An Acoustic-Based System for Early Detection of Elephants in Nyangores Forest

By:

Sandra Kahoro

113610

Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Information Technology (MSc.IT) at Strathmore University

School of Computing and Engineering Sciences, Strathmore University,
Nairobi, Kenya

June 2025

Declaration and Approval

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Sandra Kahoro

June 2025

Approval

The thesis of Sandra Gathoni Kahoro was reviewed and approved for examination by the following:

Dr. Henry Muchiri,

Senior Lecturer, School of Computing and Engineering Sciences,

Strathmore University

Dr. Julius Butime,

Dean, School of computing and Engineering Sciences,

Strathmore University

Prof. Bernard Shibwabo,

Director of Graduate Studies,

Strathmore University

Table of Contents

Abstract

Elephants greatly contribute to their ecosystem through seed dispersal, creation of water sources when it is dry through digging and even carbon sequestration. However, their population has been reducing due human-elephant conflict, poaching and habitat degradation. In Nyangores Forest, which is a part of Kenya’s greater Mau Region, frequent encounters between forest-adjacent communities and elephants often lead to injuries or fatalities on either side, exacerbating the human-elephant conflict. Elephants communicate through vocalization and one of the most common and frequently used are the low rumbles which are very low frequency sounds used by the elephants and can travel over long distances and are below human hearing level. The study aimed to develop an early detection system based on the elephant vocalizations to reduce conflict and enhance conservation. There have been many methods that have been used in the detection of elephants such as imaging with the use of camera traps which does not work well with harsh weather conditions which can affect the angles and the clarity and seismic monitoring using vibrations from their vocalizations which are felt on the ground which has limited range and high sensitivity to environmental noise. This was achieved using Rapid Application Development (RAD), integrating a machine learning model with an IoT device for sound classification and localization. The CNN model was trained and tested and attained an accuracy of 99%. The IoT device was developed using Raspberry Pi Pico which resulted in a working integrated system which can detect elephant vocalizations and send an alert demonstrating real-time functionality. The results confirmed the viability of acoustic localization for elephant detection. The system developed offers a low-cost, efficient and scalable alternative for elephant detection and monitoring. Its implementation has the potential to strengthen existing conservation efforts and infrastructure used by different agencies and bodies, improve ranger response times and integrate with deterrent mechanisms such as bee fences.

Acknowledgements

I express gratitude to God for granting me the grace and good health that He has accorded me throughout the process of writing this research thesis. I greatly appreciate my supervisor who has been of great support throughout the process and provided me with invaluable advice and knowledge. I am grateful for his guidance and feedback on the project. I am grateful to Strathmore University for the resources that I have had access to throughout the process of proposal writing which have allowed me to prepare for the research study. Lastly, I would like to recognize my family and friends for their support, input and insight on my ideas and for the emotional support.

Abbreviations/Acronyms

CNN	Convolutional Neural Networks
DCT	Discrete Cosine Transform
DNN	Deep Neural Networks
DoA	Direction of Arrival
ELP	Elephant Listening Project
FFT	Fast Fourier Transform
GCC-PHAT	Generalized Cross-Correlation with Phase Transform
HEC	Human Elephant Conflict
IoT	Internet of Things
MFCC	Mel-Frequency Cepstral Coefficient
ML	Machine Learning
MSC	Multiple Signal Classification
MT	Multitask Learning
RNN	Recurrent Neural Networks
SRP	Steered Response Power
SSL	Sound Source Localization
TDOA	Time Difference of Arrival
TF	Time Frequency

Definition of Terms

Acoustic Localization	This is a method that is used to identify the location of a sound source that is based on the time delays of arrival between multiple time-synchronized audio recordings (Acoustic Localization — Opensoundscape 0.11.0 Documentation, 2024).
Acoustic Source Localization	This is similar to acoustic localization where the location of a sound source is determined by analysing the time differences in sound arrival at many or multiple sources (Acoustic Localization — Opensoundscape 0.11.0 Documentation, 2024).
Biodiversity	This is all the different kinds of life forms present in one area which could be the variety of animals, plants, fungi and even microorganisms that work together in an ecosystem to create a balance (What Is Biodiversity and Why Is It under Threat? 2024).
Ethogram	A detailed list, inventory, catalogue or description of the behaviours or actions exhibited by different species and can be said to be a master list (Elephant Listening Project Home, 2024).
Home Range	The space used by animals for difference purposes: as a habitat, to reproduce, source of food and water (Viana et al., 2018).
Human Elephant Conflict	Negative interactions between elephants and humans such as crop raiding, property destruction or close interactions between humans and elephants that may trigger defensive behaviours and lead to injury or death for both parties (International Fund for Animal Welfare, 2023)

List of Figures

List of Tables

List of Equations

Chapter 1: Introduction

1.1 Background

Kenya's forests are more than just landscapes: they are the life support systems which play a vital role in maintaining ecological balance, regulating water flow, and sequestering carbon and elephants are a critical part of this ecosystem. The Nyangores Forest in the Mau Region is not only home to elephants but also supports a diverse range of wildlife elephants being the largest in the ecosystem and contributing the most to it. The elephant situation in Nyangores Forest is intertwined with ongoing conservation efforts and challenges that are related to human-wildlife conflict. In Nyangores Forest, a community of farmers live on the forest edge which is close to the elephant habitat. With the need for resources for both the humans and elephants, conflict is bound to come up. Farmers depend on the produce from their settled farms and elephants can be quite destructive even when just passing through and the dependence on the farm produce makes the encounters emotionally charged (*Human-Elephant Conflict in Kenya*, 2022).

Both species of elephants play a vital role in their ecosystems. Over the years, factors such as climate change, land use conversion and human population growth have been leading to human elephant conflict due to coexistence challenges between humans and elephants (Montero-Botey et al., 2024). Elephant activities such as pushing down trees as they walk and digging for water, results in the creation of habitats and water sources for smaller animals within the ecosystem (Mutero, 2021).In some ecosystems, the eating and trampling of trees which creates gaps in the vegetation, in turn ends up affecting some ecosystem services such as carbon accumulation and carbon storage within the ecosystem (Cardoso et al., 2020). Additionally, these large mammals assist in seed dispersal and nutrient cycling through their dung which encourages the growth of different vegetation all around their ecosystem.

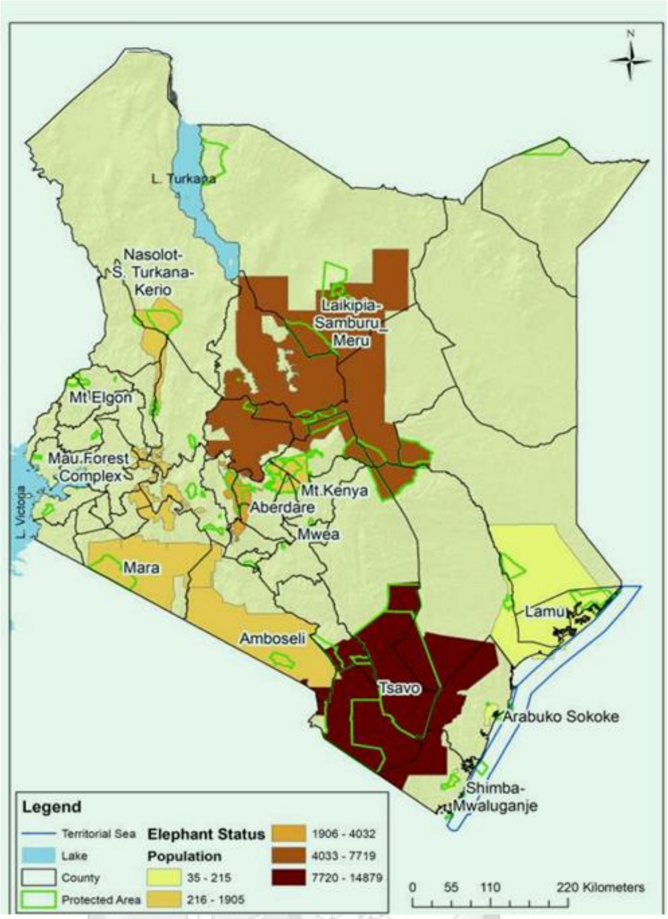


Figure 1.1: Map Showing the Elephant Range in Kenya (National Elephant Action Plan for Kenya, 2021)

Illegal hunting, unauthorized timber harvesting, and infrastructure expansion represent just a few of the numerous ongoing threats to biodiversity, with large mammals being particularly vulnerable (Bjorck et al., 2019). The elephant habitat has been rapidly decreasing over the years due to different factors such as climate change which affects their natural habitat and the expansion of human corridors especially in areas that are forest or savanna adjacent (Ramasubramanian et al., 2022). The heightened conversion of land and population growth within the different wildlife habitats: savanna or forest, has escalated human-wildlife conflicts, posing a significant concern. This is particularly critical as large, wide-ranging mammal species bear the brunt of these conflicts, with many facing critical endangerment or rapid population decline (Hahn et al., 2022).

According to Santosh et al. (2018), most human wildlife interactions prove dangerous for both parties and in this case, human elephant conflict which usually ends up in fatalities. This highlights the need for an early warning system which can help mitigate human elephant conflict. Human-elephant conflict (HEC) usually describes negative interactions between elephants and humans such as crop raiding, property destruction or close interactions between humans and elephants that may trigger defensive behaviours and lead to injury or death for both parties (International Fund for Animal Welfare, 2023). Ranasinghe et al. (2023) states that such systems are very important given the growing human population and human activities which leads to the shrinking of the elephant habitats. Usually in human-dominated areas with lots of hunting or persecution, wildlife perceives humans as a threat which leads to the animals avoiding any encounters although the animals may seek out the areas that are human-occupied in search of high-quality food (Hahn et al., 2022).

The co-existence of humans and wildlife is a fundamental aspect of our ecosystems which will shape the human experience (Ravikrishna et al., 2023). Human elephant conflict usually involves both parties that impact each other negatively and pose a threat to human and animal life. According to Ravikrishna et al. (2023), this conflict has more significant consequences for our biodiversity and ecosystem health. Currently, there are inadequate monitoring and response technologies that are being used or may also fall short especially in the remote areas or dense forest areas in Kenya which prove difficult and resource intensive to monitor and respond to. The systems set up appear to be more of a reaction to the elephant's behaviour. By implementing acoustic localization, early detection can be used and even tracking and studying of the elephants especially in the dense forests in Kenya can be effectively done. Traditional conservation efforts such as camera trapping are no longer feasible (Gomes-Selman et al., 2021).

Over the years, conservation challenges have grown with the increased human activity and different factors such as climate change varying across different elephant populations and ecosystems. Acoustic source localization technology is used to determine the location of a sound source or multiple sources by processing acoustic signals. This technology is applied to elephant monitoring due to their rumble that is a low frequency call that can be heard over a very long distance which can be picked up.

1.2 Problem Statement

In Kenya's intricate forest ecosystem, African elephants experience a multitude of challenges, including the persistent menace of poaching, habitat degradation, and escalating conflicts with human activities. The traditional methods used in conservation efforts are no longer applicable as they have certain limitations. Aerial and ground surveys are very labour intensive, costly and are inevitably affected by unfavourable or bad weather conditions

(Brickson et al., 2023). Monitoring sometimes proves inadequate or difficult, particularly in remote and densely forested regions, highlighting the necessity for the development and deployment of advanced acoustic localization techniques tailored specifically for elephant conservation.

Human elephant conflict usually involves both parties that impact each other negatively and pose a threat to human and animal life and this is the leading cause of elephant mortality (Ravikrishna et al., 2023).

1.3 Research Objectives

1. To identify the challenges in existing elephant detection and monitoring techniques in Kenya
2. To review the different machine learning models used in bioacoustics monitoring
3. To develop an early detection system for elephants using the low rumble sounds
4. To test the early detection system

1.4 Research Questions

1. What are the challenges in elephant detection?
2. What are the current technologies in use for African elephant detection?
3. How can an elephant detection system be developed?
4. How well can the system detect and transmit data

1.5 Justification

Human elephant conflict (HEC) has been known to pose significant threats to both humans and elephants themselves which sometimes leads to fatalities and economic loses. Over the years, elephants have been forced to share land, being squeezed into smaller areas and as a result, there is some conflict between the neighbouring communities and the elephants (Human -Elephant Conflict | WWF, 2024). In northern Kenya, elephants have become victims of retaliatory killings as the gravity of HEC becomes prevalent and are injured, shot or speared to death by the local communities as a defence mechanism (The Grim Reality of Human-Elephant Conflict in Northern Kenya - Save the Elephants, 2021). Different factors are aggravating the human elephant coexistence where the land where the elephants once foraged is now home to human agriculture making it difficult to access resources. Climate change is also a major factor whereas the temperatures rise, and rain patterns change, the resources become more scarce and the elephants are pushed into community land (International Fund for Animal Welfare, 2023).

Humans also face their own challenges where they have to move deeper into elephant territory such as in forest adjacent areas where they have to fetch water or collect firewood (International Fund for Animal Welfare, 2023). The elephant habitat is also increasingly fragmented by new villages, farms or highways. With all these changes, this shows that humans and elephants are closer than ever before and must coexist with minimum conflict. Elephant detection and deterrence methods are mainly implemented in mitigating HEC for forest adjacent communities such as sensor-based alarm systems which use cameras and infrared sensors and alert the communities on the presence of elephants (Ramasubramanian et al., 2022).

According to Gomes-Selman et al. (2021), acoustic monitoring is especially useful in monitoring and detection due to the low frequency elephant rumbles or calls that can usually travel over a kilometre in the dense forest. Autonomous sensing methods are transforming the fields of ecology and conservation which is a non-invasive method as compared to human observers or even collaring of the animals (Rhinehart et al., 2020). Acoustic meaning sound signals travel in all directions and elephants are very good in production of low frequency sound for long distance communication (Acoustic Communication, 2024).

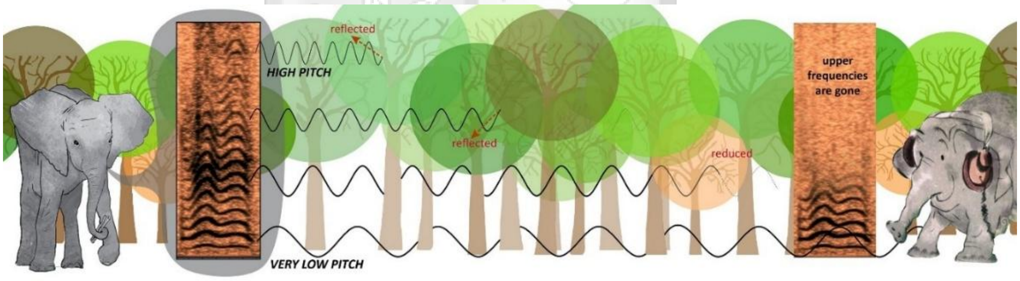


Figure 1.2 Illustration of low frequency calls (Elephant Listening Project Home, 2024)

1.6 Scope and Limitations

The Geographical scope of this study focused on Kenya’s forest ecosystems which is where the African Forest elephants reside where the Nyangores forest has been identified as part of greater Mau Forest. The research focus areas are: Acoustic Localization of African elephants which involves investigation and development of advanced methods. The data collection and analysis used existing data instead of deploying sensor networks across the forests. It was not possible to deploy the alert system to the forest Rangers and the forest adjacent communities for user testing due to the limited time factor. The research did not cover the sensor placement issues which was not considered out of scope for this study.

Chapter 2: Literature Review

2.1 Introduction

The correlation and co-existence between humans and elephants is a complex and dynamic challenge as is creating a future for both (Gross et al., 2022). Unfortunately, over the years human-elephant interactions are becoming more and more negative and even fatal sometimes in some parts of Africa and Asia (Kamdar et al., 2022). Over the years, different methods have been developed and used ranging from traditional methods to the use of technology to assist in the mitigation of HEC. In understanding the different aspects such as studies done, the different theories that are linked to human elephant conflict and the technologies used so far, the direction of the study can be more informed and have proper basis.

2.2 Empirical Literature Review

An empirical literature review is important to understand the current technologies being implemented in the field and their effectiveness in mitigating or reducing human elephant conflict. Numerous studies have tested and implemented different technologies all over the world, especially in Asia and Africa such as motion detectors, acoustic systems, collaring to detect and monitor elephant movements.

2.2.1 Elephant Movement Patterns

The movement of the African elephant is complex and is influenced by many different environmental and social factors. The patterns are known to vary depending on the habitat, seasons and in response to certain conditions or stimuli such as resource availability. According to Beirne et al. (2021) who conducted a study in Gabon observed that elephant movement patterns show variation influenced by individual behaviours and environmental factors. Tracking data from a total of 96 elephants showed that that the elephants had an annual movement distance of 2,463 kilometers and a distance range from 1,420 kilometers to 3,436 kilometers. Environmental factors such as heavy rainfall increased the distance by which the elephants moved by 59%, while human disturbance reduced the movement distance by 58% (Beirne et al., 2021). As shown by Beirne et al. (2021), the male and female also have a difference in home range where the male were observed to be 238km2 which is higher compared to the female range of 135km2. Home range is the space used by for difference purposes: as a habitat, to reproduce, source of food and water (Viana et al., 2018). The home range was seen to significantly reduce in size due to human disturbance by 63%.

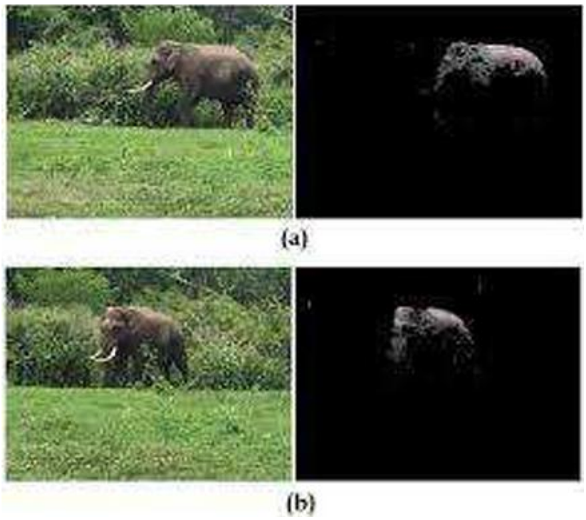
A different study on elephant movement patterns and the relation to the human factor showed that elephants move faster through community lands to avoid the humans where elephants in community land moved at speeds of 1.24km/h compared to their speed of 0.55km/h in protected areas (Adams et al., 2021). The study used collars to track four female elephants over a period of 13 months. The elephant movement patterns varied and differed by the land use and the time of day. Elephants are known to be dependent on vegetation hence depend on large and distributed areas for their resources hence their movement into cultivated land (Sai & Dewangan, 2023). Elephant movement patterns are also seen to be linked to resource availability such as fruit or water and environmental factors such as climate where the availability influences the elephants' movement into community land or farms in search of resources (Beirne et al., 2020).

2.2.2 Elephant Detection Techniques

Over the years there have been different detection and monitoring techniques used. These techniques focused more on deterrence. Some of the methods of detecting and tracking elephants are GPS trackers which are collared, monitoring using video and radio and use of wireless sensors (Sai & Dewangan, 2023).

2.2.2.1 Image Processing

There are different ways to integrate imaging techniques in elephant detection and monitoring. One method that is seen is the use of TensorFlow where Raspberry Pi 4 is integrated with an IP camera and the Raspberry controller processes the camera video feed (Ravikrishna et al., 2023). Image processing can be costly in terms of resources being used such as devices, algorithms and even data leading to slow detection. As seen in the figure below, data must be collected, and this would include large amounts of images that would be used in detection of elephants. This would mean that the camera placement would also have to be good to ensure quality imaging to ensure accuracy. This would be a shortcoming as to get proper angles, the camera would have to be placed all around the forest, such as on the ground, on the trees. With the use of cameras, it can be difficult to capture proper images in dense areas or locations.



(a)

(b)

Figure 2.1 Images of elephants during the day and at night (Ravikrishna et al., 2023)

The use of imaging had been the most widespread before the use of sensing. Photographic elephant detection is done by comparing physical characteristics of the elephants and finding any mismatches between new elephant images and existing images in the database (Weerasinghe et al., 2022). The real-time detection has been integrated with alert and warning systems. As seen in Figure 2.2, the output is seen as the elephants are detected and distinguished through a labelling process and shown with the use of differently coloured boxes (Ravikrishna et al., 2023).

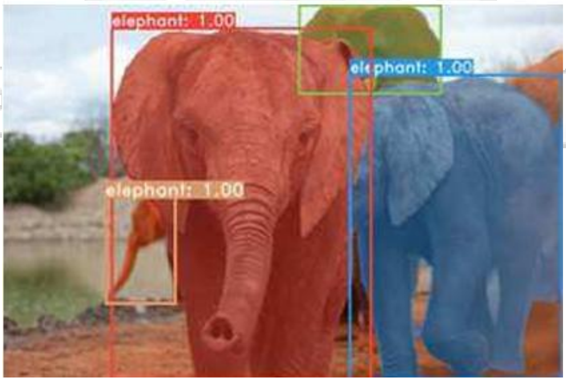


Figure 2.2 Imaging Detection Output (Ravikrishna et al., 2023)

Furthermore, counting wildlife has been done using aerial remote sensing images which is one of the popular vision tasks. This can be done using detection-based approach or regression-based methods (Singh et al., 2023). This method can be applied to detection of animals both marine and land. A disadvantage of this as an application in detection is that it avoids training and learning a detector to predict object location but directly works on counting. In its own light, this is an advantage in counting as the resources are reduced. The inputs required are not as resource intense as those for imaging. Here, dots or scalar numbers are required as inputs and not bounding box annotations (Singh et al., 2023).

Thermal imagery has also been proposed and used in animal detection. Algorithms such as CNN-based deep-learning model (YOLO v4) have been used in thermal imagery (Khatri et al., 2022). This method has also been said to have some challenges as it can be affected by the atmospheric temperature and extreme weather conditions.

