A Wavelet Transform-based Feature Extraction Pipeline for Elephant Rumble Detection

Chamath K. Silva, Vinuri Piyathilake, Chamath Keppitiyagama, Asanka P. Sayakkara, Prabash Kumarasinghe, Namal Jayasuriya, and Udayanga Sampath

***Abstract*—Elephants generate infrasonic vocalisations that traverse through the air for long distances. Utilising this phenomenon, a previous work proposed a system, called *Eloc*, to localise and track elephants in the wild. The Eloc system has been demonstrated to be accurate in calculating the location of infrasonic sources. However, it still lacks the capability to accurately distinguish elephant infrasonic calls from various other infrasonic sources using limited computing power on board. Addressing this problem, the work presented in this paper introduces an approach to distinguish elephant infrasonic calls with a high accuracy on low-resourced hardware. Firstly, a sequence of operations are performed to reduce the effect of noise in the infrasonic signal captured by an Eloc node. Secondly, a wavelet-based signal reconstruction technique is applied to extract spectral features from the infrasonic signal. Finally, the extracted features are fed to a pre-trained machine learning classifier to distinguish the infrasonic vocalisations of elephants. The experimental evaluation using Asian elephant (*Elephas Maximus Maximus*) infrasonic vocalisation datasets demonstrates that the proposed approach is capable of accurately distinguishing elephant infrasonic calls on low-resourced hardware platform of the Eloc system, with accuracy levels over 82% under varying environmental conditions.**

***Index Terms*—passive acoustic monitoring, infrasonic detection, elephant rumble detection, wavelet transform, feature extraction.**

I. Introduction

The population of elephants in the world has dramatically declined over the last few decades. At the commencement of the 20th century there were an estimated 200,000 Asian elephants, but at present, there are probably no more than 35,000 to 40,000 elephants left in the wild [1]. The human-elephant conflict (HEC) is one of the major threats to wild elephants, which has been caused by habitat loss and fragmentation. In Asia, most of the areas that form elephant habitat are in close proximity to human settlements. Therefore, conflicts often emerge between animals foraging for food and the local human population [2], causing life-threatening consequence to both humans and elephants. For instance, in the year 2019, 121 people and 405 elephants lost their lives due to the human-elephant conflict in Sri Lanka [3].

Chamath K. Silva (mbckchamathsilva@gmail.com), Vinuri Piyathilake (piyathilakev@gmail.com), Chamath Keppitiyagama (chamath@ucsc.cmb.ac.lk), Asanka P. Sayakkara (asa@ucsc.cmb.ac.lk), Prabash Kumarasinghe (jpk@ucsc.cmb.ac.lk), Namal Jayasuriya (nmj@ucsc.cmb.ac.lk), and Udayanga Sampath (uds@ucsc.cmb.ac.lk) are with the University of Colombo School of Computing, Sri Lanka

Manuscript received Month 19, 2005; revised August 26, 2015.

Traditionally, several approaches have been used to keep elephants away from human habitats and farm lands, such as burning fires, lighting firecrackers, and implementing electric fencing. Although electric fence systems are effective at controlling the movements of elephants, they have their own limitations as well; they are relatively expensive to build to cover a sufficiently large geographic area, and when built incorrectly, can cause life-threatening injuries. Furthermore, electric fences are subject to frequent breakdowns, which require labour-intensive maintenance. Frustrated by the damages caused by elephants, some farmers resort to illegal methods, such as shooting and poisoning elephants [4].

Instead of entirely relying on aggressive methods to drive elephants away after their arrival, it would be more effective to detect the presence and movements of elephants in advance. In that way, accidental encounters with elephants can be avoided, which will reduce the damage caused by HEC. The recent developments of remote tracking systems, such as radio and global positioning system (GPS) collars, enable the accurate detection and tracking of elephants over a long period of time [5]. Although such collars are effective and useful in elephant conservation research to monitor their movements, they are too expensive to be used at large scale to address the general problem of HEC. Furthermore, the invasive nature of radio collars obstruct the natural behaviour of elephants, and their behaviour constantly causes damages to the devices. Under these circumstances, more affordable and reliable elephant tracking mechanisms are necessary.

Elephants use vocalizations for both long-range and short-range communication, consisting of both lower and higher frequencies. Moreover, these vocalisations can be classified into different call types based on their physical properties [6]. There are four main types of vocalisations for Asian elephants, namely, trumpet, roar, chirp, and rumble [7], [8]. Elephants use these calls in various contexts such as when being disturbed, playing with each other, moving in the presence of other species or vehicles, and communicating within the herd. It has been shown that, elephants produce low-frequency vocalisations – known as rumbles – when communicating between herds over long distances [8]. A typical African male elephant rumble fluctuates around a minimum of 12 Hz and a female rumble at around 13 Hz [9]. In Asian elephants, these values vary between 14 Hz to 24 Hz within a call duration of 1 to 25 seconds [8] due to their smaller vocal cords compared to African elephants.

Since elephant rumbles contain low-frequency sound waves that propagate longer distances without significant power

attenuation, they can be detected from several hundred meters away [10]. Previous studies have shown the possibility of using these infrasonic rumbles of wild elephants to detect their presence and possibly localizing them. However, there are only a very few applications that have attempted to implement such a system [11], [12], especially with low-cost hardware platforms. A system that can be used to detect elephants at large scale to minimise HEC requires reliable methods to process elephant infrasonic data on low-powered devices with minimum processing and time overhead. Currently, the lack of such infrasonic data processing and analysing pipelines hinders the practical deployment of infrasonic-based elephant detection systems.

This work addresses the problem of detecting elephant rumbles on low-powered devices through a novel approach that consists a wavelet-based signal reconstruction technique. The proposed approach is aimed at deploying on the previously published *Eloc* elephant infrasonic capturing platform [12]. This study is an extension of our long-term attempt to minimize human-elephant conflicts by early detection of elephants near human habitats.

This paper makes the following contributions:

1) Shows that wavelet-based feature extraction is effective at enabling elephant rumble detection on infrasonic data.
2) Presents a complete sound processing pipeline for infrasonic elephant rumble detection, which is capable of operating with natural noisy situations, on top of the resource-limited low-cost hardware platform.
3) Evaluates the proposed sound processing pipeline on an elephant vocalisation dataset of Asian elephants, and demonstrates its effectiveness.

The paper proceeds as follows. In Section II, the related work of this domain are illustrated. In Section III , the proposed automated elephant detection approach is described in detail. The implementation and experiment setup are illustrated in Section IV, which is followed by the results of the experiments in Section V. Finally, Section VI concludes the paper.

II. Related Work

Payne et al. first discovered that elephants generate infrasonic calls known as rumbles in the range of 14 Hz to 24 Hz [13]. Subsequent research has demonstrated that these low-frequency vocalizations travel long distances up to 6 kilometres [14], [15]. This is due to the lower attenuation of low-frequency sound in contrast to high-frequency sound. When attempting to detect these infrasonic calls of elephants, environmental factors, such as the temperature of the air and the speed of the wind, are shown to have a considerable impact [16]. Furthermore, the signal-to-noise ratio (SNR) of the infrasonic capturing hardware plays a significant role in the ability to successfully detect elephant infrasonic calls.

Zeppelzauer et al. introduced a novel spectro-temporal signal enhancement method that improves the signal-to-noise ratio in acoustic recordings, enabling effective detection of infrasound in noisy environments [15]. Based on the initial work, Zeppelzauer et al. further developed an automated elephant vocalisation detection technique with the objective of building an early warning system for elephants [11]. In this approach, the input signal is framed and transformed into spectrograms using the fast fourier transform (FFT). The spectrogram is then enhanced, filtered using a Greenwood filter bank, and logarithmized. The resulting spectrogram is mapped to the cepstral domain using the discrete cosine transform (DCT) and the cepstral feature vectors are temporally aggregated and classified using a trained support vector machine (SVM). The classifier attained an 88.2% detection rate with a 13.7% false-positive rate.

