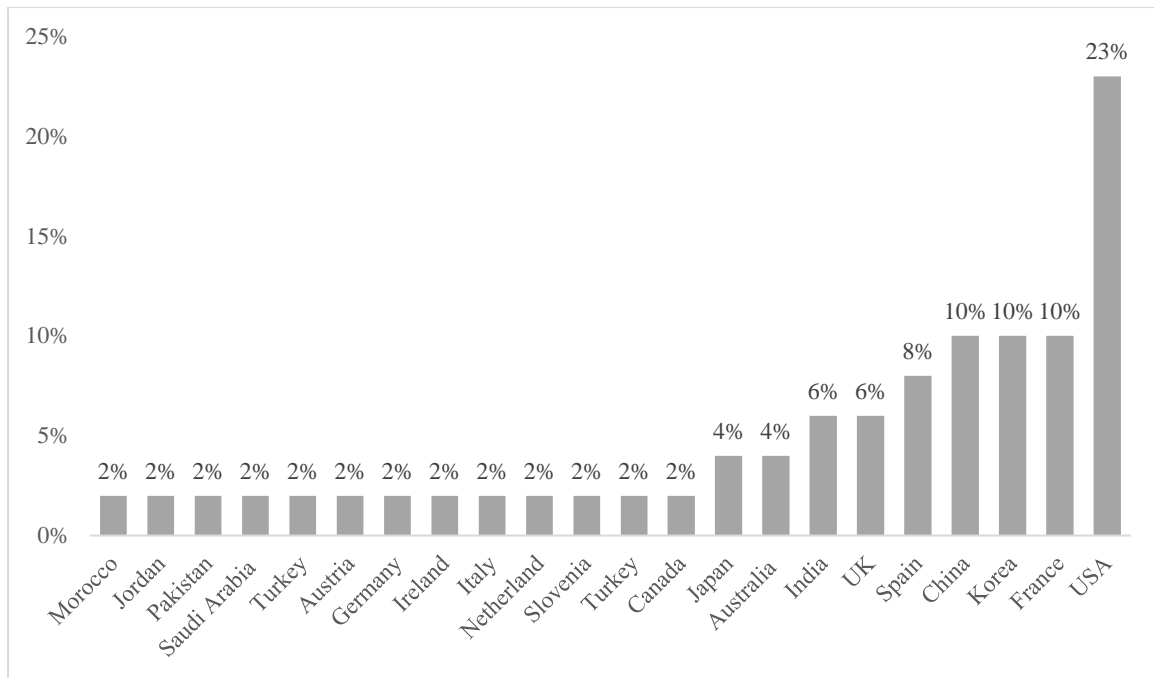


Figure 0.3: Sound classification publication trend by study country

3.6 SUMMARY OF METHODOLOGIES FROM REVIEWED ARTICLES

This section seeks to address category two of the review questions stated in Table 3.1. As earlier mentioned, forty-eight (48) articles were selected based on the criteria used. The discussions below are results obtained in line with the research questions.

For ease of reference, the selected articles have been numbered in the order in which there were selected - A1 TO A48 and will be used accordingly in further analysis (see Appendix A for a list of papers).

3.6.1 SOUND/ACOUSTIC SIGNALS CLASSIFIED AND DATA SOURCES

Sound is considered as the second most important sense after sight and it is capable of carrying information about anything in our environment (Perr, 2005). Although information from sounds are different from that obtained from radio frequency (RF), infrared (IR), and optics, it can be used for detection, classification and, localization (Hartman & Candy, 2014; Lopatka et al. 2016). Hence, the ability to differentiate sound or signal types becomes imperative as it will enable the extraction of relevant information about the sound source and the environment

(Rascon & Meza, 2017). So far, sound has been used in areas such as medical acoustics for medical diagnosis (Beach & Dunmire, 2007; Oweis et al. 2015) environmental monitoring for security surveillance (Salamon & Bello, 2017; Wu et al. 2018) and bioacoustics for prediction of natural disasters (Pandeya & Lee, 2018).

However, its application varies in land, air, and water depending on the medium of propagation, seasons, activities, and geographic location. Furthermore, it can be generated or caused by various activities including Anthrophony (sounds made or caused by humans) e.g. shipping and drilling noise, Geophony (sound from the environment) e.g. sea surface noise like the breaking of waves, ice-breaking, raindrops, and Biophony (sounds from animals) e.g. vocalizations of mammals, anurans, groupers.

This section discusses the different kinds of sounds that were classified, source of data, sample rates, duration of sound recordings per file, and availability of datasets as found in the selected articles.

From Table 3.5, it can be observed that 31 researchers were specifically interested in Biophony, 13 in Anthrophony, and the last 4 in all sound categories - anthrophony, geophony, and Biophony.

The marine mammal group which is the dominating was made up of different species of dolphins and whales (odontocetes & Mysticetes). It was observed that, some of the researchers were interested in automatically detecting, differentiating and classifying call types from different species (Guilment et al. 2018; Halkias et al. 2013; Roch et al. 2011; Shamir et al. 2014). Others were interested in classifying; vocalizations of humpback whales, whistles & pulse of dolphins, song cycles of whales and echolocation clicks of beaked whales respectively (Allen et al. 2017; LeBien & Ioup, 2018; Ou et al. 2013; Peso Parada & Cardenal-López, 2014). Cvangros et al. (2017) classified blast sound with the aim of monitoring environmental noise, classifying signals and differentiating between blast sound and non-blast sound.

Human sounds that were classified included respiratory sounds, human voice disorder, and baby cry. Aucouturier et al. (2011) described baby cry as a reflexive signal that reflects the state of a baby by conveying a message of either a need, pain, discomfort or a medical condition.

Table 0.5: Summary of classified sounds and datasets

REF	SOUND TYPE	Link to Dataset/Name of datasets	Source of datasets		D.A	Sample rate	Time(s)
			L.R	E.D			
A1	Whale	N/M	✓	x	x	96 – 192	1 - 8
A2	Birds	http://www.animalsoundarchive.org/Refsys/Statistics.Php	x	✓	✓	1-4	60
A3	Fish	SEACOUSTIC2014	x	✓	x	256	10 - 30
A4	Mysticetes calls	Mobysound.org	x	✓	✓	0:1- 8	-
A5	Birds	Birdcalls71, Flight calls datasets, Anuran dataset	x	✓	x	22.1 – 44- 1	0.5 - 320
A6	Military blast sound	LRPE, East South Central, APG, SERDP-PITT, MCBC-PITT, New York (Fort Drum)	x	✓	x	5 – 25.6	5
A7	Birds	N/M	✓	x	x	8 - 16	10

A8	Primate Calls	N/M	✓	x	x	44	3
A9	Grouper	N/M	✓	x	x	10	20
A10	Marmosets-monkey	http://home.ustc.edu.cn/~zyj008/ background_noise.wav.	✓	x	✓	44.1	0.5 - 4
A11	Marmosets-monkey	http://marmosetbehavior.mit.edu	✓	x	✓	48	0.5
A12	Red Hind Grouper	N/M	✓	x	x	10	10
A13	Mysticete calls	DEFLOHYDRO, OHAS-ISBIO, DCLDE 2015 datasets	x	✓	x	0.25	6
A14	Birds song phrases	http:// bn.birds.cornell.edu/bna/species (CAVI database)	x	✓	✓	20	3
A15	Mysticete calls	Mobysound.org	x	✓	✓	I - 4	2
A16	Odontocetes Sound	N/M	✓	x	x	192	10
A17	Humpback whale song unit	N/M	✓	x	x	22 – 44.1	3

A18	Bird song Phrase	http://taylor0.biology.ucla.edu/birdsong/ (CAVI database)	x	✓	✓	20 – 44.1	0.12 – 0.25
A19	Humpback whales	Auau Channel 2002 and French Frigate Shoals (FFS) dataset	x	✓	x	10	0.4 – 3.7
A20	Beaked whales	https://data.gulfresearchinitiative.org/	✓	x	✓	92	0.0021
A21	African gray parrot	N/M	✓	x	x	22	2.5
A22	Dolphins whistles and pulses	http://www.cemma.org/ (CEMMA database)	x	✓	✓	96	6 - 25
A23	Anuran calls	Recordings of frog vocalizations were obtained from commercially available compact discs (CD)	x	✓	N/M	44.1	8
A24	Baby cry	N/M	✓	x	x	44.1	30

A25	Livestock (sheep, cattle, & Maremma sheepdogs)	N/M	✓	x	N/M	44.1	1
A26	Aircap, Bells, Bottle, Buzzer, Case, Clap, Cup, Drum, Phone, Pump, Saw, Spray, Stapler, Tear, Toy, Whistle & Wood	http://dcase.community/challenge2018/index . And http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.357&rep=rep1&type=pdf . (Real-world computing partnership (RWCP) sound scene dataset and DCASE challenge dataset).	x	✓	✓	48	4
A27	Respiratory sound (wheezes, crackles & normal sound)	Int. Conf. on Biomedical Health Informatics (ICBHI) scientific challenge database	x	✓	N/M	N/M	1.5
A28	Heart sound	https://physionet.org/challenge/2016/ . (the Physionet database)	x	✓	✓	2	5 - 120

A29	Heart sound	https://github.com/yaseen21khan/Classification-of-heart-sound-signal-using-multiple-features/blob/master/README.md	x	✓	✓	8	-
A30	Cat (mother call, paining, resting, warning, angry, defense, fighting, happy, hunting mind, mating)	Online video sources including YouTube, Kaggle challenge database and Flickr	x	✓	N/M	16	2 - 6
A31	Anuran (mating and release call)	http://www.fonozoo.com/.	x	✓	✓	44.1	96
A32	Pet dog (barking, growling, howling & whining)	https://github.com/kyb2629/pdse .	✓	x	✓	22 – 44.1	0.24 – 1.47
A33	Anurans (mating, release, distress calls)	http://www.fonozoo.com/.	x	✓	✓	N/M	5
A34	Lung sounds	N/M	✓	x	x	N/M	
A35	Birds, frogs, wind, rain, & thunder	N/M	N/M	N/M	N/M	16	N/M

A36	People, animals, nature, vehicles, noisemakers, office, & musical instrument	http://www.findsounds.com/type_s.html (FindSounds database)	x	✓	✓	16	10
A37	Fish	http://www.fishbase.org/ http://www.dosits.org/	x	✓	✓	44.1	14
A38	Heartbeat sound (normal, murmur & extra-systole)	Dataset B- PASCAL classifying heart sounds challenge	x	✓	N/M	4	12.5
A39	Air conditioner, car horn, children playing, dog bark, drilling, engine idling, gunshots, jackhammer, siren, & street music.	https://dl.acm.org/doi/10.1145/2647868.2655045 . (Urban 8k dataset)	x	✓	✓	22.1	4
A40	Dog barking, firecrackers, rain, rooster, baby cries, sneezing, sea waves, chainsaw, helicopter, & clock sound)	https://www.karolpiczak.com/papers/Piczak2015-ESC-Dataset.pdf (ESC-10 AND ESC 50 datasets)	x	✓	✓	N/M	N/M

A41	Bird	http://www.vision.caltech.edu/vision/visualizing/2011/03/CUB/ (CUB-200-2011 standard dataset)	x	✓	✓	10	10
A42	Conversation, children shouting, walk-footsteps, crowd, hubbub, children-playing, bird, vocalization, truck-horn, motorcycle, traffic-noise, light-engine, medium-engine, engine starting, idling, silence	YouTube videos	✓	x	✓	20	10
A43	Cymbals, horn, phone, bells, kara, bottle, buzzer, metal, whistle, ring	RWCP and TIDIGITS datasets	x	✓	N/M	16 - 20	3
A44	Cat (warning, angry, defense, fighting, happy, hunting, mating, mother-call, paining, resting)	YouTube, and flicker	✓	x	N/M	N/M	N/M
A45	EEG (electroencephalogram) signals	http://www.cs.colostate.edu/eeg/	x	✓	✓		

A46	air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren, street music)	Urban- sound 8k dataset	x	✓	✓	44.1	4
A47	Respiratory sound	N/M	✓	x	x	44.1	
A48	Human voice disorders	N/M	✓	x	N/M	44.1	N/M

Hints: L.R = life recording, E.D = Existing sound dataset, D.A = data availability (✓ = available, x = not available), N/M = Not mentioned.

SAMPLE RATE, AUDIO FORMAT & 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 48 classified sounds, the dominant audio format used was the .wav format. Others included mp3 (Peso Parada & Cardenal-López, 2014; Shamir et al. 2014), ARFF (Zhang et al. 2016) and HDF5 format (Bold et al. 2019).

Furthermore, the 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 which represents the intensity of the sound at each time-frequency location. According to (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, unlike 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 (with time and frequency as the vertical and horizontal dimensions).

SOURCES OF DATA

With the aim of identifying publicly available datasets, the datasets used in the reviewed articles were divided into two categories; pre-existing sound datasets and life recordings.

- i. **PRE-EXISTING SOUND DATASETS:** This was made up of sound collected from past experiments, past projects or pre-existing sound databases, 27 datasets were obtained from this category. Out of the 27, only 17 were stated to be publicly available,

while the availability of others was either not mentioned or not available due to licensing or privacy issues.

- ii. **LIFE RECORDINGS:** This category of datasets was generated by the researchers. 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 including 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 (Oikarinen et al. 2019; Shamir et al. 2014), while in other cases, the data was 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, 19 datasets were privately generated and 5 are available to the public.

With a total of 47 mentioned data sources from both categories, only 23 are reported to be publicly available, this is a confirmation to the challenges of limited datasets faced by numerous researchers in the area of sound classification.

3.6.2 DISTRIBUTION OF CLASSIFIED SOUNDS ACCORDING TO APPLICATION DOMAIN

Considering the different type of classified sounds, the specific sound environment, and the researcher's objective for classifying the chosen sound, three broad application domains of classified sounds were identified. They include bioacoustics, medicine, and the environment (see Figure 3.4).

The application domain of bioacoustics was the most explored with a 69% occurrence rate. In this domain, researchers were predominantly interested in the classification of sounds and vocalizations of birds, mammals, and domestic animals. It was observed that some classified animal calls or vocalization with the intent of detecting and differentiating one animal species from the others (Guilment et al. 2018; Halkias et al. 2013; Roch et al. 2011). Some others were

interested in identifying and differentiating the call types of specific animals (Kim et al. 2018; Pandeya & Lee, 2018; Roch et al. 2011).

Researchers in the medical domain were predominantly interested in classifying sounds from heart and lungs related diseases. This was with a major objective of providing an automated and efficient classification and recognition system that will assist medical doctors or physicians in smart diagnosis.

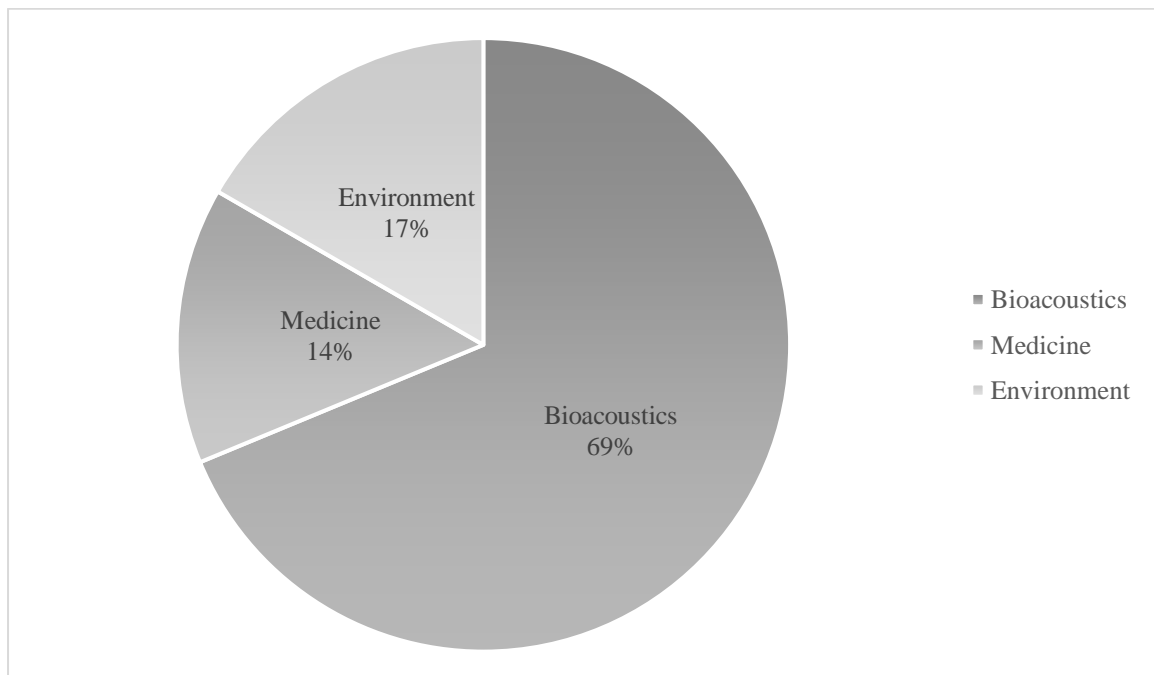


Figure 0.4: Pie Chart showing the distribution of application domains

Furthermore, they also sought to eliminate the invasive traditional computer vision methodologies like the use of medical imaging (Chen et al. 2019; Oweis et al. 2015; Vrbancic & Podgorelec, 2018). Conversely, 17% of the researchers explored sounds from the environment to automatically recognize environmental acoustics scenes as well as to precisely classify the detected sound. Environment as a domain consisted of sounds from sub-domains such as human activities, urban environment, surveillance, machinery, weather, musical instruments, etc.

3.6.3 FEATURE EXTRACTION METHODS

The classification of sound/acoustic signals as with other classification task requires the extraction of relevant features that will make the classification process more efficient and accurate. According to Wu et al. (2018), feature extraction reduces the size of data and represents the complex data as feature vectors. Additionally, the choice of features used to represent any given set of data may have a high impact rate on the classifiers as well as the classification results (Binder & Paul, 2019; Malfante et al. 2018; Oweis et al. 2015).

In order to ensure high classification accuracy, some researchers explored feature selection techniques such as the Jensen-Shannon divergence (Luque, Romero-Lemos, Carrasco, & Gonzalez-Abril, 2018), step-wise feature selection (LeBien & Ioup, 2018; Malfante et al. 2018).

Table 3.6 highlights the feature extraction methods used in the reviewed articles. The methods have been categorized according to the feature extraction methods stated by Wang & Nanda, (2012).

Table 0.6: Feature extraction methods

Methods		Reference
Time series transforms	Frequency domain (Stationary signals)	(Binder & Paul, 2019; Bourouhou et al. 2019; Fang et al. 2019; X. C. Halkias et al. 2013; Han et al. 2016; Ibrahim et al. 2018; Kim et al. 2018b; Luque, Romero-Lemos, Carrasco, & Gonzalez-Abril, 2018; Malfante et al. 2018; Parada & Cardenal-Lopez, 2014; Roch, Newport, et al. 2011; Su et al. 2019; Yan Zhang et al. 2016)

	Time-frequency (Non-stationary signals)	(Aykanat et al. 2017; H. Chen et al. 2019; Guilment et al. 2018; Malfante et al. 2018; Noda et al. 2016; Ou et al. 2013; Wu et al. 2018)
	Wavelets (Non-stationary signals)	(Aucouturier et al. 2011; Bishop et al. 2019; Qian et al. 2017b; Raza et al. 2019; Yaseen et al. 2018)
Data descriptive statistics	For sensors	(Gingras & Fitch, 2013; Oikarinen et al. 2019; Oweis et al. 2015; Ya-jie Zhang et al. 2019)
	For events	(Allen et al. 2017; Robakis et al. 2018)
Data descriptive models	Distribution models	(Briggs et al. 2012)
	Information-based models	(Giret et al., 2011; LeBien & Ioup, 2018)
	Regression models	-
	Classification/clustering models	(Aziz et al. 2019; Ibrahim et al. 2019; Kaewtip et al. 2016; Khamparia et al. 2019; Shamir et al. 2014)
Time-independent	Explicit mathematical operations	-
	Data dimension reduction	(Cvengros et al. 2012a; Tan et al. 2015; Thakur et al. 2019)

From the reviewed articles, the domain of bioacoustics was predominantly made up of marine mammals. According to Ou et al. (2013), the classification of marine mammals based on sound begins from analyzing their vocalizations. This includes sound detection from ambient noise, signal extraction, and feature analysis. Accordingly, they integrated contour extraction and

spectrogram correlation for feature extraction. Specifically, frequency contours were extracted from the spectrogram of humpback whales by applying image edge detection filters. Shamir et al. (2014) on the other hand analyzed spectrograms using Wndchrm scheme based on numerical content descriptors such as; 2D texture features (Haralick & Tamura textures), statistical distribution (mean, standard deviation, skewness, & kurtosis) and multi-scale histogram of the pixel intensities, polynomial distribution (Chebyshev coefficients & Zemike polynomials), Gabor wavelets and Radon features. While Halkias et al. (2013) on the other hand attempted to learn the underlying structures of the calls directly from the spectrogram using discriminative features. Inversely, (Binder & Paul, 2019; Roch et al. 2011) adopted the use of perceptual features as they provide better discriminative signals for the classification of inter-species of marine mammals, but most importantly, for its ability to take into consideration how a human listener would differentiate sound.

Identifying features of the Mysticetes calls is challenging because it is expected that the selected features do not only differentiate the call types but should also be able to detect any other signal within the same context (Guilment et al. 2018). Accordingly, Guilment et al. (2018) proposed a feature extraction method in which the feature vectors were digitized time-series of Mysticetes calls and extracted features were obtained from click waveforms. With the primary objective of obtaining a low false-positive rate (FPR), LeBien & Ioup, (2018) adopted a stepwise feature selection procedure that iteratively added features which minimizes a loss function. Consequently, a false positive rate of 0.001% was achieved.