2.2.2.2 Acoustic Monitoring

An acoustic sensor is used in identifying elephant calls which are low-frequency rumbles which can travel over long distances as seen in Figure 1.1 (Wireless Integrated Sensor Network: Boundary Intellect System for Elephant Detection via Cognitive Theory and Fuzzy Cognitive Maps - ScienceDirect, 2018).The frequency of the rumble ranges between 15-35Hz and has a duration of 0.5 seconds to 5 seconds. Passive acoustic monitoring (PAM) is a non-invasive method that is used for conservation strategies and provides evidence through large-scale monitoring of elusive, acoustically active species such as elephants (Conserving African Forest Elephants through Bioacoustic, 2024). Acoustic monitoring primarily uses Machine Learning (ML) and Internet of Things (IoT) to collect the data and analyse it. This means that the acoustic sounds can be captured real-time for classification which is pivotal for an early detection system. The sensors which are IoT devices are used to capture sounds while ML is used for the classification of inputs from the sensors and the results are further transmitted to the early detection system.

In the applications of acoustic source localization, the environment can be extremely challenging due to reverberation and reflections, complex noise fields, interference signals and concurrent speakers. Acoustic localization is done to estimate an elephant's location or a herd's location by quantifying the time difference of arrival (TDOA) where the time delay of arrival is also considered of its sound at each sensor (Rhinehart et al., 2020). Acoustic localization can be done for both marine and terrestrial animals. According to (Rhinehart et al., 2020), terrestrial localization relies on receiving sound directly form the source, in this case the source being an African elephant. The microphones used to pick up the sounds are omnidirectional meaning that they can receive the sounds from different directions ("Introduction to the Issue on Acoustic Source Localization and Tracking in Dynamic Real-Life Scenes," 2019).

2.2.2.3 Acoustic Localization

Acoustic source localization was first applied more than 100 years ago where an invention of acoustic telescopes was made to detect ships in foggy weather (Yangfan J et al., 2021). According to Yangfan J et al., (2021), there are two main types of acoustic localization techniques which are beamforming and holography which are used to predict sound fields. With the use of beamforming, sound sources are found by scanning potential sound source locations and determining the likelihood of an actual source at each potential location. On the other hand, in holography, the sound sources are assumed to be existent in all the potential locations and the strength of each potential sound source is estimated by finding the best match. After this, a prediction of the sound field can be made by solving a sound propagation problem.

Most applications of acoustic localization use direction of arrival (DoA) and Time of Arrival (ToA) technique. The DoA method estimates the angle from which an acoustic signal reaches a sensor. Algorithms such as beamforming and phase-mode expansions are used in this method. ToA methods, unlike DoA measures the time taken for an acoustic signal to reach a sensor. The accuracy can be affected by the positioning of the sensors. An approach to the two has been studied, where a linear system is used to model sensor data where DoA and ToA are both used by under sampling Fourier coefficients of the sensor data to optimize the method and sounds can be located over a long distance.

According to Krause et al., (2021), there are many classical Direction of Arrival estimation methods that have been established such as generalized cross-correlation with phase transform (GCC-PHAT), Steered Response Power (SRP) and subspace -based approaches. The estimation methods have changed over time with the growth of technology, machine learning is now being used in this estimation which increases their robustness. Current work focuses on Deep Neural Network (DNN) which can use diverse signal representations with the use of feature extraction being incorporated into the data training process. Convolutional Neural Networks (CNN) are used for feature extraction. Multitask learning (MT) is also a technique that has been discussed which is a machine learning technique where multiple tasks can be performed by a trained model. Multiple tasks can be performed at the same time as some of the network's layers have been shared. In a study, multitask learning of CNN has also been used where sound source localization (SSL) is considered a classification problem (Pang et al., 2019).

2.2.2.4 Seismic Monitoring

Elephants also communicate through seismic vibrations which are ground vibrations which are used to transfer information (Why Do Elephants Use Seismic Communication?, 2023). The seismic vibrations are produced with the low frequency rumbles which is also acoustic (Reinwald et al., 2021). According to Reinwald et al., (2021), seismic signals generated by vocalizing elephants can be used to locate them. The study used seismometers together with acoustic sensors which were placed in four locations where the seismic sensors were buried. GPS synchronization was also used for the accurate timing of signals. With these sensors, a stable power source is required to allow for prolonged data collection. This study concluded that the combination of seismic and acoustic localization did not add any accuracy or additional information but was rather preferable to use them individually.

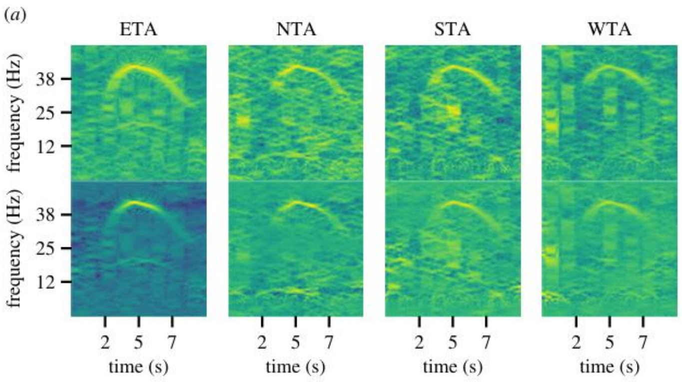


Figure 2.3 A sample of signals collected from four seismograms (Reinwald et al., 2021)

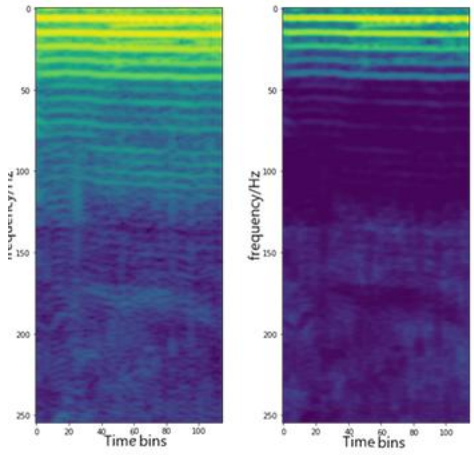
The use of seismograms also faces challenges such as the external weather conditions such as wind which can introduce other forms of seismic noise and the type of instrument or tool used to collect the sounds could also affect the outputs of the study. This means that the choice of instruments and conditions for study must be considered. The choice has to be well reasoned and justified to ensure optimum results and collection which would include noise in difficult or harsh environments where there are different forms seismic noise introduced.

2.2.3 Sound Analysis in Conservation

Sound analysis has become a very important tool in conservation technology using bioacoustics to monitor the biodiversity, assess ecosystem health and to inform management strategies and decisions. This method uses audio recording technologies to gather data on wildlife and environmental conditions with minimal disturbance making it a non-invasive method. Ranasinghe et al., (2023) focuses on an approach to detecting wild elephants using acoustic signals. This approach is to aid in curbing and reducing human-elephant conflict which is a major socio-economic issue that is experienced in Sri Lanka fueled by the increasing human population and shrinking elephant habitats.

In this paper, a robust approach is proposed to detect elephant vocalizations by using time frequency domain analysis. Magnitude spectrograms are used as input features and feature extraction is carried out using a CNN-based approach which is implemented using YAMNet model. The study opted to use spectrograms as the input feature which is the sound data representation for this study. Spectrograms first transform the audio signal into a two-dimensional (2D) representation having time on one axis and frequency on the other which enables the DNN to be able to analyze the sound signals with better precision. It is noted that most studies that are working with elephant acoustics prefer to work with spectrograms due to simplicity and ability to enhance audio components especially with low frequencies such as rumbles.

Two main datasets were used in Ranasinghe et al., (2023) study; an elephant data set that already existed then combined with recently recorded audio to have comprehensive training data which can also be tested and validated. According to Gomes-Selman et al., (2021) one of the main challenges experienced is the diversity and noise found in forest sounds. Which adds a great amount of data variability. In but in Gomes-Selman et al., (2021) study however, a very small number of elephant calls was available for training and hence decided to generate new examples by combining existing calls with new randomly sampled non-call segments. Both studies ensured hard negative samples were included during training. Noise reduction or filtering techniques can be used to mitigate this. In the Ranasinghe et al., (2023) study, a noise reduction technique was implemented to increase the accuracy of the output from the spectrograms. A median filter was applied on top of the input spectrogram creating a data-dependent mask which was then multiplied with the original spectrogram, and this resulted in reduction of noise as seen in figure 2.5 below.



(a) Before Filtering (b) After Filtering

Figure 2.4 Result of the applied filtering technique (Ranasinghe et al., 2023)

After noise reduction, a Mel-spectrogram was applied with the use of a weight matrix further accentuating the lower frequency components of the spectrogram which is crucial for accurate acoustic detection of elephants. A study carried out, implemented the use of HARKBird, a MUSIC-based method which has made significant steps towards automatic detection of sounds where overlapping sounds are separated into multiple noise-reduced recordings but a downside to this is that the audio recordings still must be manually reviewed. This is a big disadvantage as carrying out both steps would be time consuming as compared to implementation of a different noise reduction technique as previously discussed (Rhinehart et al., 2020). With elephant rumbles, the data collected highly influences the accuracy as it is a low frequency audio signal. Different techniques are used to reduce noise or even a combination of techniques which assist in the filtering or noise reduction in the audio data.

2.3 Theoretical Literature Review

The different theories that revolve around elephant conservation and human-elephant conflict are of great importance in this study. The purpose is to examine the different theories that exist, the relationships between the different mitigation strategies, the communities, the biodiversity and the challenges experienced (Karas, 2024).

2.3.1 Human-Elephant Mitigation

Human-elephant conflict (HEC) poses significant challenges for both the conservation efforts and local communities or adjacent communities. There have been several methods that have been used to mitigate the HEC in different regions and areas in both Africa and Asia. For long-term solutions for HEC, a focus on both management efforts on site-specific considerations as a well as the formulation and application of strategic plans at the landscape level (Beirne et al., 2020). Elephants are also largely known for their socio-cognitive abilities and capacity for multi-modal sensory perception and communication (Ball et al., 2022). One method suggested by Ball et al., (2022) is to consider how elephants can be provided with an option or options that affect their decision-making process which is mostly linked to resources in ways that would prevent negative interactions with the communities.

2.3.1.1 Barrier-based solutions

A different mitigation method was by a study in India that participated in the establishment of fences in 42 villages across different districts to check how effective fences were in the mitigation of human elephant conflict. From the study, it was concluded that for fences to work, they had to be well maintained and a good way to do this is through community maintainers (Kamdar et al., 2022).

2.3.1.2 Community-Based Strategies

HEC seems to not only be mitigated using one method, but the neighbouring communities are part of the stakeholders, and the mitigation strategies have to consider both parties as considering one party will not fully solve or the mitigate the issue at hand (Gross et al., 2022). Communities are also involved in conservation efforts and educated on the importance of the wildlife and their habitats and incentivised to participate in taking care of the elephants and assist in conservation efforts to make it easier to coexist with the elephants (Gross et al., 2022).

2.3.1.3 Translocation and Habitat Management

Translocation of elephants is the deliberate movement of wild elephants from one area to another which could be from one conservancy to another or one protected area to another which is usually to mitigate human elephant conflict (HEC). Over the years, Kenya has been using translocation for its elephant conservation efforts to enhance habitat sustainability. Recently, Kenya Wildlife Service translocated approximately 50 elephants from Mwea National Reserve to Aberdare National Park (Mwangi, 2024).

2.4 Algorithms

Algorithms are step by step instructions that are used to complete a task and is usually a mathematical process or function (Nikolopoulou, 2023). The choice of algorithm is critical to the system's overall effectiveness and accuracy for the early detection of elephants. The insights gained form this analysis did not only inform the design of the detection system but also contribute to a better understanding of the use of different algorithms in conservation technologies, specifically monitoring technologies.

2.4.1 Deep Learning in Acoustic Localization of elephants

Deep Learning uses several layers of neural networks known as Deep Neural Networks to make more accurate predictions or estimations. This allows for more features to be identified from the data. In most studies, deep learning has been implemented and one of the algorithms used is Convolutional Neural Network (CNN) which are used in training very large data sets. CNN can be used differently in acoustic localization such as in classification of elephant vocalizations. In a study, three different CNNs were combined to form a two-level hierarchical model which gave good results with 92% accuracy for first level, model and 100.0% accuracy for the second level model which classified low-level frequency vocalizations and 75.0% for second level model that classified the high frequency vocalizations (Ravikrishna et al., 2023).

In other studies, the CNN comes in as early as feature extraction as they are well suited for processing spectrograms (Ranasinghe et al., 2023). The CNNs in this study were trained on both clean and noisy which enabled them to be able to only extract meaningful features from the audio signal. Convolutional layers were used in both studies but a difference in application is seen. The optimization of the models is done differently in both models where one utilizes YAMNet which is developed by Google and transfer learning techniques where prior knowledge is leveraged, reducing the training time. In the other study, optimization of the model is done by changing the approach to classification of the vocalizations moving from a flat approach to a hierarchical approach. A study by Mohmmad et al., (2023), had a proposed method of using CNNs and Recurrent Neural Networks (RNNs) which brings about CRNN which combines feature extraction of CNN and LSTM for sequence prediction. CRNN is known to be more efficient, accurate and robust compared to just CNN. The research study had a good accuracy in classification of acoustic signals although the data used was from YouTube which can affect the accuracy of the data.

This study selected CNN as the primary architecture for elephant vocalization detection due to the high performance in spectrogram classification as seen in the different studies carried out and the efficiency in deployment on resource-constrained devices compared to CRNN or LSTM which require more resources in terms of computational power which make them less suited for real-time processing. Additionally, CNN is seen to provide excellent accuracy in existing wildlife acoustic studies while maintaining shorter training times and simpler architectures. While hybrid models such as CRNNs may offer marginal performance improvements under ideal lab conditions, their added complexity are not practical for this study.

2.4.2 Localization and Acoustic sources tracking

There have been recent advances in the development of acoustic signal processing methods for localization and the tracking of acoustic sources. In research carried out, CNN which is used in the estimation of directions of arrival (DoA) of multiple sound sources is mentioned where this method outperforms the multiple signal classification (MUSIC) but CRNN is also brought in as an even more robust method for DoA estimation in reverberant environments. The CRNN is further proposed for a joint acoustic event localization and detection of several overlapping sound events. The proposed technique is seen to be able to associate multiple DoAs with the respective labels and even further track this association over a time ("Introduction to the Issue on Acoustic Source Localization and Tracking in Dynamic Real-Life Scenes," 2019).

2.4.3 Sound classification based on different techniques

The use of Mel-Frequency Cepstral Coefficients (MFCC) is very popular in sound classification due to their feature engineering to resemble human being hearing (In Tan et al., 2022). However, in this study sound spectrum features were used for sound classification. CNN is also mentioned as the dominant deep learning model for sound classification due to its high accuracy. A disadvantage of MFCC is that as it resembles human hearing, it lacks the ability to properly capture high frequency or pitch and further requires mor processes to convert audio signals into sound representations. This paper mentioned that sound spectrum as a sound representation can be implemented in any application that can display frequency spectrums.

In audio data pre-processing, data preparation and feature extraction are carried out. The audio data can be transformed int MFCCs which is a numerical representation that is derived from spectrograms through computations such as time frequency representations then transforming the spectrogram into a Mel-spectrogram which is achieved by mapping a linear frequency onto a non-linear frequency scale which matches the resolution of the frequency of the human ear (Dasari et al., 2023). It is then compressed to decrease the dynamic range of the data which assists in reducing fluctuations and noise. As seen in the figure below (figure 2.4), the Mel-Spectrogram has more noise compared to MFCC.

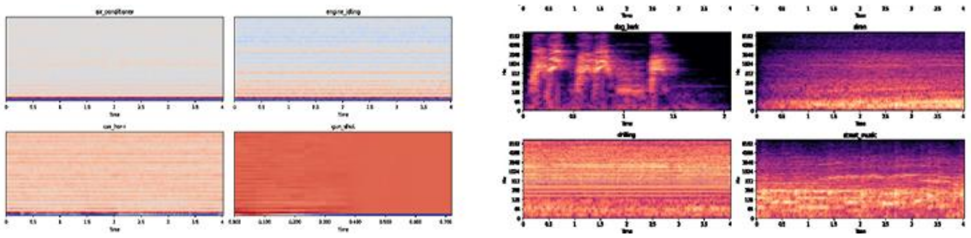


Figure 2.5 Side by side comparison of MFCC and Mel-spectrogram (Dasari et al., 2023)

Figure 2.5 shows the process of feature extraction for both MFCC and sound spectrum with the use of Fast Fourier Transform (FFT) for MFCC. For MFCC, the process is longer where the analog sound signal converted by recording devices into digital files. After, FFT is used, and a function is applied to its output and the immediate data is converted back into the time domain using Discrete Cosine Transform (DCT) or Inverse Fast Fourier Transform in order to get MFCCs.

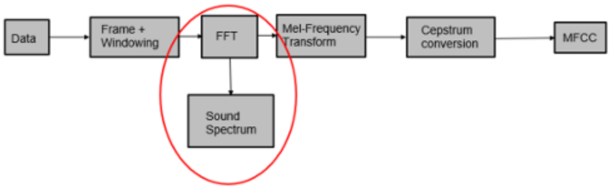
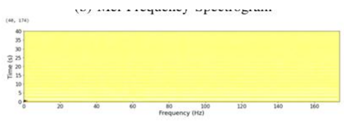
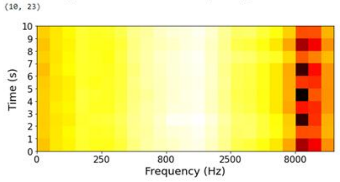


Figure 2.6 Feature Extraction of MFCC and Sound Spectrum (In Tan et al., 2022)

However, as seen in the image above, the sound spectrum is a collection of frequency bands that are the output of FFT in the conversion process. An advantage of sound spectrum that can be deduced from the two processes is that since it does not go through any logarithmic transformation, the features and range frequency of the original recording are maintained. The difference in features is illustrated in figure 2.6 which affects the precision in prediction for each model.



(c) MFCC features in spectrogram



(d) sound spectrum features in spectrogram

Figure 2.7 Difference in features in MFCC and Sound Spectrum (In Tan et al., 2022)

From this study, Sound Spectrum was seen as a viable sound representation especially with CNN due to its accurate predictions and was also able to match predictions made by models that utilized MFCC. However, sound spectrum requires complex model architectures to utilize its features and with this study, it was difficult to determine how far sound spectrum can be used due to the varied number of audio files per class. Another sound event classification technique is where acoustic feature extraction is based on graph signal. After the time frequency (TF) has been resized, it is converted into a one-dimensional vector in order to get the graph signal and divided into frames with the frame length being L_f with an overlap of $L_{overlap}$ (Liu & Wei, 2019). The local maximum of each frame is then determined so as to get the high energy spectral components to form a vector.

Equation 2.1 Getting the High energy spectral components of sound

$$M(j) = max\left(frame_j\right), j = 1, \ldots, N_f$$

Where

$$N_f = \frac{length(G) - L_f}{L_f - L_{overlap}} + 1$$

The event feature is then retrieved by comparing each component with the minimum as illustrated in equation 2.2.

Equation 2.2 Retrieval of the event feature

$$E(i) = \begin{cases} G(i), if G(i) \geq min(M) \\ min(M), otherwise \end{cases}, 1 \leq i \leq length(G)$$

According to this study Liu & Wei, (2019), the use of graph signal is an improved feature extraction method for sound classification. Th results show that the method has good accuracy and tackles the challenge of harsh or non-conducive environmental conditions. The method is easy to implement with the use of the equations after time frame matrix. Even with this method being accurate, its performance in elephant acoustic classification would have to be tested due to the low frequency of the elephant vocalizations. With large datasets, how efficient would this method be compared to CNN as there are several steps to follow to acquire the output one of which includes the use of the equation for feature extraction.

A comparison was made for different models used in sound classification. This study was specific to buffalo vocalizations where the sounds were recorded and used for the study.

In this study. Mel-frequency Cepstral Coefficients (MFCC) technique is implemented for feature extraction which captures unique frequencies and temporal dynamics of the vocalizations. Several models were employed: Support Vector Machines (SVM), Naïve Bayes, Random Forest, K-Nearest Neighbours (KNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Convolutional Recurrent Neural Networks (CRNN) and ResNet34. The models were evaluated using appropriate metrics such as accuracy, precision and recall (Sharma et al., 2023). The ability of CNN and RNN to process the raw MFCC features allows automatically or directly for more signal features such as patterns to be represented. In this study MFCC was used for feature extraction like in the developed system where the acoustics were represented as MFCCs to allow original features of the sound data to be maintained for use in pattern and sound identification. CNN and RNN had the highest accuracy.

However, the tests were carried out on a very small dataset as compared to what the models had to work with in the system. The audio recordings were also collected in a controlled environment where the cattle were separated to record each individually which is very different compared to collecting acoustic data in the forest where there many different animals, environmental noises such as wind, trees and even human beings or cars which create noise.

Table 2.1 Summary of advantages and disadvantages of sound representation techniques

Representation	**Description**	**Advantage**	**Disadvantage**
Mel-Spectrogram	Commonly used in acoustic processing	Good in the reduction of background noise	Higher frequency bands excessively lose their frequency
Multi-view spectrogram	A repetition of spectrogram	Improves precision due to the repetition especially in differentiating vocalizations and noises	With the repetition, there is vocalization overlaps hence reducing the accuracy of identification

MFCC	Resembles human hearing and is very popular in sound classification	Simple, effective and adaptable to a wide range of sounds and captures major features	lacks the ability to properly capture high frequency or pitch and further requires more processes to convert audio signals into sound representations

2.5 Architectures

System architecture is an important part in the design and development of any system. The system architecture can be defined as a blueprint or map of a system which shows the structure, the different components, interactions between the components and the behaviour (Systems Architecture, 2024). There have been several architectures used to show some of the systems that have been implemented in elephant detection employing different technologies. In the figure below, the architecture for an IoT device that uses machine learning and bioacoustics on elephant vocalizations and operates even in low visibility to warn the humans will greatly reduce HEC (Ramasubramanian et al., 2022). The device captures the elephant sounds and analyses the vocalization using AI to accurately inform the people of the oncoming danger. The device was trained on elephant sounds and can be installed in fringe areas of villages near wildlife sanctuaries. Three Convolutional Neural Networks were built and trained on modified Mel filter bank features to hierarchically classify elephant vocalizations and used a 2 level CNN model which was able to identified and classified elephant vocalizations most accurately (Ramasubramanian et al., 2022).