The study has considered a broad range of elephant vocalisations from 0 Hz to 500 Hz, where the infrasonic amounts to only a small region. Due to this reason, although they were able to detect elephant vocalisations, the detection mostly depends on the higher harmonics of the vocalisations. Consequently, this approach faces difficulties when detecting vocalisations at a longer distance, as higher frequency harmonics highly attenuate with distance. Similarly, Venter et al. has proposed another method that employs a sub-band pitch detection algorithm for automated detection of infrasound elephant call [17]. The accuracy of this method also depends on the availability of higher harmonics in elephant vocalisations.

Mohapatra et al. proposed a method to automatically detect elephant rumbles using feature extraction techniques that include the Greenwood function cepstral coefficients (GFCC) and the first three formant frequencies [18]. The GFCC features were extracted similarly to mel frequency cepstral coefficients (MFCC) extraction, while the formant frequencies were obtained using linear predictive coding (LPC). For classification, a feed-forward neural network trained with backpropagation reached 90% accuracy but had a high false-positive rate of 30%.

Bjorck et al. have made significant advancements in elephant rumble detection using passive acoustic monitoring incorporating modern neural network architecture, state-of-the-art training regimes, and data augmentation techniques [19]. The proposed system demonstrated a classification accuracy of 89.72% by utilizing the fast Fourier transform (FFT) for feature extraction and adopting a DenseNet neural network for classification. In addition, they have introduced a novel audio compression technique exclusively designed for elephant rumbles, which is useful in data transmission.

Although multiple studies exist for the acoustic detection of elephant infrasonic vocalisations as mentioned above, most of them are done on African elephants. The research aimed at detecting Asian elephant infrasonic vocalisations is very limited. Furthermore, such work does not take the cost of deployment on a large scale into account. Dabare et al. conducted a preliminary study on detecting infrasonic vocalisations of Asian elephants using low-cost hardware [20]. Their work considered the Infiltec Model INFRA-20 device, which is comparably cheaper than most commercially available infrasonic detectors (≈US $ 350). According to experimental results, this device is sensitive enough to capture elephant infrasonic vocalisations.

Sayakkara et al. proposed an infrasonic-based elephant localisation system using a low-cost hardware platform [12]. For the purpose of this system, a device called *Eloc* node

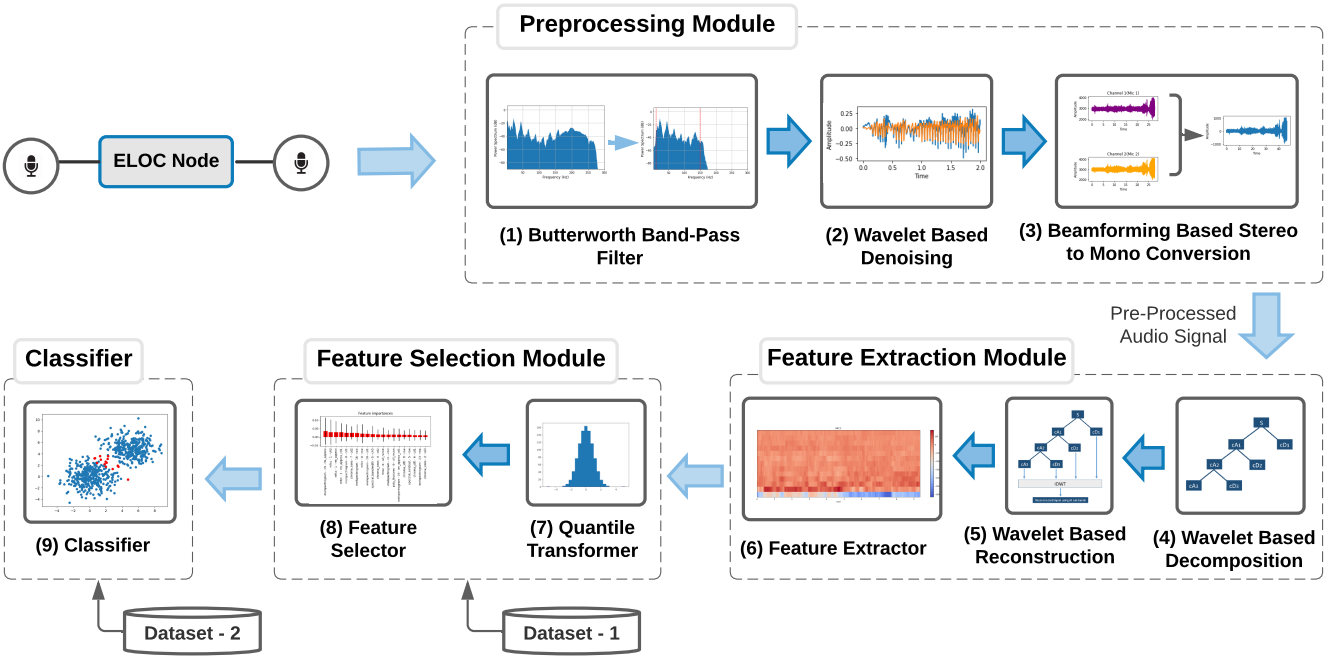


Fig. 1: High-level architecture of the proposed sound processing pipeline consisting of three modules to preprocess infrasonic data, extract features, and select best features from them before feeding into a machine learning classifier.

was developed, which consists of two Panasonic WM-61A microphones and a single board computer. This hardware device costs about US $ 73 and more sensitive to infrasonic than the previous INFRA-20 device. The proposed localisation system works by calculating the time difference of arrival (TDOA) of an elephant infrasonic call to the two microphones on an Eloc node. The direction calculation is shown to be highly accurate for angles between 30^o to 90^o. However, in order for the localisation system to work, confirming that a particular infrasonic signal originated from an elephant is necessary, i.e., rumble detection.

Jayasuriya et al. explored the potential of using a support vector machine (SVM) to distinguish elephant rumbles to be used on Eloc nodes [21]. Their work adopted mel frequency cepstral coefficients (MFCC) for feature extraction, which was originally designed to process human vocals. Their approach is shown to be most effective with data captured at a sample rate of 48 kHz; with lower frequencies, the accuracy of elephant vocalisation detection decreases. Since Eloc nodes capture data at a sample rate of 11 kHz, their approach is not sufficiently effective to be used on Eloc nodes to realise a low-cost infrasonic-based elephant detection and localisation system.

The comprehensive survey by Dan Stowell [22] outlines the progression of bioacoustics signal processing landscape. According to that study, it is evident that the recent advancements in the domain—regardless of the specific animal species being targeted—is moving towards the use of deep learning models. The author highlights the challenge of computational resource trade-off when attempting to run deep learning models on resource constrained hardware.

III. Proposed Sound-processing Pipeline

This section details the design considerations and the individual components of the proposed sound processing pipeline. Figure 1 illustrates the high-level overview of the pipeline consisting of a preprocessing, feature extraction, and feature selection modules.

A. Design Considerations

Since Asian elephant rumbles fluctuate around 14 Hz to 175 Hz (including harmonics), according to the Nyquist–Shannon sampling theorem, a sampling rate of at least 360 Hz is necessary for capturing the particular frequency range [23]. However, oversampling increases resolution, reduces noise, and helps avoid aliasing and phase distortion. Which means oversampling can improve the performance of applications which depend on information in the waveform shape of the signal.

Although oversampling has several advantages, it requires more computation power, computation time, and memory during the signal processing. Since the proposed approach intends to run on the Eloc node, the maximum recording and processing sampling rate is limited to 11 kHz. This is due to the limited computation power and memory on the Eloc nodes. Furthermore, since it is required to transmit the detected elephant rumbles to the central back-end over a cellular network [12], the oversampling will increase the operational cost. Therefore, this work considers a 600 Hz sampling rate as the balanced sampling rate between computational cost and the accuracy.

Initially, the infrasonic data should be segmented into samples of equal length. As longer data segments can pose

a computational and, consequently, an energy overhead to the low-powered elephant infrasonic detector, i.e., Eloc node, it is necessary to have shorter data segments. Therefore, a segment length of 2 seconds was selected. The segmented data goes through four main modules of the sound-processing pipeline: pre-processing module, feature extraction module, feature reduction module, and classification module. These modules and their internal functionality will be described in the following subsections.