In classifying sounds from groupers/fishes, Malfante et al. (2018) proposed an all-purpose feature extraction approach that can be used on the classification of any type of datasets except fishes. Accordingly, they used 84 general features (instead of domain-specific features), from

the time domain, frequency domain, and cepstral domain and the forward selection method to address the issue of feature selection.

Giret et al. (2011) used the extractor discovery system (EDS) to generate 11,000 features from 10 MFCC features. The acoustic features extracted from each audio signal was used to classify the calls using a C4.5 machine learning algorithm. Kaewtip et al. (2016) on the other hand, used the Mel-frequency cepstral coefficients (MFCC) as front-end features for the Hidden Markov Model Toolkit (HTK).

Sounds produced by birds are characterized by several components depending on their species. However, due to the class-specific characteristics of harmonic and percussive components of a bioacoustics sound, these two components were combined with Mel-spectrogram to produce a three-channel input for the proposed framework (Thakur et al. 2019).

MFCC as a sparse representation of the original sound was the predominant feature extracted from signals (see Table 3.8). However, Ibrahim et al. (2018) argue that they do not perform well under noisy conditions, hence they proposed the use of its optimized version. Accordingly, they used weighted MFCC (WMFCC) and weighted multiresolution features (WMRAF). Although WMFCC had better features with lower magnitudes of computational cost, due to the varying performance accuracies obtained from different species, the optimized features were concluded to be domain-specific features (Ibrahim et al. 2018). In a further study by the same authors (Ibrahim et al. 2019), sparse autoencoders (SAE) were used to learn features from the sound spectrum rather than using a particular feature extraction method (Ibrahim et al. 2019). Similarly, instead of using traditional MFCC features, Briggs et al. (2012) used mask descriptors and integrated the selected features into a single feature vector that described each segment of an audio signal in the spectrogram.

Zhang et al. (2019) used handpicked acoustic features in processing animal vocalizations and the processes of classification and detection were done distinctly. In addition to detection and

classification, attribution was included to make up a process that enabled the network to learn useful features as well as reduce likely bias consequence of the handpicked features (Oikarinen et al. 2019). Inversely, instead of using hand-picked features, features from a CNN pre-trained network was used (Bold et al. 2019; Pandeya et al. 2018).

In general, it was observed that the extracted features were either domain-specific or generic. It was also observed that, no feature extraction/selection was performed in some cases where neural networks were used for the classification. According to Vrbancic & Podgorelec, (2018), the advantage of using this approach is that no domain expert knowledge is required for classification.

3.6.4 SOUND CLASSIFICATION ALGORITHMS AND PERFORMANCE METRICS

The next step of sound analysis after feature extraction is to take the extracted features and feed them into an appropriate classifier. According to (Binder & Paul, 2019), an automatic classifier does not only identify or differentiate one sound from another, but it also reduces false detections of sound. Techniques for the various classification task are shown in Table 3.7. The predominant classifiers included Support vector machine (SVM), neural networks, k-nearest neighbor (KNN), hidden markov model (HMM), and k-means.

Table 0.7: Classification techniques used

Ref. no	Classifiers
A1	Euclidean distance
A2	Kernel-based extreme machine (KELM), Sparse-Instance-based AL, least confidence-score-based AL (LCS-AL)
A3	Random Forest & Support Vector Machine (SVM)
A4	Restricted Boltzmann machine (RBM) & sparse auto-encoder (SAE)
A5	Convolutional Neural Network (CNN) & Multilayer perceptron (MLP)
A6	Linear SVM and Radial Basis Function (RBF) SVM
A7	MIML-SVM, MIML-KNN, MIML-RBF (MIML: multi-instance multi-label)

A8	Artificial Neural Network (ANN)
A9	K-nearest neighbors (KNN), Support Vector Machine (SVM) & Sparse classifiers
A10	SVM, Deep Neural Network (DNN), Recurrent Neural Networks - Long Short-Term Memory (RNN- LSTM)
A11	Feed forward deep convolutional neural network
A12	Random ensemble of stacked autoencoders (RESAE)
A13	Sparse representation
A14	Dynamic time warping (DTW) and Hidden Markov models (HMM)
A15	Aural classifiers
A16	K-means
A17	Self-organizing map
A18	DTW-SR-2pass
A19	K-means
A20	K-means
A21	Decision tree
A22	Gaussian Mixture Model (GMM)
A23	Logistic regression (LR)
A24	Hidden Markov model (HMM)
A25	Support Vector Machine (SVM)
A26	Support Vector Machine (SVM), K-nearest neighbors (KNN)
A27	Deep residual networks (ResNets)
A28	Support Vector Machine (SVM)
A29	Support Vector Machine (SVM), Deep Neural Network (DNN)
A30	Convolutional deep belief network (CDBN)
A31	Hidden markov model (HMM)
A32	SVM, KNN, Long short-term memory-fully convolutional network (LSTM-FCN)
A33	Non-temporally aware (NTA)

A34	Convolutional Neural Network (CNN)
A35	Multi-view simple disagreement sampling (MV-SDS)
A36	SVM with linear kernels & pairwise multi-class discrimination sequential minimal optimization, logistic regression
A37	Support Vector Machine (SVM)
A38	Recurrent Neural Network (RNN)
A39	TSCNN-DS
A40	Convolutional Neural Network (CNN)
A41	CaffeNet pretrained Convolutional Neural Network (CNN)
A42	Artificial Neural Network (ANN), Recurrent Neural Networks - Long short-term memory (RNN-LSTM)
A43	Self-organizing map-Spike Neural Network (SOM-SNN)
A44	Pre-trained CNN
A45	LeNet based Convolutional Neural Network (CNN)
A46	Convolutional Neural Network (CNN)
A47	Artificial Neural Network (ANN)
A48	Deep Neural Network (DNN)

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; Cveugros et al. 2012a; Noda et al. 2016; Qian et al. 2017a; 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 (Cveugros et al. 2012a), 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

classification tasks concerning good performances on various learning tasks (Zhang et al. 2018).

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, neural networks have self-organizing and adaptive learning properties with an outstanding ability to detect trends based on the sample data (Dwivedi et al. 2019). They also have a distinctive ability to build deep architectures as well as automatically learn feature representations. Compared to conventional machine learning techniques with shallow networks that are made up of one input layer, one output layer and a hidden layer that lies in between the input/output layers, neural networks consist of several layers and has the ability to grow deeper into the network by increasing the number of hidden layers. Conversely, the difference between neural networks and deep learning depends on the depth of the model; deep learning is an application of neural networks with several layers of nodes (4 or more) between the input and output layers (Arel et al. 2010).

Recently, deep learning has enabled various applications in action detection, object recognition, speech recognition, image classification, and recognition. Findings from this systematic review indicate that deep neural networks have also been evident and effective in medical diagnosis, acoustic detection, and acoustic classification. With a widespread application in various domains, deep learning has been promoted in literature for the following reasons as stated by Wason, (2018):

1. They can filter and extract information hidden in the presence of noise
2. The algorithms train through input data to identify hidden patterns and then integrate the information obtained into visual analytics displays.
3. The algorithms can apply discrimination to data to reveal patterns and extract valuable information.

4. It can classify unstructured and structured data using methods like deep belief methods (DBM) or convolutional neural networks.
5. It mimics the human brain through artificial neural networks (ANN) and learns how to solve problems in a human-like manner.

From the reviewed articles, some of the neural networks used included, CNN, ANN, DNN, RNN, LSTM-RNN, feed-forward deep convolutional neural network (FFD-CNN), convolutional deep belief network (CDBN), and Long short-term memory-fully convolutional network (LSTM-FCN). In general, although high classification accuracies are guaranteed with neural networks, training a neural network requires huge datasets and high computational power.

Hidden markov model (HMM) is a generative model and the first segment-based approach in classification procedures (Wu et al. 2018). In sound classification, it takes a sound segment and tries to classify it as a whole without any form of framing (Luque, et al. 2018). Although it ensures realistic temporal statistics of the output (Aucouturier et al. 2011), its performance is limited due to its statistical inefficiency in modeling data that lies on a nonlinear manifold in the feature space (Ibrahim et al. 2019). Additionally, it uses sub-word features that are not suitable for non-speech sound identification since they lack the phonetic structure that speech possesses (Luque et al. 2018). Furthermore, it requires large datasets for better performance and at the same time performs badly when there is a lot of noise in the data (Kaewtip et al. 2016). From the reviewed articles, authors who used HMM was generally for segmentation.

K-NEAREST NEIGHBOR (KNN) is a supervised machine learning algorithm that finds the class to which an unknown object belongs to using majority voting of KNNs i.e. it predicts classes using the majority of nearest neighbors (Noda et al. 2016; Pandeya & Lee, 2018). In contrast to HMM, KNN is robust to noise and requires low training time but at the same time requires large memory space (Dwivedi et al. 2019).

Predominantly, 94% of the techniques used in the 48 reviewed articles were supervised machine learning techniques, while the other 6% made up unsupervised machine learning with the k-means clustering technique. Apart from traditional supervised learning models and deep learning models, other supervised machine learning techniques were explored to overcome the challenges of limited data, overfitting and lack of labeled data. They included the use of pre-trained models like VGG and CaffeNet for transfer learning, extreme learning machine (ELM), and deep metric learning (DML). ELM is a single hidden layer feedforward neural network that was used to overcome the problems of slow training speed and, over-fitting encountered by neural networks (Ding et al. 2015; Qian et al. 2017). DML was used to overcome the problem of unlabeled data (Thakur et al. 2019). 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 3.5 shows the distribution of modelling techniques used in the 48 reviewed articles; neural networks were the most used technique out of the four categories of modeling techniques identified.

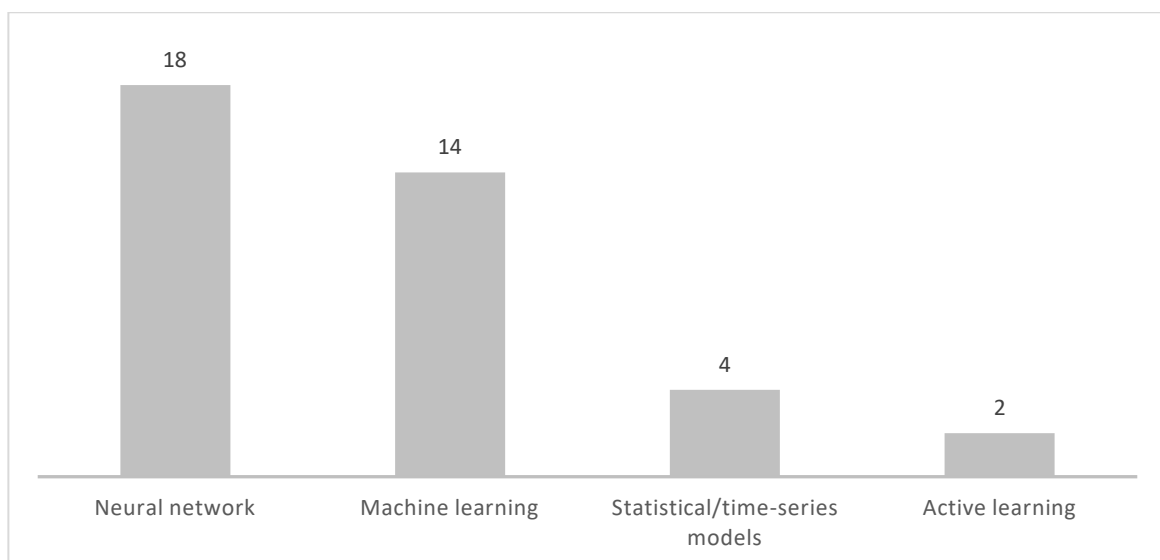


Figure 0.5: Distribution of modelling techniques

Various metrics were used to evaluate the performance of the techniques; 92% of the researchers were mostly concerned with classification accuracy. Others (8%) used the F1 score, area under curve (AUC), Sensitivity/TPR, Specificity/FPR, unweighted average recall (UAR), precision, recall and mean error rate. It was observed that the classification techniques used in the reviewed articles predominantly had good performance accuracies.

3.7 REVIEW SUMMARY

The primary objective of the systematic review was to identify methodological approaches and current algorithms used in the automatic classification of sounds. This review was restricted to Journal articles from Scopus and ASA databases and was guided by two categories of review questions which were answered accordingly.

In the first phase of the review, we identified the frequency of publications between the years 2010–2019, the distribution of journals, consistent researchers in the area of sound classification and the country origin of the various researchers. It was observed that until the year 2015 upwards, the level of research interest in sound classification was minimal. Also, researchers who published more than one article were all interested in animal sound classification. Additionally, 90% of the researchers were from European and Asian countries.

In the second phase, we identified the different types of classified sounds and their properties in terms of sample rate, audio format, datasets, and the various application domains. It was observed that in the domain of bioacoustics, researchers were mostly interested in classifying sounds from marine mammals, while the medical domain was concerned with diagnosing respiratory diseases using sound. Although different forms of environmental sound were classified, none of the articles classified natural disaster sound. This is a research gap, considering the alarming increase in natural disaster events yearly.

Considering that a major limitation to most of the studies was limited datasets or lack of annotated data, few researchers explored the techniques used for overcoming such problems in

machine learning. Three articles explored the options of transfer learning (Bold et al. 2019; Pandeya et al. 2018; and Pandeya & Lee, 2018), one used averaging methods for an ensembling of six classifiers (Pandeya & Lee, 2018), and two used cross-validation (Binder & Paul, 2019; and Roch et al. 2011). Additionally, Luque et al. (2018) used instance selection as an alternative to cross-validation.

Other research challenges identified included limited bandwidth (Binder & Paul, 2019; Luque, et al. 2018), threshold problem (Malfante et al. 2018), and lack of general applicability of classifiers (Guilment et al. 2018).

Furthermore, the feature extraction methods, classification techniques and the performance evaluation techniques used in the reviewed articles were identified. It was observed that, although a variety of feature extraction and classification techniques were used, we could not identify unique patterns in the use of these techniques to a particular application domain. However, it was observed that MFCCs were predominantly used in feature extraction for its ability to imitate the hearing properties of the human ear using a nonlinear scale of properties (Mitilineos et al. 2018; Raza et al. 2019; Turner & Joseph, 2015).

We also identified two categories of sound classification, they included detection-and-classification otherwise known as acoustic event detection and detection-by-classification otherwise known as acoustic event classification. 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, followed by localization (Lopatka et al. 2016; Temko & Nadeu, 2009).

Table 0.8: Classification Categories

Category	Domains		
Detection-and-classification	Bioacoustics	Environment	Medicine
	(Malfante et al. 2018), (Thakur et al. 2019), (Briggs et al. 2012), (Briggs et al. 2012), (Robakis et al. 2018), (Ibrahim et al. 2018), (Ya-jie Zhang et al. 2019), (Oikarinen et al. 2019), (Ibrahim et al. 2019), (Guilment et al. 2018), (Ou et al. 2013), (LeBien & Ioup, 2018), (Parada & Cardenal-Lopez, 2014), (Bishop et al. 2019), (Noda et al. 2016)	(Verma et al. 2019)	
Detection-by-classification	(Allen et al. 2017; Aucouturier et al. 2011; Binder & Paul, 2019; Bold et al. 2019; Cvengros et al. 2012b; Gingras & Fitch, 2013; Giret et al. 2011; X. C. Halkias et al. 2013; Kaewtip et al. 2016; Y. Kim et al. 2018; Luque, Romero-Lemos, Carrasco, & Gonzalez-Abril, 2018; Pandeya et al.	(Aziz et al. 2019), (Yan Zhang et al. 2016), (Han et al. 2016), (Su et al. 2019), (Wu et al. 2018), (Salamon & Bello, 2017)	(Chen et al. 2019), (Bourouhou et al. 2019), (Yaseen et al. 2018), (Aykanat et al. 2017), (Raza et al. 2019), (Khamparia et al. 2019), (Vrbancic & Podgorelec, 2018), (Oweis et al. 2015), (Fang et al. 2019)

	2018; Pandeya & Lee, 2018; Qian et al. 2018; Roch et al. 2011; Shamir et al. 2014; Tan et al. 2015)		
--	--	--	--

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 (Ren et al. 2017; Temko & Nadeu, 2009). Table 3.8 shows the researchers classification category according to the application domains earlier identified. Detection-and-classification were performed in the domains of bioacoustics and environment, while detection-by-classification cut across the three identified domains in this review.

Furthermore, it was observed that neural networks were the most used techniques for sound classification. Mitilneos et al. (2018) posits that this is due to the ability of neural networks to identify specific patterns exhibited by sound sources in its distribution of energy over frequency and time.

3.8 CHAPTER SUMMARY

In this chapter, 48 articles in the area of sound classification that were selected based on predefined criteria were reviewed.

From the reviewed articles, predominant sound application domains, feature extraction and, selection methods as well as classification techniques were identified. Two broad categories of sound classification schemes were also identified; acoustic event detection (AED) and acoustic event classification (AEC). The review also highlighted sound classification trends and limitations of existing studies.

Although this review provided methodologies and algorithms used in various domains of sound classification, we opine that the methodologies and research coverage are not exhaustive. Most importantly, we found no study on the detection of extreme events or the automatic sound classification of natural disasters. Consequently, this is one research gap amongst others mentioned in the discussions that provide us with a good justification for the relevance of this study.

Subsequent chapters will seek to address this research gap by providing methodologies and techniques for the classification of an acoustic event such as natural disaster.

Chapter Four

RESEARCH METHODOLOGY

4.1 CHAPTER OVERVIEW

This chapter will discuss the steps adopted in conducting this study. More particularly, it will discuss the design science research methodology as a research paradigm deemed appropriate for this study.

4.2 DESIGN SCIENCE RESEARCH METHODOLOGY (DSRM)

A research work should address specific issues by developing and evaluating artefacts designed to meet identified scientific or business needs (Carcary, 2011; Hevner et al. 2004; Winter, 2008). These artefacts may include but not limited to models, frameworks, methods, constructs, instantiations, and social innovations. Hevner et al. (2004) posits that the artefact must adequately coincide with the real world, it should solve a problem, and should also be able to present the steps, findings, and results clearly and concisely.

In this study, the design science research methodology (DSRM) was adopted because it ensures a relevant and rigorous research based on a set of guidelines (see Table 4.1) as proposed by (Hevner et al. 2004).

Also, the DSRM was deemed appropriate based on the aim of this study. This study is aimed at developing a model for the automatic classification of natural disasters sound as a means of providing real-time detection and warning signals to people. Accordingly, the model to be developed is an artefact that seeks to address the limitations and cover research gaps of existing natural disasters detection and warning methods. It is expected that the artefact (model) that is developed at the end of this study will be novel as it is aimed at addressing the challenges in existing natural disaster detection systems.

Table 0.1: Design Science Research (DSR) Guidelines

Guideline	Description
Design as an Artefact	DSR must produce a viable artefact.
Relevance of the Problem	The objective of a DSR is to develop technology-based solutions to important and relevant problems.
Evaluation of the Design	The utility, quality and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.
Research Contributions	An effective DSR must specify clear and verifiable contributions in the area of the design artefact or design methodologies.
Research Rigor	Design science relies upon the application of rigorous methods in both construction and evaluation of the designed artefact.
Design as a Search Process	The search for an efficient artefact requires the utilization of available means to reach the desired end but at the same time satisfying the rules in the problem domain.
Communication of Research	DSR must be presented effectively to both technology-oriented and management-oriented audience.

4.3 RESEARCH APPROACH FOR THIS STUDY

To ensure a rigorous and appropriate methodology as described by Hevner et al. (2004), this study is divided into five main phases; awareness, suggestion, development, evaluation and conclusion. This five-phase research cycle for a design science research model has been adopted in several studies including Peffers et al. (2008) and Van der Merwe et al. (2020). The research process begins with the awareness that a problem exists. Suggestions to the identified

problem are made based on existing knowledge. An attempt is made at developing an artefact (a solution to the problem). After which the artefacts are evaluated, and conclusions are drawn. The ensuing sub-sections are discussions on how each phase of the process addressed the objectives of this study.

4.3.1 AWARENESS OF THE PROBLEM

The awareness of the research problem in this study was triggered by the alarming increase in the occurrence of natural disasters as well as the false alarm signals. As mentioned in chapter one, this study aims to develop a model (artefact) that can be used for the automatic classification of acoustic events. The model will seek to be sufficiently robust to the changing ambient noise as well as low frequency sounds produced by natural disasters especially during formation. Accordingly, the result of this study will be a purposeful artefact developed to address the current existing problem.

4.3.2 SUGGESTION

In this study, the suggestion phase involved an analysis of literature related to natural disasters and sound classification. Relevant studies were selected and reviewed through a literature review and a systematic review of literature. These reviews facilitated the identification of research gaps as well as existing methodologies that can be enhanced to serve the goal of this study. More particularly, it identified the effectiveness of deep learning techniques in the classification of acoustic events. It also identified the inadequacies in the use of images, text, and numerical data to detect natural disasters and highlighted the potentials in using sound.

This phase is reported in chapters two and three and it addresses the first and second objectives of this study.