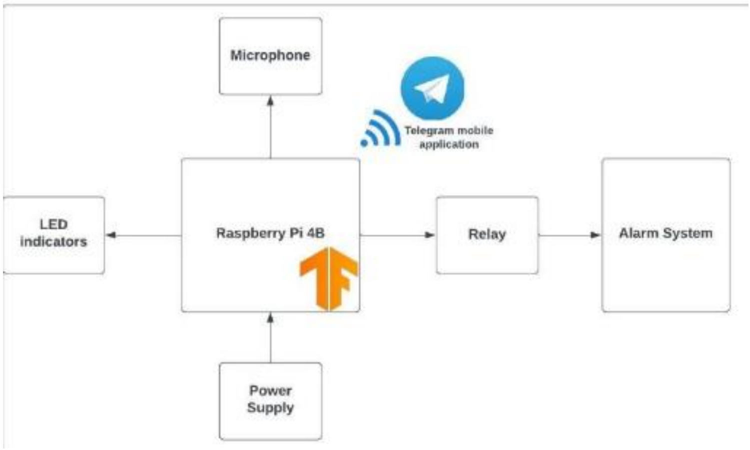


Figure 2.8 Early Elephant Detection System (Ramasubramanian et al., 2022)

From a different study, a camera was used instead of vocalizations for the elephant detection together with a deep learning algorithm. A programme script is written on the Raspberry Pi 4 to ensure seamless connectivity. The faster RCNN, embedded in the Raspberry Pi, is programmed for elephant detection using an elephant image dataset from the internet (Ravikrishna et al., 2023).

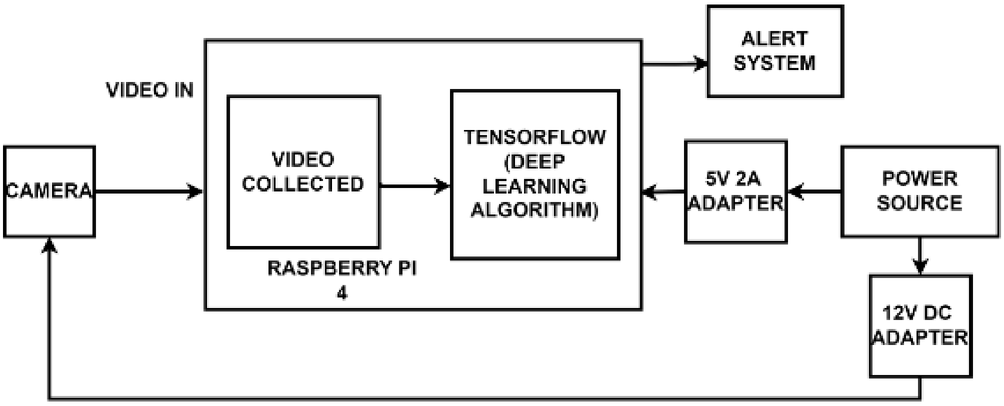


Figure 2.9 An Elephant Detection System Using TensorFlow (Ravikrishna et al., 2023)

The use of TensorFlow is what differentiates this architecture and the use of imaging for detection even with low resolution images.

2.6 Research Gap

From the literature review, most of the studies on the use of acoustic localization has been carried out in Central Africa and Asia focusing on forest environments and not in Kenya. In Kenya, collaring is mainly used and camera traps for monitoring and detection. Kenya has relied on GPS collaring and camera traps for elephant monitoring and detection where collaring provides tracking capabilities and precise location data but is a very invasive technique for the elephants and expensive due to maintenance factors. Collaring also only provides data on the collared animal limiting the capabilities of the system. On the other hand, the camera traps only detect elephants when they pass directly in front of the camera, meaning that it has a very limited range and response time for an early detection system.

The literature shows a research gap in the use of acoustic localization in Kenya. Given the challenges highlighted in table 2.1, a gap is seen robust audio feature extraction techniques that can try leverage on the strengths of the different representations while minimizing their weaknesses. From the literature, there have been success in the use of acoustic monitoring of elephants through sound classification in other regions using the low frequency rumbles that can travel over long distances and sound localization presents a non-invasive and cost-effective alternative for early detection.

2.7 Conceptual Framework

The conceptual framework was to help in modelling or structuring the research by showing relationships or the interconnectedness between concepts, variables and theories which can be used as a road map to assist in development (Singh, 2023). The conceptual framework created for the study shows the flow of processes from the elephant vocalization to the detection system. The independent variables are such as the elephant vocalization and the dependent variable is the localization of the sound, how close is the elephant or how far is it. In this case, if it is in close proximity to the community land, they are detected, and a deterrent can be used such as bee sounds which scares off the elephants.

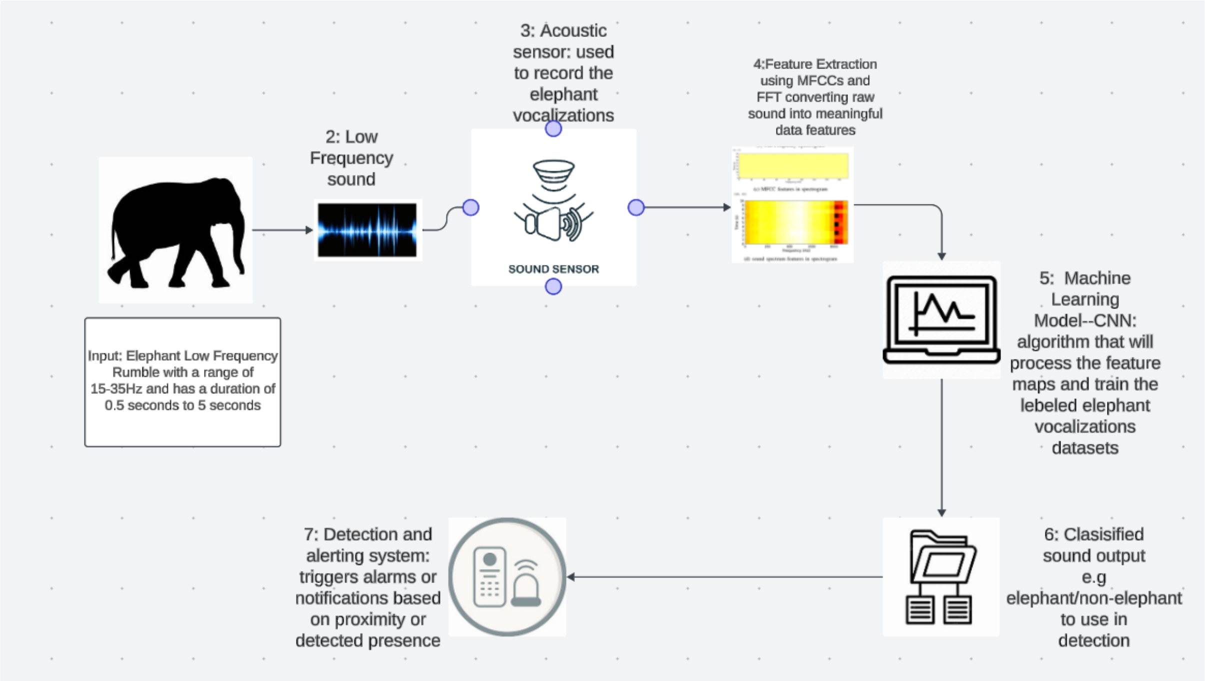


Figure 2.10 Conceptual Framework for Early Detection of Elephant Activity

Chapter 3: Research Methodology/Design

3.1 Introduction

This chapter outlines the approach taken to realize the research objectives. It provides a structured explanation of the techniques, processes and tools used to develop and implement the elephant vocalization detection and localization system. By outlining the reasoning behnd the selected methods, this chapter establishes the foundation for how the research was conducted and how the system was developed.

3.2 Research Design

For this study, a descriptive and computational research design which focuses on analysing existing elephant vocalizations data to train and test an acoustic localization detection system. by using secondary data of the vocalizations which are rumbles, the study simulates real-world conditions for detecting elephant presence through acoustic signals. The computational approach involves designing, validating machine learning models and testing the detection system.

3.3 Study Population

The study population consists of pre-recorded sounds of elephants and other environmental sounds from a different dataset. The recordings are sourced from previous studies such as Elephant Voices which has an Elephant Ethogram which is a good source of data and provide the necessary data for the system's development and FSC22 Forest sounds (Elephant Listening Project Home, 2024, ; *FSC22 Dataset*, 2024). The acoustic data included elephant vocalizations and forest sounds which are important when training the model.

3.4 Sampling Method

A purposive sampling method was used to select the relevant audio recordings from the datasets. This is a non-probability method for getting a sample where particular characteristics are preferred (Frost, 2022). Compared to probability sampling such as stratified sampling, one does not randomly select units from a population but would like the one that most fits the goal of the study and when using mixed methods research design, this is not considered a weakness. The recordings were chosen based on their clarity of the elephant vocalizations and the diversity of the sounds, specifically the forest sounds to ensure efficiency and accuracy and correct detection even in harsh conditions where there is background noise.

3.5 Sample Size – Scientific Approach

The sample size is currently not definite but was determined by the amount of acoustic data that is required for effective machine learning model training. Using Cochran’s formula, which typically suggests a sample size of 384 for large populations at 95% confidence, the use of ~300 high quality, purposefully selected samples was deemed sufficient for this proof-of-concept study, balancing data availability with practical constraints. In theory, approximately 200-300 high quality samples were enough for training and testing the system. The formula is given below:

Equation 3.1: Cochran's Formula

$$n_0 = \frac{Z^2 \cdot p \cdot (1 - p)}{e^2}$$

Where:

- n0 is the required sample size,
- Z is the Z-score corresponding to the chosen confidence level (1.96 for 95%),
- p is the estimated population proportion (0.5 for maximum variability),
- e is the margin of error (0.05 for ±5%).

The calculations are as shown below which has been adjusted to be used when the population size is known:

Equation 3.2: Cochran's Formula for Finite Population

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

In this case, N is the total population.

Z=1.96(for 95% confidence)

P=0.5(maximum variability)

e=0.05(5% margin of error)

Using this formula, the result was ~246 audio samples for the study to be statistically representative of the 686 total samples of both elephant and forest sounds, using 95% confidence and a 5% margin of error.

3.6 Data Collection Methods

The data used for this research was acquired from secondary sources which are already existing datasets from other studies which include the elephant ethogram from elephant voices and the FSC22 dataset for forest sounds. The data sources consisted of elephant vocalization datasets and the forest sounds dataset. Once collected, the audio data was be cleaned and pre-processes using audio filtering techniques as discussed in chapter 2 to isolate the relevant elephant sounds. The Elephant Ethogram is a detailed library which documents the behaviour and communication of African elephants and includes photographs, audio and video recordings (Elephant Listening Project Home, 2024).

The audio data went through pre-processing to get the right format and to filter to remove some background noises. Using MFCCs as discussed in section 2.6.4, the sound features extracted were represented after the acoustic features such as the amplitude, frequency and duration have been extracted from the cleaned audio data. A machine learning model was then be trained using the processed and labelled data which is a supervised learning technique. The algorithm that is suitable for models that utilize audio data is Convolutional Neural Networks (CNN) and 80% of the data was used for training and 20% for validation which was be compared to other machine learning models. This study relied on secondary data from reputable sources. While these datasets provided a rich basis for training and evaluating the detection model, their use introduced certain limitations such as no control over environmental conditions during data collection, including microphone placement, background noise levels or even terrain characteristics.

To validate the model, elephant vocalizations were used to evaluate performance, and the accuracy and precision were calculated to assess the model performance and ability to correctly detect and classify the sounds. Cross-validation was used to avoid overfitting.

3.7 System Development

The system development approach used in this study was Rapid Application Development (RAD). RAD and Agile are almost similar but with RAD, the quality is of key importance unlike agile where timelines are first considered. The method is also faster and cheaper but not without disadvantages especially when working with a larger team (RAD Methodology | Rapid Application Development Phases - Kissflow, 2024).

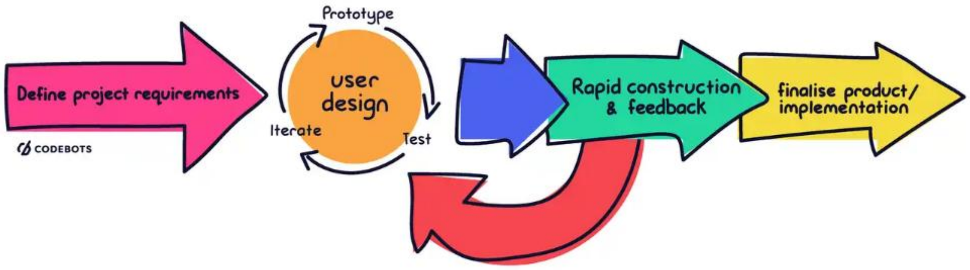


Figure 3.1 Rapid Application Development (RAD) (Chien, 2020)

The methodology selected allows for iteration where corrections can be done, and aspects amended. The development followed the following structure:

i. Data Pre-processing: The collected audio files were be cleaned, normalized and segmented to ensure that the sounds are correctly labelled.
ii. Model training: A machine learning algorithm, CNN was trained to recognize the elephant vocalizations and classify them and localize them to detect the elephants.
iii. Prototyping: In this step, a prototype of the detection system was developed using appropriate tools to ensure proper testing.
iv. Testing: The system was tested which was to also ensure that the trained model works efficiently and accurately which was done using a separate subset of the data.

3.8 Dissemination of Results

Acoustic localization of elephants in Nyangores forest is aimed at reducing negative human-elephant interactions and to encourage co-existence between the elephants and the forest adjacent communities. The outcome of this project is an early detection system that can be used by the forest rangers and the community and can also be used to collect elephant data to analyse, predict and study patterns which can further the study of elephants in the forest area. Open-access dissemination channels were used where possible to ensure that those who can benefit from the research such as forest rangers and the local communities can access and implement findings.

The findings from this research project will be shared widely to maximize its impact and promote further development in conservation technology. The dissemination strategy includes the following:

3.8.1 Online Repository

The outcomes, including datasets, algorithms and documentation will be made available in open-access repository hosted by the university or platforms such as GitHub. This ensures that other researchers and developers can access and build upon the work.

3.8.2 Scientific Publications

The results will be published in peer-reviewed journals focusing on wildlife conservation and conservation technology. This will assist in dissemination of useable results from trusted sources.

3.8.3 Collaborations

The research findings will be shared with the local authorities in Nyangores Forest who are the Kenya Forest Service which can give them actionable insights on how to mitigate human-elephant conflict or even assist in furthering the study after proof of concept which can be used to further conservation efforts in the area and assist the forest adjacent communities.

3.9 Ethical Considerations

Ensuring the ethical implementation of this research project is crucial to protecting both the environment and the stakeholders involved. The following are the ethical considerations to be integrated into the study:

i. The research process, objectives and findings will be documented clearly and shared openly with relevant stakeholders to foster trust and collaboration.
ii. The secondary data to be used will be stored securely to prevent unauthorized access.

With the current practices of the local forest adjacent communities known from the Kenya Forest Service reports on Nyangores Forest, this research considers the customs and beliefs of the communities. The solution is a non-invasive technique for monitoring of the elephants but at this stage of development, no devices or sensors will be deployed into the field as this is only a proof-of-concept study.

Chapter 4: System Analysis and Design

4.1 Introduction

This chapter discusses the system analysis and design that have been implemented in the early elephant detection system, detailing its functionality, architecture and various system components. It explores the system requirements both functional and non-functional requirements. Additionally, this chapter includes system diagrams to illustrate overall design, the interaction between system component and the data flow.

4.2 System Requirement Analysis

This was a crucial phase in the development and design of the detection system as the needs and constraints for both hardware and software. This is a critical phase in the development lifecycle where the requirements of the system were gathered and analysed (*Differences between System Analysis and System Design*, 2024).

4.2.1 Functional Requirements

These requirements describe the core features and operations of the system that make up the detection system.

Table 4.1 Functional requirements

No.	
1	The system should allow the IoT device to capture audio
2	The machine learning model should process audio recordings and classify them.
3	The system should trigger alerts when an elephant vocalization is detected.
5	The users should be able to view real-time data on dashboard
6	The system should generate reports summarizing detections and alerts

4.2.2 Non-Functional requirements

The quality attributes of the system were defined which are critical for efficient deployment of the system

Table 4.2 Non-Functional Requirements

No.	
1	The system should process real-time data with low latency
2	The system should support multiple microphones capturing sound data
3	The system's dashboard user interface should be easy to use.
5	The system should be reliable in terms of uptime for real-time monitoring

4.3 System Design Diagrams

In this section, the detailed specifications of the elephant early detection system were developed by designing the architecture, components, interfaces and data structures that utilize the requirements that were previously defined in the system requirements analysis phase ((*Differences between System Analysis and System Design*, 2024). The design diagrams were used to better understand

4.3.1 Use Case Diagram

The use case diagram shows how the different actors or users interact with the system which shows the different cations that can be performed by the system. The elephant detection system has two main actors the ranger and the IoT device. The IoT device which includes a microphone, captures the sound data which is then goes through the machine learning model after being processed into a useable format such as spectrograms and finally classified as either elephant or non-elephant. The output is then given from the classifier where if the result is elephant, an SMS alert is sent to the ranger and an alert is also sent to the user interface which can then be logged into the system for report generation. The ranger can view real-time system alerts and view the reports generated.

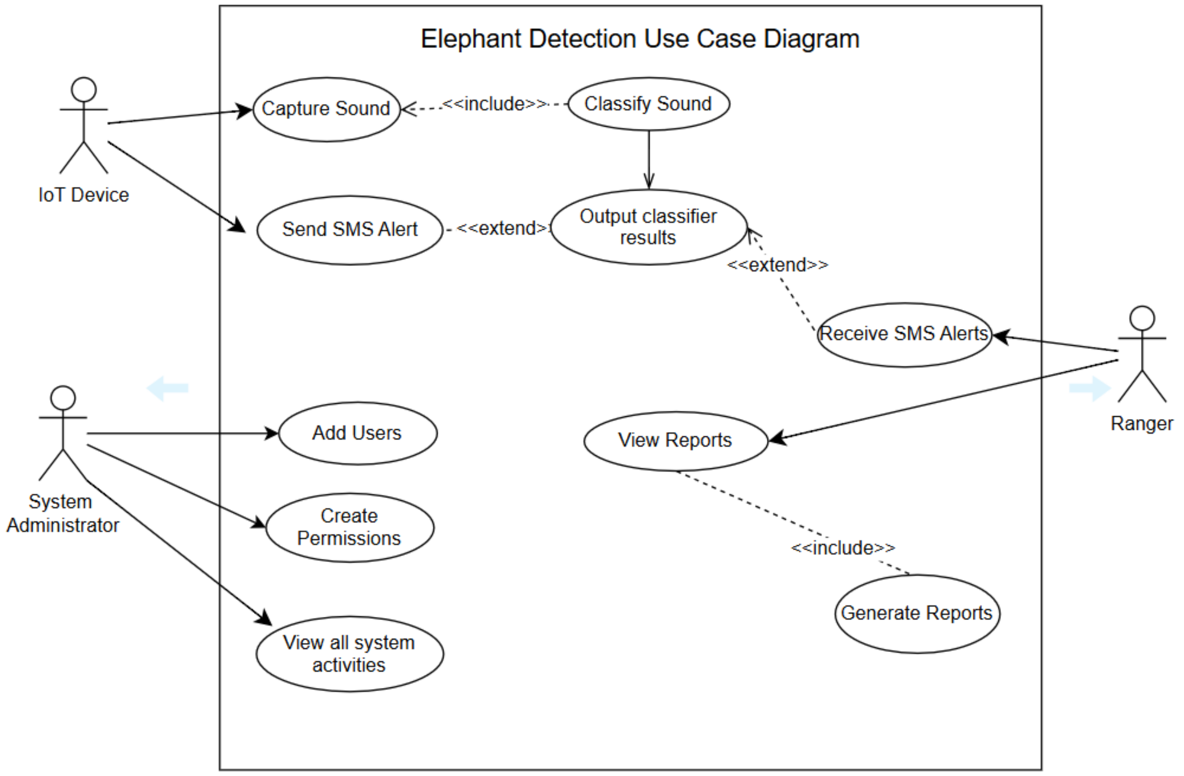


Figure 4.1 Use Case Diagram for Early Detection System

4.3.2 Sequence Diagram

The sequence diagram shows the different processes within the elephant detection system and how the different components interact to ensure functionality of the system. From the figure below, the IoT device which includes a microphone captures sound which is the main input for the machine learning model which processes the sound input and gives the output which is a classification of either elephant or non-elephant. Alerts are then sent from the IoT device to the ranger via text message. The whole process is a loop as the sound capturing is continuous as long as the microphone is recording sound.