B. Preprocessing Module

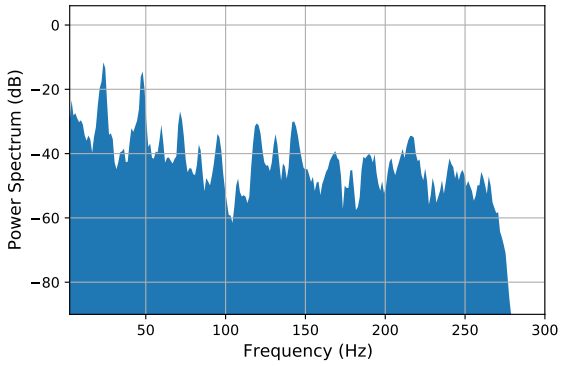
Figure 1 illustrates the preprocessing module as the first component in the pipeline. It consists of three main steps: Butterworth band-pass filter, wavelet-based denoising, and beamforming-based stereo to mono conversion.

1) Butterworth Band-pass Filter: Although the hardware-level low-pass filter in the Eloc node attenuates frequencies above 150 Hz, it will still allow frequencies up to 300 Hz to pass to a certain degree as the sample rate is 600 Hz. Since elephant rumbles fluctuate around the range of 14 Hz to 174 Hz, any signal components beyond that range contain unwanted information for elephant rumble detection. Moreover, these unwanted signals may have an adverse effect on the elephant rumble detection. Thus, a Butterworth band-pass filter [24] was applied with a low-cutoff frequency of 10 Hz, a high-cutoff frequency of 150 Hz, and a filter order of 9. Filter order 9 was determined empirically by analyzing the frequency response at several filter orders for the same sampling rate and cutoff frequencies.

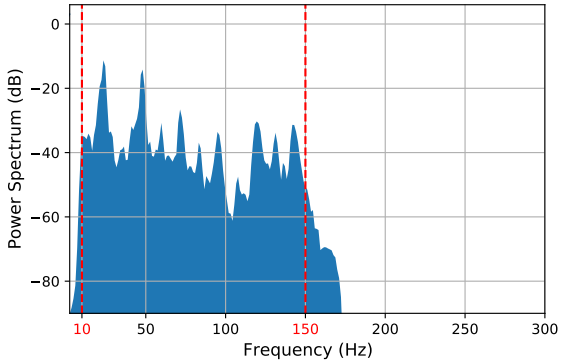
Figure 2 represents the frequency spectrum of an elephant rumble recording (a) before and (b) after applying the Butterworth band-pass filter. It is clear that the tuned band-pass filter removes the unwanted frequencies while maintaining uniform sensitivity for the required frequency range. Since the filtering process removes the frequency range from 0 Hz to 10 Hz and frequencies above 150 Hz, it will mitigate the influence of the low-frequency noise, and high-frequency noise that exists beyond the typical range of elephant rumbles.

2) Beamforming-based Stereo to Mono Conversion: As shown in Figure 3, two microphones in the Eloc node, located at a distance of 3 metres from each other, simultaneously captures the audio signal. Hence, the captured audio signals are in the stereo format. Any sound waves coming either from the front or the back, as seen in Figures 3, reach one of the microphones earlier than the other. Thus, both channels contain similar audio signals but with a slight phase difference. As the sound waves are coming from the far-field to the Eloc node, it can be assumed that the sound waves are parallel to each other from the point of view of the microphones.

However, both microphones can be affected by background noise coming from different directions to the primary sound source. Therefore a beamforming-based stereo to mono conversion approach, which minimizes the effect of such background noises, is designed. By considering the signal that reaches the setup first, i.e., reference signal, and the signal that is delayed, we can calculate the shift between the two signals in term of samples as shown in Equation 1.



(a)



(b)

Fig. 2: Frequency spectrum of an elephant rumble recording (a) before applying the band-pass filter and (b) after applying the band-pass filter.

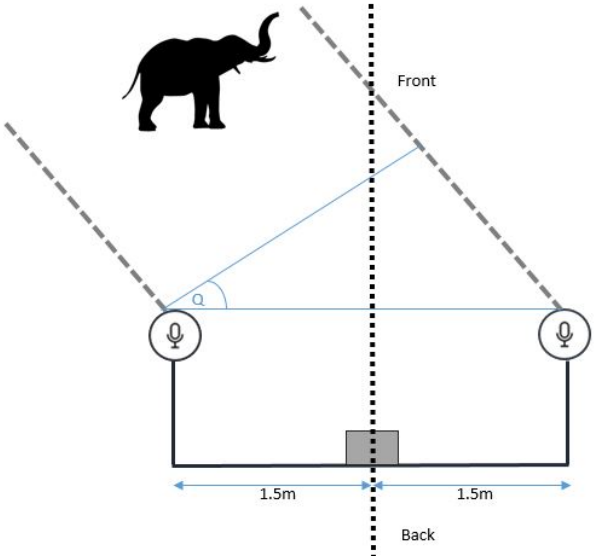


Fig. 3: Basic setup of the Eloc deployment unit where the pair of microphones (Eloc nodes) located at a 3m distance from each other and capturing data in a time-synchronized manner.

In Equation 1, n is the number of samples in the given signal segment, m is the number of shifted samples between the two signals and $C_{ref,delayed[m]}$ is the cross-correlation matrix

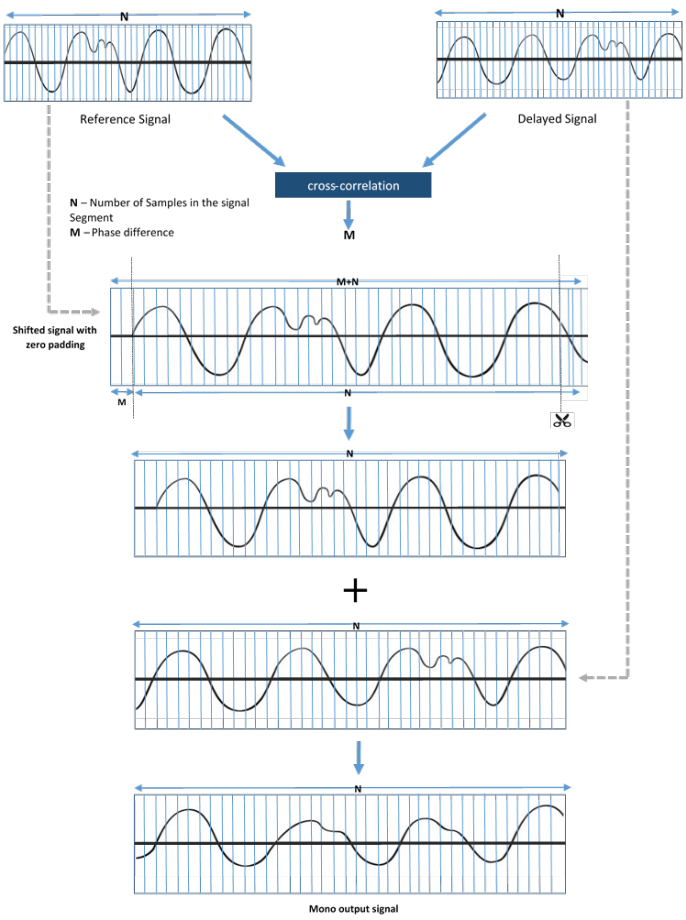


Fig. 4: The process of beamforming based stereo to mono conversion

between two signals.

$$argmax\left\{\sum_{n=0}^{n}\sum_{m=0}^{n}\|delayed\{n\}\ ref\{n+m\}\|\right\} = argmax\left(\sum_{n=0}^{n}\|C_{ref,delayed[m]}\|\right) \quad (1)$$

After calculating m, the reference signal is shifted by adding zero padding in front of the reference signal segment. Then, m frames are lifted from the tail of the reference signal segment to maintain the constant number of frames in the signal segment. Finally, the shifted reference signal and the delayed signal are merged by taking the average of the corresponding frames in the two signals. This process improves the strength of the dominant sound source while reducing the strength of the noise coming from unrelated source. A flowchart of this process of beamforming-based stereo to mono conversion is presented in Figure 4.