4.3.3 DEVELOPMENT

The development phase involves building and training a model that will automatically classify an acoustic event. Models are trained and developed in this stage using a convolutional neural

network (CNN) and a recurrent neural network (RNN). The development phase is reported in chapter five and it addresses the third objective of this study.

i. **CONVOLUTIONAL NEURAL NETWORK (CNN)**

Convolutional neural networks are the most popular type of neural networks. It is inspired by the primary visual system of the brain; hence it is particularly tailored for image recognition and classification. Typically, CNN works with two-dimensional (2D) convolution operation (Maccagno et al. 2019). A convolution is a mathematical procedure that defines a rule of how two functions (the input data and a convolution kernel) will be mixed to produce a transformed feature map (the integral) (Coşkun et al. 2017). A CNN is made up of three main layers, each of the layers performs different sets of tasks on the input data and, the layers also have different optimized parameters. The CNN layers include the convolutional layer, pooling layer and a fully connected layer as shown in Figure 4.1.

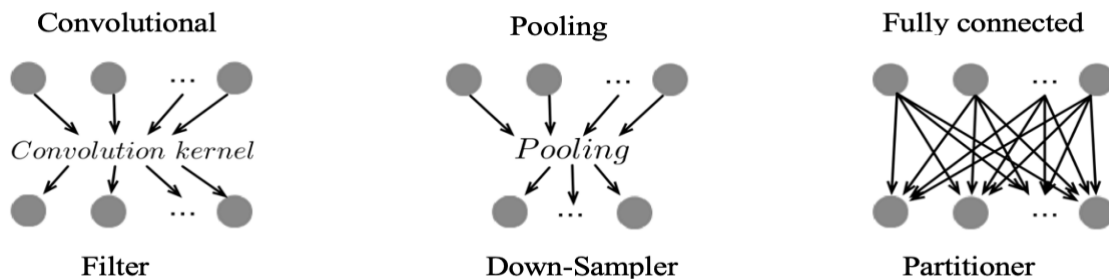


Figure 0.1: Layers of CNN (Borgne & Bontempi, 2017)

CONVOLUTION LAYER

The convolutional layer otherwise called filter is the first layer in which features are extracted from the input data. It is also where most of the user-specified parameters are in the network. The three layers of a CNN have the input say, x arranged in three dimensions, $p \times q \times r$, where p and q are the height and width of the input and r is the depth. In each convolution layer, several filters or kernels k of size exist; in the kernel k of size $m \times m \times n$, m is always smaller than the size of the input data while n can either be smaller or equal to the size of r . Furthermore, the filters which form the base of a connection convolve with the input, share the

same parameters (in terms of weight W^k and bias b^k) and then generate k feature maps (h^k) of size $p - m - 1$. The convolutional layer also calculates a dot function between the weight and the inputs, thereafter, an activation function f is applied to the output of the layers.

POOLING LAYER

Similar to the convolution layer is the pooling layer; it is used to reduce the dimensionality of the network by reducing the number of parameters if the input data is large. More particularly, it decreases the number of parameters in the network, speeds up the training process and also controls overfitting by downsampling each feature map in the network. Three basic operations are performed in the pooling layer, they include, max pooling which takes the largest value in a defined filter region, average pooling which takes the average value and sum pooling which takes the sum of all the values in the defined filter region. Generally, the pooling operations are performed over a specified contiguous region for all feature maps in the network.

FULLY CONNECTED LAYER

The fully connected layer (partitioner) is the last layer of the network, placed before the classification output of a CNN. It uses previous low-level and mid-level features to generate high-level abstraction from the data. It outputs the probability that an input belongs to a certain class for a given instance.

Furthermore, these three layers have three sets of features categorized into; action, parameters and input/output. Table 4.2 summarizes these features with respect to the CNN layers.

Table 0.2: Features of the 3-layers in a CNN

ACTION	PARAMETERS	INPUT/OUTPUT
Convolutional layers		
<ul style="list-style-type: none"> - Filters are applied to extract features. - Filters are composed of small learned kernels. - Activation functions are applied on every value of the feature map 	<ul style="list-style-type: none"> - Number of kernels. - Size of the kernels. - Activation functions. - Stride - Padding - Regularization type and value. 	<ul style="list-style-type: none"> - Input: 3D cube, previous set of feature maps. - Output: 3D cube, one 2D map per filter.
Pooling layer		
<ul style="list-style-type: none"> - Dimensionality reduction. - Extraction of maximum, average or sum of a specified region. - It uses the sliding window approach. 	<ul style="list-style-type: none"> Strides and window size. 	<ul style="list-style-type: none"> - Input: 3D cube, previous set of feature maps. - Output: 3D cube, one 2D map per filter, reduced spatial dimensions.
Fully connected layers		
<ul style="list-style-type: none"> - Combines information from final feature maps. 	<ul style="list-style-type: none"> - Number of nodes. - Activation function: uses either ReLU for aggregating 	<ul style="list-style-type: none"> - Input: flattened 3D cube, and previous set of feature maps.

- Develops final classification.	final	or SoftMax for producing a final classification.	- Output: 3D cube, and one 2D map per filter
----------------------------------	-------	--	--

ii. RECURRENT NEURAL NETWORK (RNN)

Recurrent neural network (RNN) is another popular class of neural networks that are predominant in the fields of natural language processing (NLP) and speech processing.

A major feature of a RNN that distinguishes it from other neural networks is that the network contains at least a feed-back connection which enables the network to perform temporal processing as well as learn sequential information (Pouyanfar et al. 2018). It uses the sequential characteristics of data and its patterns to make predictions.

Also, in a traditional neural network, the inputs and outputs are independent of each other, whereas a Recurrent Neural Network uses the output from a previous step to make input to the current step.

Table 0.3: Calculating the current state, activation functions and output in RNN

		Formula	Variable definitions
1.	Current state	$h_t = f(h_{t-1}, X_t)$	h_t is the current state, h_{t-1} is the previous state, X_t is the input state.
2.	Activation function	$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$	W_{xh} is the weight of input neuron, W_{hh} is the weight at the recurrent neuron.
3.	Output	$Y_t = W_{hy}h_t$	Y_t is the output, W_{hy} is the weight at the output layer.

TRAINING A RECURRENT NEURAL NETWORK

The recurrent neural network consists of the input layer ($X_0, X_1, X_2, X_3, \dots, X_t$), the hidden layers ($h_0, h_1, h_2, h_3, \dots, h_t$) and the output layers ($y_0, y_1, y_2, y_3, \dots, y_t$). The current state, activation function and output are calculated as shown in Table 4.3.

The steps for training a Recurrent Neural Network are as follows:

- i. In the input layer, load or send the initial inputs with equal weights and activation functions.
- ii. Calculate the current state using the current inputs and the previous state output.
- iii. Current state h_t , will become h_{t-1} for the second time step.
- iv. To solve a particular problem, this process keeps repeating for all the steps.
- v. Calculate the final step with the current state of the final step as well as other previous steps.
- vi. Generate the error by calculating the difference between the actual output and the output generated by the RNN model.
- vii. Steps end when the process of backpropagation occurs such that the error is backpropagated to update the weights.

4.3.4 EVALUATION

The performance of the developed models is evaluated in this phase. In this study, the cross-validation and classification metrics will be used to validate the performance of the models.

Cross-validation is a re-sampling method that is used to evaluate and validate the performance of a classification algorithm. It reduces possible bias that might result from the training/testing split on a specific dataset, and also increase model reliability as it checks to ensure that a model is not overfitting (Jacoby, 2014; Raza et al. 2019; Sayad et al. 2019; Wang & Peng, 2018). The cross-validation process involves a cross-over in successive iterations between the training and testing sets in order to validate the model. There are different cross-validation methods, they

include K-Fold cross-validation, Stratified K-Fold cross validation, leave out one, shuffle split and adversarial validation. The K-Fold cross-validation is used in this study.

The classification metrics used include accuracy, precision, recall and AUC (area under curve) score.

The evaluation of the models is reported in chapter six and it addresses the fourth objectives of this study.

4.3.5 CONCLUSION

This is the final stage of this research cycle and it is expected to serve as a knowledge base for future research. In this stage, summaries, limitations of the study, discussions based on findings are provided. The concluding phase is reported in chapter seven.

4.4 CHAPTER SUMMARY

This chapter discussed the design science research methodology as an appropriate paradigm for this study. It also discussed the five-phase research cycle, and how each phase of the cycle is related to the research objectives stated in chapter one.

Chapter Five

USING DEEP LEARNING FOR ACOUSTIC EVENT CLASSIFICATION

5.1 CHAPTER OVERVIEW

From the systematic review in chapter three, two broad categories of sound classification schemes were identified; acoustic event detection and acoustic event classification. In this chapter, the acoustic event classification (AEC) approach will be explored. Hence, the methodologies and techniques adopted in conducting this study will be discussed. It is important to note that most of the methods mentioned and adopted in literature for the classification of an acoustic event are borrowed from the speech recognition system.

The framework for this chapter is summarized in Figure 5.1. It begins with extracting the sounds of interest, pre-processing the sound, extracting relevant features, building the classification model and then model validation.

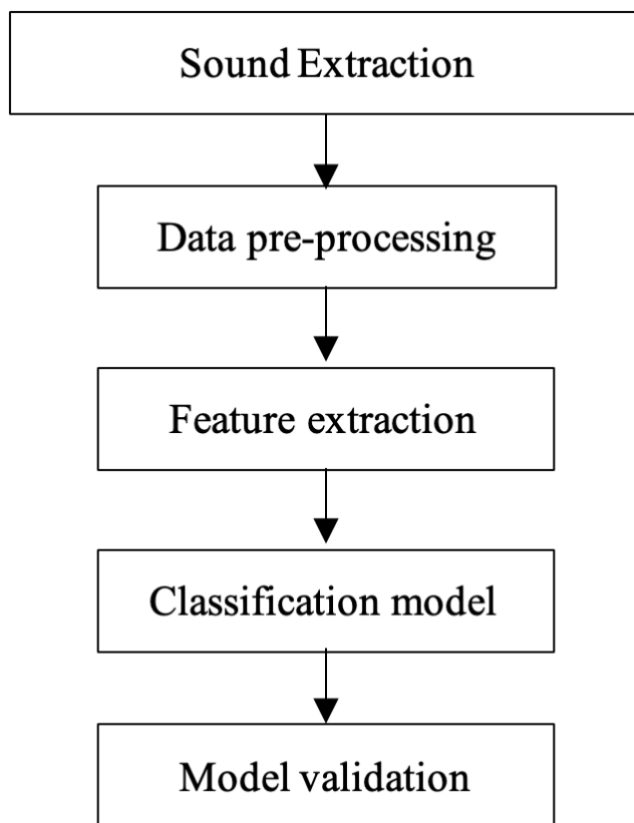


Figure 0.1: Sound classification architecture

5.2 SOFTWARE USED FOR THE EXPERIMENT

In this study, Anaconda Python and various libraries were used to build and train the neural networks as well as for feature extraction. Apart from the general python libraries used for data processing and analysis such as Numpy, Matplotlib, Scikit-Learn. Other specific libraries used for the experiment included;

- i. Keras in this experiment was used on top of TensorFlow GPU to build the neural networks.
- ii. The audio analysis library LibROSA was used to read in data and for resampling the audio files.
- iii. Python_speech_features was used for providing speech features such as Mel frequency cepstral coefficient (MFCC) and filter-banks.
- iv. Tqdm creates a progress path for nested loops.

5.3 NATURAL DISASTER SOUND DATASET

Data used for this study were extracted from the Freesound database (freesoundeffects.com/free-sounds/ambience-10005/). All the sound recordings are in WAV format. The sound format is relevant because the features extracted in this study (MFCC) supports WAV sound formats (Sasmaz & Tek, 2018).

The datasets are made up of five classes of unevenly distributed disaster sounds (class imbalance) namely, earthquake, windstorm, waves, forest fire and volcano. The five classes are made up of a total number of 244 sound recordings and total duration of 1560.52694 seconds. The pie chart in Figure 5.2 shows the distribution of the disaster sound dataset where each class is named according to the type of the disaster.

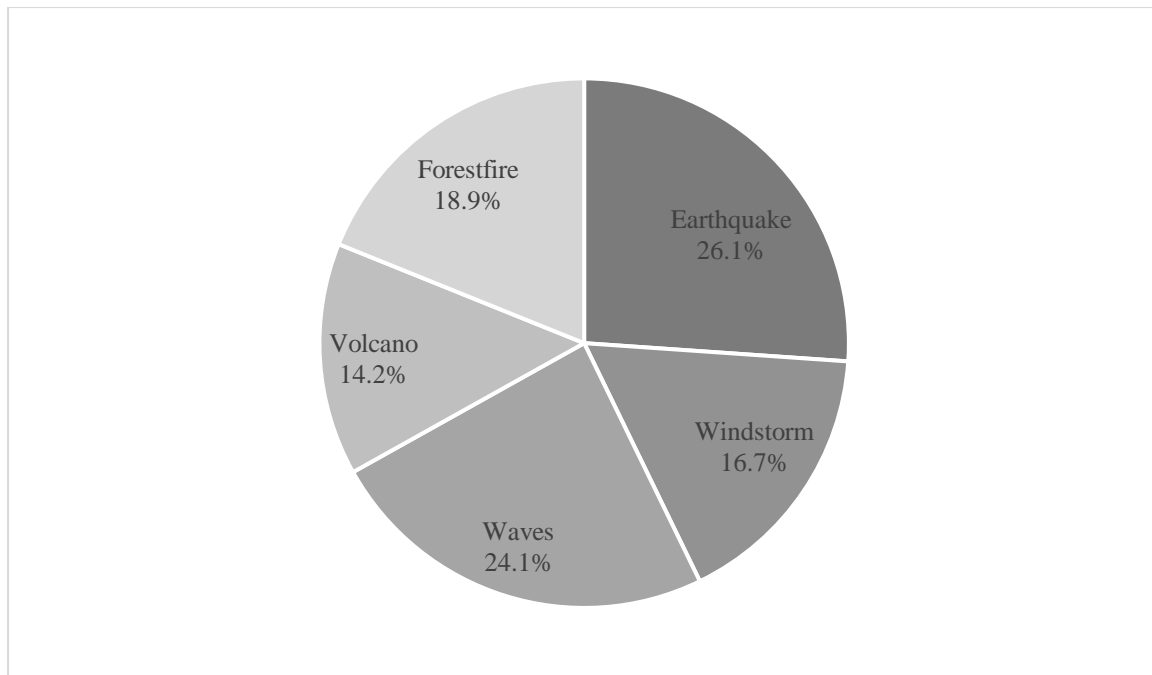


Figure 0.2: Class distribution of disaster sound dataset

5.4 SOUND PREPROCESSING AND FEATURE EXTRACTION

Sound/acoustic signals can be visualized in two forms, one as a wave plot (time-domain representation) and the other as a spectrogram (frequency domain representation). A wave plot is an amplitude versus time plot that shows the loudness of sound waves as it changes over time (see Figure 5.3). However, due to the varying amplitudes, there is the tendency that the classifiers can misjudge the magnitudes of sound intensity during learning (Kim et al. 2018b; Oikarinen et al. 2019; Pramono et al. 2017). Findings from the systematic review in chapter three showed that spectrograms, and spectral features such as Mel frequency cepstral coefficients (MFCC) were widely used features in sound classification that can enable the classifiers to learn discriminative features (Guilment et al. 2018; Halkias et al. 2013; Salamon & Bello, 2017). Spectral features generate sound waves that produce more accurate results in noisy conditions. Spectrograms on the other hand, reduces the number of trainable parameters in contrast to direct sound classification (Khamparia et al. 2019).

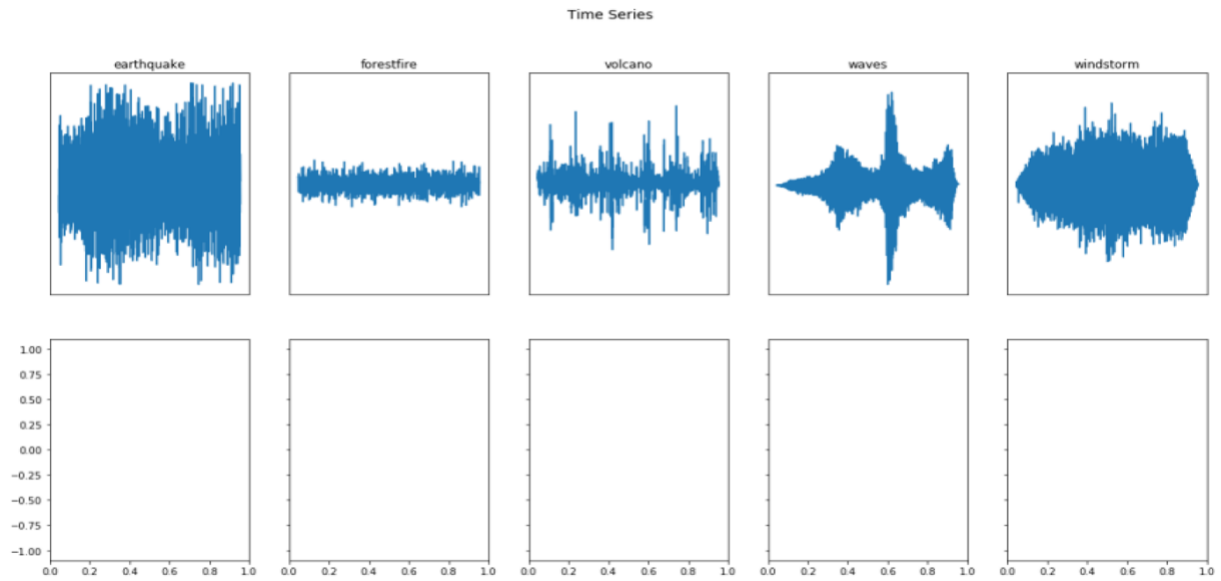


Figure 0.3: Time series representation of five random samples belonging to the five different classes of the dataset

Since this study aims to classify sounds amidst noisy conditions, spectral features (which are usually obtained by converting the time-based signal into the frequency domain using Fourier transforms) instead of spectrograms will be adopted. Figure 5.4 and 5.5 shows the visual representation of sound from each class of the dataset using MFCCs and filter bank coefficients.

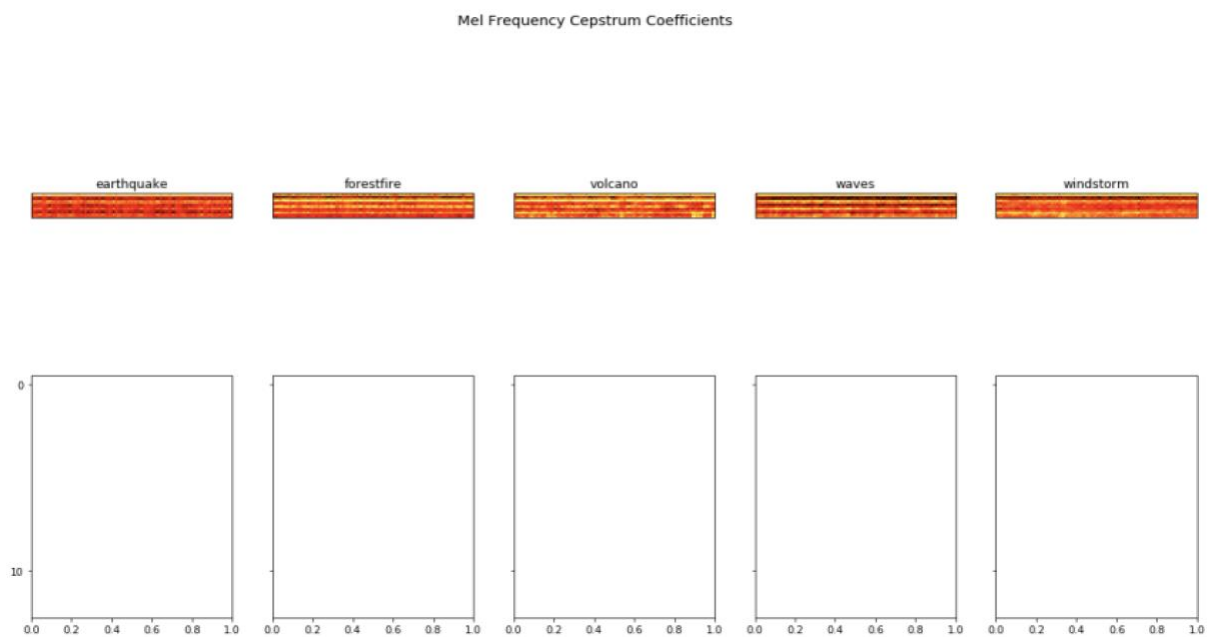


Figure 0.4: MFCC representation of five random samples belonging to the five different classes of the dataset

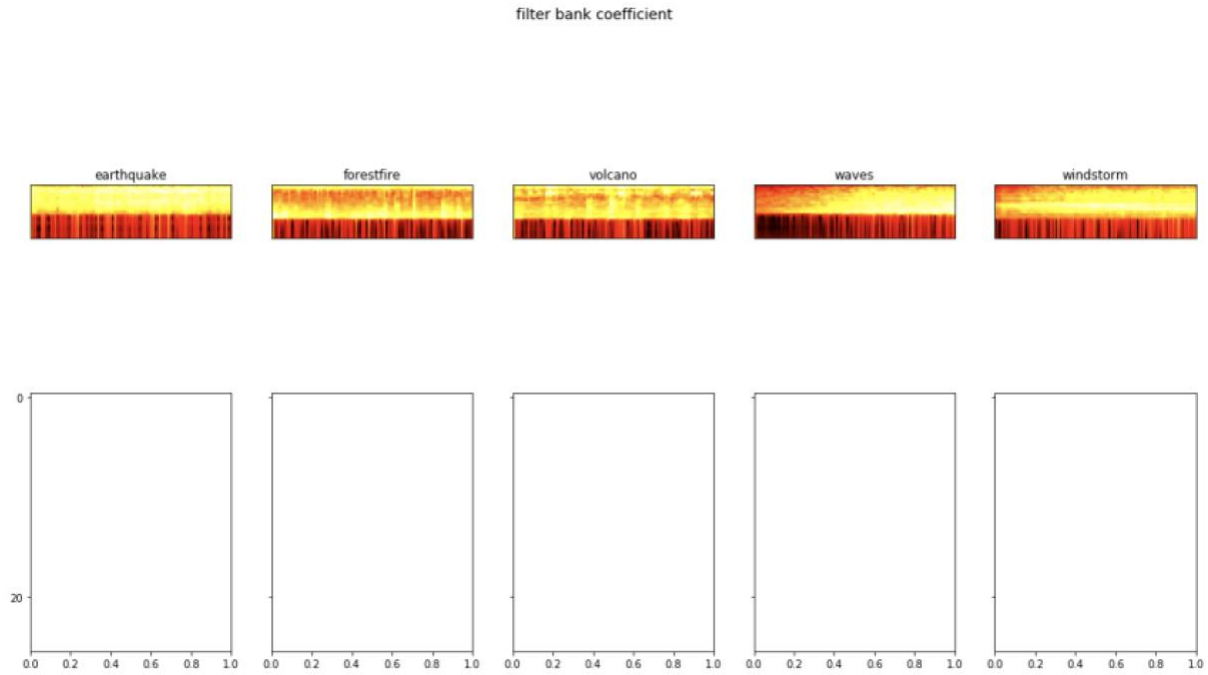


Figure 0.5: Filter bank coefficient representation of five random samples belonging to the five different classes of the dataset

5.4.1 DE-NOISING THE SIGNAL

It was observed that there were lots of dead spaces which can be said to contain irrelevant data within the audio files. To get rid of these dead spaces, we calculated and created an envelope of the signal. An envelope of a signal defines the boundary (in most cases the upper boundary) within which a signal is contained when viewed in the time domain.