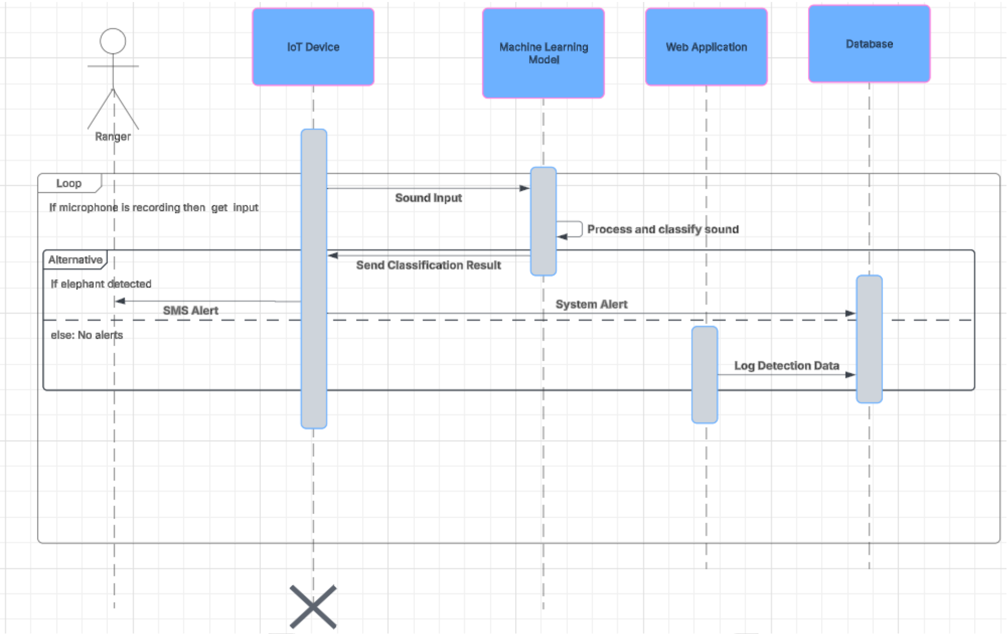


Figure 4.2 Sequence Diagram

4.3.3 Class Diagram

The figure 4.3 below shows a class diagram that was used to visually represent the structure of the elephant early detection system. The IoT device class is responsible for capturing vocalizations or sounds and sending them to the machine learning model that processes and classifies the sounds. The IoT device is linked to vocalizations showing each device records multiple vocalizations and the machine learning model has a one-to-many relationship where one model processes many captured sounds. The results from the machine learning model are in the class detection result that stores the output including the result either elephant or non-elephant and the timestamp to show date and time for future analysis. The machine learning model is also linked to detection result class and detection result is linked to vocalizations for traceability that a result is linked to a vocalization.

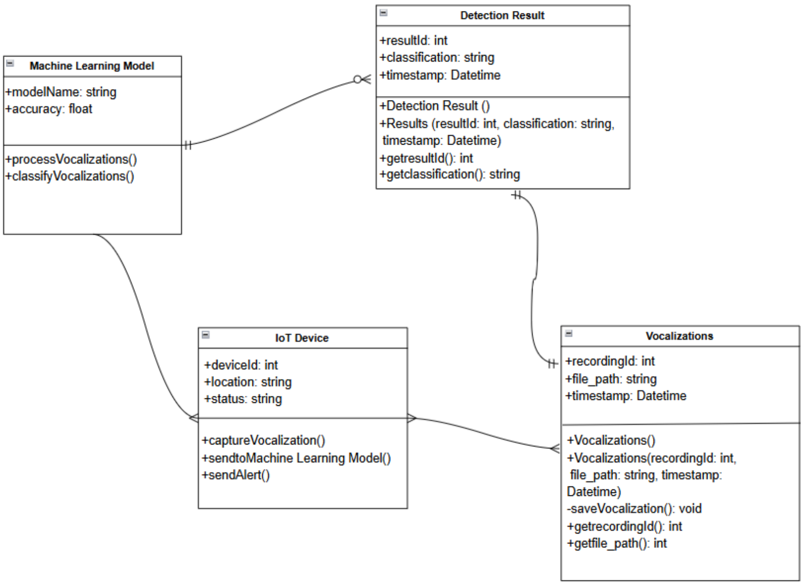


Figure 4.3 Class Diagram

4.3.4 Circuit Diagram

A circuit diagram was created to show the different component connections on the breadboard. The Raspberry Pi Pico based system for collecting and transmitting data using sensors and some communication modules. The core controller is the Raspberry Pi Pico which is the green board in the centre and acts as the brain of the system reading inputs from the microphone to sending the data to different modules. The batteries are for the power supply and are rechargeable batteries that can be re-charged using a solar panel. A SIM module which was used to send alerts to the stakeholders. The GPS Module determines the system's exact location using satellites which is useful for tracking. The sound sensor which is a microphone detects the vocalizations which are then processed by the microcontroller and the output displayed on the screen which can then trigger the alerts. The figure below shows a circuit diagram for the elephant detection system.

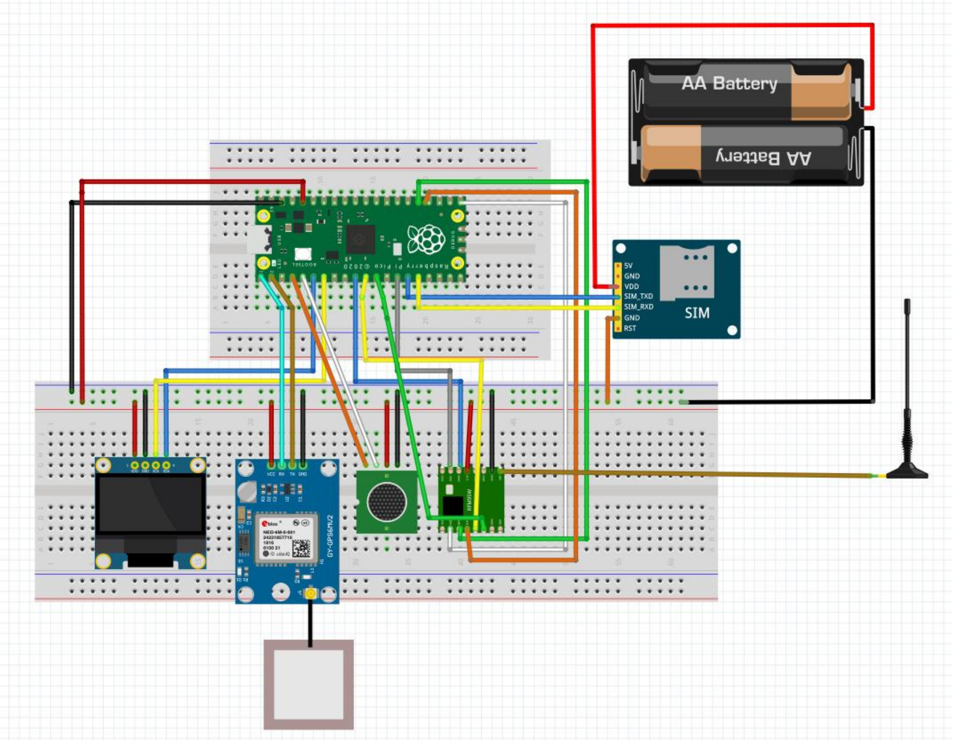


Figure 4.4 IoT Device Circuit Diagram

Chapter 5: System implementation and Testing

5.1 Introduction

This chapter greatly focuses on the implementation and testing phase of the development lifecycle of the early detection system. System and implementation and testing describe how the system was developed, including the integration of the hardware and software components. The chapter also details the testing process used to ensure the system functions correctly and meets the defined functional and non-functional requirements. Different metrics are used in this chapter to evaluate the accuracy, precision, reliability and performance of the system in detecting the elephant sounds and triggering the alerts. The results are analysed to determine the effectiveness of the entire detection system.

5.2 Development Environment

For the implementation of the early detection system, appropriate development environment was selected to ensure efficient and easy implementation of the system.

5.2.1 Hardware Requirements

Hardware	**Description**
HP Envy	Windows 11-64bit, 8GB RAM
Raspberry pi pico micro controller	Dual-core Arm Cortex-M0+ running at up to 133 MHz, 26 multifunctional GPIO pins, including 3 analog inputs

5.2.2 Software Requirements

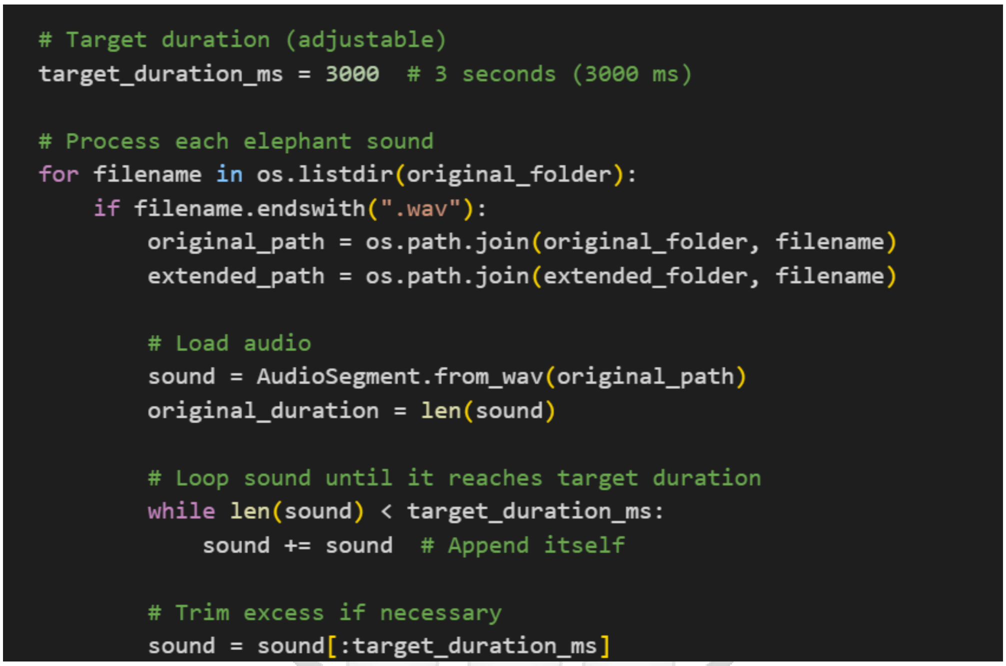
Software	**Description**
Operating System	Windows 11 Home 64-bit (10.0, build 26100)
Arduino IDE	Arduino IDE 2.0 version 2.3.4
Visual Studio Code	Vs Code version 1.98

5.3 Machine Learning Models

Machine learning models are used to automate multiple tasks to make them more accurate and efficient (*Machine Learning Models*, 2024). For the early detection system, a classification model is ideal to be used to indicate whether the sound detected are elephant or non-elephant. Each stage from data preprocessing to deployment plays an important role in optimizing the model for real time detection. In this section, it details on data preprocessing, feature extraction and model training and evaluation. The goal was to develop a robust, efficient and accurate model that can correctly detect and classify elephant sounds. For this prototype, CNN was selected due to its strong performance in processing spectrogram data and its proven success in similar acoustic detection tasks as discussed in chapter 2. CNN offers a balance between accuracy and computational efficiency, making it suitable for deployment on low-power devices such as Raspberry Pi Pico.

5.3.1 Data Preprocessing

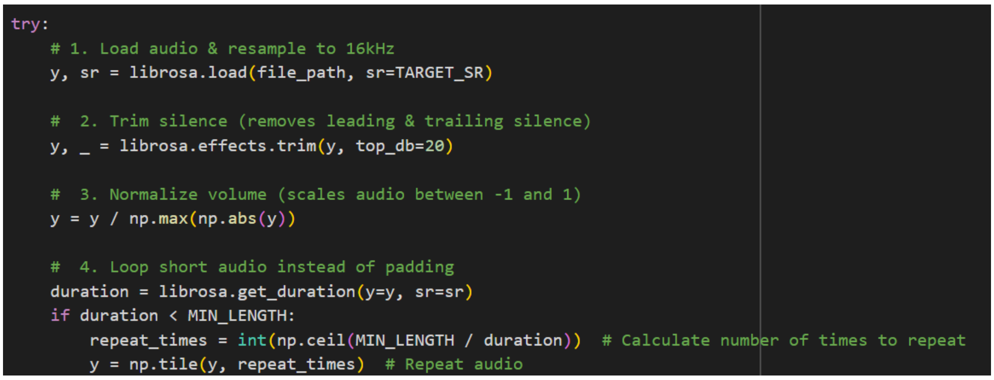
The raw audio data often contains some noise and irrelevant information that can affect the accuracy of the model and usually before training the model, the raw audio data must be cleaned. This involves noise reduction, normalization and conversion into a suitable format. To create the dataset, two different datasets were used: Elephant vocalizations from elephant voices from Soundcloud and FSC22 dataset which contains forest sounds. In the figure 5.1 below, the next step after uploading the elephant vocalizations audio files, due to their short nature, they had to be extended to allow for correct feature extraction. The method used was looping over padding which does not introduce excess noise, avoids unnatural zero-padding which can distort learning and does not remove data that may be of importance during the next steps. The audios were looped to 3 seconds which in this is case allows for feature extraction. Machine learning models require consistent input size to enhance performance or avoid shape mismatch errors.



```
# Target duration (adjustable)
target_duration_ms = 3000  # 3 seconds (3000 ms)

# Process each elephant sound
for filename in os.listdir(original_folder):
    if filename.endswith(".wav"):
        original_path = os.path.join(original_folder, filename)
        extended_path = os.path.join(extended_folder, filename)

        # Load audio
        sound = AudioSegment.from_wav(original_path)
        original_duration = len(sound)

        # Loop sound until it reaches target duration
        while len(sound) < target_duration_ms:
            sound += sound  # Append itself

        # Trim excess if necessary
        sound = sound[:target_duration_ms]
```

Figure 5.1 Extending the Elephant vocalizations

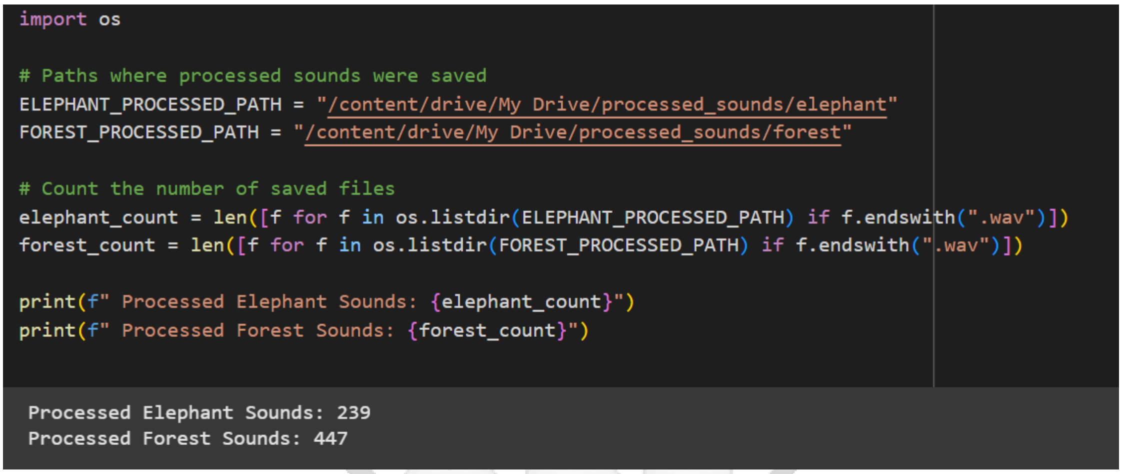
After extending the vocalizations, librosa was used to resample with a frequency 16kHz for the dataset where 16kHz was most preferred as seen in figure 5.2. This was because as discussed in chapter 2 of this document, elephants communicate using low rumbles which are infrasonic and some audible sounds which typically range between 5Hz to 10kHz. A higher sampling rate would have captured more high frequency noise and miss the low-frequency elephant sounds. Silence trimming was done to eliminate irrelevant audio that might confuse the classifier. Normalization was also done to ensure volume variation does not affect learning this is in terms of louder vocalizations or quieter vocalizations.



```
try:
    # 1. Load audio & resample to 16kHz
    y, sr = librosa.load(file_path, sr=TARGET_SR)

    #  2. Trim silence (removes leading & trailing silence)
    y, _ = librosa.effects.trim(y, top_db=20)

    #  3. Normalize volume (scales audio between -1 and 1)
    y = y / np.max(np.abs(y))

    #  4. Loop short audio instead of padding
    duration = librosa.get_duration(y=y, sr=sr)
    if duration < MIN_LENGTH:
        repeat_times = int(np.ceil(MIN_LENGTH / duration))  # Calculate number of times to repeat
        y = np.tile(y, repeat_times)  # Repeat audio
```

Figure 5.2 Using Librosa to resample to 16kHz

The dataset after combining the elephant vocalizations and the forest sounds, was imbalanced which is not ideal when training the model as it can lead to bias. The dataset was seen to have a big imbalance where the elephant vocalizations count was 239 and the sampled forest sounds was 447 audio files as seen in figure 5.3 below. This was a verification step to ensure data preparation is complete and balanced. Balanced datasets are critical for avoiding biased learning where the model might lean toward the more frequent class.



```
import os

# Paths where processed sounds were saved
ELEPHANT_PROCESSED_PATH = "/content/drive/My Drive/processed_sounds/elephant"
FOREST_PROCESSED_PATH = "/content/drive/My Drive/processed_sounds/forest"

# Count the number of saved files
elephant_count = len([f for f in os.listdir(ELEPHANT_PROCESSED_PATH) if f.endswith(".wav")])
forest_count = len([f for f in os.listdir(FOREST_PROCESSED_PATH) if f.endswith(".wav")])

print(f" Processed Elephant Sounds: {elephant_count}")
print(f" Processed Forest Sounds: {forest_count}")
```

```
Processed Elephant Sounds: 239
Processed Forest Sounds: 447
```

Figure 5.3 Dataset Class Distribution Count

From the figure below, the plot represents the imbalanced class distribution seen in figure 5.3 where the forest sounds samples were more than the elephant vocalizations.

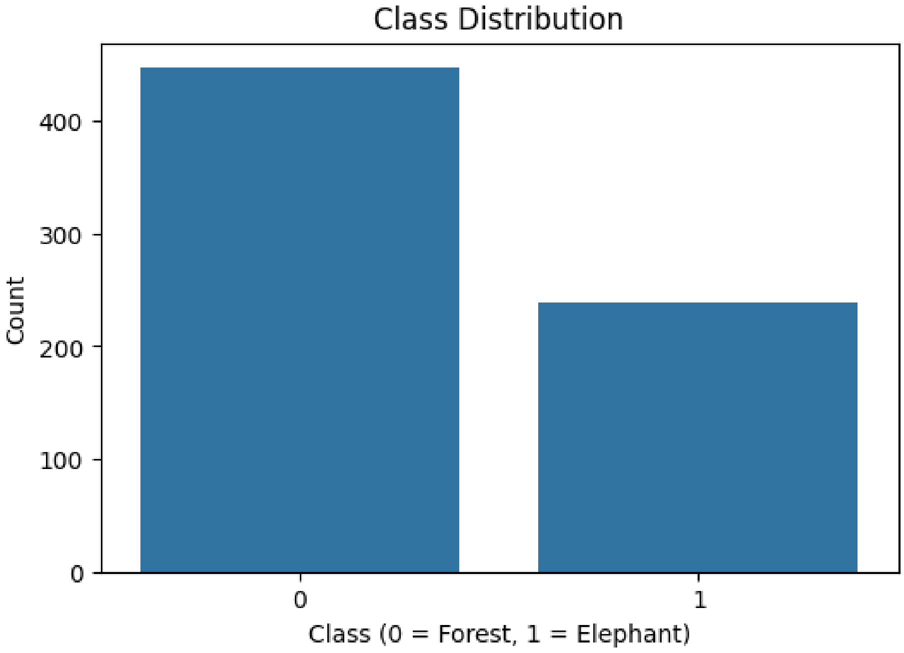
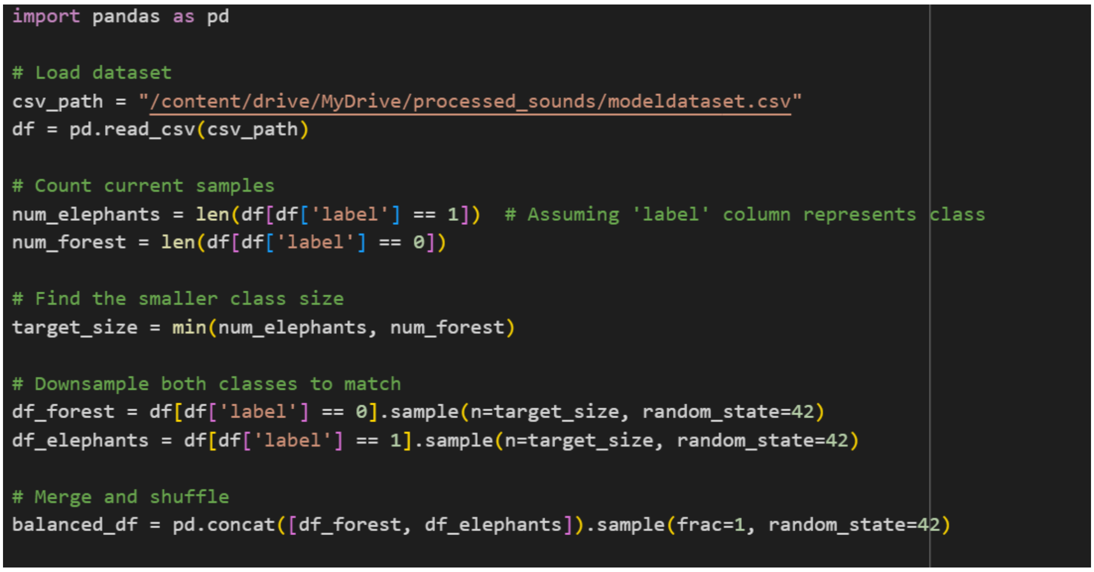


Figure 5.4 Initial Class Distribution

The figure 5.5 shows the process of down sampling the forest sounds from the forest dataset to ensure a balance and avoid any bias while training the model.



```
import pandas as pd

# Load dataset
csv_path = "/content/drive/MyDrive/processed_sounds/modeldataset.csv"
df = pd.read_csv(csv_path)

# Count current samples
num_elephants = len(df[df['label'] == 1])  # Assuming 'label' column represents class
num_forest = len(df[df['label'] == 0])

# Find the smaller class size
target_size = min(num_elephants, num_forest)

# Downsample both classes to match
df_forest = df[df['label'] == 0].sample(n=target_size, random_state=42)
df_elephants = df[df['label'] == 1].sample(n=target_size, random_state=42)

# Merge and shuffle
balanced_df = pd.concat([df_forest, df_elephants]).sample(frac=1, random_state=42)
```

Figure 5.5 Balancing Class Distribution

The class distribution was balanced by down sampling as the optimal number of audio samples as discussed in chapter 3 section 3.4 where 200-300 high quality audio samples are enough to train a machine learning model. The figure 5.6 below shows the balanced class distribution of the two classes.