C. Feature Extraction Module

The output signal segment from the pre-processing module is taken as the input for the feature extraction module. The proposed feature extraction module consists of three steps: wavelet-based signal decomposition, wavelet-based signal reconstruction, and feature extraction (see Figure 1).

1) Wavelet Analysis: Fourier transform (FT) helps to convert a time-domain signal into a frequency-domain signal in order to identify its frequency components. However, FT has a critical drawback; the time information of a signal is lost during the transformation. In other words, when looking at an FT of a signal, it is impossible to recognize when exactly a particular event has occured [25]. But, this drawback is not crucial for the analysis of stationary signals, i.e., the signals with properties that do not significantly change over time. However, most of the natural infrasonic has a short duration and frequently-changing spectral characteristics. Because of that, they are non-stationary signals. Therefore, FT is not sufficient to analyse the behavior of elephant infrasonic signals.

To overcome the problem of losing time information, short-time Fourier transform (STFT) was developed. STFT is one of the most basic forms of time-frequency representations. It involves a technique called windowing, which modifies the FT to analyse smaller sections of the signal at a time [26]. Therefore, STFT can provide information to a certain level about when and at what frequencies a signal event occurs. However, the accuracy of these information depends on the size of the fixed time window. This means, narrow window sizes result in high time resolution but poor frequency resolution. In contrast, broad window sizes results in good frequency resolution but poor time resolution. It is necessary to have a more flexible approach where the resolution in both time and frequency domains can be attained on demand.

The limitations of STFT can be overcome by using Wavelet Transform (WT). It uses short windows at high frequencies and long windows at low frequencies [27]. Wavelet transform can be classified into two main groups: Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT). CWT operates over every possible scale and translation whereas DWT uses a specific subset of scale and translation values. Among them, DWT is used to reduce the computational burden of CWT [28]. It has been widely used for analyzing non-stationary signals and provides a time-frequency representation of the signals [29]. In DWT, a signal is decomposed into low-frequency band (approximation coefficients) and high-frequency band (detail coefficients). The low-frequency band is used for further decomposition [28].

WT can often compress or denoise a signal without significant degradation. These advantages of the wavelet method have also been shown practically in the extraction of information from ultrasonic Lamb waves where the received signal is naturally noisy [30], [31]. DWT has also been widely used in the ultrasonic signal analysis as a fast algorithm to obtain the wavelet transform of a discrete time signal [32].

2) Feature Extraction from Infrasonic Data: The accuracy of classification depends on the type of the selected mother wavelet [33]. We evaluated the decomposition capability of different mother wavelets with different decomposition levels based on the energy-to-Shannon entropy-based self-evaluation criterion given in [33]. Based on this evaluation, the mother wavelet was selected to be *Daubechies 3* (db3) with a decomposition level of 3, which provide the maximum energy-to-

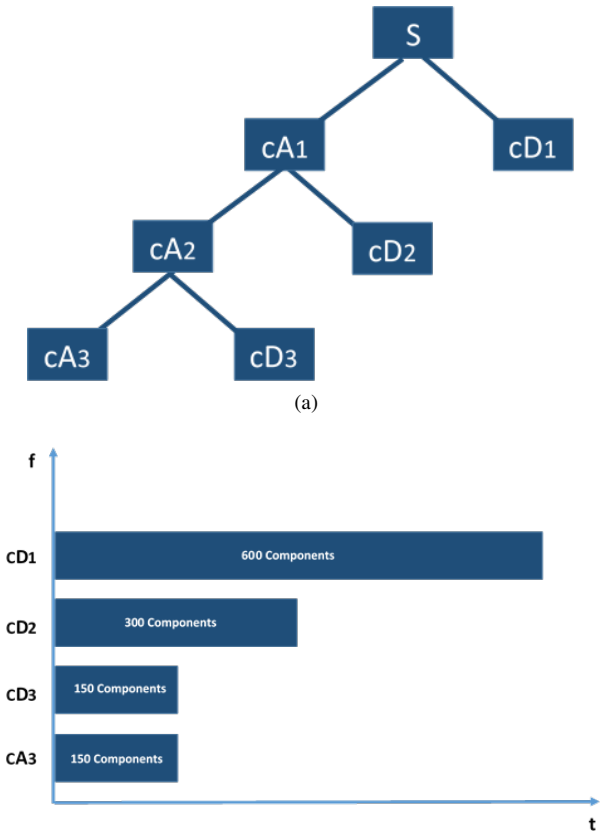


Fig. 5: (a) DWT Decomposition tree. (b) Layout of the resulting wavelet-transform vector. S is the input signal, cAn is the approximate coefficients of n^{th} level and cDn is detailed coefficients of n^{th} level.

Shannon entropy ratio for elephant rumble signal decomposition. The signal segment length is 2 seconds as mentioned previously; this includes 1200 samples, as the sample rate is 600 Hz.

Once DWT is applied to the signal segment with the decomposition level 3, signal segments are divided into three detail sub-bands cD1, cD2, and cD3. Furthermore it creates the cA3 approximation sub-band. Figure 5 (a) represents the DWT decomposition tree, while Figure 5 (b) illustrates the layout of the coefficients in the output vector. After obtaining the wavelet transform coefficients, the 7 reconstructed signal variation of the original signal were reconstructed by applying the inverse discrete wavelet transformation (IDWT) to individual wavelet coefficient sub-bands. The combinations of wavelet coefficients sub-bands are depicted in the Table I. Figure 6 represents the process of generating reconstructed signals. These reconstructed signal variations and the original signal were used for the final feature extraction process.

Since DWT decomposes the original input signal into different frequency sub-bands, reconstructed signals only contain the frequency components in the particular wavelet coefficients, which are used to reconstruct the signal. Extracting features from such reconstructed signal variations will allow extracting features from each frequency sub-band separately.

TABLE I: Description of reconstructed signals

A3	Reconstructed signal using approximation coefficient sub-bands at level 3 (cA3)
D3	Reconstructed signal using detailed coefficients sub-bands at level 3 (cD3)
D2	Reconstructed signal using detailed coefficients sub-bands at level 2 (cD2)
D1	Reconstructed signal using detailed coefficients sub-bands at level 1 (cD1)
all_recon	Reconstructed signal using cD1-cD3 detailed sub-bands and cA3 approximation sub-band
rm_cA3_cD1	Reconstructed signal using cD1 and cD2 detailed sub-bands.
rm_cA3	Reconstructed signal using cD1-cD3 detailed sub-bands

TABLE II: Feature categories

Chroma Features
Mel-Frequency Cepstral Coefficient
Root-Mean-Square (RMS) Energy
The spectral centroid
Spectral Contrast
Spectral bandwidth
Spectral-Roll-off
Zero Crossing Rate
Polynomial Features

That means noise with a particular frequency will only affect one or several of the frequency sub-bands and other sub-bands remain unaffected. Thus, combinations of features extracted from the different frequency sub-bands are more robust to noise than the features extracted from the entire frequency range.

Furthermore, DWT does not merely divide the given signal into frequency sub-bands. The iterative process of multi-resolution wavelet decomposition depends on the mother wavelet, which is used for decomposition. Mother wavelet db3 was empirically selected to match with the properties of elephant rumbles. Therefore, it has a higher tendency to pass sound waves more similar to elephant rumbles while filtering out waves significantly different than the elephant rumbles in each iteration of multi-resolution wavelet decomposition.

During the feature extraction process, 84 different spectral features under 9 main feature categories were extracted. Table II illustrates these main feature categories, while Table III depicts the composition of the extracted features. These feature composition were extracted separately from every reconstructed signal variation and the original signal. Therefore, finally it ended up with 672 (84×8) extracted features from a given signal segment. However, the entire feature vector will not be used for training and prediction purposes. The extracted feature vector is then passed on to the feature selection module to identify the features that are most effective.