This was achieved by passing in a signal with a collection rate and a specified threshold value of 0.005. To avoid getting rid of relevant data, a rolling window with a tenth of seconds (0.1seconds) size was generated over the data which uses the mean of all the signal values to identify signal values that are growing dead or fading out. Accordingly, an aggregated mean of all the values in a window was generated and used to develop a mask over the signals, for mean values greater than the specified threshold. Consequently, the dataset now consists of signals with relevant data; this can be observed by slight changes in the class distribution (see Figure 5.6).

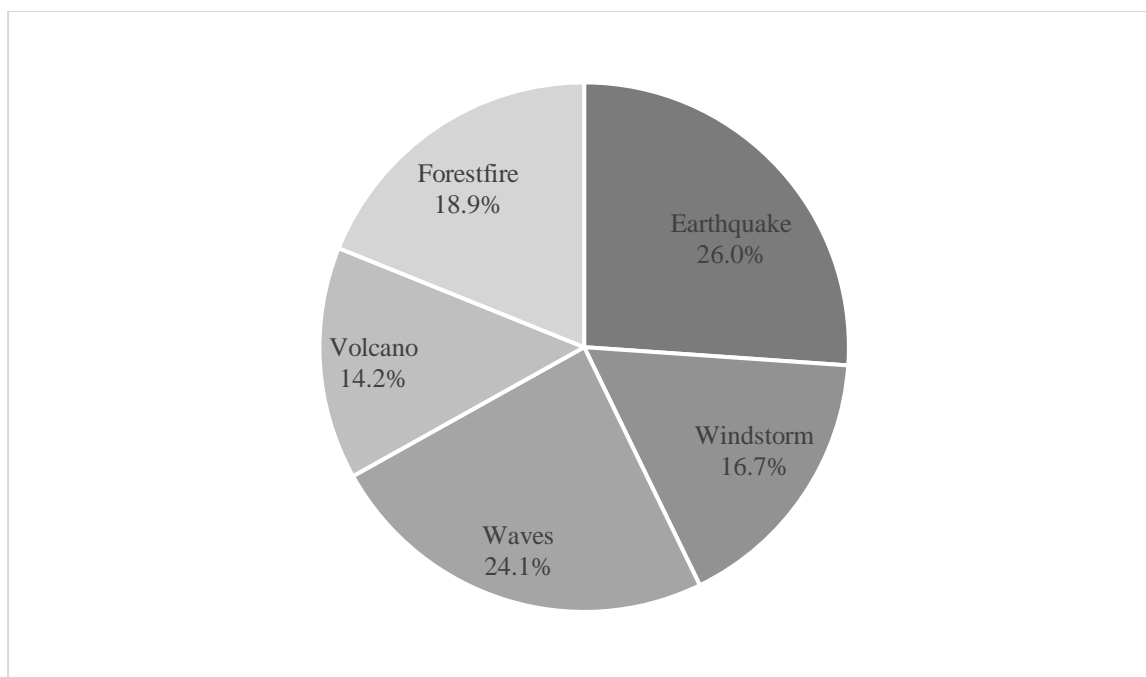


Figure 0.6: Pie-chart showing the class distribution of denoised disaster sound dataset

5.4.2 ACOUSTIC DOWN-SAMPLING

Down-sampling is a signal reduction technique that reduces the sound features by extracting more discriminative features which are in turn used for modelling (Raza et al. 2019). Considering that natural disasters generate low-frequency noise particularly at the formation stage and this study is focused on early detection of the signals generated by these disasters, it was imperative to remove any form of redundancy from the data. Hence, with a window size of 25ms, the audio signals were downsampled from 44100Hz to 16000Hz. Accordingly, a clean directory was created to store the cleaned-up audio files. Data from this directory will be used for the classification.

5.4.3 FILTER BANK-BASED FEATURE EXTRACTION METHOD

A fundamental aspect in the design of an acoustic event classification system is in the selection and extraction of appropriate signal features that will enhance the efficient differentiation between different types of sound signals. The selection of appropriate features is essential because recorded sounds are generally non-stationary signals with super-imposed background noises that originates from natural ambient noise (Mitilineos et al. 2018). Commonly used

sound features are either related to the time-domain representation (time-frequency features), to the frequency domain representation (spectral features) or to statistical features.

Since seismic activities have wide range of spectral contents (Simmonds & MacLennan, 2005), the feature extraction in this study is performed on spectral features of the audio recordings based on filter bank-based Mel Frequency Cepstral Coefficient approach. Although there are various approaches for the filter-bank feature extraction methods used in the areas of speech/sound feature extraction, Mel Frequency Cepstral Coefficients (MFCC) have been predominantly used in audio-specific feature extraction. This is mainly because it is robust to noise and yields high performance in speech signal processing (Aykanat et al. 2017; Aziz et al. 2019; Bishop et al. 2019; Chen et al. 2019; Turner & Joseph, 2015; Yaseen et al. 2018). It is also commonly used for its ability to leverage on the robustness of CNN based classifiers (Verma et al. 2019). Accordingly, MFCC method of feature extraction is adopted in this study. Due to the downsampling earlier performed, the number of MFCC features was reduced from 26 to 13 features, while FFT was reduced from 1103 to 512 (Aziz et al. 2019; Chu et al. 2009). Accordingly, the 13 MFCCs is calculated using a downsampled 512-point fast Fourier transform with 0.1s window frame length. More specifically, the fast Fourier transform (FFT) is applied to the waveforms of each window frame, then the log Mel-filter bank spectrum is extracted to obtain spectral reports of each frame. Furthermore, the obtained spectrum is tied to a Mel-frequency scale by passing it through a Mel-filter bank created by triangle filters. The logs of the outputs are used to develop the log Mel-filter bank spectrum for each frame and the MFCC is finally obtained by applying a Discrete Cosine Transform (DCT) to the filter banks (Yaseen et al. 2018).

5.5 CLASSIFICATION TECHNIQUES

Two deep learning algorithms; Convolutional Neural Networks (CNN) and Recurrent Neural Network (RNN) have been adopted in this study for the classification tasks. Before developing

the classification models, it is important that the class imbalance situation is considered. This is because imbalances may result in low performance of some classes and also make the process of building a classification model difficult (Imran et al. 2017; Ya-jie Zhang et al. 2019). Also, this study is particularly focused on detection-by-classification. Sound detection-by-classification is predominantly concerned with the choice of the window length, this window length can be any arbitrary value from half a second to several minutes depending on the task application domain (Temko & Nadeu, 2009). Accordingly, an arbitrary length of time of a tenth of a second (0.1seconds) was chosen. Then a random sampling along the length of the audio files was performed to extract chunks of 0.1second. To determine the total number of audio samples ($n_{samples}$) generated within the signal after the extraction, the total length in seconds of all the data was divided by 0.1second then multiplied the results by 2.

That is, $[n_{samples} = 2x]$, where x in the python code is defined as; $x = int(sounds['length'].sum()/0.1)$.

Consequently, this process increased the number of samples from 244 to 31194; a process generally termed data augmentation. Although the 0.1second time frame may be described as too short, it was chosen to ensure that the model can quickly discern different classifications in real-time. Finally, the datasets were split into two sets; 80% for training and 20% for testing.

5.5.1 CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is one of the most competitive neural networks applied in computer vision for image classification and recognition. In this study, the CNN was adopted specifically for the following reasons: CNNs can capture patterns across time and frequency for given input spectrograms (Maccagno et al. 2019). Also, it can make distinctions even when sound is masked in time and frequency by other noise (Salamon & Bello, 2017).

The CNN model used in this experiment consists of 4 fully connected layers. We used 16 filters built with a 3×3 convolution, all the layers have a ReLU activation function with 1×1 stride

(because of the small input space) and padding as ‘same’, a 2x2 kernel for the maxpooling, and a dropout layer to reduce overfitting on the training data. The convolutional model dimensions used in the study is shown in Figure 5.8.

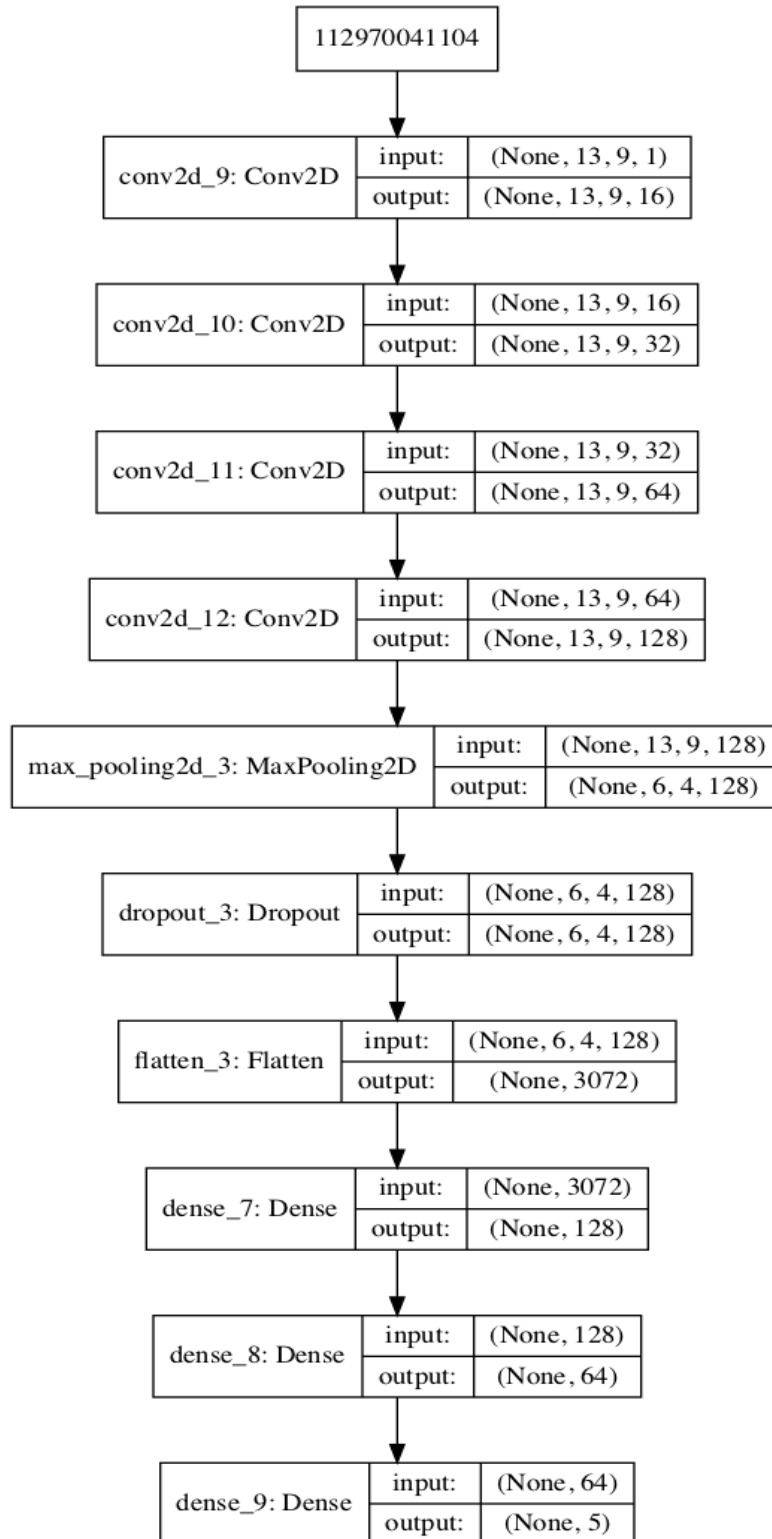


Figure 0.7: CNN model dimensions

5.5.2 RECURRENT NEURAL NETWORK

Recurrent neural network (RNN) is another type of deep neural network that learns the significant features from an upcoming data sequence, stores it in memory cells and then predicts the next steps based on the stored up features (Verma et al. 2019). RNN is used in this study because it performs well on time series data by producing a 1D array consisting of frequency values (Raza et al. 2019; Verma et al. 2019). In order words, RNNs are used to model features that change over time. Generally, RNNs are efficient in tasks that involve sequential inputs such as speech and language, yet it is limited by the inability to store the learned features for a long time (Lecun et al. 2015). Accordingly, the use of LSTM to mitigate this challenge has been widely adopted (Verma et al. 2019).

This study adopts Raza et al. (2019)'s RNN model architecture presented in Figure 5.8.

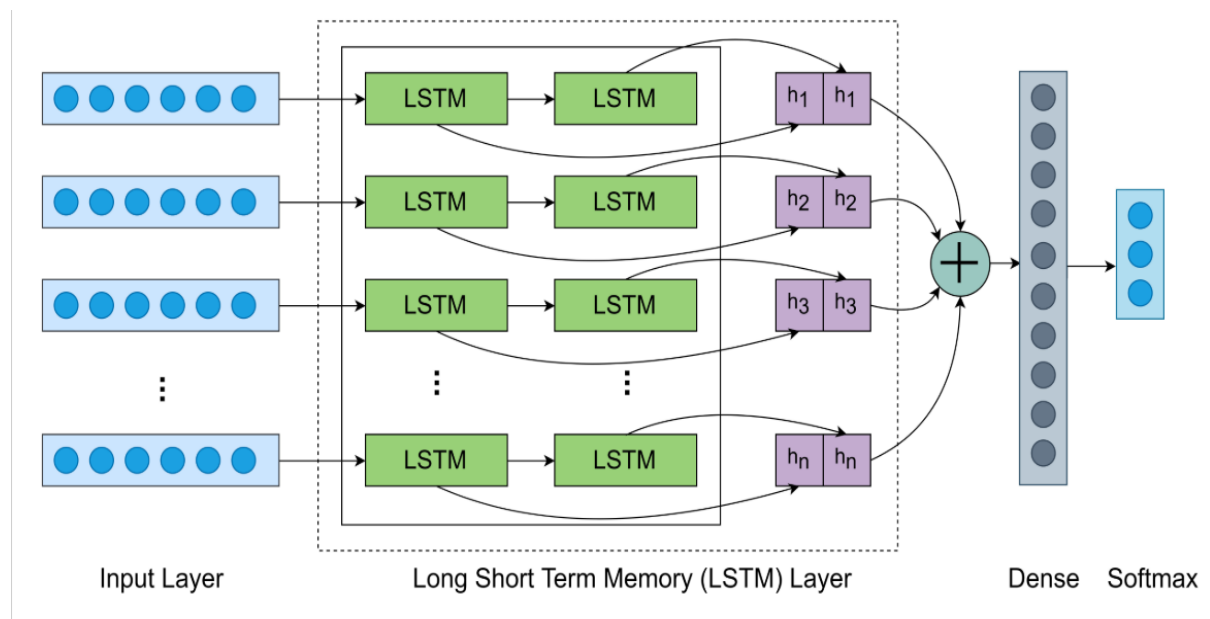


Figure 0.8: RNN-LSTM Architecture (Raza et al. 2019)

The RNN model in this study is LSTM based with a data shape of (n, time, feat), one recurrent layer, a 0.5 Dropout, a time distributed fully connected layer with 64 neurons and a ReLU activation, flatten, and a SoftMax activation. Since sequences are returned, the time distributed is carried down from layer to layer. Hence, more parameters can be created while enabling

deeper modelling. The RNN-LSTM model dimensions used in this study is shown in Figure 5.9.

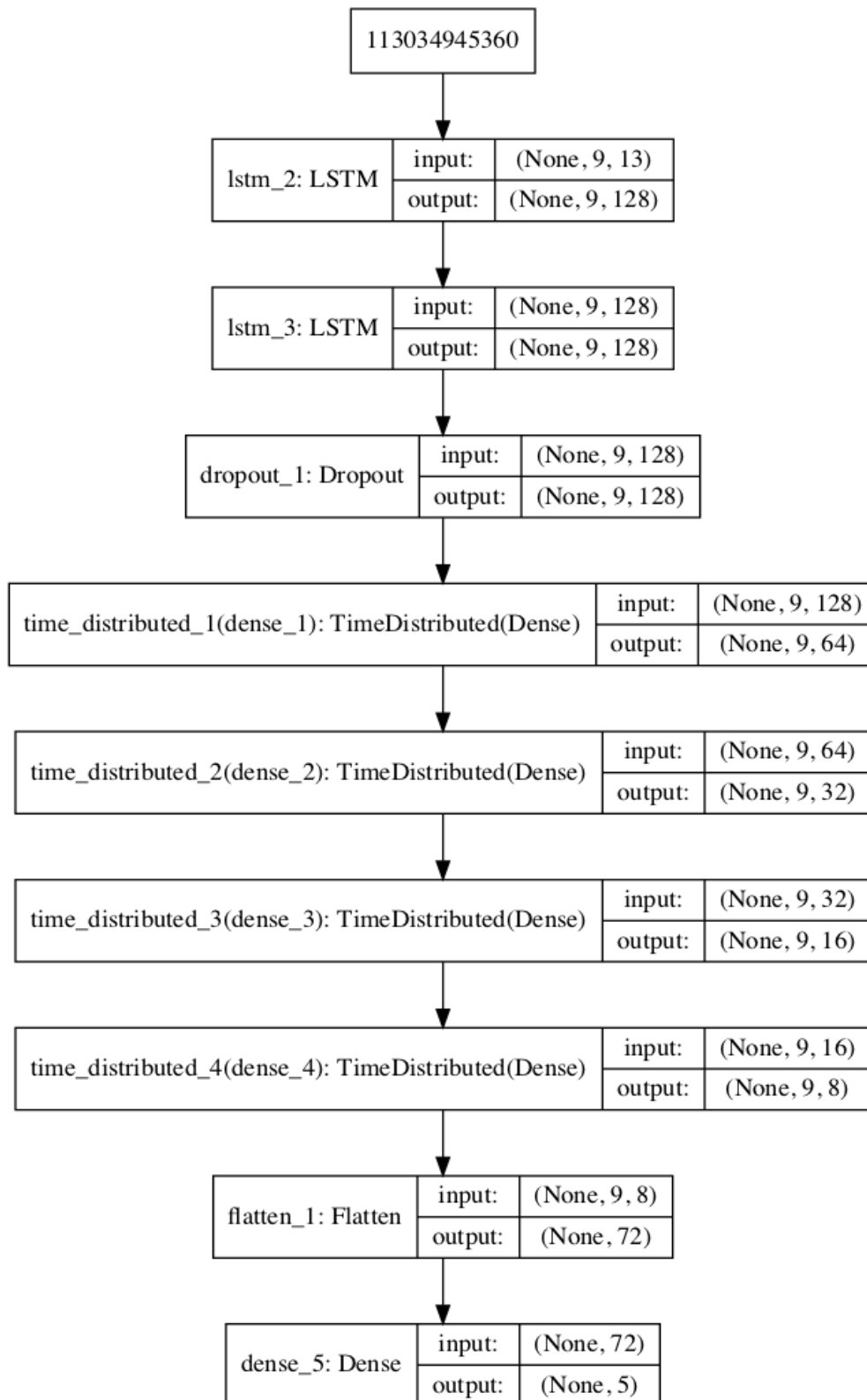


Figure 0.9: RNN-LSTM model dimensions

5.6 CHAPTER SUMMARY

This chapter provided the steps involved in the classification of sound. It is worthy to note that all the methodologies described in this chapter were fully automated with little or no form of human intervention.

After sound extraction from Freesound database, the methodology was made up of three main steps: pre-processing, feature extraction and classification. Data preprocessing entailed denoising and downsampling the audio recordings as well as augmenting the data size. The filter bank-based Mel Frequency Cepstral Coefficient approach was used for feature extraction, after which the dataset was split into train and test sets. Classification models namely CNN and RNN-LSTM were then built and applied to the preprocessed data to enable predictions.

The next chapter will evaluate the performance of the two models, numerical results will be evaluated in terms of accuracy, precision, recall and area under curve (AUC) score.