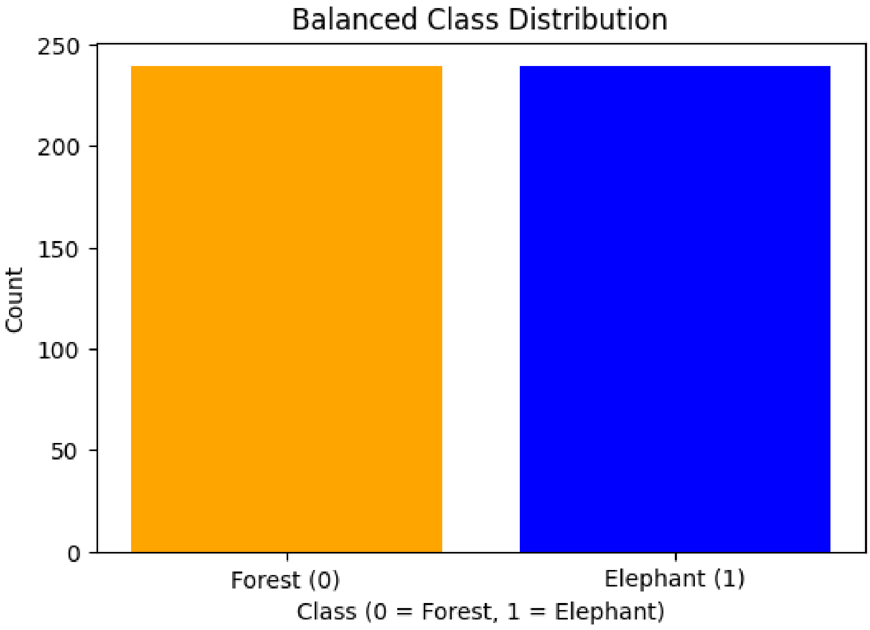


Figure 5.6 Balanced Class Distribution

5.3.2 Exploratory Data Analysis

In this step, exploratory data analysis was carried out to check the differences in the features in the elephant vocalizations and the forest vocalizations. The main output in the exploratory data analysis (EDA) were the waveforms of the elephant vocalizations and forest sounds which clearly shows the difference between the two especially in terms of frequency. From the figure 5.7 below, shows the elephant vocalization waveform which shows the amplitude over time and as seen, there is some noise resent but not significant. The second plot which is a spectrogram which is a different representation of the audio where warmer colours red and orange show higher intensity while the purple and black colours represent lower intensity. The low and mid-range frequencies are dominant which shows the range of elephant vocalizations. The bright yellow and red areas on the plot show higher amplitude which is strong sound signals, and the dark purple and black areas show lower amplitude which is weaker audio signals. The most bottom bit of the plot shows the strongest sound signals showing the low frequency of elephant calls.

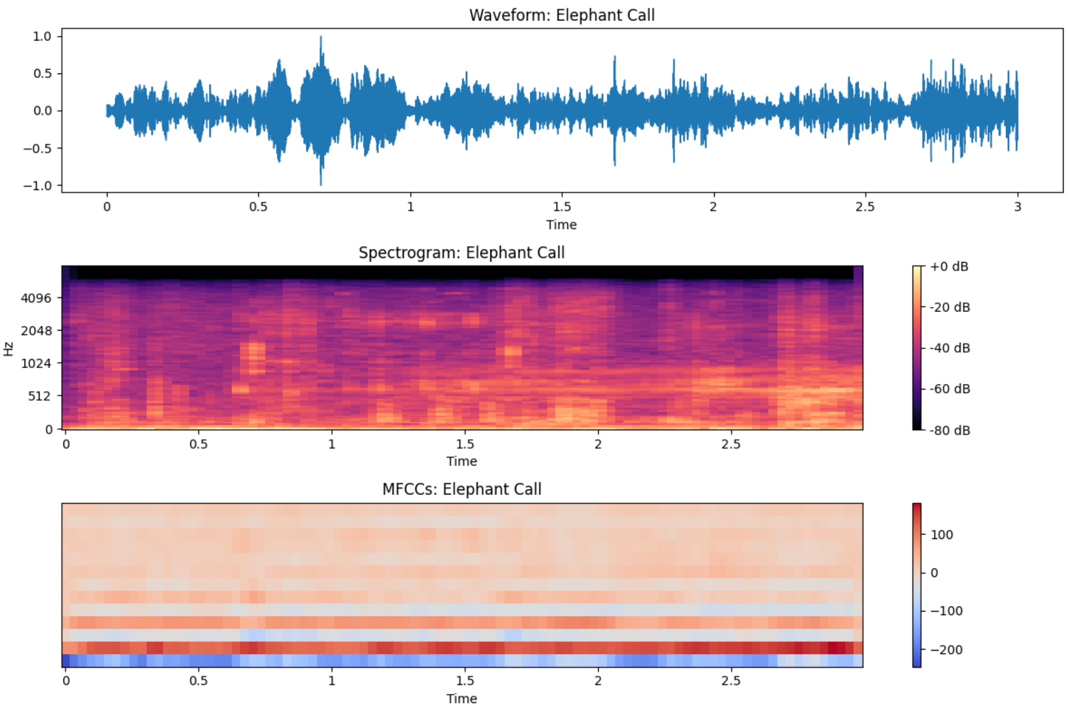


Figure 5.7Elephant Vocalization Representations

The third plot in the figure 5.7 above, represents Mel-Frequency Cepstral Coefficients (MFCCs) which show the spectral characteristics of the sound and is widely used as discussed in chapter two in speech and bioacoustics analysis. The MFCC breaks down the frequency spectrum into coefficients that represent different sound characteristics. The plot is temporal as it shows the progression of the sound over time. The y-axis represents the different coefficients that have been extracted from the sound. The red and orange areas on the plot indicate strong contributions from a specific or certain frequency feature while the blue areas which are very light in colour also indicate very weak intensity meaning weaker or negative contribution.

A different representation carried out was of the forest sounds before and after cleaning by removing the background noise from the forest sounds. The figure 5.8 below shows the waveform representation and the spectrogram representation. The original forest sound waveform has high amplitude bursts, which indicate environmental noise which is also seen in the spectrogram where the brighter spots indicate noise which may be potential unwanted sounds. After data cleaning which was done on the forest sounds dataset, the waveform and the spectrogram representations changed where now the amplitude reduced meaning less background noise while in the spectrogram, some of the white spots reduced meaning that he background noise was reduced.

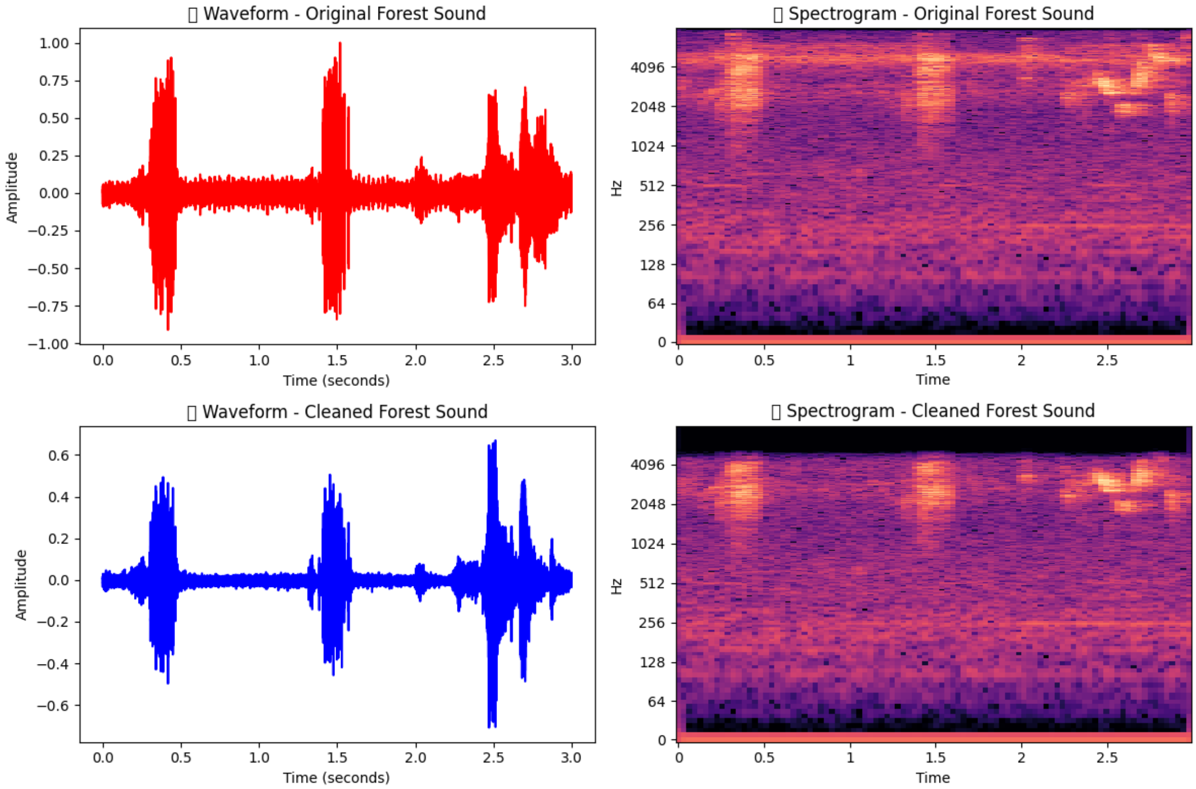


Figure 5.8 Forest Sounds Representations

The waveform visualization in figure 5.8 and figure 5.9 show the difference in the sounds and assist in understanding the audio structure of the audio data, identification of different patterns in vocalizations and in the detection of background noise or even silence in the audio.

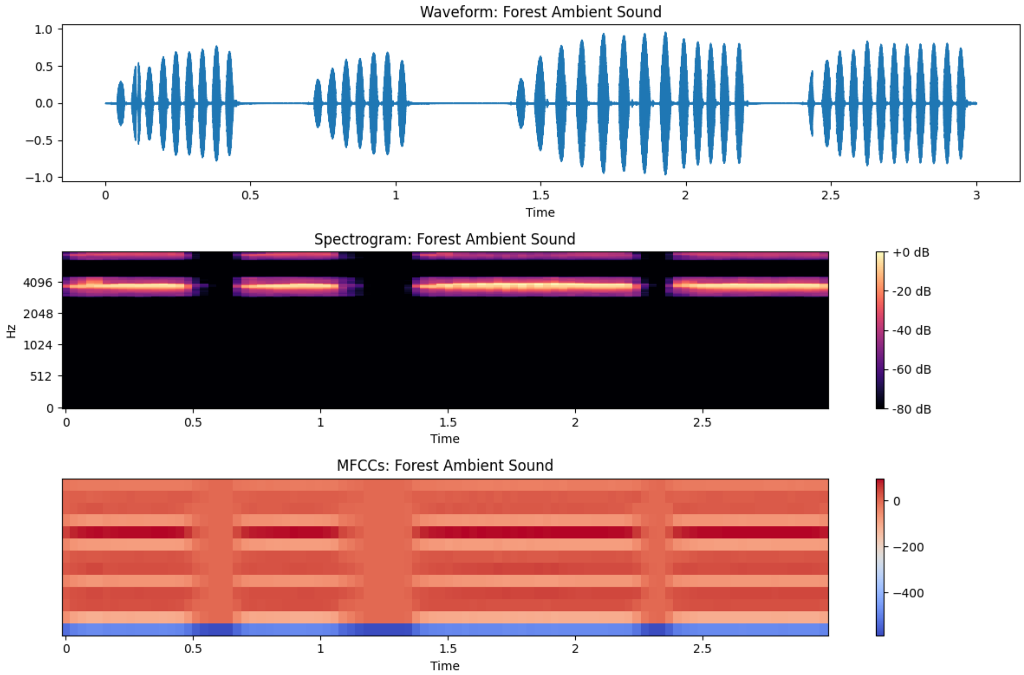


Figure 5.9 Forest Sample Representations

5.3.3 Feature Extraction

In this step, the raw audio is transformed into meaningful data as discussed in chapter two that can be used to train and test a machine learning model. For my study, key features were extracted: Mel Frequency Cepstral Coefficients (MFCCs), Mel Spectrograms, chroma STFT and spectral centroid which capture the unique characteristics of the elephant vocalizations in a forest environment. The features reduce data complexity while preserving essential sound patterns which improves the model's ability to accurately classify the different vocalizations. Different features were extracted for each model iteration to allow for comparisons on these features and how they affect the model performance.

Figure 5.10 represents one of the features, the Mel spectrogram which is a time-frequency representation of an audio signal that generally shows how energy is distributed across different frequencies. For the study, the features are required as they are used in the machine learning model. In the representation below, the x-axis is time ins econds which represents the progression of the sound over time which is 3 seconds long as discussed in the data preprocessing step. The y-axis it shows the frequency but using a mel scale meaning that the lower frequencies have finer resolution mimicking the human hearing. The lower frequencies which are at the bottom represent bass sound while higher frequencies at the top represent treble sounds. In the red areas on the decibels scale, the sounds are louder while the blue represents softer sounds.

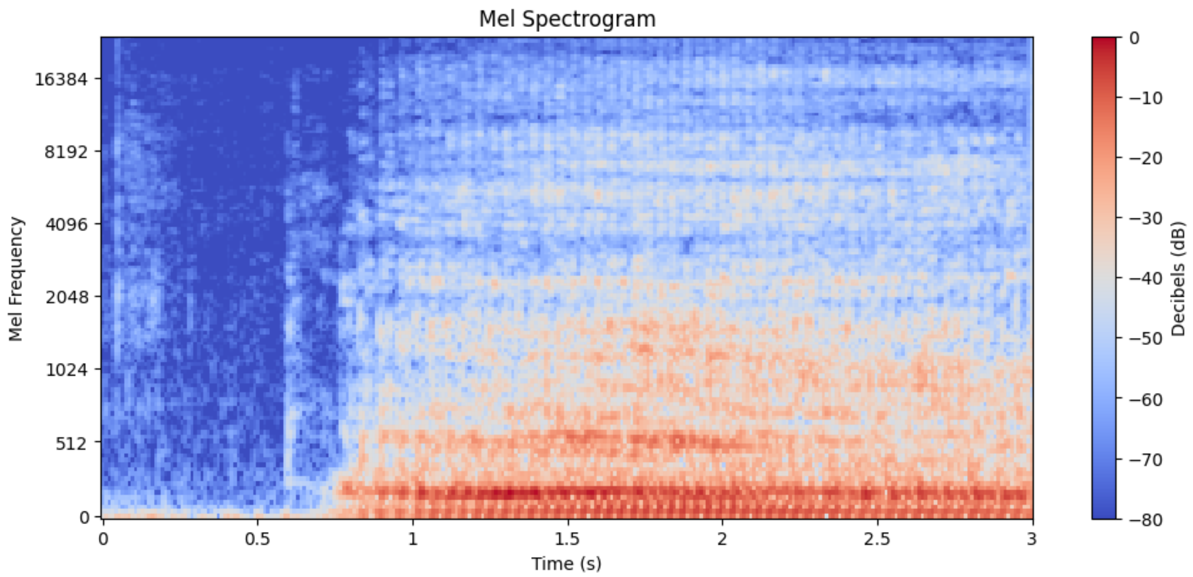


Figure 5.10 Mel Spectrogram

The next feature that was extracted was chroma features which represent the pitch classes present in sound over time. The pitch classes on the y-axis are musical notes and shows which notes are dominant represented by the colour intensity where the red areas show the strong presence of the pitch and blue shows weak presence. This feature is important for sound classification as it detects tonal patterns in sounds which is useful in identifying specific calls. Moreover, chroma features work well with machine learning models as they provide a compact way to analyse pitch related information from the sound.

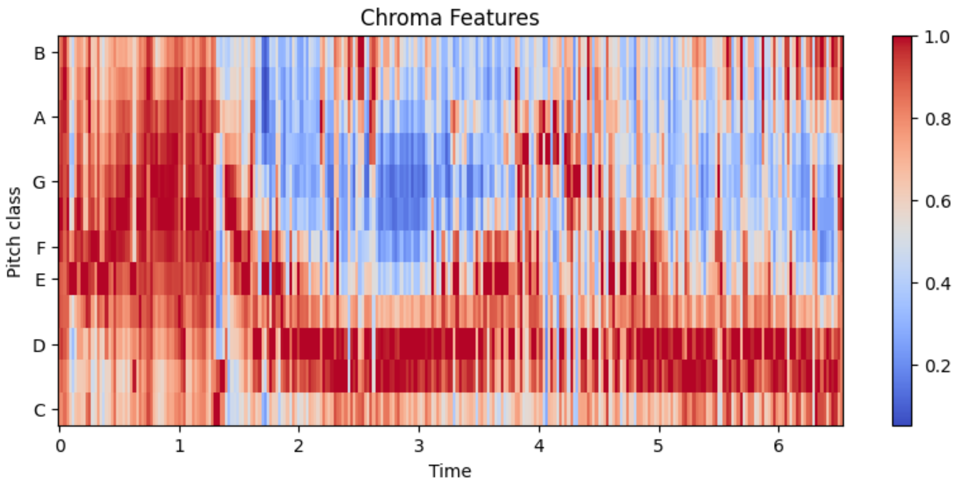


Figure 5.11 Chroma Features

Zero-Crossing Rate (ZCR) was also one of the features extracted and was visualized by the plot shown in figure 5.12. It measures how regularly the audio signal crosses the zero amplitude. This is often does the audio signal change from positive to negative or negative to positive. The histogram below shows that most sounds in the data set have a low ZCR meaning that the sounds are likely to be smooth and continuous due to the low ZCR.

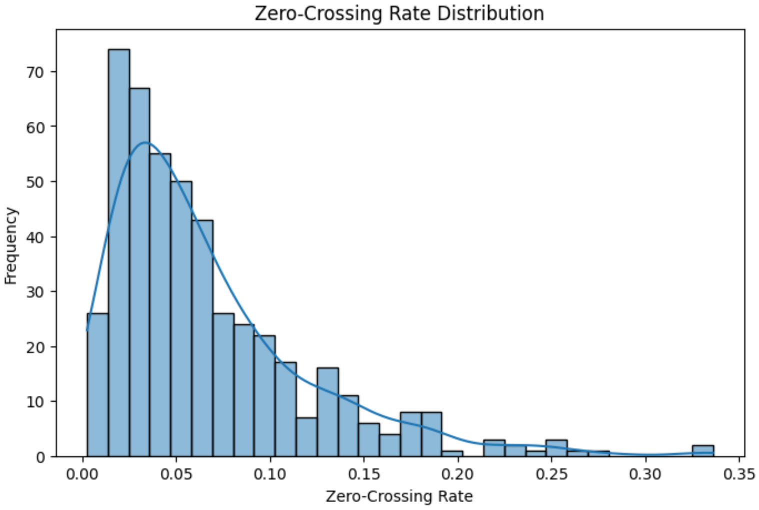


Figure 5.12 Zero-Crossing Rate

Root Mean Square (RMS) energy distribution feature was also extracted which measures the intensity of a sound over time. A high RMS indicates louder sounds while lower values indicate quieter sounds. From the graph below, the histogram shows the distribution of RMS energy values, and the smooth line represents the probability density function which shows the overall trend of how energy is distributed within the sound. The distribution in the curve shows that most of the sounds in the dataset used are low energy with very few instances of loud noises.

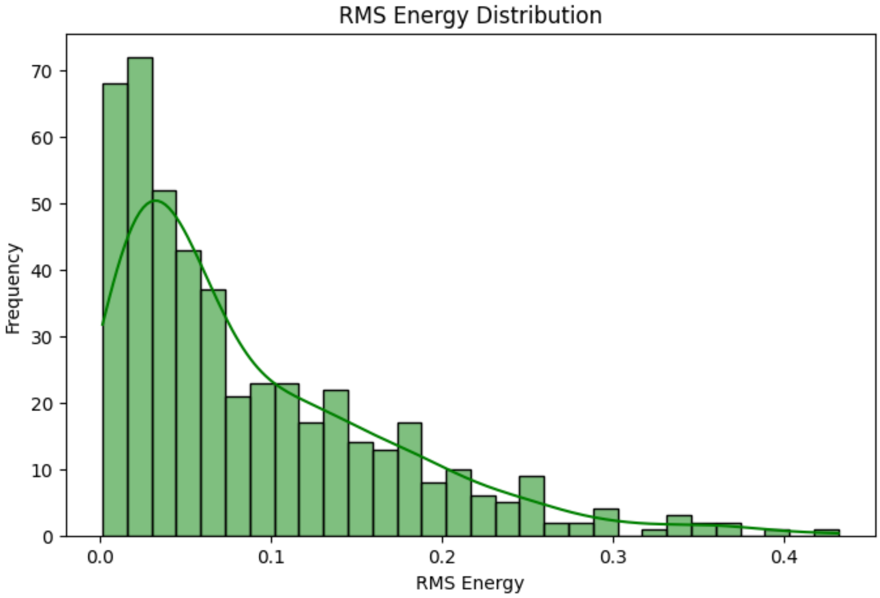


Figure 5.13 Root Mean Square (RMS) Energy Distribution

A box plot was used to show the distribution of extracted audio features after standard scaling which is z-score normalization which transforms the features. From the box plot it is seen that the data is centred around 0 and the spread of the features is standardized preventing features with large magnitudes from dominating the model and the extremes are outliers which are represented by dots. The normalization was done as most of the machine learning algorithms work better when the data is normalized which also allows for faster and more stable training of machine learning models.