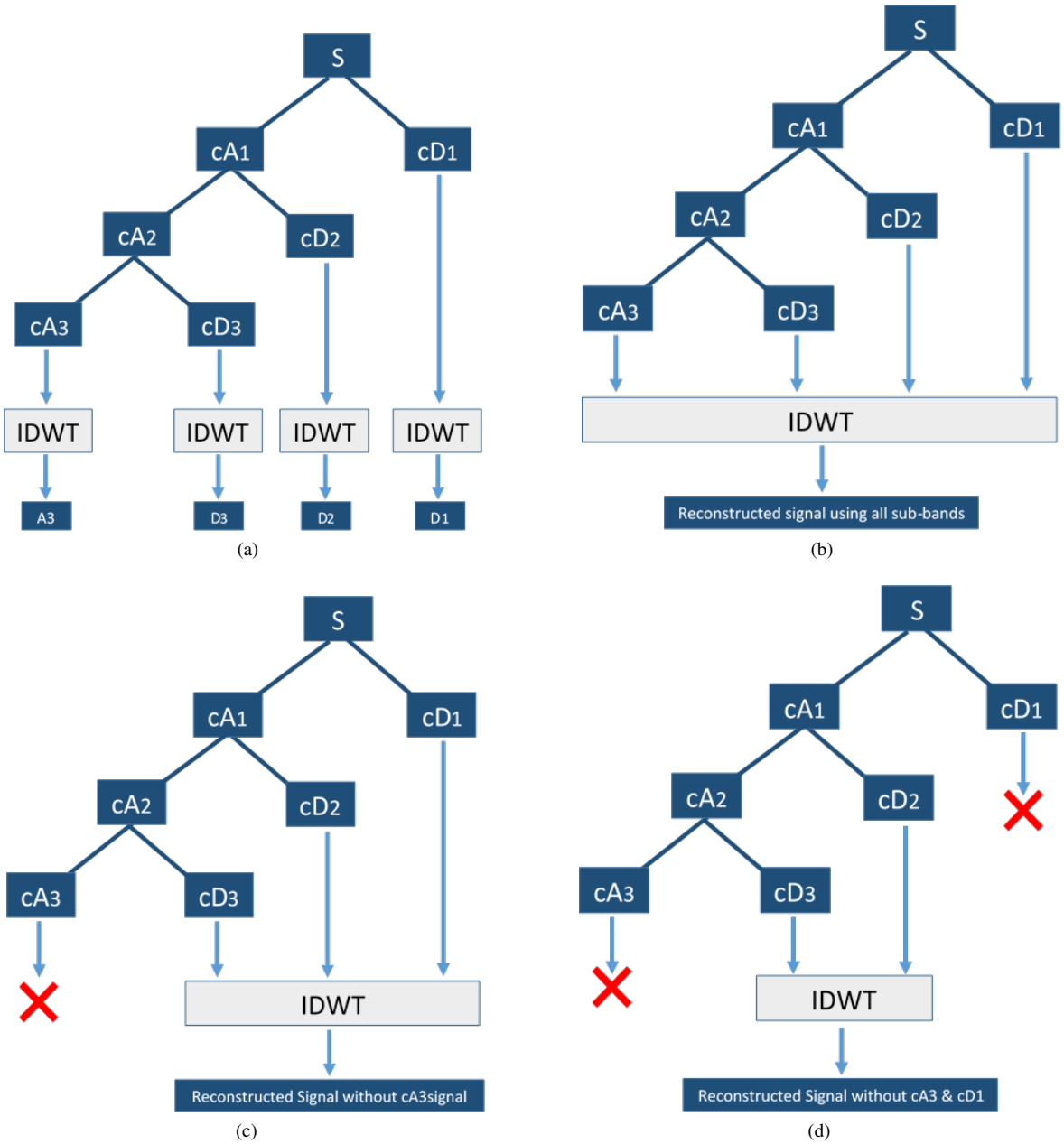


Fig. 6: Process of generating reconstructed signals.(a) Signal reconstruction using individual sub-bands, (b) Signal reconstruction using all sub-bands, (c) Signal reconstruction without cA3, and (d) Signal reconstruction without cA3 and cD1

D. Feature Selection Module

Feature selection module assists in creating an accurate predictive model. Proposed feature selection module consists of two sub-modules: quantile transformer and feature selector (see Figure 1). The output feature vector from the feature extraction module is taken as the input for the feature selection module.

1) Quantile Transformer: Many machine learning algorithms are designed with an assumption that all features vary on comparable scales while each feature takes values near zero. In contrast, features in the feature vector provided by the feature extraction module have different scales. Moreover, because of the nature of the input signals, it has a higher tendency to contain large number of outliers. This is because, it is impossible to capture the exact elephant rumble in the natural environment. These two characteristics can decrease the predictive performance of many machine learning algorithms. Also, unscaled data can slow down the training and prediction process of many machine learning algorithms.

The proposed approach uses quantile transformer [34] to bring the extracted features into a comparable scale while minimizing the effect of outliers. The quantile transformer transforms the features to follow a uniform distribution, using quantile information. This means, for a given feature, the quantile transformer tends to flatten out the most frequent value. The quantile transformer applies a nonlinear transformation to

TABLE III: Feature categories

Features	Number of values per frame
Chroma features	36
MFCC features	5
RMS Feature	1
Spectral centroid	1
Spectral bandwidth	1
Spectral contrast	6
Spectral roll-off	1
Zero Crossing Rate	1
Melspectrogram features	30
Polynomial features	2
Total number of features	84

TABLE IV: Supporting libraries.

LibROSA	python package for music and audio analysis
PyWavelets	Open Source wavelet transform software for the Python programming language
scikit-learn	Open source tools for data mining and data analysis
resampy	python module for efficient time-series resampling
pandas	Open source library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language

each feature independently by using the cumulative density function of a feature. Therefore, quantile transformer should be tuned with the features extracted from the training dataset.

2) Feature Selector: For classification with small training samples and high dimensionality, feature selection plays a vital role in avoiding over-fitting problems and improving classification performance. Feature selection will remove irrelevant, correlated, and redundant features and choose a robust subset of features that has the highest relevance to classification. Proposed feature selection module has been built based on the feature ranking with recursive feature elimination [35] and cross-validated selection of the optimal number of features.

Since this is a wrapper feature selection approach, the feature selection process depends on the given external classifier and the feature vector obtained from the training dataset with corresponding class labels. As SVMs have been widely used and demonstrated to be effective in detecting infrasonic elephant vocalisations [11], [12], [21], this work also uses an SVM as the classifier. The feature selection module is tuned based on recursive feature elimination and cross-validated with the SVM.

IV. Implementation and Experiment Setup

A. Implementation

The proposed feature extraction pipeline was implemented using Python 3.5 with the libraries shown in Table IV. Wider availability of the supporting libraries and the ability for the same implementation to run on both the Desktop computer (test environment), and the Eloc node (production environment) are the main reasons to select Python as the programming language. All of the features mentioned in Table II are extracted using the feature extraction methods in Python LibROSA [36] library. These feature extraction methods take the audio file, sample rate, frame length, and hop length (shift between frames) as the input arguments. However, some of the LibROSA functions require some degree of tuning based on the dataset and specific frequency range that is to be analysed. These feature extraction methods are configured to analyse the infrasonic as follows.

- **chroma_cqt and chroma_cens**
  These two methods require the minimum frequency to analyse and the number of octaves required to analyse this minimum frequency as parameters. Since elephant rumbles fluctuate around 14 Hz to 24 Hz (with harmonics from 14 Hz to 175 Hz), 10 Hz was selected as the minimum frequency with 4 octaves. It results in the analysis of a frequency range from 10 Hz to 160 Hz (minimum frequency 10 Hz, 1st octave at 20 Hz, 2nd octave at 40 Hz, 3rd octave at 80 Hz, and 4th octave at 160 Hz)
- **melspectrogram and mfcc**
  These two methods require the minimum and maximum frequency range that has to be analyzed. To match with the frequency range of elephant rumbles, 10 Hz was selected as the minimum frequency and 160 Hz was selected as the maximum frequency.
- **spectral_contrast**
  This method requires the frequency cut-off for the first frequency sub-band and the number of frequency bands as input. As the minimum frequency cut-off, 5 Hz was selected. Furthermore, six frequency sub-bands were considered, which result in the analysis of a frequency range from 0 Hz to 160 Hz. ([0 Hz - 5 Hz] [5 Hz - 10 Hz] [10 Hz - 20 Hz] [20 Hz - 40 Hz] [40 Hz - 80 Hz] [80 Hz - 160 Hz])

Audacity 2.2.1, an open source, cross-platform audio software for multi-track recording and editing, was used for the experiments and for analytical purposes during this study.