Chapter Six

EVALUATION OF DEEP LEARNING TECHNIQUES

6.1 CHAPTER OVERVIEW

This chapter presents a comparison of the results obtained from the experiment conducted in chapter five. It discusses the model validation techniques and the metrics used in the classification of an acoustic event. Results are displayed as bar charts, tables and confusion matrix (5x5 contingency table).

6.2 MODEL VALIDATION

Model validation is performed after the model has been trained, it aims to find the model with the best performance. More particularly, model validation is the process whereby a trained model is evaluated with a separate portion of the same dataset commonly referred to as the testing data (Gingras & Fitch, 2013; Lebien & Ioup, 2018; Sayad et al. 2019). Amongst the different validation techniques, this study will use cross-validation (Krishna et al. 2018; Pandeya & Lee, 2018; Su et al. 2019) and classification metrics (Luque, Romero-Lemos, Carrasco, & Gonzalez-Abril, 2018; Sayad et al. 2019).

6.2.1 CROSS-VALIDATION

In the K-fold cross-validation, the entire training dataset is divided into K-subsets such that in each iteration, all the subsets in the datasets are trained except one subset 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. Previous studies have argued that, the value of K is predominantly either 5 (5-fold cross-validation) or 10 (10-fold cross-validation). However, the 10-fold cross-validation has been adopted by several researchers for its ability to produce better performance of model hyper-parameters (Chen et al. 2017; Davis & Suresh, 2019; Han et al. 2016; Pandeya & Lee, 2018; Thakur et al.

2019). Thus, the 10-fold cross-validation technique is adopted in this study to evaluate the performance of the models. Table 6.1 shows an illustration of the 10-fold cross-validation process.

Table 0.1: 10-Fold cross-validation

Split 1	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 1
Split 2	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 2
Split 3	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 3
Split 4	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 4
Split 5	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 5
Split 6	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 6
Split 7	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 7
Split 8	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 8
Split 9	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 9
Split 10	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Metric 10

Training data

Test data

Maintaining the initial 80% and 20% training/test splits, the training datasets (80%) was split into 10 folds; such that for each iteration, 9 of 10 folds of sound recordings are selected for the training the model, then the trained model is tested on the remaining one-fold (holdout set). This process is repeated 10 times with a different holdout set in each iteration.

Figure 6.2 shows the classification accuracy and average accuracy obtained from the 10-fold cross validation performed on the training set. As shown in Figure 6.1, the minimum classification accuracy obtained from the training set was 99.16% and 97.05% for CNN and RNN-LSTM respectively. The highest accuracy of 100% was achieved at the 9th fold for both classifiers. CNN and RNN-LSTM had an average accuracy of 99.85% and 99.23% respectively on the train set.

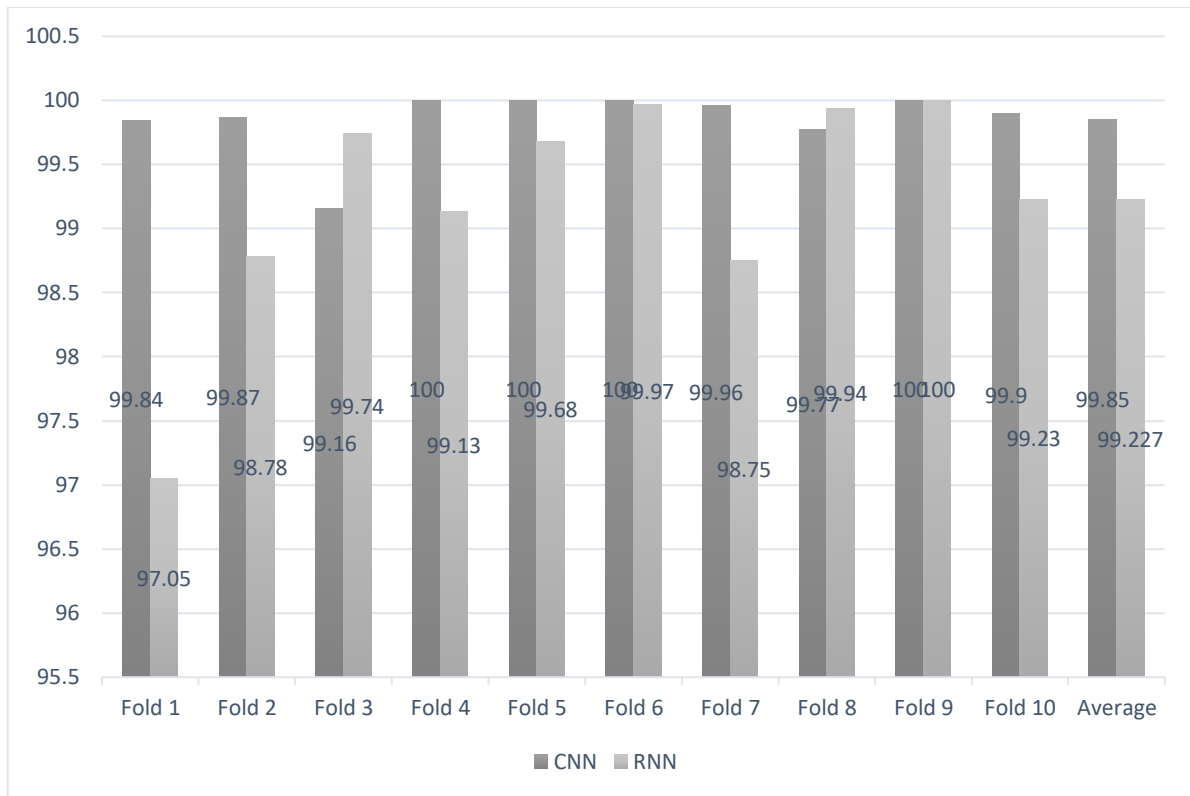


Figure 0.1: Classification Accuracy and average accuracy of the 10-folds

Furthermore, the performance of the 10-fold cross-validation model was validated on the remaining 20% (unseen data) of the dataset. Figure 6.2 shows a bar chart of the classification accuracy obtained from the test set.

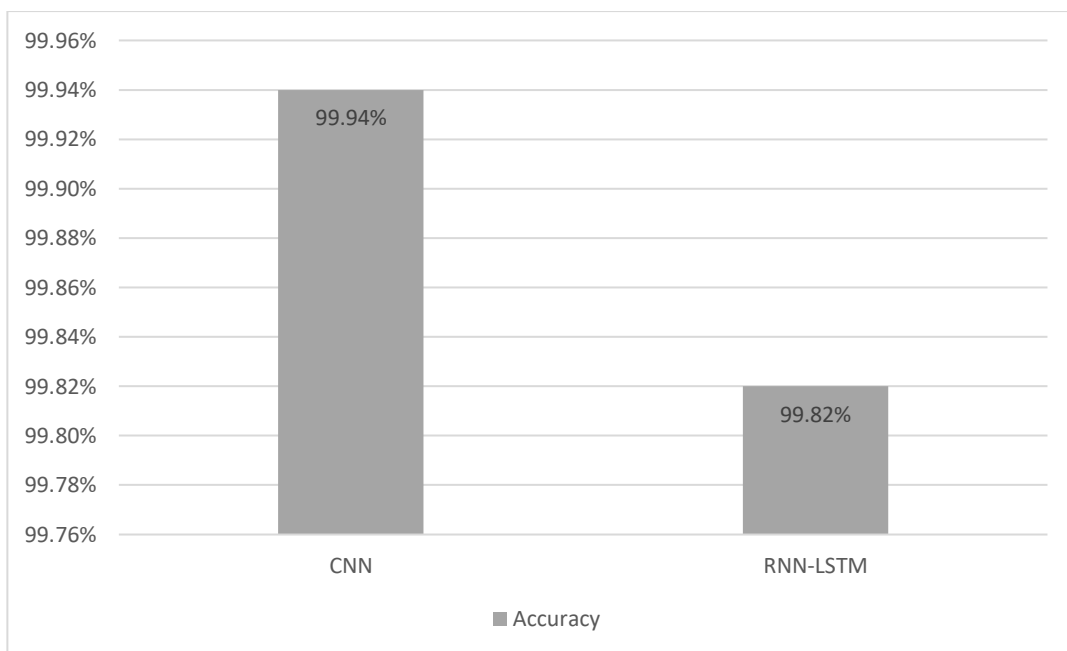


Figure 0.2: Accuracy obtained from 10-fold cross-validation for CNN and RNN-LSTM

Comparing the results from Figure 6.1 (train validation) to the test validation of Figure 6.2, it can be observed that the model performs better on unseen data as it achieved higher accuracies. CNN performed better on the test set by 0.09% and RNN-LSTM performed better by 0.59%.

6.2.2 CLASSIFICATION METRICS

Classification metrics are a set of metrics generally used to evaluate the performance of a model using the test datasets. Twenty percent (20%) of the natural disaster sound dataset was assigned for testing the performance of the models. The metrics used in this study for the evaluation of the CNN and RNN-LSTM models include Confusion Matrix, Accuracy, Precision, Recall, and area under curve (AUC).

i. CONFUSION MATRIX

A confusion matrix is a table that provides a detailed breakdown of the correct (true positives) and incorrect (errors) classifications for each class in a dataset. It allows the visualization of the performance of a model by tabulating the values of the actual and predicted classes as columns and rows respectively. Four terms are commonly associated with confusion matrix, they include

- i. true positives (TP) is obtained when both the actual class and the predicted class is true.
- ii. true negatives (TN) is obtained when both the actual class and the predicted class is false. The total number of true negatives for a certain class is the sum of all the columns and rows excluding that class's column and row.
- iii. false positives (FP) is obtained when the actual class is false, and the predicted class is true. The total number of false positives for a class is the sum of values in the corresponding column excluding the true positive (TP).
- iv. and false negatives (FN) is obtained when the actual class is true, and the predicted class is false. The total number of false negatives for a class is the sum of values in the corresponding row excluding the true positive (TP).

Furthermore, various performance measures or metrics can be calculated based on values from the confusion matrix. The performance of a model can be obtained using the following measures:

- i. Accuracy is the measure of correctly predicted instance to the total instances. It is calculated as; $\frac{(TP+TN)}{TP+FP+FN+TN}$
- ii. Error rate, EER is calculated as; $\frac{(FP+FN)}{(TP+TN+FN+FP)}$
- iii. True Positive Rate (TPR) also known as sensitivity or recall is a measure of the ability of a prediction model to correctly select instances. TPR is calculated as; $\frac{TP}{(TP+FN)}$
- iv. True Negative Rate (TNR) also known as specificity is a measure of negative instance correctly predicted. TNR is calculated as $\frac{TN}{(FP+TN)}$
- v. False Positive Rate (FPR) is the portion of negative samples that are predicted as positive. FPR is calculated as $1 - \textit{specificity}$.
- vi. False Negative Rate (FNR) is the portion of positive samples that are predicted as negative.
- vii. Precision also known as positive predictive value (PPV) is the measure of positive instances that are actually positive. PPV is calculated as; $\frac{TP}{(TP+FP)}$.

In this study, in order to compare the actual classes with the predicted, confusion matrix was generated for each classifier used in this study. Table 6.2 and 6.3 shows the confusion matrix for CNN and RNN-LSTM. Recall from the pie chart in Figure 5.6 that the class distribution was imbalanced; hence the varying numbers of test splits in the different classes of the confusion matrix.

Table 0.2: Confusion matrix showing CNN predictions

	Earthquake	Forestfire	Volcano	Waves	Windstorm
Earthquake	1648	0	0	0	1
Forestfire	0	1219	0	0	0
Volcano	0	0	852	0	0
Waves	0	1	0	1473	0
Windstorm	0	0	0	1	1044

The confusion matrix in Table 6.2 shows predictions made on the test dataset of natural disaster sound using CNN. The test dataset is made up of a total of 6239 sound recordings, out of which 6236 instances were correctly predicted. Confusions were observed in the following instances; earthquake as windstorm (1), waves as forestfire (1), and windstorm as waves (1). Forestfire and volcano were correctly predicted at all instances.

Similarly, Table 6.3 shows the confusion matrix for predictions made using the RNN-LSTM model. Out of the 6,239 sound recordings, 8 errors were recorded across the five classes.

Earthquake and waves were correctly predicted at all instances. Regarding forest fire, 1171 recordings were correctly predicted, while being confused for volcano and waves in 5 and 1 instances respectively. Volcano and windstorm were both confused as earthquake in 1 instance.

Table 0.3: Confusion matrix showing RNN-LSTM predictions

	Earthquake	Forestfire	Volcano	Waves	Windstorm
Earthquake	1637	0	0	0	0
Forestfire	0	1171	5	1	0
Volcano	1	0	893	0	0
Waves	0	0	0	1462	0

Windstorm	1	0	0	0	1068
-----------	---	---	---	---	------

ii. **ACCURACY, PRECISION AND RECALL**

Based on the confusion matrix, classification metrics such as accuracy, precision, and recall for both CNN and RNN-LSTM were computed. It was observed that for each of the classifiers, accuracy, precision and recall produced the same results (see Figure 6.3). As shown in Figure 6.3, although CNN outperformed RNN-LSTM, both classifiers had good accuracy results (99.95% and 99.87%).

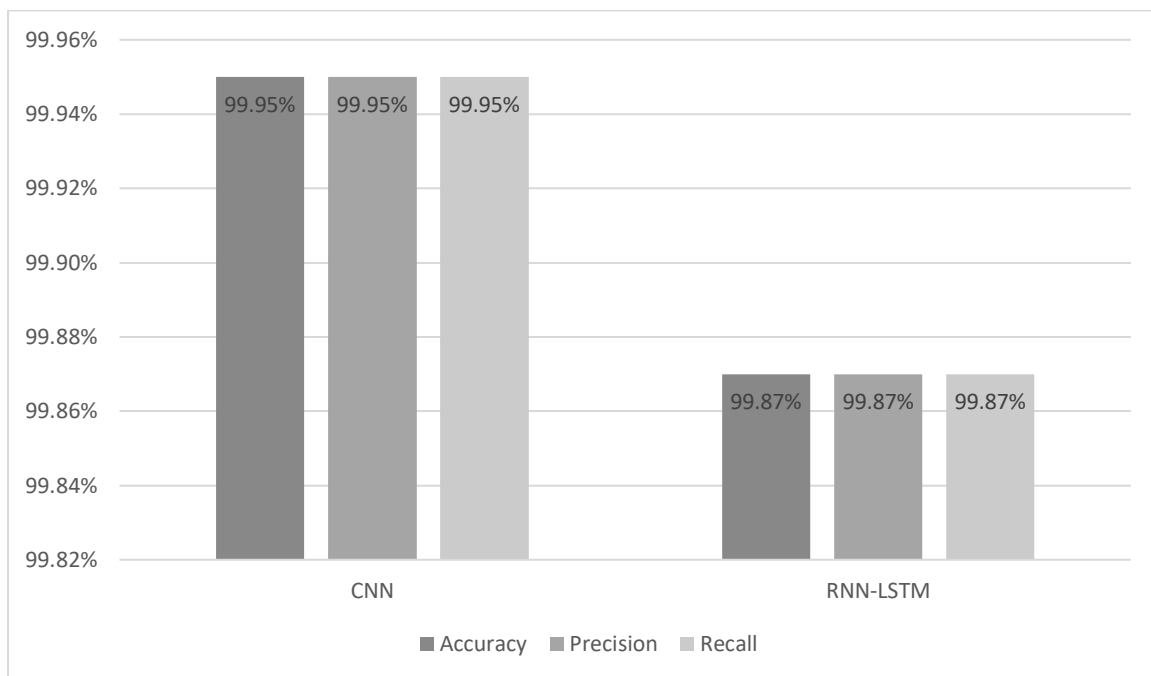


Figure 0.3: Classification accuracy for CNN and RNN-LSTM

iii. **AUC- ROC (AREA UNDER CURVE – RECIEVER OPERATING CHARACTERISTICS) CURVE**

Due to the dominating effect of the majority class in an imbalanced dataset, classification metrics such as accuracy, precision and recall are not sufficient for evaluating the performance of a model (Luque, Romero-Lemos, Carrasco, & Gonzalez-Abril, 2018; Weng & Poon, 2008; Yang et al. 2015). Thus, the AUC-ROC as a visualization tool for comparing classification

models has been used to mitigate the dominating effect of dataset imbalance (Ling et al. 2003; Weng & Poon, 2008; Yang et al. 2015). As shown in Figure 5.2 and 5.6, the datasets used in this study is imbalanced, therefore this study will also adopt the AUC-ROC as a classification metrics. The AUC-ROC in Figures 6.4 and 6.5 plots the false positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis for CNN and RNN-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 is able to distinguish between classes. It can be observed from the AUC plots in Figures 6.4 and 6.5 that both models are equal to one, implying that they both have good measures of separability of the disaster sounds (Weng & Poon, 2008; Yang et al. 2015). Additionally, both models had AUC-ROC score of 0.999.

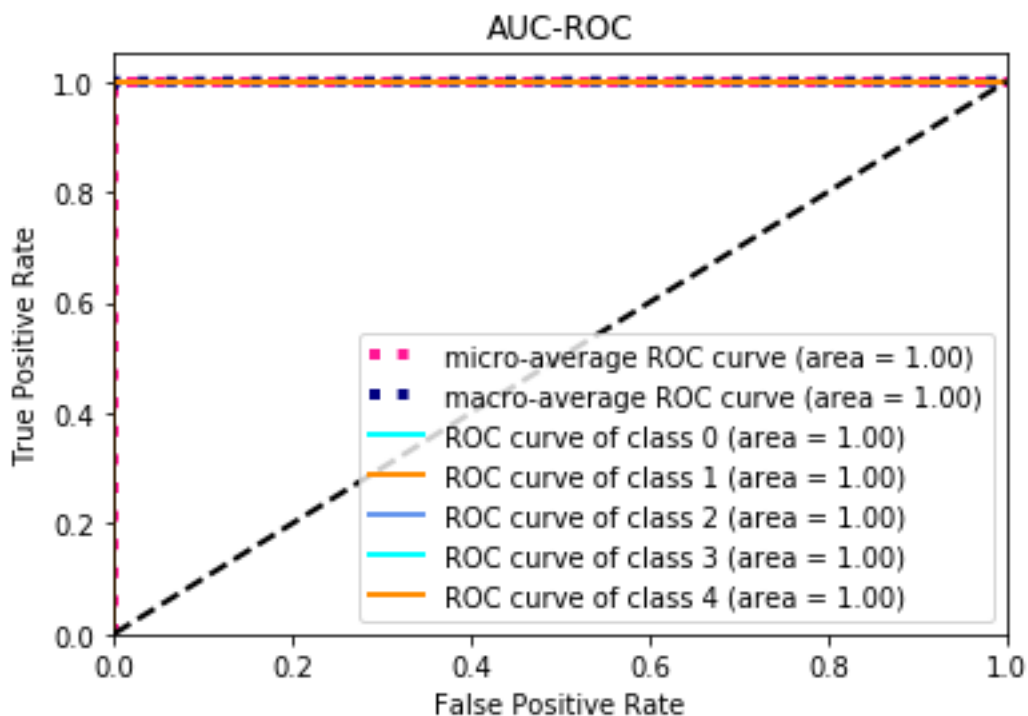


Figure 0.4: AUC-ROC for CNN Model

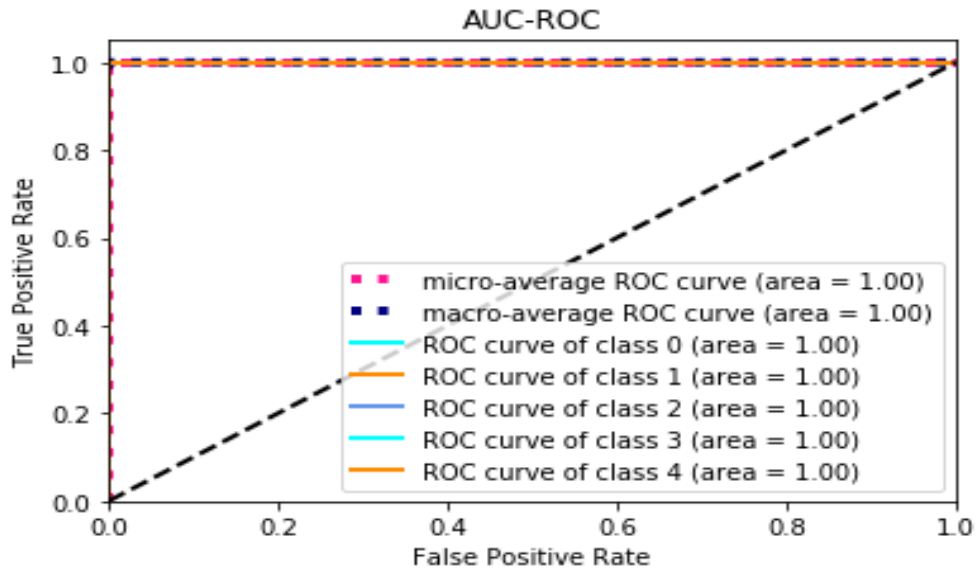


Figure 0.5: AUC-ROC for RNN-LSTM model

6.3 TESTING THE VALIDITY OF THE MODELS IN REAL-TIME CLASSIFICATION OF DISASTER SOUNDS

Recall that a time frame of a tenth of a second (0.1seconds) was selected primarily with the aim of achieving real-time classification of disaster sound (see section 5.6). To test the real-time validity of the model, the time frame was increased from 0.1second to 0.2seconds and 0.4seconds. Results are shown in Figure 6.6.

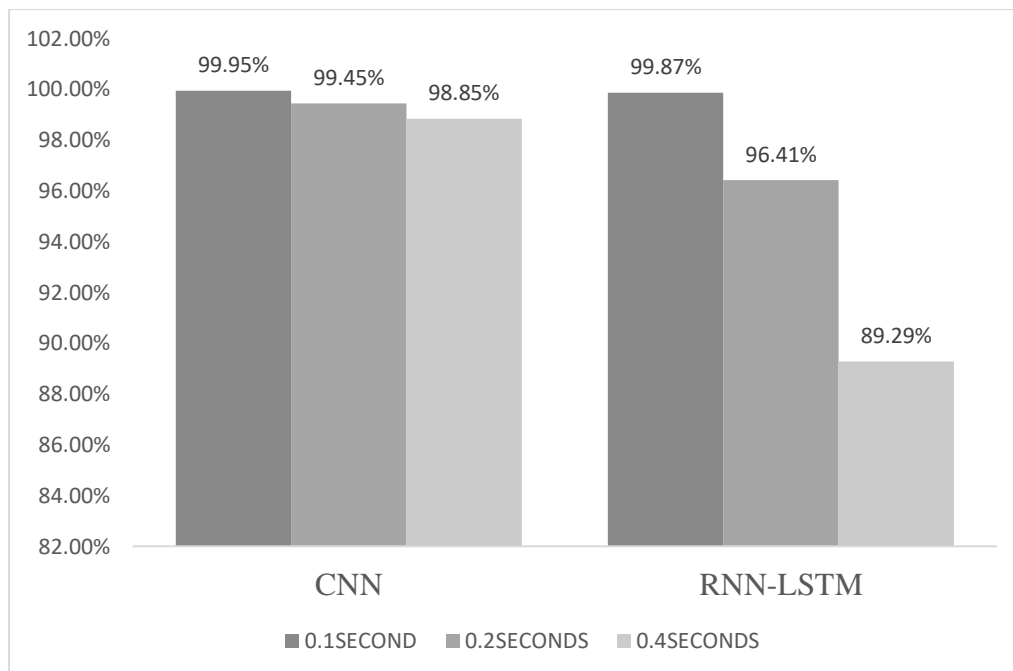


Figure 0.6: Chart showing accuracy score comparison for initial and increased time frames.

From Figure 6.6, it can be observed that, an increase in the time frame resulted in a decrease in the classification accuracies for both CNN and RNN-LSTM. This indicates that the model performs best at automatically classifying disaster sound.

Conversely, it can be argued that the low classification accuracy of the 0.2seconds and 0.4seconds time frame can be attributed to the fact that, the increase in time frame reduced the total number of sound samples, and consequently also reduced the number of test sets from 6239 to 3120 for the 0.2seconds, and from 6233 to 1560 for the 0.4seconds time frame.

Accordingly, the 0.2 and 0.4seconds time frame was maintained while the number of sound samples was augmented ($n_samples$) by:

DA1. Multiplying it by 4; $n_{samples} = 4x$,

Where, $x = \text{int}(\text{sounds}['length'].\text{sum}()/0.2)$

instead of [$n_{samples} = 2x$].

DA2. multiplying it by 6; [$n_{samples} = 6x$],

Where, $x = \text{int}(\text{sounds}['length'].\text{sum}()/0.2)$

instead of [$n_{samples} = 2x$].

[Note: the formula is as written in python code.]

Consequently, at DA1; number of test sets increased to 6239 (this is equal to the initial number of test sets) and at DA2; number of test sets increased to 9358 (higher than the initial test sets).

The classification accuracy of the $4x$ and $6x$ augmentation with 0.2seconds time frame and the initial $2x$ with 0.1second time frame is shown in Figure 6.7.

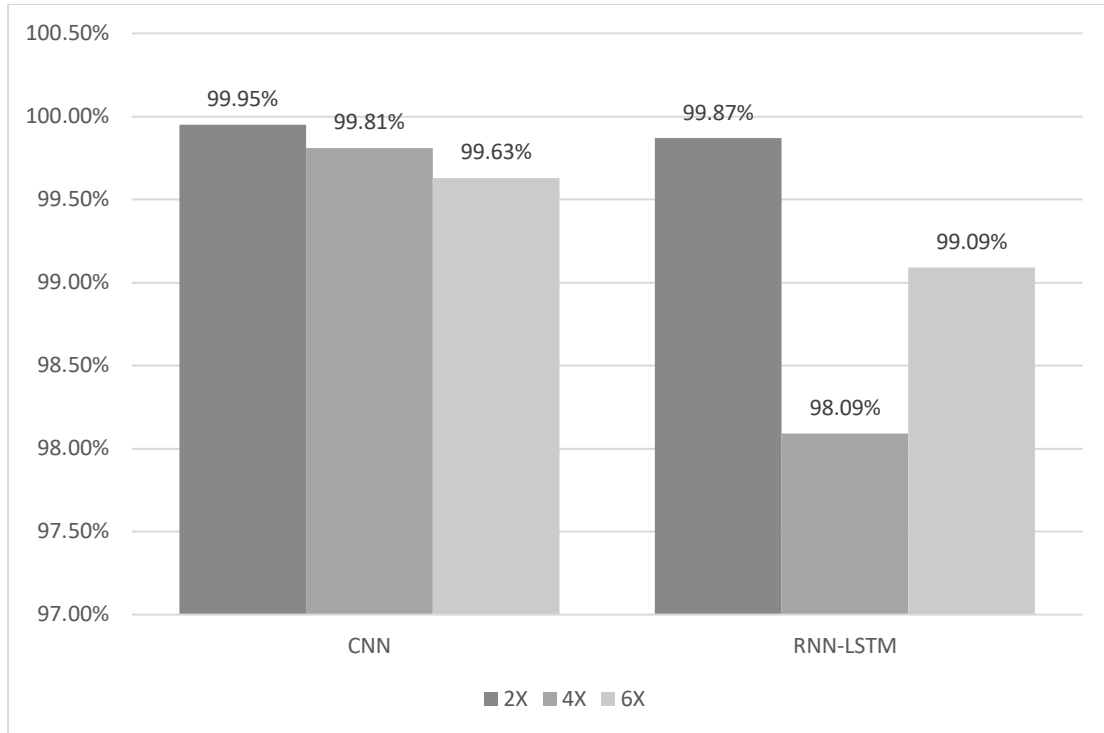


Figure 0.7: Chart showing the classification accuracy for the real time model (2x) and augmented dataset (4x, 6x).

As shown in Figure 6.7, although CNN had higher accuracies at all time instances compared to RNN-LSTM, again both classifiers performed better in real-time classification as compared to higher time frames.

Therefore, it can be concluded that both models can effectually and automatically detect-by-classification a natural disaster sound in less time.

6.4 CHAPTER SUMMARY

In this chapter the performance of the proposed deep learning models, convolutional neural network (CNN) and recurrent neural network with long short-term memory (RNN-LSTM) were validated on the 20% test dataset using the classification metrics and 10-fold cross-validation. It was observed that in all instances of the model validation process, CNN outperformed RNN-LSTM with the highest accuracy of 99.95%. Also, classification metrics such as precision and recall had the same results with the accuracy in both models. Contingency tables (confusion matrix) were used to show the level of accurate/inaccurate predictions

between the five different classes of natural disaster sound. It was observed that, while CNN had 3 incorrect instances in the confusion matrix, RNN-LSTM had 6 incorrect instances. Since the datasets used in this study was imbalanced, the AUC-ROC as a more robust classification metrics (especially for extreme events like natural disasters (Chan, 2020)) was also explored. Furthermore, the robustness of the model was tested in terms of automatic classification of natural disaster sound by increasing the time frame from 0.1second to 0.2seconds and 0.4seconds with augmented datasets. Results showed that, CNN and RNN-LSTM consistently maintained a higher classification accuracy even when compared with other studies. Results of the classification metrics for the best performing models are summarized in table 6.4.

Table 0.4: Result summary of classification metrics

Model	Accuracy	Precision	Recall	AUC score
CNN	99.95%	99.95%	99.95%	0.999
RNN-LSTM	99.87%	99.87%	99.87%	0.999

Chapter Seven

CONCLUSION

7.1 CHAPTER OVERVIEW

This thesis is a work in the area of acoustic event classification. As stated in chapter one, this study aimed to develop an automatic natural disaster sound classification model using deep learning techniques. Due to complexities in natural disaster sound, its varying amplitudes and frequencies, the detection-by-classification (Acoustic Event Classification (AEC)) approach was adopted instead of detection-and-classification (Acoustic Event Detection (AED)). In this method of classification, the task of detecting a natural disaster sound automatically translates to the task of classification.

This chapter summarizes and concludes this thesis.

7.2 THESIS SUMMARY

Natural disasters are in no doubt a phenomenon that can neither be prevented nor stopped, however, studies have shown that its effects on life and properties can be mitigated by predicting a disaster before it occurs or by detecting a disaster as soon as it occurs (Alam et al. 2019; Gupta & Doshi, 2018; Wisner & Adams, 2002).

To this end, this study sought to achieve a set of four research objectives.

Chapter one highlighted a set of research problems, as well as the research aim and objectives.

To achieve the set aim and objectives, we started by conducting a literature review on studies related to natural disasters. Through the literature review in chapter two, research trends and methods proposed by researchers to mitigate the effects of natural disasters was identified.

Trends in predicting, detecting, and managing natural disasters were also identified. It was observed that researchers were predominantly interested in post-disaster management strategies. Tools such as machine learning and data mining techniques were used to analyze

historical and meteorological data of natural disasters. The datasets used for the analysis were either in the form of text, images, or numerical data. Several research gaps including the lack of studies on the use of sound to differentiate one natural disaster type from the other (using AI techniques) was identified. The lack of literature in this domain justified the need for a systematic review.

Accordingly, a systematic review of literature on sound classification was conducted in chapter three. This was the first main contribution of this thesis. In the review, 48 articles from the Scopus and ASA databases were selected based on predefined criteria. This review aimed to identify research trends and methodologies in the area of sound classification. Most importantly, this review sought to investigate existing literature on the use of sound for various classification tasks as well as the application domains. Although substantive evidence in the use of sound to classify acoustic events in the domains of bioacoustics, medicine and, the environment was found, no study on the use of sound to classify an acoustic event such as natural disasters were identified. However, algorithms and machine learning techniques that can be adopted in this study was identified. Furthermore, two broad categories of differentiating sound events were classified, they included acoustic event detection (AED) also known as detection-and-classification, and acoustic event classification (AEC) also known as detection-by-classification. Findings indicated that this study falls in the second category; acoustic event classification (AEC). It was also observed that neural networks (deep learning), achieved higher classification accuracies compared to other classification techniques used. Hence, this study discussed the convolutional neural networks (CNN) and recurrent neural networks (RNN) as deep learning techniques that will be used for the classification of natural disaster sound. With these reviews, research objectives one and two were achieved.

In chapter four, the design science research methodology (DSRM) was explored as an appropriate research paradigm for this study. The DSRM primarily seeks to create an artefact

that solves a real-world problem through a set of rigorous steps that must be reported clearly and concisely. Since this study is aimed at developing a model for the automatic sound classification of natural disasters, the developed model is deemed as an appropriate artefact.

The sequence of steps taken, and the experiments conducted in classifying the natural disasters sound was presented in a framework and described accordingly in chapter five. The dataset was made up of five classes of natural disaster sound namely earthquake, forestfire, volcano, waves and windstorm; all downloaded from the Freesound database. Denoising was done to get rid of irrelevant sound signals using a signal envelope and a threshold value of 0.005, while down sampling reduced the sampling rate from 41000Hz to 16000Hz. Furthermore, based on the filter bank-based Mel frequency cepstral coefficient (MFCC) approach, discriminative spectral features were extracted from the sound recordings.

Preparing and developing an automatic disaster sound classification model was the third objective of this study. Hence, models were prepared to divide the sound samples in the five classes into a 0.1second window frame as well as to augment the data size. This process increased the number of sound recordings from 244 to 15596. Finally, 80% of the preprocessed sound was trained using CNN and RNN-LSTM.

Model validation is the next and compulsory process after training a model, it entailed testing the performance of the models on the remaining 20% (holdout) data using two validation techniques; the classification metrics (accuracy, precision, recall and AUC scores) and 10-fold cross-validation.

Using the accuracy, precision, recall, and AUC scores, a comparison of results obtained from the two selected methods of validating both models showed that convolutional neural networks (CNN) consistently performed better than recurrent neural networks (RNN). Hence, the fourth objective of this study was achieved.

This study to the best of our knowledge is the first to classify natural disasters based on acoustic signals/sound. Accordingly, this brings an advancement to both modeling techniques and disaster detection-by-classification.

7.3 DISCUSSIONS

In this section the performance of the CNN and LSTM-RNN acoustic event classification model is compared with other reported studies that used either CNN and/or RNN. The best performing models from the various validation processes will be used for the comparison. Table 7.1 shows the comparison of the results of previous studies with this study. It highlights the classification technique, classification category, type of sound, input acoustic features, or acoustic features representation and classification metrics. Generally, in comparison to previous studies, it was observed that the TensorFlow GPU (graphical processing unit) was a commonly used open-source library for the classification experiments.

7.3.1 CLASSIFICATION CATEGORY

The automatic classification of an acoustic event is primarily focused on classifying and or differentiating environmental sounds into one of a set of identified classes (Pooja & Usha, 2015; Ren et al. 2017). The goal of this study was to develop a model that will automatically classify natural disaster sounds as early as possible. Hence, the acoustic event classification (AEC) approach was adopted. With acoustic event classification, the task of detecting a sound automatically translates to classifying the sound (Aykanat et al. 2017; Raza et al. 2019; Temko & Nadeu, 2009). In contrast to studies that performed acoustic event detection (AED), performing acoustic event classification (AEC) truncates the three-phase rigor of AED (Lopatka et al. 2016) which involves detection, segmentation and localization as performed in studies such as (Oikarinen et al. 2019; Thakur et al. 2019; Zhang et al. 2018). The AEC approach is also not faced with problem of overlapping segments which the AED is predominantly faced with (Temko & Nadeu, 2009).

7.3.2 INPUT ACOUSTIC FEATURES

Most of the studies found, adopted the image-based approach by using the spectrogram of the sound for the classification of the sounds of interest. Although the use of spectrograms as a time-frequency representation of sound has been reported to reduce the number of trainable parameters compared to direct sound classification (Huzaifah, 2017; Khamparia et al. 2019; Zhang et al. 2019), Mitilineos et al. (2018) argues that the image-based approach results in huge feature spaces. Furthermore, the low power quantization areas of spectrograms are affected by noisy conditions (Pooja & Usha, 2015). Instead of using spectrograms, MFCCs are used in this study for their classification effectiveness at reduced data rates (Wyse, 2017).

Furthermore, CNN which is predominantly used for image classification works with two-dimensional image filters with shared weights across both axes (Wyse, 2017). However, this is not the case with using spectrograms of sounds as images, because the axes of a spectrogram do not carry the same information as with a typical image (Ren et al. 2017; Rothmann, 2019; Wyse, 2017). Ren et al. (2017) argues that using spectrograms for sound classification is currently not sufficient as existing approaches do not capture the texture information appropriately.

Mel frequency cepstral coefficient (MFCC) on the other hand are commonly used acoustic features in speech/sound recognition and classification for its ability to represent signal information accurately (Y. Kim et al. 2018; Luque, Romero-Lemos, Carrasco, & Barbancho, 2018; Sengupta et al. 2016; Yaseen et al. 2018). However, due to the non-stationary nature of acoustic signals, Su et al. (2019) posit that adopting MFCC as a single feature for classifying environmental sounds may be insufficient for capturing relevant information about an acoustic event. Thus, this study leveraged on the strength of CNN as an image-based classifier in

addition to the MFCC as a spectral (frequency-domain) feature for the classification (Sasmaz & Tek, 2018; Verma et al. 2019).

7.3.3 CLASSIFICATION PERFORMANCE

Compared to other studies shown in Table 7.1, the CNN and RNN-LSTM models used in this study had the highest classification accuracies with the shortest sound duration of 0.1 seconds.

In the model validation stage, it was observed that CNN performed slightly better than RNN-LSTM. Convolutional neural networks 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. Furthermore, it also affirms the argument by Maccagno et al. (2019) and Salamon & Bello, (2017) that; CNNs are also efficient in sound classification as they can identify patterns across inputs in a time-frequency spectrogram as well as differentiate one sound event from another even when the sound of interest is masked in noise. RNN-LSTM, on the other hand, achieved good performance in terms of accuracy, precision, and recall. Even though the results were lower than that of CNN, it further affirms that RNNs perform well on sound and time-series data (Raza et al., 2019; Verma et al., 2019). However, it was observed from the confusion matrix that the RNN-LSTM model was more prone to confusing one disaster type for another, a situation also known as false positives. Although this may be as a result of acoustic similarities between one disaster type and another, this situation must be further investigated. To overcome bias in the classification due to imbalance dataset, this study further adopted the AUC-ROC. With all the class predictions being equal to one, it can be concluded that the developed models are best fit for classifying natural disaster sound.

Table 0.1: Comparison of study approaches with other studies.

Reference	Technique	Classification category	Type of Sound	Input Acoustic Features	No of sound recordings	Duration (seconds)	Accuracy (%)	F1-Score
(Aykanat et al. 2017)	CNN	AEC	Respiratory sound	Spectrogram images	17,930	N/M	86	-
(Khamparia et al. 2019)	CNN	AEC	Environmental sound	Spectrogram images	2,400	N/M	77	-
(Salamon & Bello, 2017)	CNN	AEC	Environmental sound	Spectrogram images	8,732	4	85	-
(Sasmaz & Tek, 2018)	CNN	AEC	Animal sound	MFCC	875	N/M	75	-
(Thakur et al. 2019)	CNN	AED	Bird sound	Spectrogram images	10,208	0.5 to 320	-	0.94
(Oikarinen et al. 2019)	CNN	AED	Marmoset sound	Spectrogram images	15,970	N/M	99	0.81
(Verma et al. 2019)	RNN-LSTM	AED	Environmental sound	MFCC	52,845	10		0.83
(Raza et al. 2019)	RNN-LSTM	AEC	Heartbeat sound	Spectrogram images	322	12.5 & 27.8	80.80	-
(Zhang et al. 2018)	RNN-LSTM	AED	Marmoset sound	Mel-filter bank spectrum	20000	5	92	-
This study	CNN	AEC	Natural disaster sound	MFCC	31194	0.1	99.95	99.95
This study	RNN-LSTM	AEC	Natural disaster sound	MFCC	31194	0.1	99.87	99.87

7.4 LIMITATION OF THE STUDY

Although this study contributes to knowledge, there are some limitations and drawbacks that need to be highlighted.

It is worthy to note that all the processes in developing the disaster sound classification model were fully automated and without expert knowledge. This may have posed certain limitations to this study. They are discussed below.

7.4.1 DATASETS

It would have been appropriate to test and compare the model's performance on various historical natural disasters datasets analysis as well as datasets from different scenes of a natural disaster. However, due to the unavailability of free disaster sound datasets, this study could not explore this option.

7.4.2 DENOISING THE SIGNAL

To denoise the signals, a signal envelope with a threshold value of 0.005 was created. This comes with its disadvantages considering that the threshold value was arbitrarily chosen. Although only the upper boundary was considered, a too high threshold will imply that only well-defined sounds will be detected and the rest undetected, and a too low threshold will result in jumbling up sound recordings into a class (confusing one sound type for the other). An ideal situation would entail using a different threshold for each class of the disaster sound dataset, this could not be achieved because it requires expert knowledge (Malfante et al. 2018).

7.4.3 FEATURE EXTRACTION

Sounds have a unique pattern of changing through time and extracting these dynamic features enables the identification of one sound type from the other (Karbasi et al. 2011). Although the spectral features used in this study were suitable for the sound classification task, Huzaifah, (2017) and Mitilineos et al. (2018) argue that using spectral features alone is not sufficient because they are unable to provide time-based progression information of acoustic signals due

to their non-stationary nature. While a more robust classification system would be obtainable by combining the spectral features with time-frequency features (since they can extract non-stationary signals), this approach is computationally expensive (Mitolinos et al. 2018; Ou et al. 2013).

7.5 RECOMMENDATION

The terms detection and classification are in the context of this study inseparable. However, the decision to choose between the task of detection-and-classification (AED), or detection-by-classification (AEC) is the determining factor for the automatic detection of an acoustic event such as natural disasters.

In this study, the latter was adopted. While the use of sound for classifying a natural disaster fills in the gap left by satellite images or numerical data as posited by researchers such as (Aziz et al. 2019; Panagiota et al. 2011), having a robust model for real-time classification is the starting point. We, therefore, recommend that a model such as the CNN or RNN models should be integrated into an application that can be installed in mobile devices, and IoT devices.

7.6 FUTURE WORK

A future work will be testing and comparing the model's performance with different disaster sound datasets from different locations.

Secondly, it was observed that both models, misclassified one disaster sound with the other in the various instances of predicting the five classes. While this may be due to the acoustical similarities of the natural disasters, it is however worth investigating.

Based on the identified limitation of this study concerning the feature extraction technique, a comparative study of time-frequency features and spectral features using the CNN and RNN-LSTM classifiers will be considered for future study. A hybrid CNN RNN approach may also be considered.

References:

- Alam, F., Ofli, F., & Imran, M. (2019). Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria. *Behaviour and Information Technology*, 3001(May 2019). <https://doi.org/10.1080/0144929X.2019.1610908>
- 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. <https://doi.org/10.1121/1.4982040>
- Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep Machine Learning—A New Frontier. *Ieee, November*, 13–18.
- Asnaning, A. R., & Putra, S. D. (2018). Flood Early Warning System Using Cognitive Artificial Intelligence: The Design of AWLR Sensor. *2018 International Conference on Information Technology Systems and Innovation, ICITSI 2018 - Proceedings*, 165–170. <https://doi.org/10.1109/ICITSI.2018.8695948>
- 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. <https://doi.org/10.1121/1.3641377>
- Aykanat, M., Kılıç, Ö., Kurt, B., & Saryal, S. (2017). Classification of lung sounds using convolutional neural networks. *Eurasip Journal on Image and Video Processing*, 2017(1). <https://doi.org/10.1186/s13640-017-0213-2>
- Aziz, S., Awais, M., Akram, T., Khan, U., Alhussein, M., & Aurangzeb, K. (2019). Automatic scene recognition through acoustic classification for behavioral robotics. *Electronics (Switzerland)*, 8(5). <https://doi.org/10.3390/electronics8050483>

- BBC News. (2018). *False earthquake warning panics Japan*. BBC.
<https://www.bbc.com/news/world-asia-42582113>
- Beach, K., & Dunmire, B. (2007). Medical Acoustics. In *Rossing T. (eds) Springer Handbook of Acoustics*. Springer Handbooks. Springer, New York, NY.
https://doi.org/https://doi.org/10.1007/978-0-387-30425-0_21
- 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. <https://doi.org/10.1121/1.5097593>
- Binkhonain, M., & Zhao, L. (2019). A review of machine learning algorithms for identification and classification of non-functional requirements. *Expert Systems with Applications: X*, 1. <https://doi.org/10.1016/j.eswax.2019.100001>
- Bishop, J. C., Falzon, G., Trotter, M., Kwan, P., & Meek, P. D. (2019). Livestock vocalisation classification in farm soundscapes. *Computers and Electronics in Agriculture*, 162(April), 531–542. <https://doi.org/10.1016/j.compag.2019.04.020>
- 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. <https://doi.org/10.1587/transinf.2018EDP7383>
- Borgne, Y.-A. Le, & Bontempi, G. (2017). *Deep learning techniques-Overview*. May. <https://doi.org/10.13140/RG.2.2.33519.84643>
- Bourouhou, A., Jilbab, A., Nacir, C., & Hammouch, A. (2019). Heart sounds classification for a medical diagnostic assistance. *International Journal of Online and Biomedical Engineering*, 15(11), 88–103. <https://doi.org/10.3991/ijoe.v15i11.10804>
- Boustan, P. L., Kahn, M. E., Rhode, P. W., & 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*.

<http://www.nber.org/papers/w23410>

- 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. <https://doi.org/10.1121/1.4707424>
- Calvet, L., Lopeman, M., De Armas, J., Franco, G., & Juan, A. A. (2017). Statistical and machine learning approaches for the minimization of trigger errors in parametric earthquake catastrophe bonds. *Sort*, *41*(2), 373–391. <https://doi.org/10.2436/20.8080.02.64>
- Cao, X., Zhang, X., Yu, Y., & Niu, L. (2017). Deep learning-based recognition of underwater target. *International Conference on Digital Signal Processing, DSP*, 89–93. <https://doi.org/10.1109/ICDSP.2016.7868522>
- Carcary, M. (2011). Design science research: The case of the IT capability maturity framework (IT CMF). *Electronic Journal of Business Research Methods*, *9*(2), 109–118.
- Chan, C. (2020). *What is a ROC Curve and How to Interpret It | Displayr*. <https://www.displayr.com/what-is-a-roc-curve-how-to-interpret-it/>
- Chappell, C. (2019). Natural disasters cost \$91 billion in 2018, according to federal report. *CNBC*. <https://www.cnbc.com/2019/02/06/natural-disasters-cost-91-billion-in-2018-federal-report.html>
- 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*, *7*(April), 32845–32852. <https://doi.org/10.1109/ACCESS.2019.2903859>
- Chen, W., Shirzadi, A., Shahabi, H., Ahmad, B. Bin, Zhang, S., Hong, H., & Zhang, N. (2017). A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County,

- China. *Geomatics, Natural Hazards and Risk*, 8(2), 1955–1977.
<https://doi.org/10.1080/19475705.2017.1401560>
- Chu, S., Narayanan, S., & Kuo, C. C. J. (2009). Environmental sound recognition with timeFrequency audio features. *IEEE Transactions on Audio, Speech and Language Processing*, 17(6), 1142–1158. <https://doi.org/10.1109/TASL.2009.2017438>
- Coskun, M., YILDIRIM, Ö., UÇAR, A., & DEMIR, Y. (2017). An Overview of Popular Deep Learning Methods. *European Journal of Technic*, 7(2), 165–176.
<https://doi.org/10.23884/ejt.2017.7.2.11>
- Cvengros, R. M., Valente, D., Nykaza, E. T., & Vipperman, J. S. (2012a). Blast noise classification with common sound level meter metrics. *The Journal of the Acoustical Society of America*, 132(2), 822–831. <https://doi.org/10.1121/1.4730921>
- Cvengros, R. M., Valente, D., Nykaza, E. T., & Vipperman, J. S. (2012b). Blast noise classification with common sound level meter metrics Blast noise classification with common sound level meter metrics. *The Journal of the Acoustical Society of America*, 822. <https://doi.org/10.1121/1.4730921>
- 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.
<https://doi.org/10.1109/RAICS.2018.8635051>
- Ding, S., Zhao, H., Zhang, Y., Xu, X., & Nie, R. (2015). Extreme learning machine: algorithm, theory and applications. *Artificial Intelligence Review*, 44(1), 103–115.
<https://doi.org/10.1007/s10462-013-9405-z>
- Domingo, C. (2012). Journal of Network and Computer Applications An overview of the internet of underwater things. *Journal of Network and Computer Applications*, 35(6), 1879–1890. <https://doi.org/10.1016/j.jnca.2012.07.012>

- Doxani, G., Siachalou, S., Mitraka, Z., & Patias, P. (2019). Decision making on disaster management in agriculture with sentinel applications. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(3/W8), 121–126. <https://doi.org/10.5194/isprs-archives-XLII-3-W8-121-2019>
- Dubey, S., Dahiya, M., & Jain, S. (2018). Application of Distributed Data Center in Logistics as Cloud Collaboration for handling Disaster Relief. *Proceedings - 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages, IoT-SIU 2018*, 1–11. <https://doi.org/10.1109/IoT-SIU.2018.8519865>
- Duggar, E., Li, Q., & Praagh, A. Van. (2016). *Understanding the Impact of Natural Disasters : Exposure to Direct Damages Across Countries* (Issue November).
- Dwivedi, A. K., Imtiaz, S. A., & Rodriguez-Villegas, E. (2019). Algorithms for automatic analysis and classification of heart sounds-A systematic review. *IEEE Access*, 7(c), 8316–8345. <https://doi.org/10.1109/ACCESS.2018.2889437>
- Epelbaum, T. (2017). *Deep learning: Technical introduction*. <http://arxiv.org/abs/1709.01412>
- Evans, M. (2011). Natural disasters. In *Virginia Quarterly Review* (Vol. 93, Issue 1).
- 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. *APSIPA Transactions on Signal and Information Processing*, 8(2019), 1–11. <https://doi.org/10.1017/ATSIP.2019.7>
- Furquim, G., Filho, G. P. R., Jalali, R., Pessin, G., Pazzi, R. W., & 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), 1–20. <https://doi.org/10.3390/s18030907>
- 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 2012), 547–559.

- Giordano, B. L. (2005). Everyday listening: an annotated bibliography. In D. Rocchesso & F. Fontana (Eds.), *The Sounding Object: Vol. 6 PART B* (pp. 1–12).
<https://doi.org/10.1115/GT2005-69036>
- Giret, N., Roy, P., Albert, A., Pachet, F., Kreuzer, 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.
<https://doi.org/10.1121/1.3531953>
- Gopaldaswami, R. (2018). A Study on the Correlation of Physiological and Psychological Health Hazards in Human Habitats with Seismicity, Mountain Air Turbulence and Environmental Infrasound. *Open Journal of Earthquake Research*, *07*(02), 69–87.
<https://doi.org/10.4236/ojer.2018.72005>
- Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A., & Chakraborty, B. (2018). A review on application of data mining techniques to combat natural disasters. *Ain Shams Engineering Journal*, *9*(3), 365–378. <https://doi.org/10.1016/j.asej.2016.01.012>
- Greenhalgh, T. (1997). How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *Bmj*, *315*(7109), 672–675.
<https://doi.org/10.1136/bmj.315.7109.672>
- 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. <https://doi.org/10.1121/1.5055209>
- Gupta, S., & Doshi, L. (2018). An Acknowledgement Based System for Forest Fire Detection via Leach Algorithm. *Proceedings - 2017 International Conference on Computational Intelligence and Networks, CINE 2017*, 17–21. <https://doi.org/10.1109/CINE.2017.16>
- Halkias, X. C., 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. <https://doi.org/10.1121/1.4821203>

- Halkias, X., Paris, S., & Glotin, H. (2013). *Classification of mysticete sounds using machine learning techniques*. 3496(2013). <https://doi.org/10.1121/1.4821203>
- 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. <https://doi.org/10.1371/journal.pone.0162075>
- Hartman, W. M., & Candy, J. V. (2014). Acoustic Signal Processing. *Springer Handbook of Acoustics*, December. <https://doi.org/10.1007/978-1-4939-0755-7>
- Hassiotis, C. (2018). *Infrasound Can Detect Tornadoes an Hour Before They Form*.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). *Design Science in Information Systems Research* (Vol. 28, Issue 1, pp. 75–105).
- <https://www.bbc.com/news/technology-40366816>. (2017). *California earthquake alarm sounded - 92 years late*. BBC.
- Huzaifah, M. (2017). *Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Networks*. 1–5. <http://arxiv.org/abs/1706.07156>
- Hwang, C. J., Kush, A., & Kumar, A. (2018). Multihop ad hoc networks for disaster response scenarios. *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018*, 810–814. <https://doi.org/10.1109/CSCI46756.2018.00162>
- Ibrahim, A. K., Chérubin, L. M., Zhuang, H., Schärer Umpierre, M. T., Dagleish, F., Erdol, N., Ouyang, B., & Dagleish, 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. <https://doi.org/10.1121/1.5022281>
- 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*, 2155. <https://doi.org/10.1121/1.5126861>
- Imran, M., Alam, F., Ofli, F., & Aupetit, M. (2017). *Enabling Rapid Disaster Response Using Artificial Intelligence and Social Media*. 1–12.
- Ivić, M. (2019). Artificial intelligence and geospatial analysis in disaster management. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(3/W8), 161–166. <https://doi.org/10.5194/isprs-archives-XLII-3-W8-161-2019>
- Jacoby, C. B. (2014). *Automatic Urban Sound Classification Using Feature Learning Techniques*. <https://steinhardt.nyu.edu/scmsAdmin/media/users/ec109/MTT-14-01-013.pdf>
- Joshi, N. (2019). How AI Can And Will Predict Disasters. *Forbes*.
- 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. <https://doi.org/10.1121/1.4966592>
- Kansal, A., Singh, Y., Kumar, N., & Mohindru, V. (2016). Detection of forest fires using machine learning technique: A perspective. *Proceedings of 2015 3rd International Conference on Image Information Processing, ICIIP 2015*, 241–245. <https://doi.org/10.1109/ICIIP.2015.7414773>
- Karbasi, M., Ahadi, S. M., & Bahmanian, M. (2011). Environmental Sound Classification using Spectral Dynamic Features. *IEEE ICICS*, 2–7. <https://doi.org/10.1109/ICICS.2011.6173513>
- Khalaf, M., Hussain, A. J., Al-Jumeily, D., Baker, T., Keight, R., Lisboa, P., Fergus, P., & Al Kafri, A. S. (2018). A Data Science Methodology Based on Machine Learning Algorithms

- for Flood Severity Prediction. *2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings*, 1–8. <https://doi.org/10.1109/CEC.2018.8477904>
- Khalaf, M., Hussain, A. J., Al-Jumeily, D., Fergus, P., & Idowu, I. O. (2015). Advance flood detection and notification system based on sensor technology and machine learning algorithm. *2015 22nd International Conference on Systems, Signals and Image Processing - Proceedings of IWSSIP 2015*, 105–108. <https://doi.org/10.1109/IWSSIP.2015.7314188>
- 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. *IEEE Access*, 7(January), 7717–7727. <https://doi.org/10.1109/ACCESS.2018.2888882>
- Kim, S., Lee, W., Park, Y. S., Lee, H. W., & Lee, Y. T. (2017). Forest fire monitoring system based on aerial image. *Proceedings of the 2016 3rd International Conference on Information and Communication Technologies for Disaster Management, ICT-DM 2016*, 5–10. <https://doi.org/10.1109/ICT-DM.2016.7857214>
- 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. *Sensors (Switzerland)*, 18(11). <https://doi.org/10.3390/s18114019>
- Krishna, D., Marcelino, P., Doxani, G., Siachalou, S., Mitraka, Z., Patias, P., Gingras, B., Fitch, W. T., Union, I. T., LeBien, J., Ioup, J., Gomes, L., Vale, Z., Han, W., Coutinho, E., Ruan, H., Li, H. H. H., Schuller, B., Yu, X., ... Erdol, N. (2018). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *The Journal of the Acoustical Society of America*, 8(5), 1–11. <https://doi.org/10.1109/LSP.2017.2657381>
- LeBien, J. G., & Ioup, J. W. (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. <https://doi.org/10.1121/1.5047435>
- 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*, 387, 3278–3282. <https://doi.org/10.1121/1.5047435>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, H., Fei, X., & He, C. (2018). Study on Most Important Factor and Most Vulnerable Location for a Forest Fire Case Using Various Machine Learning Techniques. *Proceedings - 2018 6th International Conference on Advanced Cloud and Big Data, CBD 2018*, 298–303. <https://doi.org/10.1109/CBD.2018.00060>
- Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: A statistically consistent and more discriminating measure than accuracy. *IJCAI International Joint Conference on Artificial Intelligence*, 519–524.
- Lopatka, K., Kotus, J., & Czyzewski, A. (2016). Detection, classification and localization of acoustic events in the presence of background noise for acoustic surveillance of hazardous situations. *Multimedia Tools and Applications*, 75(17), 10407–10439. <https://doi.org/10.1007/s11042-015-3105-4>
- Luque, A., Romero-Lemos, J., Carrasco, A., & Barbancho, J. (2018). Non-sequential automatic classification of anuran sounds for the estimation of climate-change indicators. *Expert Systems with Applications*, 95, 248–260. <https://doi.org/10.1016/j.eswa.2017.11.016>
- Luque, A., Romero-Lemos, J., Carrasco, A., & Gonzalez-Abril, L. (2018). Temporally-aware algorithms for the classification of anuran sounds. *PeerJ*, 2018(5), 1–40. <https://doi.org/10.7717/peerj.4732>
- Maccagno, A., Mastropietro, A., Mazziotta, U., Lee, Y., & Uncini, A. (2019). A CNN Approach for Audio Classification in Construction Sites. *WIRNAt: Vietri Sul Mare (SA)*,

Italy, June.

- 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. <https://doi.org/10.1121/1.5036628>
- 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*. <https://doi.org/10.1155/2018/5828074>
- Mone, G. (2007). Earth Speaks in an Inaudible Voice. *Discover Magazine*, *August*.
- Monroe-Kane, C. (2019). 20 Seconds Makes All The Difference : How Sound Waves Help Us Understand Earthquakes Geophysicist Ben Holtzman On Using Sound Recordings To Study Earthquakes ' Past ,. *WISCONSIN Public Radio (Npr), a Service of the Wisconsin Educational Communications Board and the University of Wisconsin-Madison*. <https://www.wpr.org/20-seconds-makes-all-difference-how-sound-waves-help-us-understand-earthquakes>
- Mousa, M., Zhang, X., & Claudel, C. (2016). Flash Flood Detection in Urban Cities Using Ultrasonic and Infrared Sensors. *IEEE Sensors Journal*, *16*(19), 7204–7216. <https://doi.org/10.1109/JSEN.2016.2592359>
- Muir, T. G., & Bradley, D. L. (2016). Underwater Acoustics: A Brief Historical Overview Through World War II. *Acoustics Today*, *12*(3), 40–48. <http://acousticstoday.org/wp-content/uploads/2016/09/Underwater-Acoustics.pdf>
- Nasanbat, E., Lkhamjav, O., Balkhai, A., Tsevee-Oirov, C., Purev, A., & Dorjsuren, M. (2018). A spatial distribution map of the wildfire risk in Mongolia using decision support system. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(3W4), 357–362. <https://doi.org/10.5194/isprs-archives-XLII-3-W4-357-2018>

- 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). <https://doi.org/10.3390/app6120443>
- 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. <https://doi.org/10.1121/1.5087827>
- Okamoto, K., Mochida, T., Nozaki, D., Wen, Z., Qi, X., & Sato, T. (2018). Content-Oriented Surveillance System Based on ICN in Disaster Scenarios. *International Symposium on Wireless Personal Multimedia Communications, WPMC, 2018-Novem*, 484–489. <https://doi.org/10.1109/WPMC.2018.8712852>
- 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(January).
- 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. <https://doi.org/10.4103/2319-4170.137773>
- 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. <https://doi.org/10.1016/j.bbe.2013.07.001>
- Panagiota, M., Jocelyn, C., & Erwan, P. (2011). State of the art on Remote Sensing for vulnerability and damage assessment on urban context. *Grenoble, France: URBASIS Consortium, March*. <https://doi.org/10.1097/MD.00000000000008031>
- 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.

<https://doi.org/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. <https://doi.org/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(June), 3371–3381. <http://dx.doi.org/10.1121/1.4876439>
- Peffer, K., Tuunanen, T., & Rothenberger, M. A. (2008). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, August 2014. <https://doi.org/10.2753/MIS0742-1222240302>
- Perlman, D. (2013). THE RUMBLE OF DESTRUCTION / Infrasonic sound. *Hearst Communication, Inc.* <https://www.sfgate.com/news/article/THE-RUMBLE-OF-DESTRUCTION-Infrasonic-sound-too-2632570.php>
- Perr, J. (2005). Basic acoustics and Signal Processing. *LinuxFocus.Org*, 271. <http://linuxfocus.org>
- Peso Parada, P., & Cardenal-López, A. (2014). Using Gaussian mixture models to detect and classify dolphin whistles and pulses. *The Journal of the Acoustical Society of America*, 135(6), 3371–3380. <https://doi.org/10.1121/1.4876439>
- Pooja, K. J., & Usha, L. (2015). Robust Sound Event Recognition using Subband Power Distribution Image Feature. *International Journal of Engineering Research and Technology*, V4(05), 1116–1121. <https://doi.org/10.17577/ijertv4is051087>
- Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., & Iyengar, S. S. (2018). A Survey on Deep Learning. *ACM Computing Surveys*, 51(5), 1–36. <https://doi.org/10.1145/3234150>
- Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious

- respiratory sound analysis: A systematic review. In *PLoS ONE* (Vol. 12, Issue 5).
<https://doi.org/10.1371/journal.pone.0177926>
- Qian, K., Zhang, Z., Baird, A., & Schuller, B. (2017a). 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. <https://doi.org/10.1121/1.5004570>
- Qian, K., Zhang, Z., Baird, A., & Schuller, B. (2017b). Active learning for bird sounds classification. *Acta Acustica United with Acustica*, *103*(3), 361–364. <https://doi.org/10.3813/AAA.919064>
- Qian, K., Zhang, Z., Baird, A., & Schuller, B. (2018). Active learning for bird sound classification via a kernel-based extreme learning machine. *The Journal of the Acoustical Society of America*, *1796*(2017). <https://doi.org/10.1121/1.5004570>
- Rascon, C., & Meza, I. (2017). Localization of sound sources in robotics: A review. *Robotics and Autonomous Systems*, *96*, 184–210. <https://doi.org/10.1016/j.robot.2017.07.011>
- Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S., & On, B. W. (2019). Heartbeat sound signal classification using deep learning. *Sensors (Switzerland)*, *19*(21), 1–15. <https://doi.org/10.3390/s19214819>
- Ren, J., Jiang, X., Yuan, J., & Magnenat-Thalmann, N. (2017). Sound-Event Classification Using Robust Texture Features for Robot Hearing. *IEEE Transactions on Multimedia*, *19*(3), 447–458. <https://doi.org/10.1109/TMM.2016.2618218>
- Resch, B., Usländer, F., & Havas, C. (2018). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science*, *45*(4), 362–376. <https://doi.org/10.1080/15230406.2017.1356242>
- Robakis, E., Watsa, M., & Erkenswick, G. (2018). Classification of producer characteristics in primate long calls using neural networks. *The Journal of the Acoustical Society of*

America, 144(1), 344–353. <https://doi.org/10.1121/1.5046526>

- 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. *The Journal of the Acoustical Society of America*, 129(January), 467–476. <https://doi.org/10.1121/1.3514383>
- Rothmann, D. (2019). *What 's wrong with CNNs and spectrograms for audio processing? Sounds are "transparent."* 1–9.
- Rubinstein, G. (2008). ON SOUNDS EMITTED BY INANIMATE OBJECTS IN RUSSIAN. *The Slavic and East European Journal*, 52(4), 561–588. <https://www.jstor.org/stable/40651272>
- 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. <https://doi.org/10.1109/LSP.2017.2657381>
- Sasmaz, E., & Tek, F. B. (2018). Animal Sound Classification Using A Convolutional Neural Network. *UBMK 2018 - 3rd International Conference on Computer Science and Engineering*, 625–629. <https://doi.org/10.1109/UBMK.2018.8566449>
- Sayad, Y. O., Mousannif, H., & Al Moatassime, H. (2019). Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*, 104(September 2018), 130–146. <https://doi.org/10.1016/j.firesaf.2019.01.006>
- Sengupta, N., Sahidullah, M., & Saha, G. (2016). Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine*, 75, 118–129. <https://doi.org/10.1016/j.combiomed.2016.05.013>
- Sermet, Y., & Demir, I. (2018). An intelligent system on knowledge generation and communication about flooding. *Environmental Modelling and Software*, 108(August 2017), 51–60. <https://doi.org/10.1016/j.envsoft.2018.06.003>

- 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. <https://doi.org/10.1121/1.4861348>
- Simmonds, J., & MacLennan, D. (2005). Underwater Sound. *Fisheries Acoustics: Theory and Practice, 1945*, 20–69.
- Singhvi, A., Saget, B., & Lee, J. (2018). *What Went Wrong With Indonesia's Tsunami Early Warning System*. The New York Times.
- Soule, B. (2014). Post-crisis analysis of an ineffective tsunami alert: The 2010 earthquake in Maule, Chile. *Disasters*, *38*. <https://doi.org/10.1111/disa.12045>
- Stojanovic, M., & Beaujean, P. P. J. (2016). Acoustic communication. *Springer Handbook of Ocean Engineering*, 359–386. https://doi.org/10.1007/978-3-319-16649-0_15
- 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. <https://doi.org/10.3390/s19071733>
- 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). <https://doi.org/10.1121/1.4906168>
- Tarasconi, F., Farina, M., Mazzei, A., & Bosca, A. (2017). The role of unstructured data in real-time disaster-related social media monitoring. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017, 2018-Janua*, 3769–3778. <https://doi.org/10.1109/BigData.2017.8258377>
- Temko, A., & Nadeu, C. (2009). Acoustic Event Detection and Classification. In *Computers in the Human Interaction Loop* (Issue December). <https://doi.org/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. <https://doi.org/10.1121/1.5118245>
- Tierney, K. J. (2019). Businesses and Disasters: Vulnerability, Impact, and Recovery. In H. Rodríguez, E. L. Quarantelli, & R. R. Dynes (Eds.), *Handbook of Disaster Research* (pp. 275–296). Springer. <https://doi.org/10.1093/oxfordhb/9780190274481.013.35>
- Tobergte, D. R., & Curtis, S. (2013). Environmental Health in Emergencies. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>
- Turner, C., & Joseph, A. (2015). A Wavelet Packet and Mel-Frequency Cepstral Coefficients-Based Feature Extraction Method for Speaker Identification. *Procedia Computer Science*, *61*, 416–421. <https://doi.org/10.1016/j.procs.2015.09.177>
- UNDP. (2012). Disaster Risk Reduction and Recovery. *United Nations Development Programme (UNDP) FAST FACTS*. www.undp.org/cpr
- Vallimeena, P., Nair, B. B., & Rao, S. N. (2018). Machine Vision Based Flood Depth Estimation Using Crowdsourced Images of Humans. *2018 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2018*, 1–4. <https://doi.org/10.1109/ICCIC.2018.8782363>
- Van der Merwe, A., Gerber, A., & Smuts, H. (2020). *Guidelines for Conducting Design Science Research in Information Systems* (pp. 163–178). https://doi.org/10.1007/978-3-030-35629-3_11

- Van Dulmen, S., Sluijs, E., Van Dijk, L., De Ridder, D., Heerdink, R., & Bensing, J. (2007). Patient adherence to medical treatment: A review of reviews. *BMC Health Services Research*, 7, 1–13. <https://doi.org/10.1186/1472-6963-7-55>
- 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. *Noise Mapping*, 6(1), 52–71. <https://doi.org/10.1515/noise-2019-0005>
- Vrbancic, G., & Podgorelec, V. (2018). Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks. *Elektronika Ir Elektrotehnika*, 24(4), 1–7. <https://doi.org/10.5755/j01.eie.24.4.21469>
- Wallemacq, P. (2015). *The Human Cost of Natural disasters*.
- Wang, T., & Nanda, S. (2012). Feature Extraction Methods & Application. *GE Global Research and GE Power & Water*, 41. <http://www.gis.usu.edu/~doug/RS5750/PastProj/FA2002/KelliTaylor.pdf>
- Wang, Y., & Peng, H. (2018). Underwater acoustic source localization using generalized regression neural network. *The Journal of the Acoustical Society of America*, 143(4), 2321–2331. <https://doi.org/10.1121/1.5032311>
- Wason, R. (2018). Deep learning: Evolution and expansion. *Cognitive Systems Research*, 52, 701–708. <https://doi.org/10.1016/j.cogsys.2018.08.023>
- Weng, C. G., & Poon, J. (2008). A new evaluation measure for imbalanced datasets. *Conferences in Research and Practice in Information Technology Series*, 87, 27–32.
- Wieland, M., Liu, W., & Yamazaki, F. (2016). Learning change from Synthetic Aperture Radar images: Performance evaluation of a Support Vector Machine to detect earthquake and tsunami-induced changes. *Remote Sensing*, 8(10). <https://doi.org/10.3390/rs8100792>
- Wilson, J. D., & Makris, N. C. (2006). Ocean acoustic hurricane classification. *The Journal of the Acoustical Society of America*, 119(1), 168–181. <https://doi.org/10.1121/1.2130961>

- Winter, R. (2008). Design science research in Europe. *European Journal of Information Systems*, 17(5), 470–475. <https://doi.org/10.1057/ejis.2008.44>
- Wisner, B., & Adams, J. (2002). Environmental health in emergencies and disasters. In *World Health Organization* (Vol. 62, Issue 5). <https://doi.org/10.1007/s00393-003-0515-x>
- 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 and Communication Disorders*, 53(3), 446–467. <https://doi.org/10.1111/1460-6984.12371>
- Wu, J., Chua, Y., Zhang, M., Li, H., & Tan, K. C. (2018). A spiking neural network framework for robust sound classification. *Frontiers in Neuroscience*, 12(NOV), 1–17. <https://doi.org/10.3389/fnins.2018.00836>
- Wyse, L. (2017). Audio Spectrogram Representations for Processing with Convolutional Neural Networks. *Proceedings of the First International Workshop on Deep Learning and Music Joint with IJCNN*, 1(1), 37–41. <http://arxiv.org/abs/1706.09559>
- Yang, J., Wang, Y. X., Qiao, Y. Y., Zhao, X. X., Liu, F., & Cheng, G. (2015). On Evaluating Multi-class Network Traffic Classifiers Based on AUC. *Wireless Personal Communications*, 83(3), 1731–1750. <https://doi.org/10.1007/s11277-015-2473-4>
- Yaseen, Son, G. Y., & Kwon, S. (2018). Classification of heart sound signal using multiple features. *Applied Sciences (Switzerland)*, 8(12). <https://doi.org/10.3390/app8122344>
- Zhang, L., Wang, D., Bao, C., Wang, Y., & Xu, K. (2019). Large-scale whale-call classification by transfer learning on multi-scale waveforms and time-frequency features. *Applied Sciences (Switzerland)*, 9(5), 1–11. <https://doi.org/10.3390/app9051020>
- 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.

<https://doi.org/10.1121/1.5047743>

Zhang, Ya-jie, Huang, J., Gong, N., Ling, Z., & Hu, Y. (2019). Automatic detection and classification of marmoset vocalizations using deep and recurrent neural networks. *The Journal of the Acoustical Society of America*, 478(2018).

<https://doi.org/10.1121/1.5047743>

Zhang, Yan, Lv, D., & Zhao, Y. (2016). Multiple-view active learning for environmental sound classification. *International Journal of Online Engineering*, 12(12), 49–54.

<https://doi.org/10.3991/ijoe.v12i12.6458>

APPENDIX A: PRIMARY STUDIES USED FOR THE SYSTEMATIC REVIEW

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. Viperman, J. S. Cvengros, R. M. Valente, D. & Nykaza, E. T. (2017). Blast noise classification with common sound level meter metrics Blast noise classification with common sound level meter metrics. 822(2012). 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. Dalgleish, F. Erdol, N. Ouyang, B. & Dalgleish, 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 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
A15.	Binder, C. & Paul, H. (2019). <i>Range-dependent impacts of ocean acoustic propagation on automated classification of transmitted bowhead and humpback whale vocalizations</i> . 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). <i>Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training data a</i>). 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.	Peso Parada, P. & Cardenal-López, 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(6), 3371–3380. https://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 vocalisation 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. Alhusein, 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 a 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
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.	https://doi.org/10.1121/1.3641377 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
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

APPENDIX B: PYTHON CODES FOR LOADING THE DATA

```

import os
from tqdm import tqdm
import pandas as pd
import numpy as np
from scipy.io import wavfile
from python_speech_features import mfcc, logfbank
from matplotlib import pyplot as plt
from path import Path
import librosa
import librosa.display
#waveforms
def plot_signals(signals):
    fig, axes = plt.subplots(nrows=2, ncols=5, sharex=False, sharey=True, figsize=(20,5))
    fig.suptitle('Time Series', size=16)
    i = 0
    try:
        for x in range(2):
            for y in range(5):
                axes[x,y].set_title(list(signals.keys())[i])
                axes[x,y].plot(list(signals.values())[i])
                axes[x,y].get_xaxis().set_visible(False)
                axes[x,y].get_yaxis().set_visible(False)
                i += 1
    except IndexError:
        pass
#fft
def plot_fft(fft):
    fig, axes = plt.subplots(nrows=2, ncols=5, sharex=False, sharey=True, figsize=(20,5))
    fig.suptitle('Fourier Transforms', size=16)
    i = 0
    try:
        for x in range(2):
            for y in range(5):
                data = list(fft.values())[i]
                Y, freq = data[0], data[1]
                axes[x,y].set_title(list(fft.keys())[i])
                axes[x,y].plot(freq, Y)
                axes[x,y].get_xaxis().set_visible(False)
                axes[x,y].get_yaxis().set_visible(False)
                i += 1
    except IndexError:
        pass
#fbc
def plot_fbank(fbank):
    fig, axes = plt.subplots(nrows=2, ncols=5, sharex=False, sharey=True, figsize=(20,5))
    fig.suptitle('Filter Bank Coefficients', size=16)
    i = 0
    try:
        for x in range(2):
            for y in range(5):
                axes[x,y].set_title(list(fbank.keys())[i])
                axes[x,y].imshow(list(fbank.values())[i], cmap='hot', interpolation='nearest')
                axes[x,y].get_xaxis().set_visible(False)
                axes[x,y].get_yaxis().set_visible(False)
                i += 1
    except IndexError:

```

```

    pass
#mfcc
def plot_mfccs(mfcc):
    fig, axes = plt.subplots(nrows=2, ncols=5, sharex=False, sharey=True, figsize=(20,5))
    fig.suptitle('Mel Frequency Cepstrum Coefficients', size=16)
    i = 0
    try:
        for x in range(2):
            for y in range(5):
                axes[x,y].set_title(list(mfcc.keys())[i])
                axes[x,y].imshow(list(mfcc.values())[i], cmap='hot', interpolation='nearest')
                axes[x,y].get_xaxis().set_visible(False)
                axes[x,y].get_yaxis().set_visible(False)
                i += 1
    except IndexError:
        pass
def calc_fft(y, rate):
    n = len(y)
    freq = np.fft.rfftfreq(n, d=1/rate)
    Y = abs(np.fft.rfft(y)/n)
    return (Y, freq)
#define envelope
def envelope(y, rate, threshold):
    mask = []
    y = pd.Series(y).apply(np.abs)
    y_mean = y.rolling(window=int(rate/10), min_periods=1, center=True).mean()
    for mean in y_mean:
        if mean > threshold:
            mask.append(True)
        else:
            mask.append(False)
    return mask
#load data
sounds = pd.read_csv('sounds.csv')
sounds.set_index('filename', inplace=True)
for f in sounds.index:
    rate, signal = wavfile.read('dataset/'+f)
    sounds.at[f, 'length'] = signal.shape[0]/rate
classes = list(np.unique(sounds.label))
class_dist = sounds.groupby(['label'])['length'].mean()
fig, ax = plt.subplots()
ax.set_title('Class Distribution', y=1.10)
ax.pie(class_dist, labels=class_dist.index, autopct='%1.1f%%', shadow=False,
startangle=90)
ax.axis('equal')
plt.show()
sounds.reset_index(inplace=True)
signals = {}
fft = {}
fbank = {}
mfccs = {}
for c in classes:
    wav_file = sounds[sounds.label == c].iloc[0,0]
    signal, rate = librosa.load('dataset/'+wav_file, sr=44100)
    mask = envelope(signal, rate, 0.0005)
    signal = signal[mask]
    signals[c] = signal
    fft[c] = calc_fft(signal, rate)
    bank = logfbank(signal[:rate], rate, nfilt=26, nfft=1103).T
    fbank[c] = bank

```

```

    mel = mfcc(signal[:rate], rate, numcep=13, nfilt=26, nfft=1103).T
    mfccs[c] = mel
plot_signals(signals)
plt.show()
plot_fft(fft)
plt.show()
plot_fbank(fbank)
plt.show()
plot_mfccs(mfccs)
plt.show()
#audio downsampling
if len(os.listdir('clean')) == 0:
    for f in tqdm(sounds.filename):
        signal,rate = librosa.load('dataset/'+f, sr=16000)
        mask = envelope(signal, rate, 0.0005)
        wavfile.write(filename='clean/'+f, rate=rate, data=signal[mask])

```

APPENDIX C: PYTHON CODES FOR MODEL PREPARATION/PREDICTION

```

from tqdm import tqdm
import pandas as pd
import numpy as np
from scipy.io import wavfile
from python_speech_features import mfcc
from matplotlib import pyplot as plt
from keras.models import load_model
from keras.utils import to_categorical
from keras.utils.vis_utils import plot_model
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score, confusion_matrix,
roc_auc_score
from sklearn.metrics import roc_curve
#defining a function to plot the ROC curves
def plot_roc_curve(fpr, tpr):
    plt.plot(fpr, tpr, color='blue', label='ROC')
    plt.plot([0,1], [0,1], color='orange', linestyle='--')
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title('Receiver Operating Characteristics (ROC) Curve')
    plt.legend()
    plt.show()
#Add comment
def build_rand_feat():
    X = []
    y = []
    _min, _max = float('inf'), -float('inf')
    for _ in tqdm(range(n_samples)):
        rand_class = np.random.choice(class_dist.index, p=prob_dist)
        file = np.random.choice(sounds[sounds.label==rand_class].index)
        rate, wav = wavfile.read('clean/'+file)
        label = sounds.at[file, 'label']
        rand_index = np.random.randint(0, wav.shape[0]-config.step)
        sample = wav[rand_index:rand_index+config.step]
        X_sample = mfcc(sample, rate,
            numcep=config.nfeat, nfilt=config.nfilt, nfft=config.nfft).T
        _min = min(np.amin(X_sample), _min)
        _max = max(np.amax(X_sample), _max)
    X.append(X_sample if config.mode == 'conv' else X_sample.T)

```

```

    y.append(classes.index(label))
X, y = np.array(X), np.array(y)
X = (X - _min) / (_max - _min)
if config.mode == 'conv':
    X = X.reshape(X.shape[0], X.shape[1], X.shape[2], 1)
elif config.mode == 'time':
    X = X.reshape(X.shape[0], X.shape[1], X.shape[2])
y = to_categorical(y, num_classes=5)
return X, y
class Config:
    def __init__(self, mode='conv', nfilt=26, nfeat=13, nfft=512, rate=16000):
        self.mode = mode
        self.nfilt = nfilt
        self.nfeat = nfeat
        self.nfft = nfft
        self.rate = rate
        self.step = int(rate/10)
#load data
sounds = pd.read_csv('sounds.csv')
sounds.set_index('filename', inplace=True)
for f in sounds.index:
    rate, signal = wavfile.read('clean/'+f)
    sounds.at[f, 'length'] = signal.shape[0]/rate
classes = list(np.unique(sounds.label))
class_dist = sounds.groupby(['label'])['length'].mean()
#creating a class balance by extracting 100ms(0.1s) from each audio recording
n_samples = 2 * int(sounds['length'].sum()/0.1)
prob_dist = class_dist / class_dist.sum()
choices = np.random.choice(class_dist.index, p=prob_dist)
fig, ax = plt.subplots()
ax.set_title('Class Distribution', y=1.10)
ax.pie(class_dist, labels=class_dist.index, autopct='%1.1f%%', shadow=False, startangle=90)
ax.axis('equal')
plt.show()
config = Config(mode='conv')
if config.mode == 'conv':
    X, y = build_rand_feat()
    y_flat = np.argmax(y, axis=1)
    X_train, X_val, y_train, y_val = train_test_split(X, y, test_size = 0.2, random_state = 0)
    input_shape = (X.shape[1], X.shape[2], 1)
elif config.mode == 'time':
    X, y = build_rand_feat()
    y_flat = np.argmax(y, axis=1)
    X_train, X_val, y_train, y_val = train_test_split(X, y, test_size = 0.2, random_state = 0)
    input_shape = (X.shape[1], X.shape[2])
if config.mode == 'conv':
    model = load_model('conv_model.h5')
    model.summary()
    plot_model(model, to_file='conv_model.png', show_shapes=True, show_layer_names=True)
elif config.mode == 'time':
    model = load_model('rnn_model.h5')
    model.summary()
    plot_model(model, to_file='RNN_model.png', show_shapes=True, show_layer_names=True)
#plot the roc curve for the model
y_pred = model.predict(X_val)
print(roc_auc_score(y_val, y_pred, average='micro'))
auc = roc_auc_score(y_val, y_pred, average='micro')
fpr, tpr, thresholds = roc_curve(y_pred, y_val)
plot_roc_curve(fpr, tpr)
y_pred = y_pred.argmax(axis=1)

```

```

y_val = y_val.argmax(axis=1)
#Evaluation Measures
print(accuracy_score(y_val, y_pred))
print(recall_score(y_val, y_pred, average='micro'))
print(precision_score(y_val, y_pred, average='micro'))
f1_score = f1_score(y_val, y_pred, average='micro')
print(f1_score)
confusion_matrix = confusion_matrix(y_val, y_pred)
#print(f1_score)
print(confusion_matrix)

```

APPENDIX D: PYTHON CODES FOR 10-FOLD MODEL VALIDATION

```

# KFold Cross Validation approach
kf = KFold(n_splits=10,shuffle=True)
kf.split(X)
# Initialize the accuracy of the models to blank list. The accuracy of each model will be appended to
this list
accuracy_model = []
# Iterate over each train-test split
for train_index, test_index in kf.split(X):
    # Split train-test
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    # Train the model
    model.fit(X_train, y_train, epochs=10, batch_size=32,
              shuffle=True,
              class_weight=class_weight)
    # Append to accuracy_model the accuracy of the model
    accuracy_model.append(accuracy_score(y_test.argmax(axis=1),
model.predict(X_test).argmax(axis=1), normalize=True)*100)
if config.mode == 'conv':
    model.save('conv_model.h5')
elif config.mode == 'time':
    model.save('rnn_model.h5')
y_pred = model.predict(X_val)
y_pred = y_pred.argmax(axis=1)
y_val = y_val.argmax(axis=1)
#Evaluation Measures
print(accuracy_score(y_val, y_pred))
print(recall_score(y_val, y_pred, average='micro'))
print(precision_score(y_val, y_pred, average='micro'))
#f1_score = f1_score(y_test, y_pred, average='micro')
confusion_matrixs = confusion_matrix(y_val, y_pred)
#print(f1_score)
print(confusion_matrixs)

```

APPENDIX E: PYTHON CODES FOR AUC-ROC

```

# Compute ROC curve and ROC area for each class
fpr = dict()
tpr = dict()
roc_auc = dict()
n_classes = len(classes)
for i in range(n_classes):
    fpr[i], tpr[i], _ = roc_curve(y_val[:, i], y_prob[:, i])

```

```

roc_auc[i] = auc(fpr[i], tpr[i])
# Compute micro-average ROC curve and ROC area
fpr["micro"], tpr["micro"], _ = roc_curve(y_val.ravel(), y_prob.ravel())
roc_auc["micro"] = auc(fpr["micro"], tpr["micro"])
# First aggregate all false positive rates
all_fpr = np.unique(np.concatenate([fpr[i] for i in range(n_classes)]))
# Then interpolate all ROC curves at this points
mean_tpr = np.zeros_like(all_fpr)
for i in range(n_classes):
    mean_tpr += interp(all_fpr, fpr[i], tpr[i])
# Finally average it and compute AUC
mean_tpr /= n_classes
fpr["macro"] = all_fpr
tpr["macro"] = mean_tpr
roc_auc["macro"] = auc(fpr["macro"], tpr["macro"])
# Plot all ROC curves
plt.figure()
lw=2
plt.plot(fpr["micro"], tpr["micro"],
         label='micro-average ROC curve (area = {0:0.2f})'
         ".format(roc_auc["micro"]),
         color='deeppink', linestyle=':', linewidth=4)
plt.plot(fpr["macro"], tpr["macro"],
         label='macro-average ROC curve (area = {0:0.2f})'
         ".format(roc_auc["macro"]),
         color='navy', linestyle=':', linewidth=4)
colors = cycle(['aqua', 'darkorange', 'cornflowerblue'])
for i, color in zip(range(n_classes), colors):
    plt.plot(fpr[i], tpr[i], color=color, lw=lw,
             label='ROC curve of class {0} (area = {1:0.2f})'
             ".format(i, roc_auc[i]))
plt.plot([0, 1], [0, 1], 'k--', lw=lw)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('AUC-ROC')
plt.legend(loc="lower right")
plt.show()
print(roc_auc_score(y_val, y_pred, average='micro'))

```