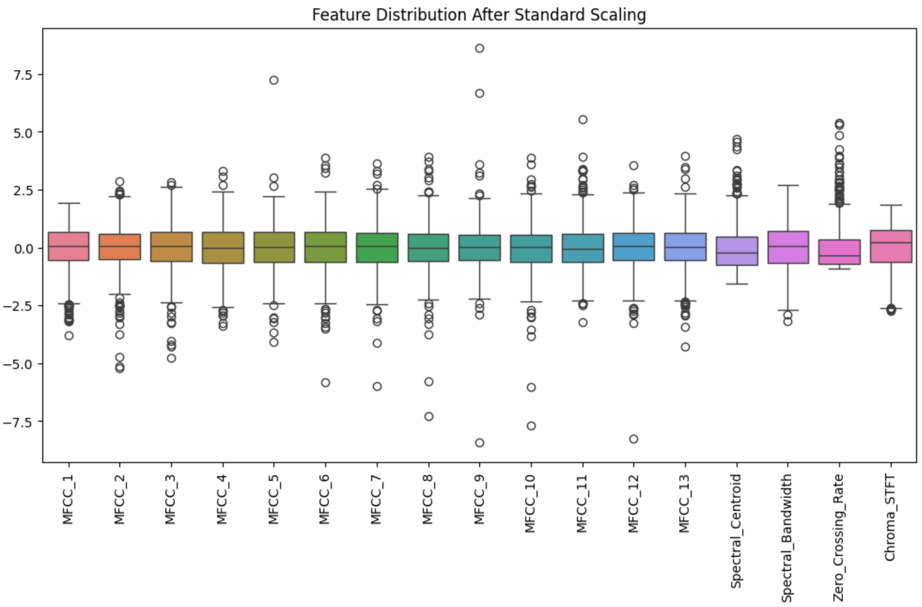


Figure 5.14 Box plot showing Feature Distribution after standard scaling

5.3.4 Feature Selection

After feature extraction, not all features are used as this can easily lead to overfitting of the machine learning model being trained. Different feature selection methodologies were applied to ensure proper selection of features for the model development.

5.3.4.1 Correlation Matrix

A correlation matrix was used for feature selection where the correlation between different features in the dataset. The diagonal dark red shows self-correlation meaning that each feature is perfectly correlated with itself. The colour intensity represents the type of correlation such as the red colour shows a strong positive correlation while bleu which is closer to -10 represents a strong negative correlation and the light blue which is closer to zero shows a weak or even no correlation. The indices 0 to 14 in the heat map correspond to the features extracted before.

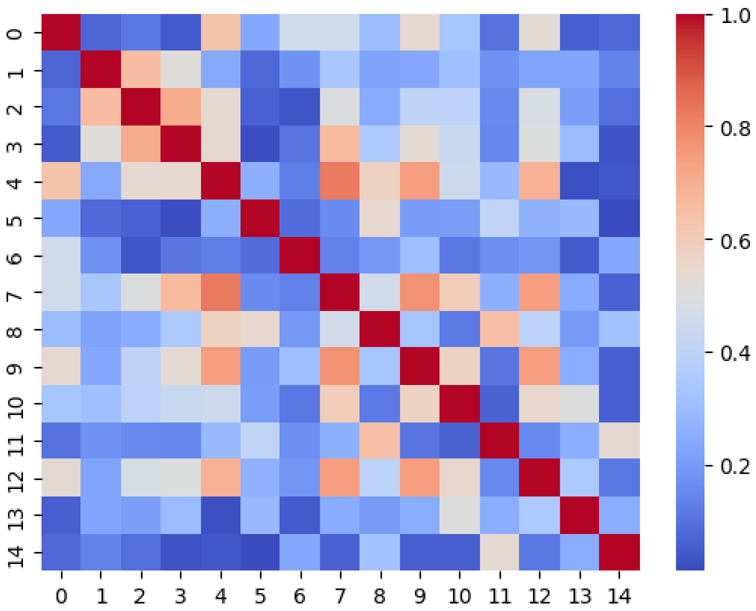


Figure 5.15 Correlation matrix for Feature Selection

5.3.5 Model Training

After the data preprocessing and feature extraction, the machine learning algorithms were implemented in training the model. For this research, I trained several models to compare their accuracy and overall performance.

5.3.5.1 Convolution Neural Network (CNN)

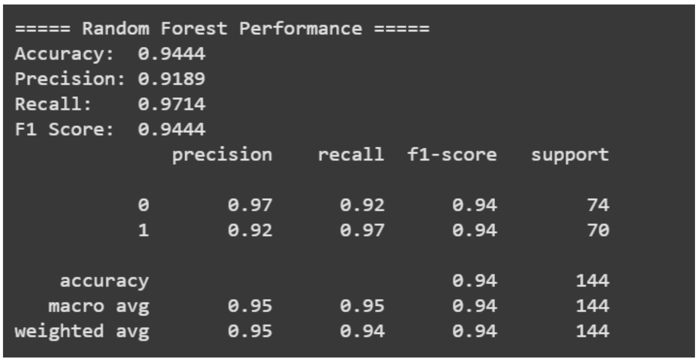
CNN was trained and tested and validated using TensorFlow keras where for the CNN to work, a one-dimension model (1D) was used with the tabular data as seen in figure below. The model was seen to have an overall high accuracy of 98%.

```
3/3 ━━━━━━━━━━━━━━━━━━━━ 0s
Test Accuracy: 0.9896
```

Figure 5.16 CNN(1D) Model

5.3.5.2 Random Forest

The random forest model achieved an accuracy of 94.44% which is lower than the CNN model with a precision of 91.89% and recall of 97.14% which indicated a strong classification ability.

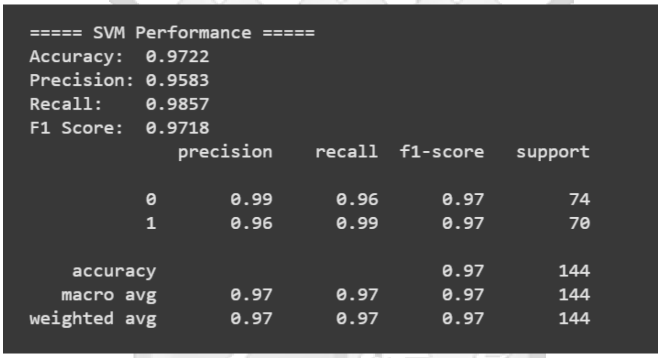


```
===== Random Forest Performance =====
Accuracy:  0.9444
Precision: 0.9189
Recall:    0.9714
F1 Score:  0.9444
              precision    recall  f1-score   support

           0       0.97      0.92      0.94        74
           1       0.92      0.97      0.94        70

    accuracy                           0.94       144
   macro avg       0.95      0.95      0.94       144
weighted avg       0.95      0.94      0.94       144
```

Figure 5.17 Random Forest Model performance

5.3.5.3 Support Vector Machine (SVM)

The SVM model achieved a very high accuracy of 97.22% and had a gap between the precision and the recall as seen in figure 5.19below.

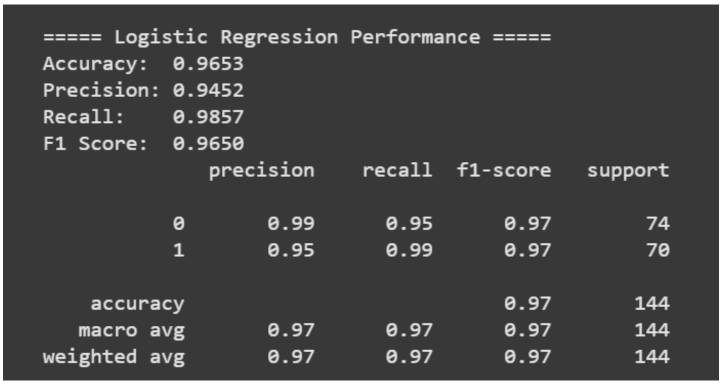


```
===== SVM Performance =====
Accuracy:  0.9722
Precision: 0.9583
Recall:    0.9857
F1 Score:  0.9718
              precision    recall  f1-score   support

           0       0.99      0.96      0.97        74
           1       0.96      0.99      0.97        70

    accuracy                           0.97       144
   macro avg       0.97      0.97      0.97       144
weighted avg       0.97      0.97      0.97       144
```

Figure 5.18 Support Vector Machine Model Performance

5.3.5.4 Logistic Regression

The logistic regression model had a good accuracy of 96% which is also close to the CNN model with a high precision rate and balanced f1 score.

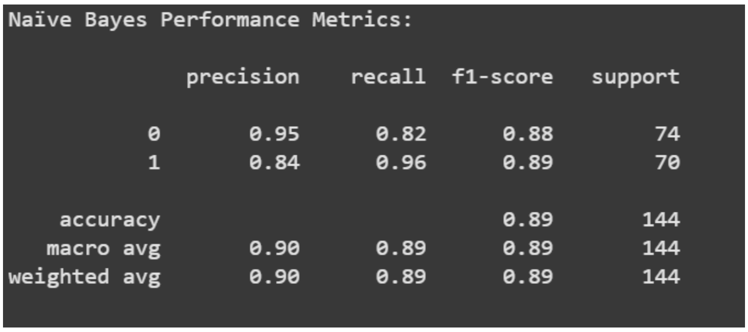


```
===== Logistic Regression Performance =====
Accuracy:  0.9653
Precision: 0.9452
Recall:    0.9857
F1 Score:  0.9650
              precision    recall  f1-score   support

           0       0.99      0.95      0.97        74
           1       0.95      0.99      0.97        70

    accuracy                           0.97       144
   macro avg       0.97      0.97      0.97       144
weighted avg       0.97      0.97      0.97       144
```

Figure 5.19 Logistic Regression Model Performance

5.3.5.5 Naïve Bayes

In the figure 5.21 below, Naïve Bayes model is seen to achieve an accuracy of 89% which is the lowest performance among the models which may be due to its naïve independence assumption.



```
Naïve Bayes Performance Metrics:

              precision    recall  f1-score   support

           0       0.95      0.82      0.88        74
           1       0.84      0.96      0.89        70

    accuracy                           0.89       144
   macro avg       0.90      0.89      0.89       144
weighted avg       0.90      0.89      0.89       144
```

Figure 5.20 Naive Bayes Model Performance

5.3.6 Summary of Model Accuracies

Table 5.1 Accuracy summary

Algorithm	**Accuracy**
Convolution Neural Network (1D CNN)	98%
Support Vector machine (SVM)	97%
Logistic Regression	96%
Random Forest	94%
Naïve Bayes	88%

5.3.7 Testing and Evaluating the model

After training the CNN model, testing and evaluation was done to check the performance of the model using different metrics as the CNN model achieved the overall highest accuracy compared to the other trained models.

5.3.7.1 Comparing Training and Validation Loss

The figure below shows a plot which describes the progression of the model over epochs. In the loss curve, both the training and validation loss decrease significantly in the first few epochs although validation loss flattens after a few epochs and only slightly fluctuates while training loss keeps decreasing steadily. This means that the trained model is learning well since the loss is decreasing but the slight fluctuations seen in validation loss suggest minor over fitting. In the accuracy curve, the training accuracy starts low, approximately 65% but quickly rises to almost 100% while the validation accuracy follows closely and finally plateaus at approximately 98%. This means that the model generalizes well because the gap between the validation accuracy and training accuracy is small which also suggests minimal overfitting.

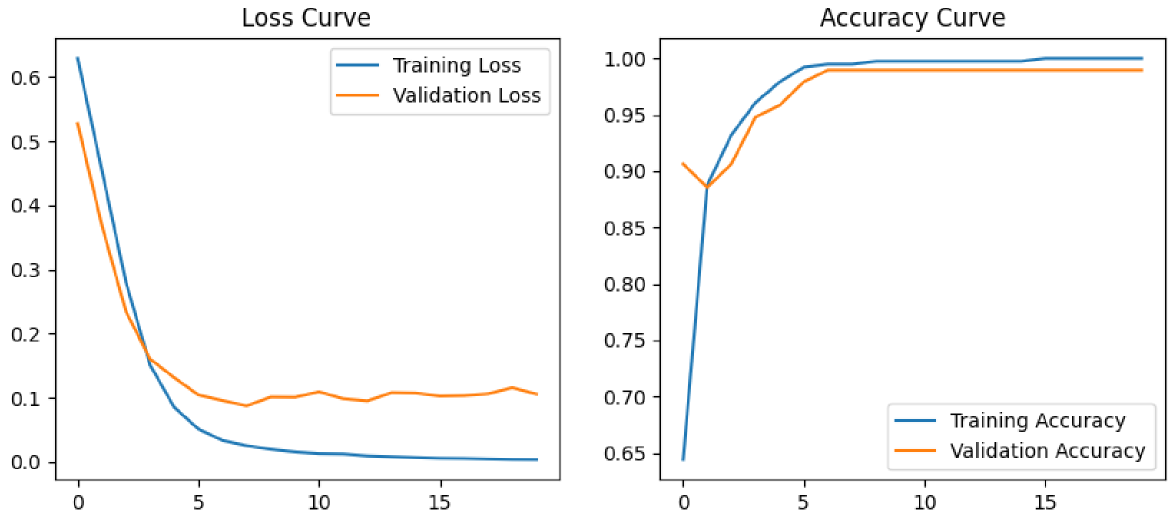


Figure 5.21 Training and Validation Loss

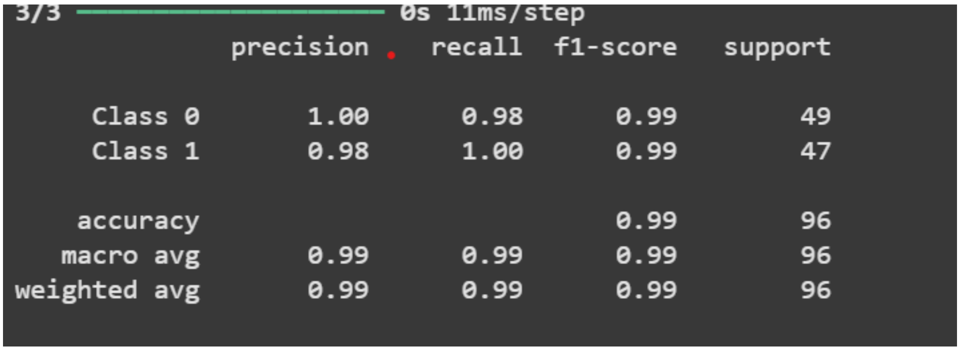
5.3.7.2 Confusion Matrix

A confusion matrix is used to evaluate the model’s performance on a binary classification task that is either 0 or 1. From the confusion matrix, where 0 is non-elephant and 1 is elephant.

Table 5.2 Confusion Matrix Summary

Actual/Predicted	**Predicted Class 0**	**Predicted Class 1**
Actual class 0	48 True Negative	1 False positive
Actual Class 1	0 False Negative	47 True positive

From the matrix, key metrics are derived.



```
3/3 ━━━━━━━━━━━━━━━━━━━━ 0s 11ms/step
              precision    recall  f1-score   support

     Class 0       1.00      0.98      0.99        49
     Class 1       0.98      1.00      0.99        47

    accuracy                           0.99        96
   macro avg       0.99      0.99      0.99        96
weighted avg       0.99      0.99      0.99        96
```

Figure 5.22 Overall Performance Metrics Summary

The figure below is a representation of figure 5.23 as a plot.

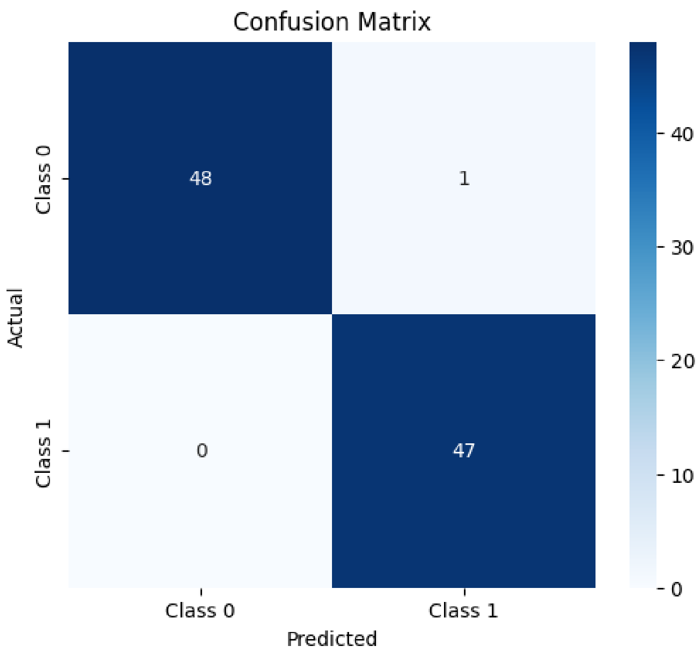


Figure 5.23 Confusion Matrix Plot

From the figure 5.23, each row represents a class which is either elephant or non-elephant which are 1 and 0 respectively. The precision for class 0 was 100% meaning correct classification and the precision for class 1 was 98%. The recall shows the percentage of actual positives that were correctly classified. From the diagram, 98% of actual class 0 instances were detected while 100% of actual class 1 instances were detected. The high precision and recall mean that the model is making very few mistakes and detecting almost all true positives. The F1 score, is the mean of precision and recall balancing both. The accuracy, macro average and weighted average show the overall model performance.

5.4 IoT Device

5.4.1 Components

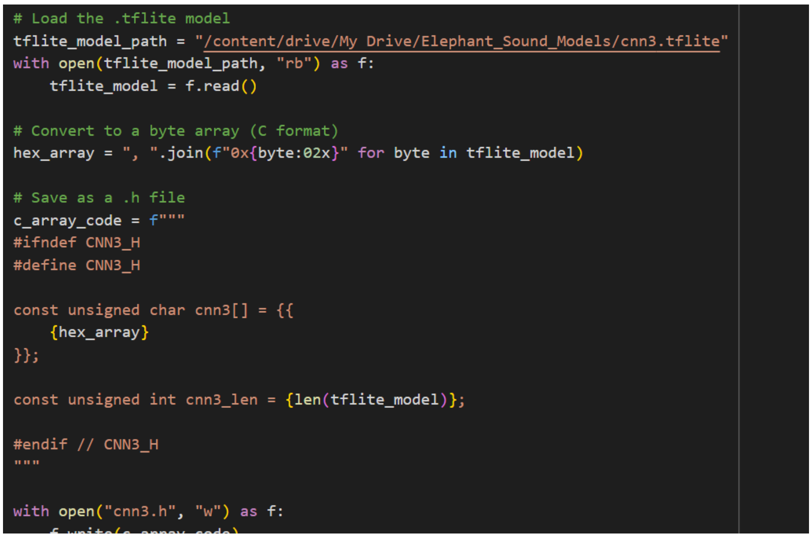
The IoT Device has several interconnected components which for the hardware of the early detection system which includes microphones to detect the elephants and a GPS component for the location coordinates. The table 5.3 below shows the list of components and the specifications for each component used for the IoT device.

Table 5.3 IoT Components

Component	Specification
Raspberry Pi Pico	Dual-core Arm Cortex-M0+ running at up to 133 MHz, 26 multifunctional GPIO pins, including 3 analog inputs
PDM Mic	Voltage Range - 1.8V to 3.3V
Neo-6M GPS Module	Power Supply 3V to 5V, antenna dimensions 23mm x 30 mm
GSM SIM 900	Supported Frequencies 850/900/1800/1900 MHz (Quand Band)
OLED Display	SSD1306 0.96 inch 12C OLED Display
YT 150X86 Solar Panel	Maximum power 150W, Maximum Power Voltage 18.1 V, Maximum Power current 8.29A
3.7V rechargeable battery	Nominal voltage 3.7V, capacity 2600 mAh
RFM-95 LoRa Module-868MHz	Frequency Range – 868MHz , Modulation: LoRa,

5.5 Integration of the Detection System

The machine learning model was integrated onto the Raspberry Pi Pico by converting the tensorflow light weight model into a c header file which makes it lighter for the micro controller. The figure below shows the code snippet of the code used to convert the tflite machine learning model to a .h model which has a c array file.



```
# Load the .tflite model
tflite_model_path = "/content/drive/My Drive/Elephant_Sound_Models/cnn3.tflite"
with open(tflite_model_path, "rb") as f:
    tflite_model = f.read()

# Convert to a byte array (C format)
hex_array = ", ".join(f"0x{byte:02x}" for byte in tflite_model)

# Save as a .h file
c_array_code = f"""
#ifndef CNN3_H
#define CNN3_H

const unsigned char cnn3[] = {{
    {hex_array}
}};

const unsigned int cnn3_len = {len(tflite_model)};

#endif // CNN3_H
"""

with open("cnn3.h", "w") as f:
    f.write(c_array_code)
```

Figure 5.24 Converting tflite model to C header file

The lightweight code was then loaded onto the Raspberry Pi pico on the Arduino IDE and the different components coded to get an output on the OLED screen. The machine learning classification model is loaded onto the microcontroller as seen in the figure below as a package to be used.

```
/* Includes ---------------------------------------------------------------- */
#include <SKahoro-project-1_inferencing.h>
#include <PDM.h>
```

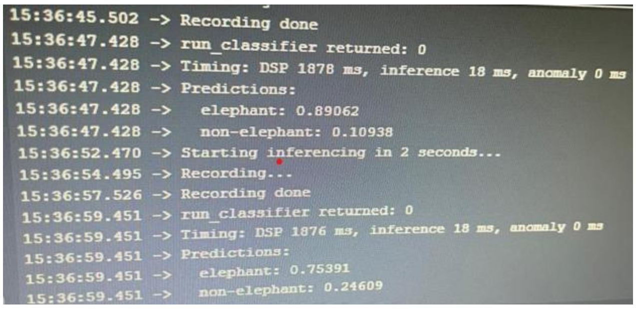
Figure 5.25 Loading the C Header File

After the C header file is loaded, the PDM microphone is then configured by inferencing as seen in the figure 5.26n below.

```
static bool microphone_inference_start(uint32_t n_samples) {
  inference.buffer = (int16_t *)malloc(n_samples * sizeof(int16_t));

  if (inference.buffer == NULL) {
    return false;
  }

  inference.buf_count = 0;
  inference.n_samples = n_samples;
  inference.buf_ready = 0;

  // configure the data receive callback
  PDM.onReceive(pdm_data_ready_inference_callback);

  PDM.setBufferSize(2048);
  delay(250);

  PDM.setDIN(dataPin);
  PDM.setCLK(clkPin);

  // initialize PDM with:
  // - one channel (mono mode)
  if (!PDM.begin(1, EI_CLASSIFIER_FREQUENCY)) {
    ei_printf("ERR: Failed to start PDM!");
    microphone_inference_end();
    return false;
  }

  // optionally set the gain, defaults to 24
  // Note: values >=52 not supported
  //PDM.setGain(40);

  return true;
}
```

Figure 5.26 PDM Microphone Configuration

The display screen was configured to display the results and send some outputs as seen in the figure below.