B. Dataset

To train and evaluate the proposed approach, this work uses a comprehensive elephant vocalisation dataset [23]. It contains 5592 different elephant sound recordings that were made in Sri Lanka, with 48 kHz sampling rate and 32-bit resolution; the dataset has been manually annotated by an expert in bioacoustics. There are 14 types of elephant calls in the data. We only use rumble recordings among these 14 call types. For negative samples, a set of unclassified sound clips from the same dataset were used; they are essentially the background noise in the environment where the dataset was originally captured.

Since the proposed approach targets the detection of elephant rumbles captured by the Eloc deployment unit, we constructed the synthesized dataset by replaying the above

elephant rumble recordings and negative dataset using a subwoofer. Earlier work has shown that subwoofers can replay elephant sounds that include fundamental frequency components in the infrasonic range with sufficient output power to emulate a real elephant [20]. For the experiments of this work, the Eloc node was placed at a fixed location at 10 metres away from the subwoofer to capture the replayed data.

Therefore, the two datasets used by this study is as follows:

- Dataset-1: Replayed elephant rumbles and negative dataset captured using Eloc node.
- Dataset-2: A collection of originally-recorded elephant rumbles and other non-elephant sounds captured in the same field has used as the positive and negative datasets. [23]

1) Training and testing datasets:

- Dataset-1 is divided into two parts for classifier evaluation and feature selection module tuning. (75 positive samples and 75 negative sample were used for the training while 25 positive and 25 negative samples used for testing.) – This dataset is not used for the classifier training.
- Dataset-2 is divided into two parts for classifier training and testing. (104 positive samples and 110 negative sample were used for the training while 75 positive and 100 negative samples used for testing)
- Each data is processed with a 1-second window with overlapping. (0.6-second shift between consecutive windows).

2) Testing dataset with noise: First, we normalized the testing samples in dataset-2 to -15dB; then noise samples are normalized to -10dB, -15dB, -25dB, -35dB, and -45dB. Finally, normalized noise clips and testing sample are mixed programmatically to obtain the desired level of SNR_{dB}; -5dB, 0dB, 10dB, 20dB and 30dB respectively. SNR_{db} is expressed in equation (2) where P is the average power, $P_{signal,dB} = 10\log_{10}(P_{signal})$ and $P_{noise,dB} = 10\log_{10}(P_{noise})$.

$$SNR_{dB} = P_{signal,dB} - P_{noise,dB} \tag{2}$$

C. Quantile Transformer Tunings

A combination of features extracted from both datasets have been used for the quantile transformer tuning process. (Feature selection module tuning portion from the dataset-1 and classifier training portion form the dataset-2)

D. Feature Selector Tuning

In contrast with the quantile transformer tuning process, here we only consider the features extracted from dataset-1 (replayed elephant rumbles and negative dataset captured using Eloc node). Because even though the proposed approach uses a very high-quality dataset (captured using domain-specific device) during the training process, it should be able to classify the sound signal captured by the Eloc node (which is not sensitive as the domain-specific device and signal can be slightly distorted because of low-cost hardware components).

The underlying assumption is, any feature subset which is robust enough to classify the replayed version of elephant rumbles captured by the ELOC node, should be able to classify the direct elephant rumble captured by Eloc node or any device more sensitive than the Eloc node. The validity of this assumption will be discussed in the evaluation section.

E. Classifier Training

The SVM classifier is trained using training portion of dataset-2. Since this dataset was recorded using a sample rate of 48 kHz, it is converted firstly to the 600 Hz sample rate. Then, all the steps described previously from preprocessing to feature selection are applied to this dataset excluding the beamforming-based stereo to mono conversion step because these recordings are already in the mono format. Finally, hyper parameters of the classifier have been tuned based on the cross-validation score with different parameter combinations.

V. Experimental Results and Discussion

A. Effect of Noise Reduction

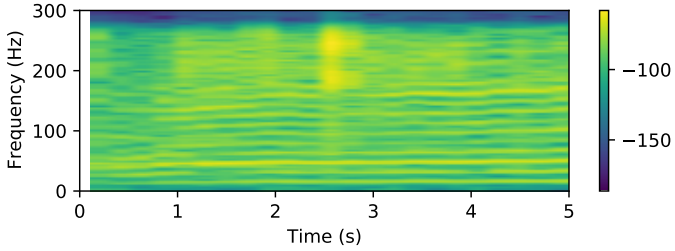


Fig. 7: Rumble recording before noise reduction.

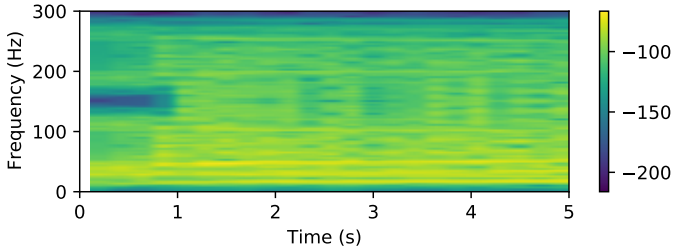


Fig. 8: Rumble recording after the noise reduction.

Figures 7 and 8 illustrate the spectrograms of a short segment of an elephant rumble recording before and after the denoising process, respectively. As it can be observed, sudden high-frequency and low-duration noise components have been reduced considerably. The noise removal process seems to affect the original rumble patterns (contours) to a certain degree as well by causing minor distortions. However, as shown in the later subsections, these distortions have not affected the expected rumble detection accuracy; this is because the machine learning models are trained on the noise-removed data.

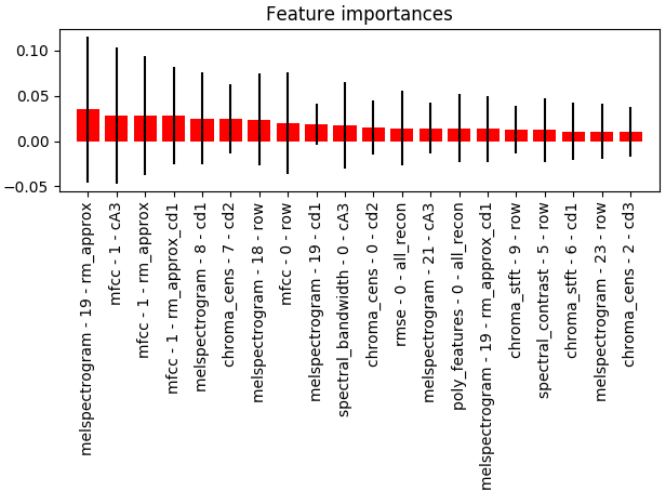


Fig. 9: The feature importance of top 20 features. The red bar represents feature important of the forest, along with their inter-tree variability.

B. Individual Feature Importance

To analyse the significance of individual features toward the classification, the importance of features using a forest of tree method was considered, which is provided in [34]. Figure 9 represents the feature importance of top 20 features. As can be seen, most of the features extracted from the reconstructed variations of a signal have a higher importance than the features extracted from the raw signal, proving the validity of extracting features on top of the DWT.

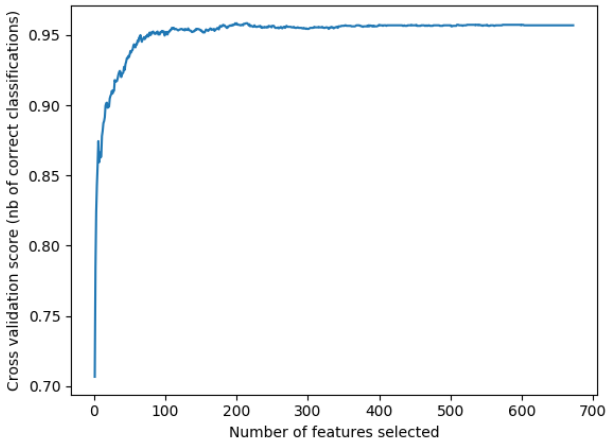
C. Feature Selection for the Classifier Training

As described earlier, optimal combinations of features are selected based on the recursive feature elimination score and cross-validation score. Figure 10 represents the cross-validation score variation with the number of features selected for features extracted from (a) dataset-1 and (b) dataset-2. According to the figure, it is clear that to classify the elephant rumbles captured by the domain-specific device (dataset-2), only 30 features are required. However, to classify the replayed version of elephant rumbles (dataset-1) 196 features are necessary. Low sensitivity and sound wave distortion due to the low-cost hardware can be assumed as the primary reasons for that.

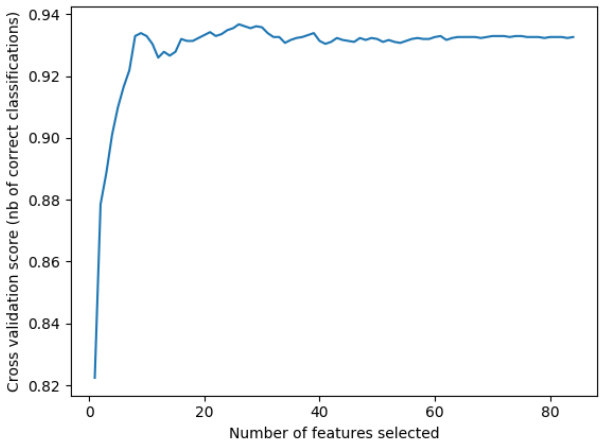
D. Classifier Evaluation

The accuracy of classification on both datasets was evaluated. Table V and VI represent the classification accuracy of dataset-1 and dataset-2 respectively. It is evident that, the proposed approach achieves 93% classification accuracy with dataset-2 and 82% classification accuracy with dataset-1. Thus, it could be stated that it performs well (82%) in detecting replayed versions of elephant rumbles.

As mentioned earlier, this work only considers the frequency range from 10 Hz to 150 Hz. Because of that, the approach depends on the fundamental infrasonic components and the



(a)



(b)

Fig. 10: Cross-validation score variation with the number of features selected for features extracted from (a) dataset-1 and (b) dataset-2

first few harmonics of elephant rumbles Furthermore, these results were achieved with the 600 Hz sampling rate for signal recording and processing, thus proving that the proposed approach is efficient enough to operate on top of the resource-limited hardware platform provided by Eloc nodes.

To evaluate the classification performance under the noisy conditions, experiments with 5 noise types were carried out under different SNR levels. Figures 11, 12, and 13 represent the detection accuracy under different noise types. The software code for the experimental evaluations of this work are available to be used freely as a Github repository [1].

VI. Conclusion and Further Work

This study presented a complete sound processing pipeline for infrasonic elephant rumble detection under noisy natural

[1]https://github.com/vinuri-s/A-Robust-Feature-Extraction-Pipeline-for-Detecting-Elephant-Rumbles

TABLE V: Prediction performance with dataset-1

	Precision	Recall	F1-score
Negative	0.79	0.94	0.84
Positive	0.92	0.69	0.79
Avg/Total	0.84	0.84	0.82

TABLE VI: Prediction performance with dataset-2

	Precision	Recall	F1-score
Negative	0.91	0.99	0.95
Positive	0.99	0.84	0.90
Avg/Total	0.94	0.93	0.93

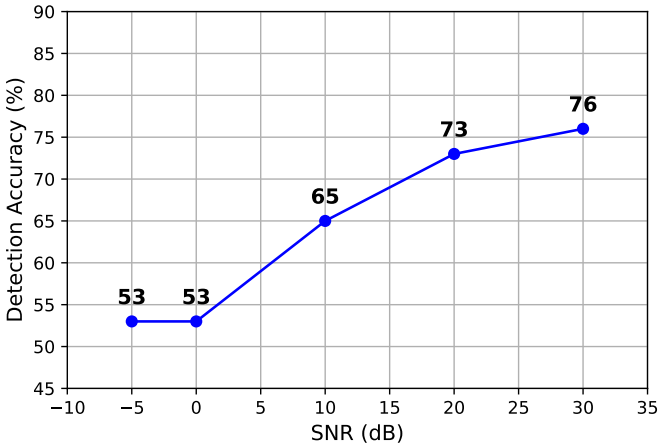


Fig. 11: Detection accuracy with white noise under different SNR

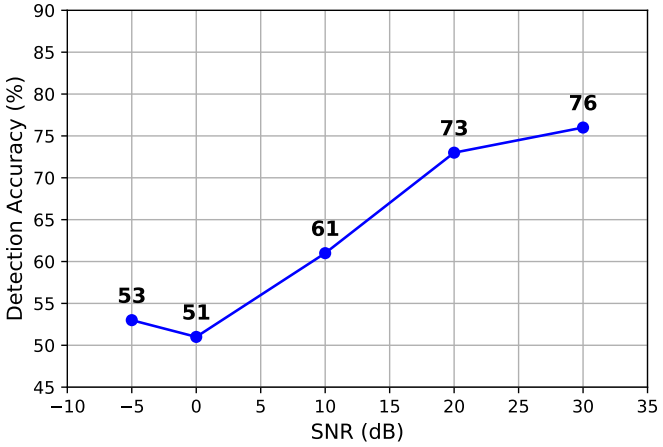


Fig. 12: Detection accuracy with pink noise under different SNR

environments, using the resource-limited and low-cost Eloc hardware platform. According to the evaluation results, it is evident that elephant infrasonic calls can be accurately detected on low-resourced hardware. This study also contributed to the domain of digital signal processing by exploring the wavelet-based feature extraction for processing and automatic

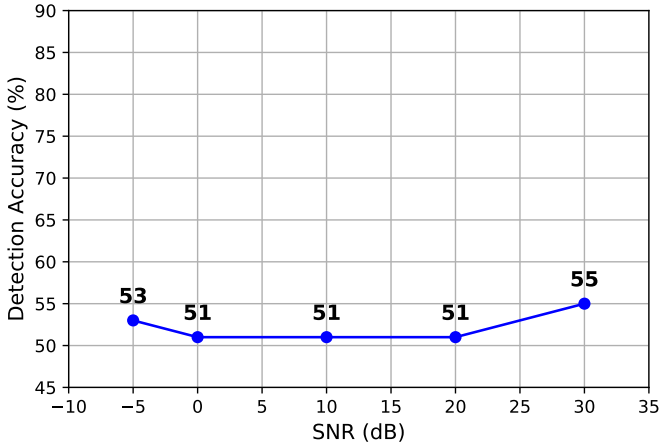


Fig. 13: Detection accuracy with petrol engine noise under different SNR

detection of infrasonic data.

This approach has been tested on the detection of replayed versions of elephant rumbles under the influence of artificially generated noise. Furthermore, the proposed approach only attempts to identify whether the dominant infrasonic signal of a given time period was emitted by an elephant or not. Other strong infrasonic sources with similar frequency patterns in the vicinity may have an adversarial effects on the rumble detection process. Therefore, comprehensive testing and tuning of the proposed method in natural environments is fundamentally required in the future.

For the experiments of this work, the replayed dataset was produced by placing an Eloc node at a fixed location to record data played by a subwoofer. The distance between the Eloc node and the infrasonic source, i.e., subwoofer, is an important parameter to be considered when evaluating the effectiveness of the proposed sound-processing pipeline. In the future, this aspect will be evaluated in detail.

Elephants generate a variety of vocalisations, from the infrasonic range to the human audible range. However, the proposed approach of this work only considers a single vocalisation type, i.e., rumbles, for the detection of elephants. This was due to that infrasonic travels farther distances in contrast to higher-frequency sounds. But, it would be worthwhile in the future to explore the potential of utilising and combining other types of elephants vocalisations to enhance the capability of elephant detection.

ACKNOWLEDGMENT

The authors would like to thank Dr. Shermin de Silva for her generous contribution of elephant infrasonic dataset used in this work.

This research was partially supported by grants from the Swedish Research Council, Sweden and the Rufford Foundation (grant number: 35179-1), UK.

REFERENCES

[1] Endangered Species Journalist, "Elephant: an Endangered Species," bagheera.com, Accessed: 2017-12-29. [Online]. Available: http://www.bagheera.com/inthewild-van_anim_elephant.htm

[2] World Wildlife Fund, "Issues: Human-elephant conflict - WWF," http://wwf.panda.org, Accessed: 2017-12-29. [Online]. Available: http://wwf.panda.org/what_we_do/endangered_species/elephants/asian_elephants/areas/issues/elephant_human_conflict

[3] T. Prakash, A. Wijeratne, and P. Fernando, "Human-elephant conflict in sri lanka: patterns and extent," *Gajah Journal of Asian Elephant Specialist Group*, vol. 51, pp. 16–25, 2020. [Online]. Available: https://www.asesg.org/PDFfiles/2020/51-16-Prakash.pdf

[4] A. Campos-Arceiz, S. Takatsuki, S. K. K. Ekanayaka, and T. Hasegawa, "The human-elephant conflict in southeastern Sri Lanka: type of damage, seasonal patterns, and sexual differences in the raiding behaviour of elephants," *Gajah*, vol. 31, pp. 5–14, 2009.

[5] "CCRSL.ORG | Tracking Elephants." [Online]. Available: http://ccrsl.gadola.com/Programs/tracking-elephants

[6] J. D. WOOD, B. MCCOWAN, W. R. L. JR., J. J. VILJOEN, and L. A. HART, "Classification of african elephant loxodonta africana rumbles using acoustic parameters and cluster analysis," *Bioacoustics*, vol. 15, no. 2, pp. 143–161, 2005. [Online]. Available: https://doi.org/10.1080/09524622.2005.9753544

[7] S. Nair, R. Balakrishnan, C. S. Seelamantula, and R. Sukumar, "Vocalizations of wild asian elephants (elephas maximus): Structural classification and social context," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2768–2778, 2009.

[8] J. Prince, "Surveillance and tracking of elephants using vocal spectral information," vol. 03, pp. 664–671, 05 2014.

[9] L. Seneviratne and G. Rossel, "Elephant Infrasound Calls as a Method for Electronic Elephant Detection." *Channels*, pp. 1–7, 2004.

[10] K. Marten and P. Marler, "Sound transmission and its significance for animal vocalization - I. Temperate habitats," *Behavioral Ecology and Sociobiology*, vol. 2, no. 3, pp. 271–290, sep 1977.

[11] M. Zeppelzauer, S. Hensman, and A. S. Stoeger, "Towards an automated acoustic detection system for free-ranging elephants," *Bioacoustics*, vol. 24, no. 1, pp. 13–29, jan 2014.

[12] A. Sayakkara, N. Jayasuriya, T. Ranathunga, C. Suduwella, N. Vithanage, C. Keppitiyagama, K. De Zoysa, K. Hewage, and T. Voigt, "Eloc: Locating Wild Elephants using Low-cost Infrasonic Detectors," in *13th International Conference on Distributed Computing in Sensor Systems*, 2017.

[13] K. B. Payne, M. Thompson, and L. Kramer, "Elephant calling patterns as indicators of group size and composition: The basis for an acoustic monitoring system," *African Journal of Ecology*, vol. 41, no. 1, pp. 99–107, 2003.

[14] C. Dissanayake, R. Kotagiri, M. Halgamuge, B. Moran, and P. Farrell, "Propagation constraints in elephant localization using an acoustic sensor network," in *ICIAFS 2012 - Proceedings: 2012 IEEE 6th International Conference on Information and Automation for Sustainability*. IEEE, sep 2012, pp. 101–105.

[15] M. Zeppelzauer, A. S. Stöger, and C. Breiteneder, "Acoustic detection of elephant presence in noisy environments," in *Proceedings of the 2nd ACM international workshop on Multimedia analysis for ecological data - MAED '13*. New York, New York, USA: ACM Press, 2013, pp. 3–8.

[16] M. M, M. Devaki, and R. Scholar, "Elephant Localization And Analysis of Signal Direction Receiving in Base Station Using Acoustic Sensor Network," *International Journal of Innovative Research in Computer and Communication Engineering (An ISO Certified Organization)*, vol. 3297, no. 2, 2007.

[17] P. J. Venter and J. J. Hanekom, "Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings," *Biosystems Engineering*, vol. 106, no. 3, pp. 286–294, jul 2010.

[20] P. Dabare, C. Suduwella, A. Sayakkara, D. Sandaruwan, C. Keppitiyagama, K. De Zoysa, K. Hewage, and T. Voigt, "Listening to the Giants," in *Proceedings of the 6th ACM Workshop on Real World*

[18] A. S. Mohapatra and S. S. Solanki, "An automatic method to detect the presence of elephant," in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, May 2014, pp. 1515–1518.

[19] J. Bjorck, B. H. Rappazzo, D. Chen, R. Bernstein, P. H. Wrege, and C. P. Gomes, "Automatic Detection and Compression for Passive Acoustic Monitoring of the African Forest Elephant," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 476–484, Jul. 2019. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/3820
Wireless Sensor Networks - RealWSN '15. ACM Press, 2015, pp. 23–26.

[21] N. Jayasuriya, T. Ranathunga, K. Gunawardana, and C. Silva, "Poster Abstract : Resource-Efficient Detection of Elephant Rumbles," pp. 17–19, 2017.

[22] D. Stowell, "Computational bioacoustics with deep learning: A review and roadmap," *PeerJ*, vol. 10, p. e13152, 2022.

[23] S. de Silva, "Acoustic communication in the asian elephant, elephas maximus maximus," *Behaviour*, pp. 825–852, 2010. [Online]. Available: https://www.jstor.org/stable/27822154

[24] G. Bianchi and R. Sorrentino, *Electronic filter simulation & design*. McGraw-Hill, 2007.

[25] "Neural classification of lung sounds using wavelet coefficients," *Computers in Biology and Medicine*, vol. 34, no. 6, pp. 523–537, sep 2004.

[26] J. S. Lim and A. V. Oppenheim, "Advanced topics in signal processing," *Advanced topics in signal processing*, pp. 289 – 337, 1987.

[27] "Wavelets and Signal Processing," *IEEE Signal Processing Magazine*, vol. 8, no. 4, pp. 14–38, 1991.

[28] A. E. Villanueva-Luna, A. Jaramillo-nuñez, D. Sanchez-lucero, C. M. Ortiz-lima, J. G. Aguilar-soto, A. Flores-gil, and M. May-alarcon, "De-Noising Audio Signals Using MATLAB Wavelets Toolbox," in *Engineering Education and Research Using MATLAB*, 2011, pp. 25–54.

[29] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13 475–13 481, sep 2011.

[30] S. Legendre, D. Massicotte, J. Goyette, and T. K. Bose, "Wavelet-transform-based method of analysis for lamb-wave ultrasonic NDE signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 3, pp. 524–530, jun 2000.

[31] O. Rioul and M. Vetterli, "Wavelets and Signal Processing," *IEEE Signal Processing Magazine*, vol. 8, no. 4, pp. 14–38, oct 1991.

[32] C. Heil and D. Walnut, "Continuous and discrete wavelet transforms*," *Society for Industrial and Applied Mathematics*, vol. 31, no. 4, pp. 628–666, dec 1989.

[33] H. He, Y. Tan, and Y. Wang, "Optimal base wavelet selection for ECG noise reduction using a comprehensive entropy criterion," *Entropy*, vol. 17, no. 9, pp. 6093–6109, sep 2015.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2012.

[35] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research (JMLR)*, vol. 3, no. 3, pp. 1157–1182, 2003.

[36] B. Mcfee, C. Raffel, D. Liang, D. P. W. Ellis, M. Mcvicar, E. Battenberg, and O. Nieto, "librosa: Audio and Music Signal Analysis in Python," *PROC. OF THE 14th PYTHON IN SCIENCE CONF*, no. Scipy, pp. 1–7, 2015.