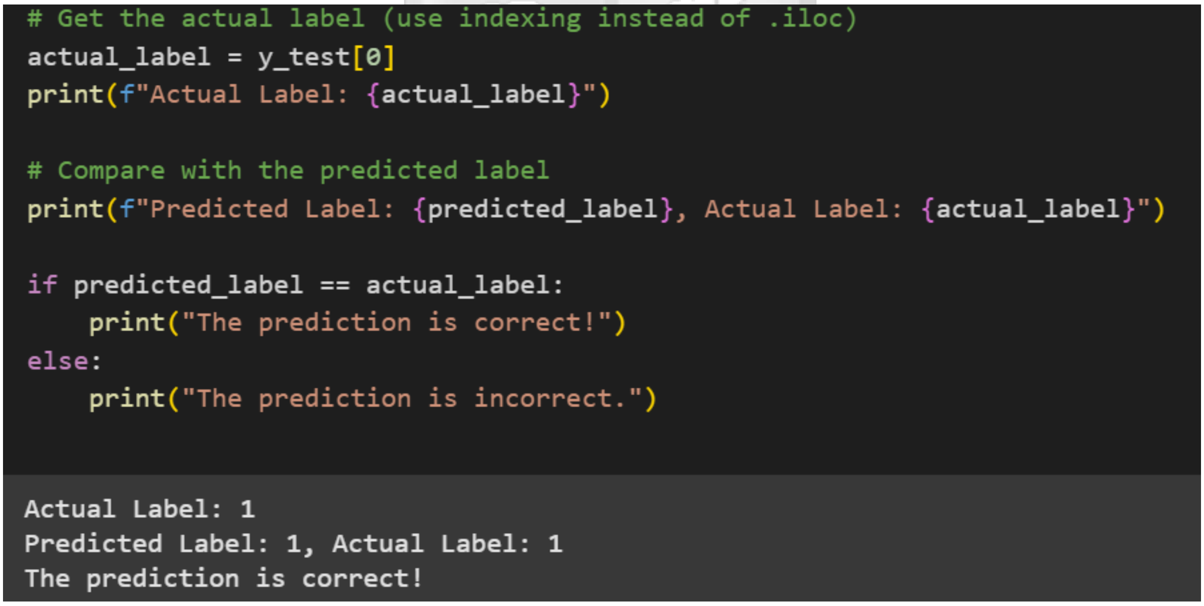
```
15:36:45.502 -> Recording done
15:36:47.428 -> run_classifier returned: 0
15:36:47.428 -> Timing: DSP 1878 ms, inference 18 ms, anomaly 0 ms
15:36:47.428 -> Predictions:
15:36:47.428 ->   elephant: 0.89062
15:36:47.428 ->   non-elephant: 0.10938
15:36:52.470 -> Starting inferencing in 2 seconds...
15:36:54.495 -> Recording...
15:36:57.526 -> Recording done
15:36:59.451 -> run_classifier returned: 0
15:36:59.451 -> Timing: DSP 1876 ms, inference 18 ms, anomaly 0 ms
15:36:59.451 -> Predictions:
15:36:59.451 ->   elephant: 0.75391
15:36:59.451 ->   non-elephant: 0.24609
```

Figure 5.27 OLED Display Output

5.6 System Testing Results

5.6.1 Model Testing

The model was first tested on the development space where it was tested on sample data to check the predictions as shown in the figure 5.28 below. From the image, it is seen that the actual label was one which is elephant, and the predicted output was also one which showed a correct output.

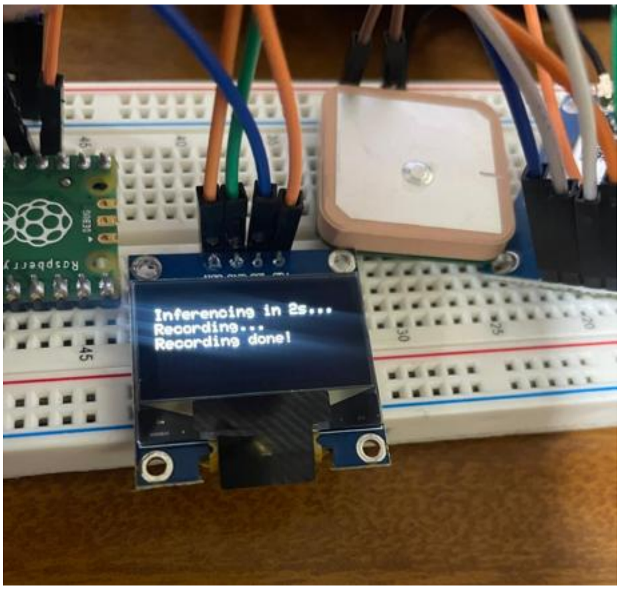


```
# Get the actual label (use indexing instead of .iloc)
actual_label = y_test[0]
print(f"Actual Label: {actual_label}")

# Compare with the predicted label
print(f"Predicted Label: {predicted_label}, Actual Label: {actual_label}")

if predicted_label == actual_label:
    print("The prediction is correct!")
else:
    print("The prediction is incorrect.")
```

```
Actual Label: 1
Predicted Label: 1, Actual Label: 1
The prediction is correct!
```

Figure 5.28 Prediction Output Test

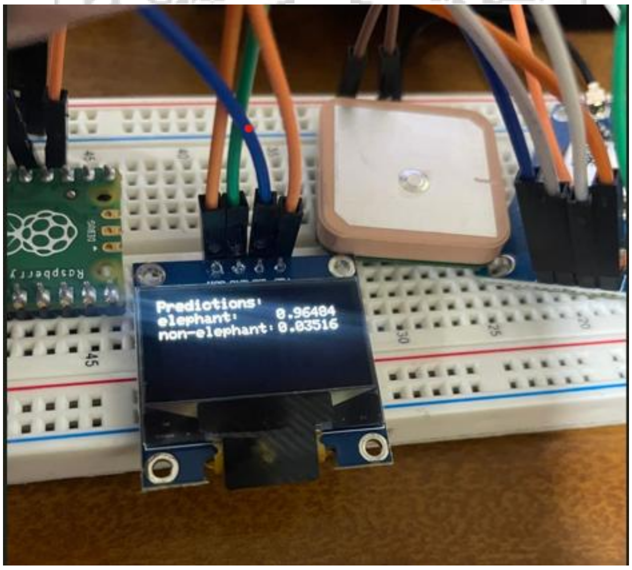
After integration, the machine learning model was tested on the IoT device to show the output on the screen. The figures below show the output on the OLED display screen from recording the audio to processing and giving the prediction result.



```
Inferencing in 2s...
Recording...
Recording done!
```

Figure 5.29 Processing Sound Input

After processing, the output is then printed next on the screen to show the output as in figure 5.29 above and the prediction output shown in the figure 5.30 below. In real-world deployments, several environmental variables can significantly affect the accuracy and performance of acoustic detection systems. Ambient forest noise such as wind, bird calls or river sounds may mask or distort elephant sounds. The sensitivity and placement of microphones also influence the quality of captured signals as does terrain interference such as trees or uneven ground which can reflect or absorb sound waves. These factors were not directly measured in this study, as it focused on training and testing using secondary data in a controlled set up.



```
Predictions:
elephant:     0.96484
non-elephant: 0.03516
```

Figure 5.30 Prediction Output on OLED Display

Chapter 6: Discussions and Results

6.1 Introduction

In this chapter, the results and findings are discussed where comparisons are done form the implementation and testing of the early elephant detection system. The system's performance is analyzed while comparing the results with existing detection methods.

6.2 Trained Model Performance Comparison

As seen in chapter 5, CNN performed better than Random Forest in overall accuracy, precision, recall and f1-score. Although random forest had a more balanced precision-recall trade off, CNN had better class distinction as seen in figure 5.15 and figure 5.16. The Support Vector Machine model achieved an accuracy of 97.22% shown in figure 5.17 which is high and outperformed the random forest and very close to the CNN model. SVM had a slightly lower recall rate than precision rate in one of the classes and a strong recall and lower precision in the other class which would bring about false positives and negatives. Logistic regression trailed closely behind SVM with an accuracy of 96.53% and maintained a balance between precision at 94.52% and recall at 98.57% compared to random forest which performed slightly lower compared to SVM and logistic regression. Naïve Bayes performed the worst with the lowest accuracy of 89% and lowest recall, f1 score and precision.

Table 6.1Performance Metrics Overview

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
CNN	99%	99%	99%	99%
SVM	97%	95%	98%	97%
Logistic Regression	96%	94%	98%	96%
Random Forest	94%	91%	97%	94.44%
Naïve Bayes	89%	90%	89%	89%

From the performance summary, the CNN model outperformed all traditional models which suggests that deep learning is effective in capturing complex patterns within the dataset. Among the traditional models, SVM had the best accuracy with a good balance between accuracy, precision and recall which would make it the next best alternative to the CNN deep learning model. From the study, CNN had some false positives in different trained models which for the early detection system is better than to have compared to false negatives meaning that an elephant is not detected.

To improve statistical robustness of the results, this study incorporated cross-validation, training and validation loss over epochs to ensure consistency. In addition to overall accuracy, as seen in chapter 5, performance metrics such as precision, recall and F1 score were calculated. To evaluate robustness under realistic conditions, environmental noise such as wind, human conversation and rustling leaves was artificially introduced to the test set. The model maintained high performance of over 93% under these simulated field conditions, which suggested it may tolerate moderate acoustic interference typical of forest environments.

6.3 Comparison with Existing Work

The trained models are compared with the existing models discussed in the literature review of this study in terms of the different performance metrics.

6.3.1 Model Performance Metrics Comparison

For this study, several models were trained one deep learning model and four traditional models to compare the performance and efficiency. The model that performed best form the trained models was a one dimensional CNN model that had a very high accuracy of 99% as compared to a different study where three CNNs were combined to have a two-level model which had an accuracy of 92% for the first level and 99% accuracy for the low frequency second level classifier and an accuracy of 75% for the second level model that classified high frequency vocalizations (Ravikrishna et al., 2023). This also goes to show that an accuracy of 99% is possible especially with deep learning models which tend to have higher accuracy compared to the traditional models. For my feature extraction, MFCCs were preferred as sound spectrum although preferred for CNN models, requires complex models to use its features (In Tan et al., 2022).

A different study on bioacoustics monitoring of frogs applied CNN where the best accuracy achieved was approximately 94% which was a filtered experiment using filtered spectrogram which greatly reduces background noise (Xie et al., 2022).According to Sharma et al., (2023), the CNN model had a validation accuracy of 89.36% which was the best compared to the other models trained which were CRNN, SVM, RNN, and ResNet34. Similar to this study, the CNN model demonstrated higher accuracy in audio sample classification. The feature extraction used MFCCs for their audio features where all 13 coefficients were used like the features used for this study with two additional features as discussed in the implementation of the system.

Table 6.2 Model Accuracy Comparison with Existing Studies

Author & Year	**Classification Model**	**Accuracy**
Ravikrishna et al., 2023	CNN	99%
Sharma et al., 2023	CNN	89.36%
Xie et al., 2022	CNN	94%

The machine learning model trained for this study achieved a high accuracy which could be due to the dataset used which was combination of clear elephant vocalizations and forest sounds from a different dataset which provided quality training, testing and validation data. The feature extraction was also done differently where MFCCs, zero crossing rate, chroma features and some spectral features were extracted as discussed in the implementation in chapter 5. With the features extracted, the original audio characteristics are preserved and allows for better model performance.

6.3.2 Architectural and Hardware Considerations

In different studies, the system architectures applied are different in one way or another. This is in terms of the hardware used in the IoT device and the implementations used. Figure 2.8 and figure 2.9 in chapter two of this study were used as a guide to coming up with an architecture for the early detection system. From the two architectures, it is seen that raspberry pi 4 and raspberry pi 4B are the microcontrollers utilized for the two systems. For this study, raspberry pi was the micro controller of choice due to different considerations such as power consumption, real-time processing, the cost and the size which are factors critical for deployment in remote and constrained environments like forests.

Table 6.3 Raspberry Pi 4 vs. Raspberry Pi Pico for Early Detection of Elephants System (Pico-Series Microcontrollers - Raspberry Pi Documentation, 2025)

Feature	**Raspberry Pi 4**	**Raspberry Pi Pico**	**Advantages (Pico over 4)**
Processor	Quad-core Cortex-A72, 1.5GHz	RP2040 (Dual core Cortex M0+ 133MHz)	Lower power consumption
Operating System (OS)	Linux-based (Raspberry Pi Os, Ubuntu)	None(runs MicroPython, C/C++)	Faster boot time and no OS overhead

RAM	2GB, 4GB, 8GB LPDDR4	264KB SRAM	More power efficient and optimized fro lightweight tasks
Storage	microSD card (expandable)	None (uses flash memory)	No need for an SD card
Power Consumption	Higher (~600- 700 mA, 5V/3A)	Very low (~100mA, 1.8V-5.5V)	Energy efficient, ideal for battery powered applications
Real-Time processing	No(runs a full OS, potential delays)	Yes (deterministic timing)	Minimal latency for real-time sensor applications
Size	larger	Very small	Compact for embedded systems
Cost	~$35+	~$4	Cost effective, easy for large scale deployment

From the summary, Raspberry Pi Pico is ideal for low-power, real-time applications where energy efficiency and embedded processing are critical which makes it the preferred choice for the elephant detection system in remote areas while the Raspberry Pi 4 is more suited for applications that require AI, networking and multitasking as it is more powerful but consumes significantly more power. Raspberry Pi Pico cannot run full ML frameworks like TensorFlow as seen with the Pico 4; it supports TensorFlow Lite which allows for lightweight ML model inference when models are properly quantized. The trained model was integrated into Raspberry Pi Pico after conversion into a lightweight model; a C header file since it has limited computational power and memory. The model.h file was then integrated into Pico using C++ on Arduino IDE.

6.3.3 Dataset Comparisons

For this study, two datasets were used elephant vocalizations from Elephant Voices on SoundCloud and FSC22 dataset for forest sounds from Kaggle. The audio files were .wav which made it easy to process but with varying lengths. The elephant vocalizations were labelled as 1 where 1 is elephant and the forest sounds labelled as 0 where 0 is non-elephant. The extracted features included MFCCs, spectrograms and waveform-based features. In most studies, the data used for training and testing was primary data as compared to the data used for this study which was secondary. In one study, vocalizations were captured over a period of three months with a total of 235 audio files with an almost 50-50 split when labelled (Sharma et al., 2022). The number of audios used in this study is close to the number used in this research study. The optimal number of audio files ranges from 200-300 audio samples.

Relying on secondary data as was done in this study, is a practical choice however it comes with several inherent limitations:

a) Lack of control over data collection conditions in terms of environmental context, microphone placement, sensitivity and calibration which might differ from a real-life scenario.
b) Limited metadata where most secondary datasets lack full metadata such as timestamp accuracy, species context and one might not know how the data was pre-processed where other species vocalizations or human-made noise may have been filtered out.

Chapter 7: Conclusions, Recommendations and Future Works

7.1 Conclusions

This study successfully designed, developed and tested an elephant early detection system that integrates IoT components with machine learning (ML) for real-time detection where when an elephant is detected at a certain location and distance, an alert is sent to the rangers with the details and exact location who can act before the elephants arrive. The system's main objective was to detect elephants based on their vocalizations and send SMS alerts to rangers with exact GPS coordinates before elephants approach the human settlements. This early warning provides rangers with actionable time to respond and mitigate human-elephant conflict (HEC).

A lightweight Convolutional Neural Network (CNN) model was trained after data preprocessing and feature extraction. The model was optimized for edge deployment and integrated into a Raspberry Pi Pico microcontroller. Despite the hardware's constraints, the model achieved high accuracy in distinguishing elephant vocalizations from other ambient forest sounds. From the testing phase, the system was validated, and the following results were observed:

i. The CNN model achieved an accuracy of ~99% indicating robust detection performance.
ii. The integration with the GSM module successfully sent real-time SMS alerts upon detection, verifying the end-to-end and real-time functionality.

The study also compared the methodology applied to this study with other existing studies. Conventional detection methods such as collaring and camera traps were seen to be cost-prohibitive, invasive and limited in spatial and temporal scope. In contrast, the developed system offers a low-cost, non-invasive, scalable alternative to detection and monitoring. The comparison with existing models demonstrated the effectiveness of using edge computing for conservation efforts. A comprehensive review of different detection approaches and technologies applied was carried out through a comprehensive literature review, highlighting the potential of bioacoustics monitoring of wildlife for real-time monitoring.

7.2 Recommendations

i. The model should be tested in the real-world environment such as a forest setting to assess the performance in varied conditions.

ii. There should be more diverse elephant vocalization samples, including in different environments or terrains and background noise levels which can enhance model robustness.

7.3 Suggestions for Future Research

Future research studies can explore:

i. Integration of the detection system with existing elephant deterrent methods such as bee fences.
ii. Behavioural analysis can be integrated into the acoustic localization to allow for better interactions between humans and elephants.
iii. Monitoring features can be integrated into the system to allow for long-term tracking which can complement existing tracking methods and assist in better decision making.

7.4 Limitations and Delimitations

7.4.1 Limitations

i. The use of secondary datasets presents limitations in terms of the context of data collection and the processing of the done carried out.
ii. The testing of the system was conducted in a controlled environment rather than in a dynamic real-world environment where everything varies.
iii. The Raspberry Pi Pico's limited processing power constrained the complexity of the machine learning model where larger models such as deep neural networks require more RAM.

7.4.2 Delimitations

i. The system was designed mainly for early detection and not long-term tracking.
ii. The study focused solely on elephants excluding all other wildlife species.

References

Acoustic communication. (2024). https://www.elephantvoices.org/elephant-communication/acoustic-communication.html

Adams, T. S., Chase, M. J., & Leggett, K. E. (2021). *Elephant movements in different human land-uses in Chobe District, Botswana*. *62*.

Ball, R., Jacobson, S. L., Rudolph, M. S., Trapani, M., & Plotnik, J. M. (2022). Acknowledging the Relevance of Elephant Sensory Perception to Human–Elephant Conflict Mitigation. *Animals*, *12*(8), Article 8. https://doi.org/10.3390/ani12081018

Beirne, C., Houslay, T. M., Morkel, P., Clark, C. J., Fay, M., Okouyi, J., White, L. J. T., & Poulsen, J. R. (2021). African forest elephant movements depend on time scale and individual behavior. *Scientific Reports*, *11*(1), 12634. https://doi.org/10.1038/s41598-021-91627-z

Beirne, C., Meier, A. C., Brumagin, G., Jasperse-Sjolander, L., Lewis, M., Masseloux, J., Myers, K., Fay, M., Okouyi, J., White, L. J. T., & Poulsen, J. R. (2020a). Climatic and Resource Determinants of Forest Elephant Movements. *Frontiers in Ecology and Evolution*, *8*, 96. https://doi.org/10.3389/fevo.2020.00096

Beirne, C., Meier, A. C., Brumagin, G., Jasperse-Sjolander, L., Lewis, M., Masseloux, J., Myers, K., Fay, M., Okouyi, J., White, L. J. T., & Poulsen, J. R. (2020b). Climatic and Resource Determinants of Forest Elephant Movements. *Frontiers in Ecology and Evolution*, *8*, 96. https://doi.org/10.3389/fevo.2020.00096

Bjorck, J., Rappazzo, B. H., Chen, D., Bernstein, R., Wrege, P. H., & Gomes, C. P. (2019). *Automatic Detection and Compression for Passive Acoustic Monitoring of the African Forest Elephant* (arXiv:1902.09069). arXiv. http://arxiv.org/abs/1902.09069

Brickson, L., Zhang, L., Vollrath, F., Douglas-Hamilton, I., & Titus, A. J. (2023). Elephants and algorithms: A review of the current and future role of AI in elephant monitoring.

Journal of The Royal Society Interface, *20*(208), 20230367. https://doi.org/10.1098/rsif.2023.0367

Cardoso, A. W., Malhi, Y., Oliveras, I., Lehmann, D., Ndong, J. E., Dimoto, E., Bush, E., Jeffery, K., Labriere, N., Lewis, S. L., White, L. T. J., Bond, W., & Abernethy, K. (2020). The Role of Forest Elephants in Shaping Tropical Forest–Savanna Coexistence. *Ecosystems*, *23*(3), 602–616. https://doi.org/10.1007/s10021-019-00424-3

Conserving African forest elephants through bioacoustic methods – with Bobbi Estabrook – Tuesday, February 20th – By Zoom Only – York County Audubon. (2024, February 20). https://www.yorkcountyaudubon.org/events/conserving-african-forest-elephants-through-bioacoustic-methods-with-bobbi-estabrook-tuesday-february-20th-by-zoom-only

Ear to the ground: Locating elephants using ground vibrations | University of Oxford. (2021, July 14). https://www.ox.ac.uk/news/2021-07-14-ear-ground-locating-elephants-using-ground-vibrations

Frost, J. (2022, October 16). *Purposive Sampling: Definition & Examples*. Statistics By Jim. https://statisticsbyjim.com/basics/purposive-sampling/

Gomes-Selman, J. M., Demir, N., Wrege, P., & Paepcke, A. (2021). Curriculum Learning to Handle Extreme Class Imbalance for Acoustic Modeling of Forest Elephant Calls. *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1022–1028. https://doi.org/10.1109/ICMLA52953.2021.00168

Gross, E. M., Pereira, J. G., Shaba, T., Bilério, S., Kumchedwa, B., & Lienenlüke, S. (2022). Exploring Routes to Coexistence: Developing and Testing a Human–Elephant Conflict-Management Framework for African Elephant-Range Countries. *Diversity*, *14*(7), 525. https://doi.org/10.3390/d14070525

Hahn, N. R., Wall, J., Denninger-Snyder, K., Goss, M., Sairowua, W., Mbise, N., Estes, A. B., Ndambuki, S., Mjingo, E. E., Douglas-Hamiliton, I., & Wittemyer, G. (2022). Risk perception and tolerance shape variation in agricultural use for a transboundary elephant population. *Journal of Animal Ecology*, *91*(1), 112–123. https://doi.org/10.1111/1365-2656.13605

Hart, J., Gobush, K., Maisels, F., Wasser, S., Okita-Ouma, B., & Slotow, R. (2021). African forest and savannah elephants treated as separate species. *Oryx*, *55*(2), 170–171. https://doi.org/10.1017/S0030605320001386

Human—Elephant Conflict | WWF. (2024). https://wwf.panda.org/discover/knowledge_hub/endangered_species/elephants/human_elephant_conflict/

International Fund for Animal Welfare. (2023, July 20). *Human-elephant conflict: What it is and why it's a major threat*. IFAW. https://www.ifaw.org/international/journal/human-elephant-conflict-major-threat

Kamdar, A., Baishya, H. K., Nagendra, H., Ratnam, J., Smith, D., & Sekar, N. (2022a). Human–elephant conflict mitigation as a public good: What determines fence maintenance? *Ecology and Society*, *27*(3). https://doi.org/10.5751/ES-13271-270324

Kamdar, A., Baishya, H. K., Nagendra, H., Ratnam, J., Smith, D., & Sekar, N. (2022b). Human–elephant conflict mitigation as a public good: What determines fence maintenance? *Ecology and Society*, *27*(3), art24. https://doi.org/10.5751/ES-13271-270324

Karas, L. (2024). *LibGuides: Literature Review: Types of Literature Reviews*. https://uscupstate.libguides.com/c.php?g=627058&p=6601225

Khatri, K., S, A. C., & D'Souza, J. M. (2022). Detection of Animals in Thermal Imagery for Surveillance using GAN and Object Detection Framework. *2022 International*

Conference for Advancement in Technology (ICONAT), 1–6. https://doi.org/10.1109/ICONAT53423.2022.9725883

Krause, D., Politis, A., & Kowalczyk, K. (2021). Comparison of Convolution Types in CNN-based Feature Extraction for Sound Source Localization. *2020 28th European Signal Processing Conference (EUSIPCO)*, 820–824. https://doi.org/10.23919/Eusipco47968.2020.9287344

Montero-Botey, M., Kivuyo, E., Sitati, N., & Perea, R. (2024). Deforestation and water availability as main drivers of human-elephant conflict. *Global Ecology and Conservation*, *54*, e03068. https://doi.org/10.1016/j.gecco.2024.e03068

Mutero, R. (n.d.). *The importance of elephants on the ecosystem*. Sabi Sabi. Retrieved October 1, 2024, from https://www.sabisabi.com/discover/topics/the-importance-of-elephants-on-the-ecosystem

Mwangi, N. (2024, October 4). 50 Elephants Translocated from Mwea To The Aberdare. *Nairobi Wire*. https://nairobiwire.com/2024/10/50-elephants-translocated-from-mwea-to-the-aberdare.html

Nikolopoulou, K. (2023, August 9). *What Is an Algorithm? | Definition & Examples*. Scribbr. https://www.scribbr.com/ai-tools/what-is-an-algorithm/

Pang, C., Liu, H., & Li, X. (2019). Multitask Learning of Time-Frequency CNN for Sound Source Localization. *IEEE Access*, *7*, 40725–40737. https://doi.org/10.1109/ACCESS.2019.2905617

Ramasubramanian, C., Lokiah, S., Viswanath, Y., & Jamthe, S. (2022). Averting Human-Elephant Conflict using IoT and Machine Learning of Elephant Vocalizations. *2022 IEEE 8th World Forum on Internet of Things (WF-IoT)*, 1–6. https://doi.org/10.1109/WF-IoT54382.2022.10152220

Ranasinghe, U., Abeyrathne, S., Samaranayake, L., Weerakoon, T., Harischandra, N., & Dissanayake, G. (2023). Enhanced Frequency Domain Analysis for Detecting Wild Elephants in Asia using Acoustics. *2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS)*, 140–145. https://doi.org/10.1109/ICIIS58898.2023.10253540

Ravikrishna, S., Kumar, C. S. S., Sharan, V., Kumar, V. S., & Logeshwar, S. (2023). Elephant Detection and Alarm System Using TensorFlow. *2023 International Conference on Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM)*, 408–413. https://doi.org/10.1109/iTechSECOM59882.2023.10435232

Reinwald, M., Moseley, B., Szenicer, A., Nissen-Meyer, T., Oduor, S., Vollrath, F., Markham, A., & Mortimer, B. (2021). Seismic localization of elephant rumbles as a monitoring approach. *Journal of the Royal Society Interface*. https://doi.org/10.1098/rsif.2021.0264

Rhinehart, T. A., Chronister, L. M., Devlin, T., & Kitzes, J. (2020). Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecology and Evolution*, *10*(13), 6794–6818. https://doi.org/10.1002/ece3.6216

Sai, C. C., & Dewangan, A. (2023). *AUTOMATIC ELEPHANT DETECTION SYSTEM USING MACHINE LEARNING AND IOT TECHNIQUES*. *8*(6).

Santosh Kumar, S., Sushmitha, M., Sirisha, P., Shilpa, J., & Roopashree, D. (2018). Sound Activated Wildlife Capturing. *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2250–2253. https://doi.org/10.1109/RTEICT42901.2018.9012357

Singh, S. (2023, July 20). *What is a Conceptual Framework and How to Make It (with Examples) | Researcher.Life*. https://researcher.life/blog/article/what-is-a-conceptual-framework-and-how-to-make-it-with-examples/

Singh, T., Gangloff, H., & Pham, M.-T. (2023). Object Counting from Aerial Remote Sensing Images: Application to Wildlife and Marine Mammals. *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 6580–6583. https://doi.org/10.1109/IGARSS52108.2023.10282150

Systems Architecture: Everything You Need to Know When Assessing Systems Architecture Skills. (n.d.). Retrieved October 6, 2024, from https://www.alooba.com/skills/concepts/systems-architecture/

The grim reality of human-elephant conflict in northern Kenya—Save the Elephants. (2021, August 7). https://www.savetheelephants.org/news/the-grim-reality-of-human-elephant-conflict-in-northern-kenya/

Viana, D. S., Granados, J. E., Fandos, P., Pérez, J. M., Cano-Manuel, F. J., Burón, D., Fandos, G., Aguado, M. Á. P., Figuerola, J., & Soriguer, R. C. (2018). Linking seasonal home range size with habitat selection and movement in a mountain ungulate. *Movement Ecology*, *6*(1), Article 1. https://doi.org/10.1186/s40462-017-0119-8

Weerasinghe, G., Karunanayaka, K., Kumarasinghe, P., Perera, D., Trenado, C., Zoysa, K. D., & Keppitiyagama, C. (2022). Identification of Previously Unseen Asian Elephants using Visual Data and Semi-Supervised Learning. *2022 22nd International Conference on Advances in ICT for Emerging Regions (ICTer)*, 019–024. https://doi.org/10.1109/ICTer58063.2022.10024068

What is the Research Methodology? Definition, Types, and Examples | Egyptian Culture Center. (2024, May 2). https://eccceg.com/what-is-the-research-methodology/

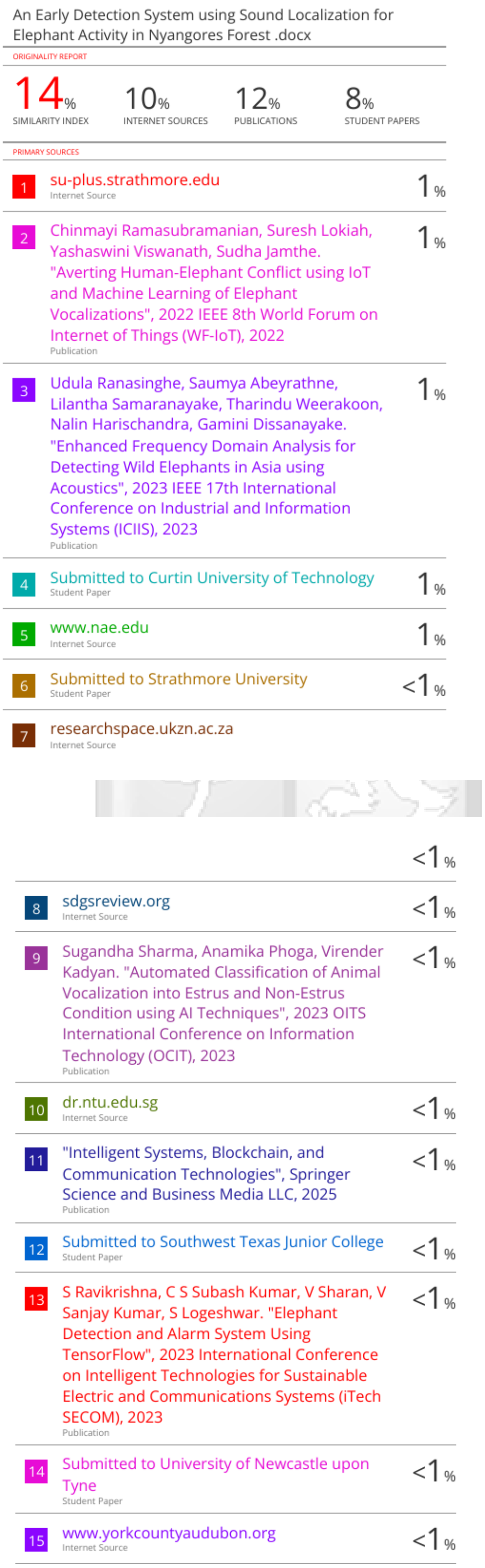
Why Do Elephants Use Seismic Communication? - Mpala. (2023, April 29). https://mpala.org/why-do-elephants-use-seismic-communication/

Wireless Integrated Sensor Network: Boundary Intellect system for elephant detection via cognitive theory and Fuzzy Cognitive Maps—ScienceDirect. (2018, June). https://www.sciencedirect.com/science/article/abs/pii/S0167739X17302170

Yangfan J, L., Stuart, B., & Patricia, D. (2024, June 14). *Acoustic Source Localization Techniques and Their Applications*. NAE Website. https://nae.edu/255823/Acoustic-Source-Localization-Techniques-and-Their-Applications

Appendices

Appendix A: Similarity Report



An Early Detection System using Sound Localization for Elephant Activity in Nyangores Forest .docx

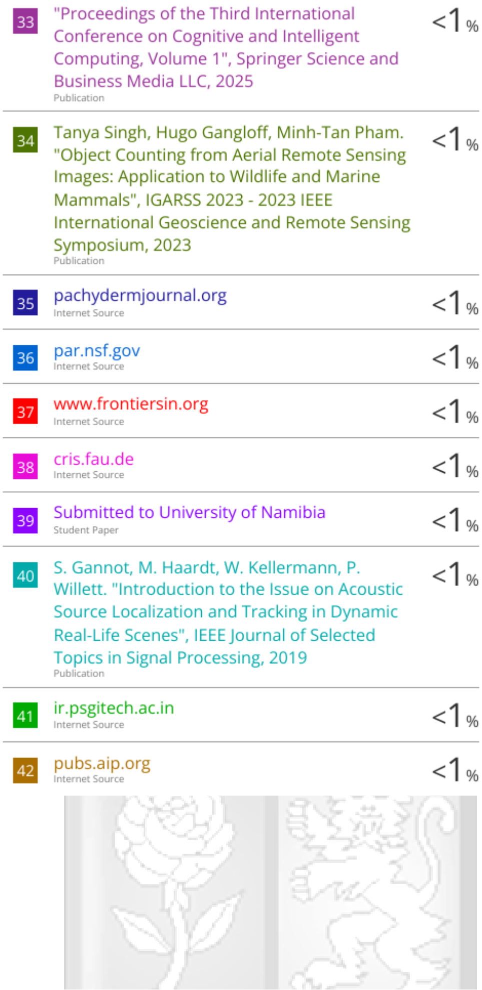
ORIGINALITY REPORT

14%	10%	12%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

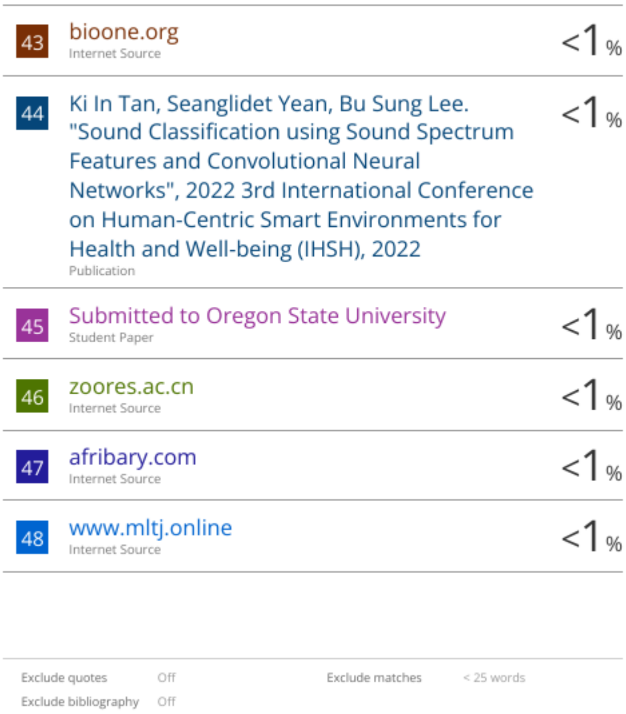
PRIMARY SOURCES

1	su-plus.strathmore.edu Internet Source	1%
2	Chinmayi Ramasubramanian, Suresh Lokiah, Yashaswini Viswanath, Sudha Jamthe. "Averting Human-Elephant Conflict using IoT and Machine Learning of Elephant Vocalizations", 2022 IEEE 8th World Forum on Internet of Things (WF-IoT), 2022 Publication	1%
3	Udula Ranasinghe, Saumya Abeyrathne, Lilantha Samaranayake, Tharindu Weerakoon, Nalin Harischandra, Gamini Dissanayake. "Enhanced Frequency Domain Analysis for Detecting Wild Elephants in Asia using Acoustics", 2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS), 2023 Publication	1%
4	Submitted to Curtin University of Technology Student Paper	1%
5	www.nae.edu Internet Source	1%
6	Submitted to Strathmore University Student Paper	<1%
7	researchspace.ukzn.ac.za Internet Source	<1%
8	sdgsreview.org Internet Source	<1%
9	Sugandha Sharma, Anamika Phoga, Virender Kadyan. "Automated Classification of Animal Vocalization into Estrus and Non-Estrus Condition using AI Techniques", 2023 OITS International Conference on Information Technology (OCIT), 2023 Publication	<1%
10	dr.ntu.edu.sg Internet Source	<1%
11	"Intelligent Systems, Blockchain, and Communication Technologies", Springer Science and Business Media LLC, 2025 Publication	<1%
12	Submitted to Southwest Texas Junior College Student Paper	<1%
13	S Ravikrishna, C S Subash Kumar, V Sharan, V Sanjay Kumar, S Logeshwar. "Elephant Detection and Alarm System Using TensorFlow", 2023 International Conference on Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM), 2023 Publication	<1%
14	Submitted to University of Newcastle upon Tyne Student Paper	<1%
15	www.yorkcountyaudubon.org Internet Source	<1%

16	Submitted to Griffth University	Student Paper	<1 %
17	Submitted to Southern New Hampshire University - Continuing Education	Student Paper	<1 %
18	Yinggang Liu, Ying Wei. "Low-Complexity Sound Event Classification Based on Graph Signal in Noisy Environments", 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2019	Publication	<1 %
19	Siva Krishna Dasari, Sarath Kumar Kella, Raghava Manda. "Enhancing Environmental Sounds Classification through Deep Learning Techniques", 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2023	Publication	<1 %
20	Submitted to Chester College of Higher Education	Student Paper	<1 %
21	futureearth.org	Internet Source	<1 %
22	H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Computer Science Engineering", CRC Press, 2024	Publication	<1 %
23	Submitted to Breda University of Applied Sciences	Student Paper	<1 %
24	Submitted to The College of New Jersey	Student Paper	<1 %
25	Submitted to Daystar University	Student Paper	<1 %
26	Jonathan M Gomes-Selman, Nikita Demir, Peter Wrege, Andreas Paepcke. "Curriculum Learning to Handle Extreme Class Imbalance for Acoustic Modeling of Forest Elephant Calls", 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021	Publication	<1 %
27	Martin Hofmann, Steffen Kiel, Lara M Kösters, Jana Wäldchen, Patrick Mäder. "Inferring Taxonomic Affinities and Genetic Distances Using Morphological Features Extracted from Specimen Images: A Case Study with a Bivalve Data Set", Systematic Biology, 2024	Publication	<1 %
28	tjet.udsm.ac.tz	Internet Source	<1 %
29	research.usq.edu.au	Internet Source	<1 %
30	Submitted to Cranfield University	Student Paper	<1 %
31	Nils Poschadel, Stephan Preihs, Jürgen Peissig. "Investigations on higher-order spherical harmonic input features for deep learning-based multiple speaker detection and localization", EURASIP Journal on Audio, Speech, and Music Processing, 2025	Publication	<1 %
32	innspub.net	Internet Source	<1 %



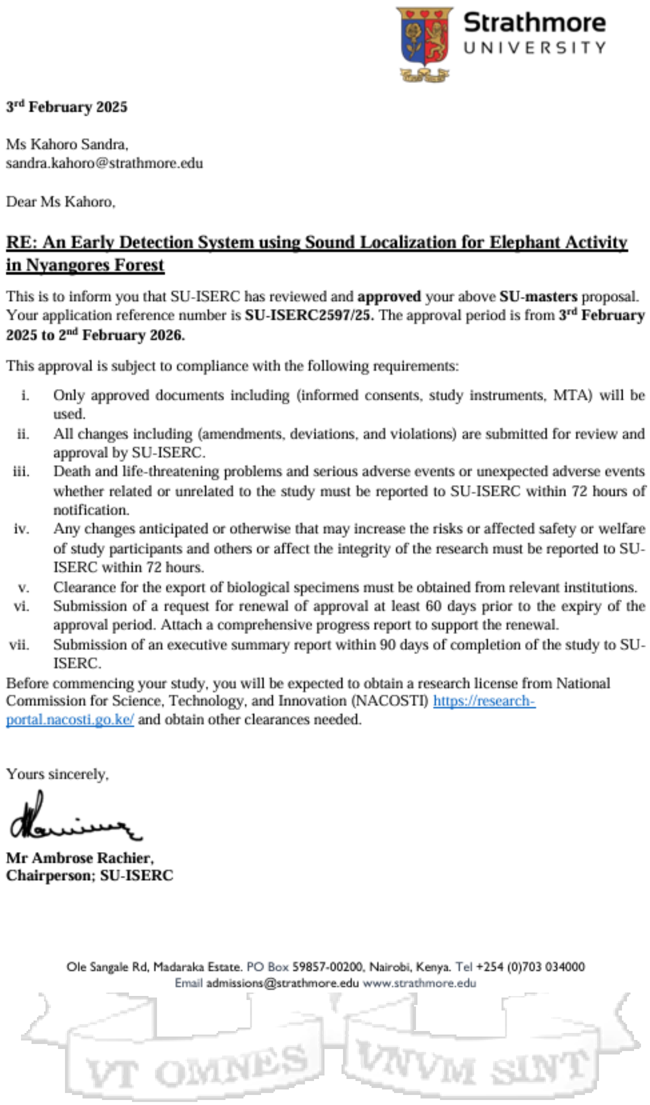
33	"Proceedings of the Third International Conference on Cognitive and Intelligent Computing, Volume 1", Springer Science and Business Media LLC, 2025 Publication	<1 %
34	Tanya Singh, Hugo Gangloff, Minh-Tan Pham. "Object Counting from Aerial Remote Sensing Images: Application to Wildlife and Marine Mammals", IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, 2023 Publication	<1 %
35	pachydermjournal.org Internet Source	<1 %
36	par.nsf.gov Internet Source	<1 %
37	www.frontiersin.org Internet Source	<1 %
38	cris.fau.de Internet Source	<1 %
39	Submitted to University of Namibia Student Paper	<1 %
40	S. Gannot, M. Haardt, W. Kellermann, P. Willett. "Introduction to the Issue on Acoustic Source Localization and Tracking in Dynamic Real-Life Scenes", IEEE Journal of Selected Topics in Signal Processing, 2019 Publication	<1 %
41	ir.psgitech.ac.in Internet Source	<1 %
42	pubs.aip.org Internet Source	<1 %



43	bioone.org Internet Source	<1 %
44	Ki In Tan, Seanglidet Yean, Bu Sung Lee. "Sound Classification using Sound Spectrum Features and Convolutional Neural Networks", 2022 3rd International Conference on Human-Centric Smart Environments for Health and Well-being (IHSH), 2022 Publication	<1 %
45	Submitted to Oregon State University Student Paper	<1 %
46	zoores.ac.cn Internet Source	<1 %
47	afribary.com Internet Source	<1 %
48	www.mltj.online Internet Source	<1 %

Exclude quotes	Off	Exclude matches	< 25 words
Exclude bibliography	Off		

Appendix B: Ethical Clearance Confirmation



Strathmore
UNIVERSITY

3rd February 2025

Ms Kahoro Sandra,
sandra.kahoro@strathmore.edu

Dear Ms Kahoro,

RE: An Early Detection System using Sound Localization for Elephant Activity in Nyangores Forest

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2597/25.** The approval period is from **3rd February 2025 to 2nd February 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu