Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-021-03141-9

ORIGINAL RESEARCH

The IoT based embedded system for the detection and discrimination of animals to avoid human–wildlife conflict

Surya Thangavel[1] · Chitra Selvi Shokkalingam[1]

Received: 21 November 2020 / Accepted: 11 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
Human–wildlife conflict (HWC) is one of the major crises in the valparai region of Anamalai Tiger Reserves (ATR). It is essential to reduce the HWC to save people from the wildlife and also to protect wildlife. In this paper, we propose an automated unsupervised system for the identification and classification of animals from their acoustic signal. The environment sound signals are captured using a microphone and the audio is stored in a.wav file and is sent to a base station through a radio frequency (RF) network. This system is processed with three steps (i) from the received audio signal initially, animal identification is done by extracting features of an animal signal using Mel frequency cepstral coefficient (MFCC) and classification of animal is performed by radial basis function (RBF) neural network, (ii) age estimation (calf/adult) is performed by autocorrelation, (iii) elephant state of mind (SOM) is detected by extracting features of an elephant acoustic signal using gammatone frequency cepstral coefficient (GFCC) and classification of various sounds of elephant are performed by support vector machine (SVM). Based on this, an early warning message which contains an animal type, age (calf/adult), elephant SOM, global positioning system (GPS) tracks its location information and all this information will be transmitted via global system for mobile communication (GSM) to the forest authorities, local communities, radio station or local channels indicating that an animal movement is near to forest border areas. The results were fed into a separate web page using the internet of thing (IoT).

Keyword Animal acoustic monitoring · Animal recognition · Machine learning techniques · Internet of things (IoT) · Early warning and monitoring module

1 Introduction

There have been very few interrogations so far on animal acoustics, which is part of the sound of the environment. Generally speaking, animals used to produce sounds for interacting with other animals and for their living activities like movement, combat with other animals, feeding, etc. Animal acoustic monitoring is very effective for real-time animal detection under various climatic conditions. The applied animal voice gives useful information (Luque et al. 2016). Sound travels for variable distances depending on frequency, sound source location, background noise, sound pressure level, and also varying climatic conditions (Darras et al. 2016).

In addition to recognizing animals, acoustic monitoring functions also extract complete information such as discriminating age groups, genders, social groups or kin groups, etc. High-dimensional features representing the entire spectral envelope, such as the Mel-frequency cepstral coefficient (MFCC) give 75% correct classification and 74% correct classification of the Greenwood frequency cepstral coefficient (GFCC) function (Stoeger et al. 2014). In the existing work, the authors (Darras et al. 2016; Stoeger et al. 2014; Bjorck et al. 2019; Lee et al. 2006; Zeppelzauer and Stoeger 2015) tried, animal recognition, age estimation with single species mainly processed with elephant audio samples and monitored elephant but in our work, we have taken audio samples of multiple species and processed further.

In this proposed framework, we took audio samples of Elephant, Leopard, and Bear. These audio samples are stored in an external read-only memory (ROM) connected to the

✉ Surya Thangavel
suryasri.tm@gmail.com

✉ Chitra Selvi Shokkalingam
chitraselvieee@gmail.com

[1] Anna University-University College of Engineering, Thirukkuvalai, Nagapattinam, India

Published online: 29 March 2021

ARM processor. Based on these audio samples we have demonstrated animal identification, age estimation, elephant state of mind (SOM) detection with log date, time, and location information (latitude and longitude), and this information will be updated automatically into a separate web page using internet of things (IoT), where the forest officials can utilize the information.

In this paper, an animal detection and discrimination system are developed. The following are the major objectives of this work.

1. To discriminate animals from their acoustic signals for animal recognition. Based on animal recognition, our proposed system also identifies whether it is a single or group of elephants. In addition to this, age is also estimated for each animal.
2. Moreover, an elephant state of mind (SOM) is also detected for a single adult elephant to identify whether it is stressed or relaxed.
3. The classification accuracy level achieved for animal recognition is 98 and 97% for elephant SOM detection.

The purpose of this system of recognition and discrimination of animal vocalization which helps to sort out specific animals to find out which animal produces sound. Different animals varying with different vocal frequencies and thus may achieve a higher detection accuracy of specific animals. This system is very useful for security purposes in national parks and wildlife sanctuaries. Animal identification with its acoustic is beneficial for environmental monitoring and biological study, especially in detecting and locating animal movements. When dealing with animal acoustic the fusion of the noise model is crucial to abundant noise interference present in the environmental sound. This interfering noise immensely decreases the classification accuracy, specifically if the noise characteristics vary across the dataset.

1.1 Study area

Human–wildlife conflict is high in estate areas such as Valparai, Manambolly, Ulandy, and Pollachi ranges of Anamalai Tiger Reserves (ATR). As the secondary data collected from the forest department shows that the conflict is more in these four ranges. Almost eight species were involved in the conflict with local people, out of which the top three species are Elephant, Leopard, and Bear. So, we have taken these three animal species as the target animal for our proposed work. In the study area, nearly 48 human deaths occurred because of various conflict issues with wild animals from 2007 to 2019. It is necessary to control HWC to save people and wildlife. Early detection of animal roaming near human settlements can help in reducing conflict.

1.2 Data collection

The animal sounds were collected from the forest department and also manually downloaded from various websites and we have done initial filtering and transformation to make sure that all the audios are in a similar format. Many web sources have permitted us to download the files in.mp3 format and then we have used the audacity software to convert.mp3 files to.wav file format of 16 bits per second (bps) with the sampling rate of 44.1 kHz. The audio signals are stored temporarily in a read-only memory (ROM). We can also retrieve these audio signals to store permanently using an external ROM or in a personal computer (PC). In this research work, we have considered Elephant, Leopard, and Bear audio samples for animal detection and discrimination.

2 Literature review

Emotional state from voices due to various factors such as (i) the first factor shows the way of pronunciation of the word, (ii) the second factor carries information related to the emotion of the speaker and (ii) the third factor contains basic features and the identities of the speaker like gender, age, and body size (Borden et al. 1994). A comprehensive analysis in the domain of speech emotion recognition (SER) from 2007 to 2017 is provided in Mustafa et al. (2018) and they have suggested the aspects of SER which gives a combination of databases, speech features, and classifier for classification of emotion which influence the recognition accuracy of speech emotion recognition (SER) system. Table 1 summarizes the literature survey of our proposed work.

Kaya et al. (2017) proposed recognition of children's emotional state, via speech and by using three emotional states such as comfort/discomfort and neutral. Recognition of child's age and gender by listeners. Kuchibhotla et al. (2016) proposed the two-stage feature selection method for speech emotion recognition using regularized discriminant analysis (RDA) and support vector machine (SVM), sequential floating forward selection (SFFS) among these classifier SFFS gives the best emotion recognition using acoustic features (Ogawa and Hori 2017) applied deep bidirectional recurrent neural networks (DBRNNs) for error detection and estimating accuracy in automatic speech recognition (ASR). Anni and Sangaiah (2018) proposed boundary intellect (BI) system for the detection of an elephant using fuzzy cognitive maps and cognitive theory.

Table 1 Summary of literature survey

Name of the author	Year	Journal name	Mertis	Achieved accuracy
Angela S. Stoeger et al.	2012	Bioacoustics	✓ The author proposed a novel sound visualization technique to investigate whether rumbles are produced using the mouth and/or trunk in these specific contexts ✓ Used a smoothed spectral representation based on linear predictive coding (LPC) for both rumble variants and machine learning	✓ The classification system yielded an accuracy of 75%
D. Gutierrez-Galan et al.	2017	Neurocomputing	Animal behavior recognition, classification, and monitoring system based on a wireless sensor network	✓ Achieved accuracy 82.11%
Johan Bjork et al.	2019	Cornell University-Computer science (Sound)	The author used machine learning techniques to the analysis and compress audio signals in the context of monitoring elephants in sub-Saharan Africa	✓ The classification accuracy on the test set for the segmentation task ✓ ✓ Average accuracy for LSTM = 69.74% Conv.LSTM = 91.71%
Kharibam Jilenkumari Devi and Khelchandra Thongam Lee, Chang-Hsing, et al.	2019 2006	Journal of Ambient Intelligence and Humanized Computing Pattern Recognition Letter	Proposed a new automatic speaker recognition system for speech signals Proposed automatic identification of animals from their sound using averaged Mel-frequency cepstral coefficients (MFCC) and linear discriminant analysis (LDA)	✓ Achieved 95.96% Accuracy ✓ Classification accuracy of 99% with LDA ✓ Classification accuracy of two further classifiers (support vector machine and nearest neighbour classifier). Both classifiers yielded accuracies above 97% similar to LDA
Patrick J. Clemins and Michael T. Johnson	2003	IEEE International Conference on Acoustics, Speech and Signal Processing	Proposed automatic identification of animals from their sound using averaged Mel-frequency cepstral coefficients (MFCC) and hidden Markov model (HMM)	✓ The accuracy achieved for African elephant vocalization type classification is 83.8% ✓ Speaker identification accuracy 88.1% ✓ The average classification accuracy is 96.8%
Slavomir Matuska et al.	2014	AASRI Procedia	Proposed a novel method for object recognition based on hybrid local descriptors	✓ Success rate 86%
Susannah J. Buchan et al.	2019	Bioacoustics	Proposed an automated detection and discrimination of blue whale vocalizations by passive acoustic monitoring with the help of Hidden Markov Model and executed in the machine learning platform, Kaldi speech processing toolkit	✓ Achieved 85.3% accuracy for detection and classification
T. Thomas Leonid, r. Jayaparvathy	2020	Journal of Ambient Intelligence and Humanized Computing	Feature extraction methods with the principal component analysis (PCA) reduces the feature data dimension to improve the performance of the system. support vector machine (SVM) classifier is used as the predictive model	✓ Classification of animal audio achieves an accuracy of 98% ✓ It is observed that the proposed approach achieved an average accuracy of 93.32%
Yongming Huang et al.	2017	Journal of Ambient Intelligence and Humanized Computing	Proposed a robust speech emotion recognition using sub-band spectral centroid weighted packet cepstral coefficients (W-WPCC)	✓ Achieved classification accuracy of 86.60% using the feature fusion method

Table 1 (continued)

Name of the author	Year	Journal name	Mertis	Achieved accuracy
Zeppelzauer and Angela s. Stoeger	2015	BMC Res Notes	Provide an integrated overview for a future early warning and monitoring system for humans who regularly experience serious conflict with elephants	✓ The proposed method without signal enhancement achieves an 88.2% detection rate and 24.4% false-positive rate ✓ The proposed method with signal enhancement achieves an 88.2% detection rate and 13.7% false-positive rate

2.1 Various measures utilized for animal detection

Anni and Sangaiah (2015) and Prabu (2016) used a seismic sensor to identify the vibration caused by the elephant and the results were shown in graph (seismogram). A seismogram can show different sized waves for various animals. Elephants can be easily detected because of their weight and size. If it is detected, it alerts the forest officials. Poshitha et al. (2015), Zeppelzauer et al. (2015) proposed an early detection and warning of an elephant using cell phones for faster communication between farmers, local forest officials to assist in cooperation in driving the problematic elephants away from their location. Likewise, satellite tracking of an elephant with radio-collared assists in early warning of problematic elephants and herds (Venkataraman et al. 2005). It responds aggressively and spoils the radio collar tags. Wood et al. (2005) used a geophone sensor to monitor the footfalls of elephants and with their spectral content of footfall, the author discriminated among species with an accuracy rate of 82%.

Some techniques which are discussed above will not be suitable for all the climatic conditions. In our proposed work, we used the acoustic signal of animals for detection and discrimination which results in greater accuracy even in varying climatic conditions.

3 System architecture of animal detection and discrimination using vocal spectral content of animals

In this section, Fig. 1 demonstrates the system architecture for animal detection and discrimination. In this architecture, five different components namely Camera (SPRIS night vision infrared camera), Wi-fi device, Base station, switch and cloud server, global system for mobile communication (GSM) and global positioning system (GPS) are used for animal detection and discrimination. Among the five components, the first component is the camera: some camera has audio recording options inbuilt or else we can use a microphone to record the audio signal with a range of about 6 m (with clear voice) up to 50 m and it can also record sound produced at above 50 m only if the intensity of sound is high. Elephants can be easily detected because they can produce enormous sound with the fundamental frequency range (Roar—305–6150 Hz; Rumble—10–173 Hz; Trumpet—405–5879 Hz), which can travel to a distance of several kilometers. Elephant sound can be recorded by the device which is far away from the recording device. There is a high probability of elephant detection even it is far away from the recording device. The received environmental signal forwards it to the nearby Wi-fi device to which it is connected.

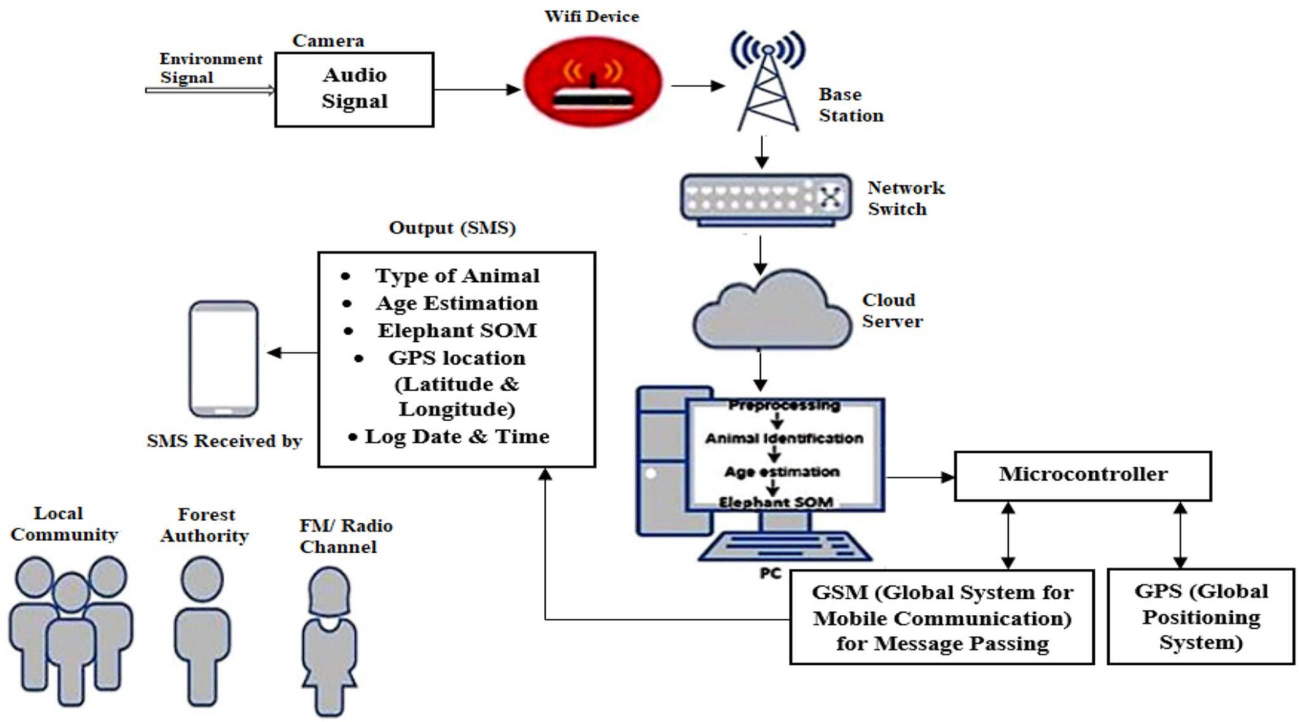


Fig. 1 System overview of the proposed work

Each Wi-fi device forwards the received audio signal to the remotely available cloud server through a base station. The cloud servers are built, hosted, and dispatched through a cloud computing environment and can be remotely accessed through an online interface (Internet). It is mainly used for storing processed information in the cloud. The cloud server will be connected with a PC in the range office. After receiving the environmental input signal initially pre-processing will be done for eliminating the unvoiced signal, disruption signals, and high amplitude signals. The next step is to obtain the various features such as a vocal tract pattern, pitch, and power from the received continuous input audio signal which helps in identifying features of the target animal by using feature extraction techniques.

The classification process is involved in discriminating the animal with its acoustic signal by implementing the classification algorithm. Once, an animal has identified an intimate message in form of a short message service (SMS) will be sent with the help of GSM (as shown in Fig. 9a) which contains the type of animal, its age, and elephant SOM information with log date, time, and location information (Latitude and Longitude) using GPS will be sent to the Forest authorities, local communities, and radio station/local TV channels and also all this information received will be placed in a single location through the internet of things (IoT) (as shown in Fig. 9b).

Internet of things (IoT) can frame a network with the physical things or objects that make use of embedded technologies to team up with other objects or things, sense, and cooperate. IoT based applications need fault tolerance, scalability and which is very tough to implement in computing environments and centralized system. Whereas it is possible in distributing system which involves (i) fault tolerance in computing (ii) communication between computational entities via message passing etc. Hence, the distributed system is used in our application where data is processed in one physical location and needed by another location.

The distributed system is more reliable than a centralized system. Because in the centralized system if any failure occurs in the database the functionality of the system comes to halt. Whereas in the distributed if it faces any failure in the database, it continues functioning the system with reduced performance. Hence, we have implemented distributed system in our proposed framework. This distributed system architecture has a collection of interconnected databases, this can spread across different locations that can communicate through a computer network.

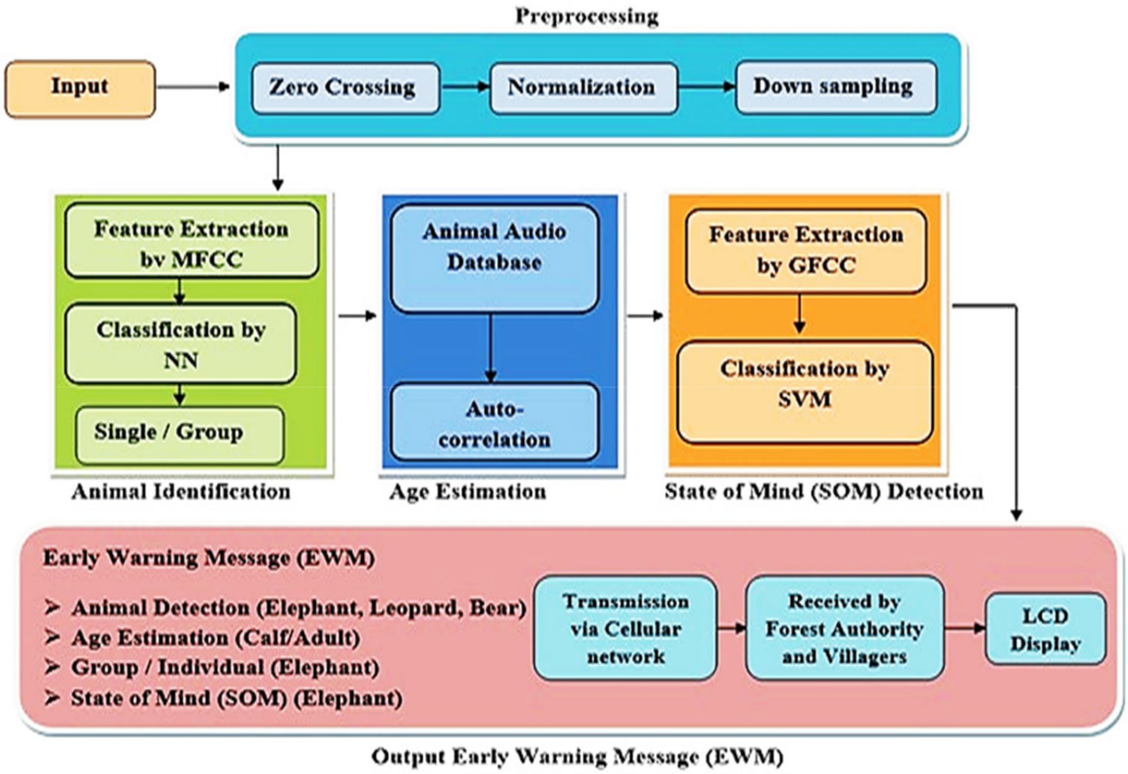


Fig. 2 Block diagram for an animal detection and discrimination system

4 Block diagram for animal detection and discrimination using its vocal spectral content

In this section, a typical block diagram for animal detection and discrimination is depicted in Fig. 2. The environmental sound signal is procured with a distortion-free audio recorded camera. Based on this sound signal, the animal identification process, Age estimation, and SOM detection is performed. Elephant, Leopard, and Bear, create more conflict in the study area so we have chosen these three animals as target species for detection. The major process involved in this work is animal identification, elephant SOM, and age estimation. Initially, pre-processing is done to remove unwanted noise present in the received audio signal. After pre-processing signal to noise ration (SNR) is calculated to estimate the quality of the audio signal.

After receiving the pre-processed audio signal Mel frequency cepstral coefficient (MFCC) Feature extraction process and radial basis function (RBF) neural network (RBFNN) based classification model is performed for animal identification. After identifying animals, it will also classify whether it is a single or group of elephants in case the detected species is an elephant. Because only adult male roam single whereas adult female elephant roams with its herds. After identifying the animal, auto-correlation is done for performing age estimation for all three animals. If the animal detected is an elephant, SOM is also determined to identify whether it is stressed or relaxed for that we used the gammatone frequency cepstral coefficient (GFCC) feature extraction technique and support vector machine (SVM) classification algorithm. If the received acoustic signal does not match with any of these three animals (Elephant, Leopard, and Bear) in the animal identification phase, the audio signal will be dropped in the initial classification stage of the animal identification process.

The following are the reasons for implementing two different feature extraction and classification algorithm for animal identification, age estimation, and elephant state of mind (SOM) detection.

(i) Target classes are different for this application like animal detection, elephant state of mind (SOM), and age estimation.
(ii) We require more dominant features and if we use them all together in one classifier overfit problem occurs and the possibilities for the wrong estimation are more.
(iii) The network will get more complex and the selection of the target class become more confusing to

the network. If we use different/separate classifiers for each application we can improve the performance and accuracy.

(iv) Also, we can add more target classes in the future. Hence, we implemented two different feature extraction and classification algorithms applied for animal detection and elephant SOM detection.

4.1 Pre-processing

In the pre-processing stage, it removes unnecessary data such as an unvoiced signal or high amplitude signals, and the result is passed to the Feature extraction process. Three processes were involved in pre-processing stage: (i) zero-crossing rate is applied here for classifying the voiced/unvoiced signal. (ii) Normalization is done for adjusting the amplitude values at the peak point of the input audio signal to the target value. (iii) Down sampling or subsampling is the process of diminishing the sampling rate of the input audio signal.

4.1.1 Signal to noise ratio (SNR)

- Signal to noise ratio is calculated to verify the quality of the audio signal after pre-processing.
- SNR is estimated to determine the signal strength and to measure the quantity of background noise (Gragido et al. 2013) present in the input audio signal and to find the true signal or extracting the useful information from the input signal. If the quality of the audio signal is poor, then it cannot be processed further and it will be dropped.
- SNR is defined as the ratio of the signal power to the background noise power.

$$SNR = \left(\frac{P_{Signal}}{P_{Noise}}\right) \tag{1}$$

$$SNR_{dB} = 10log_{10}(SNR) \tag{2}$$

$$P_{Signal} = Power\ of\ a\ signal$$

$$P_{Noise} = Power\ of\ background\ noise$$

4.2 Feature extraction using Mel frequency cepstral coefficient (MFCC)

Differences in animals can be easily varied by their acoustic. Feature extraction involves analyzing the audio signal to recognize the animal acoustic. Here we have employed MFCC feature extraction because it has a good feature vector and gives a better response compared to other feature extraction algorithms like linear predictive coding (LPC), principal component analysis (PCA), to process the speech or signal. Because of the linear computation nature, LPC parameters will not be suitable for this work. MFCC is a common and effective technique for feature extraction in signal processing. MFCC has 39 features and this feature count helps to learn more information present in the audio signal.

Transforming the input data into several sets of features is known as feature extraction. Feature extraction deals with simplifying the number of resources needed to describe a huge set of data accurately. In the feature extraction process, the appropriate information is extracted from the input data to execute the task with reduced representation rather than using the entire input data (Boussaid and Hassine 2018). MFCC extracts the features from the input audio signal and converts it into a certain series of feature vector coefficients. These features include only the information needed to recognize the received audio signal. The process of this algorithm is used to extract the optimal number of MFCC features without affecting the recognition accuracy of the system. MFCC coefficients are extracted by considering two voice samples per animal. Among the two voice samples, one is stored as a template in the database and the other one is given as the real-time input.

Algorithm 1: Pseudocode for MFCC feature extraction
Input: Pre-processed animal audio sounds
Output: MFCC Feature Vector
Step 1: Initialize parameters
Step 2: Animal audio signals are split into several frames;
Step 3: Applying the Hamming window for each frame to avoid the discontinuity of the signal;
Step 4: Fast Fourier Transform (FFT) is applied to each frame to obtain the spectrum;
Step 5: Compute matrix for Mel-filter bank;
Step 6: Spectrums are transformed into Mel-Spectrum;
Step 7: Discrete Cosine Transform (DCT) is applied to obtain an MFCC feature vector for each frame;
end function.

Step 1: Pre-emphasis improves the energy of the input signal at a higher frequency (Yildiz and Arslan 2018). Pre-emphasis is applied to a signal x in the first-order filter is shown in Eq. (3).

$$y(z) = x(z) - \alpha x(z-1), 0.95 < \alpha < 0.97 \tag{3}$$

x(z) = input signal; y(z) = output signal; α = filter coefficient. Now the signal will be in time domain after pre-emphasis.

Step 2: Frame blocking is splitting the signal into several frames by standard 25 ms. For each frame 12, MFCC feature coefficients are extracted. The voice signal is divided into N samples.

$$y(z) = x(z)w(z), 0 \leq z \leq N_z - 1) \tag{4}$$

$w(z)$ = Hamming window applied to the input signal; z-discrete data.

Step 3: Hamming window is applied to each frame to avoid the discontinuity of the input signal.

$$w(z) = (1 - \alpha) - \alpha cos\left(\frac{2n\pi}{N - 1}\right) \tag{5}$$

where, $0 \leq n \leq N - 1$, $L = N - 1$; L = window length, $\alpha = 0.46164$.

$$w(z) = 0.54 - 0.46cos\left(\frac{2n\pi}{N - 1}\right) \tag{6}$$

where, $0 \leq n \leq N - 1$, n = 0,.., N − 1; n = number of samples N = total number of samples in each frame.

Step 4: FFT is applied to all frames. FFT is applied for transforming time to the frequency domain. FFT is implemented on N_z samples as shown in Eq. (7). D_z is the frame and D_j is the retrieved complex number (Yildiz and Arslan 2018). The output signal can be obtained as a periodogram or spectrum. We compute FFT to acquire the magnitude frequency response for each frame.

$$D_j = \sum_{z=0}^{N_{z-}1} D_z e^{\frac{-j2\pi kz}{N_z}}, where\, k = 0,1 \ldots N_z - 1 \tag{7}$$

$$Mel(f) = 2595log_{10}(1 + \frac{f}{700}) \tag{8}$$

Mel(f) = Mel-frequency; *f = frequency in Hertz.*

Step 5: Finally, applying inverse fast Fourier transform (IFFT) on all frames for transforming frequency domain to the time domain and MFCC values were attained using Eq. (9). MFCC feature vectors were obtained for each frame in selected specific MFCC size.

$$C_n = \sum_{n=1}^{k} (logD_j)cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{k}\right] \tag{9}$$

j = 0,1…K
k = Number of desired Mel cepstrum coefficients
C_n *= Final MFCC coefficients*
$logD_j$*=log energy output of filter bank*

4.3 Classification of an animal using radial basis function (RBF) neural network for classification

The radial basis function neural network (RBFNN) is applied for the classification of target animals (Elephant, Leopard, and Bear) and the difference in acoustics among the three animals varies and can be classified easily with the limited dataset. RBFNN gives better classification accuracy compared with other classification algorithms (as shown in Fig. 10a).

A simple radial basis function neural network (RBFNN) algorithm has been used for animal classification tasks mainly for its strong tolerance to the input noise and good generalization. RBFNN has only one hidden layer, it uses radial basis function (RBF) as a nonlinear activation function and it is robust for prediction. Training is the representation of the network with some sample data and adjusts the weights for a better approximation of the desired function. The weights in the RBF network are the most important factor in determining its function. Some random numbers are usually set for initial weight and then they are adjusted/corrected during the training phase. The hidden layer contains several nodes that use radial basis function (RBF) represented as ϕ(r) shown in Eq. (11), it is a nonlinear activation function (Wu et al. 2012).

Each RBF neuron enumerates the similarity among the input and the prototype vector that are taken from the training dataset. When the input is more similar to the prototype vector which results closer to 1. For finding the similarity it uses a gaussian similarity function. The RBF network uses the linear optimization method for achieving a global optimal solution to the weights adjusted in the minimum mean square error (MSE) sense. We can choose the range of feature vector values to form the MFCC value (50 values) and we can select and perform fine-tuning (Wu et al. 2012) of the Gaussian function for RBF. RBF is used for boosting our classifier.

(i) Here we used one input layer, one hidden layer, and one output layer.
(ii) Initialize weights: some random numbers are usually set for initial weight.
(iii) Present a pattern and target output: radial basis function network (RBFN)-neural network will predict the weight. Then we can feed input forward towards the network.
(iv) Check the error—how far our model output varies with actual output.
(v) Check whether an error is minimum or huge. If it is huge then the weights are adjusted. For each unit error at that node is calculated in the hidden layer N. Error at previous hidden layer N − 1 is calculated.
(vi) These calculated errors are used to adjust the weights so that we can minimize errors at each output unit.

In this work, the following parameters in radial basis functions are used (Wu et al. 2012).

f(x) shows the classifier output and it is expressed as,

$$f(x) = \sum_{j=1}^{m} w_j \phi(r), j = 1,2 \ldots ..m \quad (10)$$

$$\phi(r) = e^{\frac{-r^2}{2\sigma^2}}, \text{Gaussian} \quad (11)$$

$$r = (x - \mu)^2 \quad (12)$$

$f(x)$—*output*

w_j—*Weighting vector*

$\phi(r)$—*Radial basis fucntion*

σ—*Variance*

r—*distance from the data pointx*

x—*input pattern*

μ—*mean*

4.4 Output layer

The output of the neuron contains a set of nodes. Every output node estimates a sort of score for the correlated category. The classification results are made by keeping the input to the category with the maximum score. This score is calculated by taking a weighted sum of the activation values from each RBF neuron. Here, we mean weighted sum is that an output node correlated a weight value with each RBF neuron, and then multiply the activations of the neurons with the weight before adding it to the total sum of the responses as shown in Eq. (10). Each output calculates the score for the various categories, every output has its own set of weights. The output node will give positive weight to the RBF neuron that belongs to its category and for others, it gives a negative weight. The neural network has two phases, one is the training phase and another one is the testing phase. Here we have trained 235 targeted animal sounds and 120 testing animal sound patterns.

4.5 Training and testing phases

The number of input nodes and output nodes is based on the training set in hand. If too many nodes are there in the hidden layer, then the computation that the algorithm has to deal with also increases. After reducing this function for the training set, a new unknown input is passed to the network and we expect it to interpolate with the network. Then, the network has to identify whether a new input is similar to the trained patterns and produced the same output. It is necessary to monitor the progress of the neural network during its training phase. If the acquired results are not improving then some corrections/adjustments to the model might be required. Table 2 shows the number of TRAINING AND TESTING samples of target animals.

4.6 Age estimation

Then the next step after identifying the animal species is to determine the animal age estimation whether it is calf/aged. It is necessary to determine elephant age because of some of the following reasons.

When elephant calves get separated from their mother elephant or if the elephant calf comes out from the forest and roam near to the human settlements then the villagers try to drive the elephant back to the forest by shouting, firing crackers, producing noise, or beating drums, etc. (Lenin and Sukumar 2011). The elephant calf was not so strong enough to run fast and hence it goes away from its place. The calf elephant shows its emotion by producing anguished cries when it gets stuck anywhere or when it needs any help. For these reasons to secure the elephant calf from the endangered situation and the elephant to be saved by the forest authorities. To handle this problem along with the elephant detection we have done an age estimation of the elephant also to save the elephant calf and to reunite it with its mother elephant/herds. Autocorrelation is done for age estimation where adult elephants and calves have a difference in their pitch while producing sounds.

4.6.1 Autocorrelation

The autocorrelation is the correlation among values of the process at various time series, as a function of the two-time lag or two times (Broersen 2006). Autocorrelation is the representation of the degree of similarity among a given time series and its lagged version over certain successive time

Table 2 Number of training and testing samples

Animal species	Number of training samples	Number of testing samples
Elephant	100	40
Elephant Roar	100	60
Elephant Rumble	100	60
Elephant Trumpet	100	60
Elephant Cry	100	60
Leopard	70	40
Bear	65	40

intervals. It is done by calculating the correlation among two various time series, autocorrelation makes use of the same time series twice (i) one is its original form (ii) once lagged with one or more period. It measures the relationship between the current value to its past value. The computation of autocorrelation results in range, from 1 to − 1 (− 1 to 0 indicates negative autocorrelation; 0–1 indicates positive autocorrelation). The definition of autocorrelation can be represented as Eq. (13),

$$R(s,t) = E\frac{\left(X_t - \mu_t\right)\left(X_s - \mu_s\right)}{\sigma_t \sigma_s} \quad \text{(Original form − autocorrelation between time s and t)} \tag{13}$$

Let X_t is a value of the process at time t. E is the expected value operator, mean μ_t, and variance σ_t.

The above expression (Eq. 13) is not well defined for all the time series. Because the variance σ_t maybe 0 or infinite. If the function is defined its value lies between [− 1, 1] (1 shows perfect correlation; − 1 shows perfect anticorrelation). The autocorrelation relies only on the lag between t and s. The correlation depends on the time-distance among the pair of values, not on their position in time. The autocorrelation can be defined as a function of time-lag and this would be an even function of lag $\tau = s - t$. The autocorrelation expressed as a function of time-lag $R(\tau)$,

$$R(\tau) = E\left[\frac{\left((X_t - \mu)\left(X_{t+\tau} - \mu\right)\right)}{\sigma^2}\right] \tag{14}$$

Autocorrelation defined with the function of time-lag between t and s (lagged version) is shown in Eq. (14). If it is an even function then it can be stated as, $R(\tau) = R(-\tau)$.

Autocorrelation is done for reducing the effect of noise on the autocorrelation of the animal voice signal by subtracting and estimating its effect on noisy animal voice signals. Adult and calf animals have a difference in their pitch while producing sounds. The fundamental frequency in the voiced signal is defined as pitch, by the function of autocorrelation we can determine the pitch in a discrete vocal content of an animal. It roughly approximates the fundamental frequency.

4.7 Elephant state of mind (SOM)

In this section, the need of detecting the elephant state of mind (SOM) is explored. An elephant produces a variety of sounds to express its state of mind (SOM) or to communicate with other elephants. Emotions can be conveyed by different means such as behavior, action, responses, etc. The need of detecting the SOM of an elephant is to identify whether it is stressed or relaxed. Here we have taken four elephant calls like Roar, Rumble, Trumpet, and cry for analyzing its state of mind. If it produces roar sound with a frequency range from 305 to 6150 Hz it means it is stressed due to various factors such as high level of distress, aggression, protest, etc., and the probability of an attack against human is more and also there is a chance of creating massive damage to people. Therefore, we need to alert the local communities to be safe. Whereas, SOM for calf elephant was not detected because the calf elephant does not create massive damage to the local communities so we stop at recognizing calf or adult.

If the system detects that it is a single adult elephant then the next process is finding SOM so that we can alert people to be on the safer side. Based on call tone, behavioral context, and measurement, they have interpreted the meanings and differences within each call type context. Elephants use vocalization in a wide range of situations. Oikarinen et al. (2018) evaluated various call type classification of marmoset monkeys such as Twitter, Trill, Chatter, and Phee. Translocation results change in elephant vocalizations with low frequency, known as rumbles. Raised pitch is a sign of stress in both man and animals. Changes in vocal parameters may be considered to indicate some variation in emotion, excitement level, or both (Viljoen et al. 2015).

4.7.1 Feature extraction using gammatone frequency cepstral coefficient (GFCC)

GFCC is employed here for extracting the features of elephant vocal content. Gammatone filter is a linear approximation to its filtering function (Wang and Zhang 2019) and is used in the time domain which is in form of a cascade. Initially, the equivalent rectangular bandwidth scale (ERB) is converted into a normal frequency scale. The centre frequency array is indexed using the channel by estimating the lower and upper bound of ERB.

Algorithm 2: Pseudocode for GFCC feature extraction
Input: Pre-processed animal audio sounds
Output: GFCC Feature Vector
Step 1: Initialize parameter
Step 2: Framing- Input animal audio signals are split into several frames;
Step 3: Input signal is passed through a 64-channel filter bank (Huapeng Wang & Cuiling Zhang 2019, B.Ayoub et al. 2016) that consists of an array of Band Pass Filter (BPF)
Step 4: Rectify the filter response for each channel and decimate to 100 Hz.
Step5: Convert the signal from the time domain to the frequency domain by T-F representation.
Step 6: Apply Gammatone filter

$$g(t) = a(t)^{n-1} e^{-2\pi bt} \cos(2\pi f_c t + \phi)$$

Step 7: Apply log for finite series of the dataset

$$Out1 = log(g(t)) \tag{16}$$

Step 8: Apply DCT for convolution computation and compression of an audio signal.

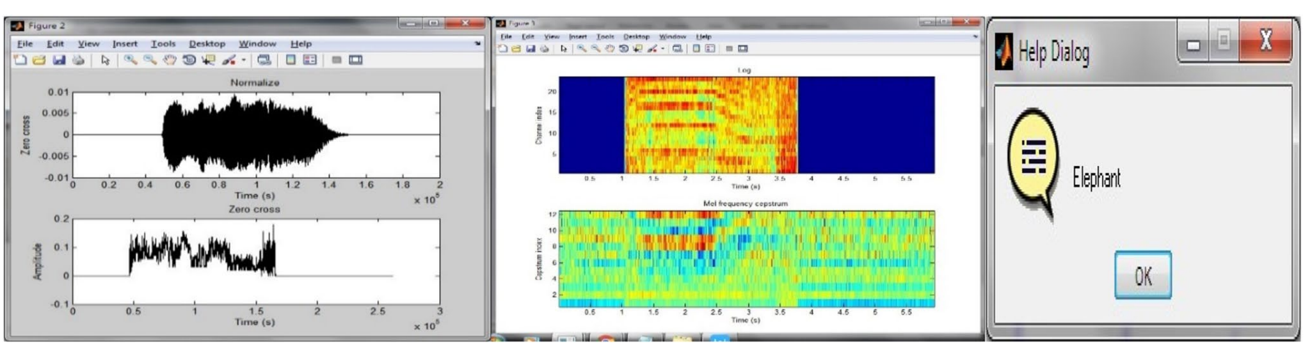
$$GFCC = dct(Out1) \tag{17}$$

4.7.2 Classification of elephant SOM using support vector machine (SVM)

SVM is a nonlinear classifier (Padmanabhan and Premkumar 2015) that can able to predict whether the input belongs to what type of class. However, the unknown events (Zhao et al.2017) such as competing noise produced from rain, wind, or any other interference and vocalization of other than these three animals (Elephant, Leopard, and Bear) acoustic signal will be dropped in the initial stage itself. Keen et al. (2014) explores an additional option is an "unknown" class when we do field recordings.

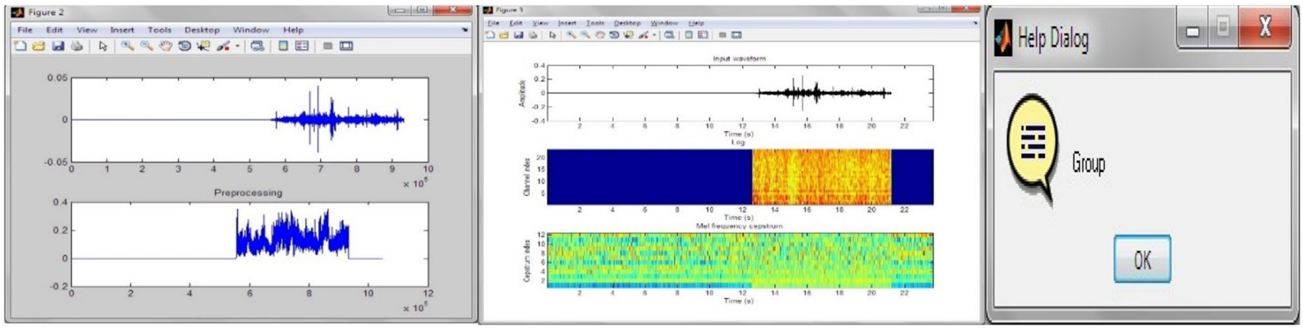
The main task of the SVM classifier is to segregate training data in the high dimensional space with the help of kernel function and it finds an optimal hyperplane with a high margin among data of two different classes. SVM acts as a predictive model for elephant SOM detection. It is a very difficult task to define the state of mind of the elephant with its sound. In this task, we have trained Roar, Rumble, Trumpet, and Cry sound but the difference among these four sounds have some slight variation. Also, we have a limited dataset for Roar, Rumble, Trumpet, and Cry sound. With the limited dataset and low difference among the input, RBFNN does not give maximum accuracy whereas SVM achieves greater accuracy with the limited dataset. SVM suits well for classification of these four classes compared to Radial Basis Function Network (RBFNN) (as shown in Fig. 10a, b) even if the sound difference among these four classes is low. Intraclass classifications in any of these three animals like elephant state of mind detection is a difficult task for RBFNN. Because the network will become more complex to classify. Hence, it needs more hidden layers for better classification accuracy and the operating time it takes will be more.

Here we have chosen the support vector machine (SVM) classifier for the classification of four different states of an elephant such as Roar, Rumble, Trumpet, and Cry. Because of two reasons we have chosen the SVM classifier among some other classification algorithms (Hsu and Lin 2002; Leonid and Jayapervathy 2020). Firstly, it is insensible to a problem that arises due to overfitting. Secondly, it can achieve high accuracy with smaller or limited training samples. It is a robust algorithm for classification applied for many applications like Intrusion detection, pattern recognition, and face recognition, etc. SVM is a dependable and fast classification algorithm that classifies well even with a limited number of the dataset to analyze.



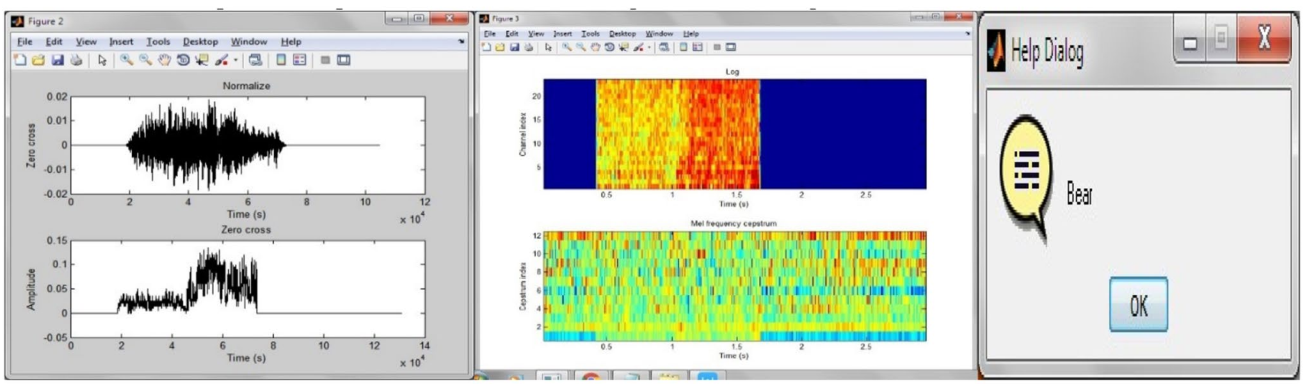
(a) (b) (c)

Fig. 3 Results for elephant detection: **a** normalized input, **b** MFCC, **c** output



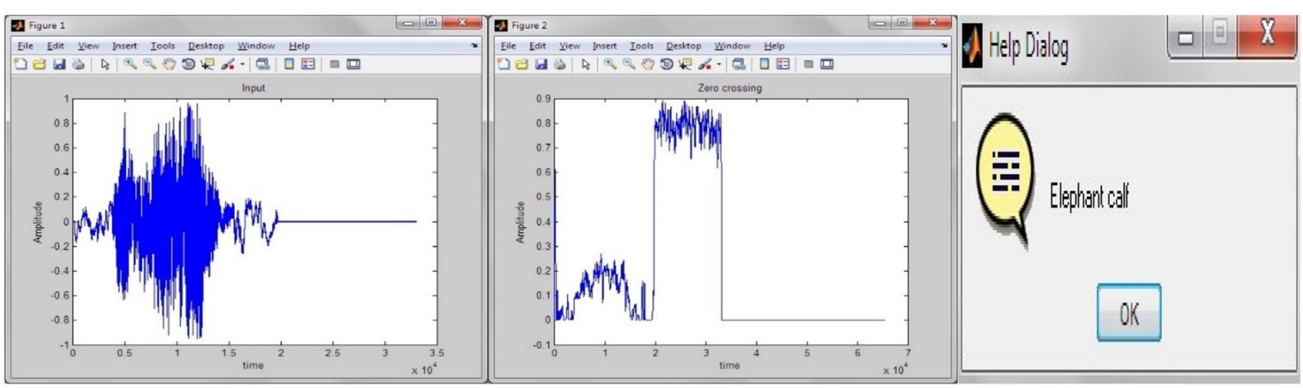
(a) (b) (c)

Fig. 4 Results for elephant group estimation: **a** normalized input, **b** MFCC, **c** output



(a) (b) (c)

Fig. 5 Results for bear detection: **a** normalized input, **b** MFCC, **c** output



(a) (b) (c)

Fig. 6 Results for elephant calf detection: **a** input, **b** zero crossing, **c** output

Reaction time (RT) is used to recognize different states in stress and three classification algorithms were applied for stress recognition such as a Naïve Bayes Classifier, Decision Tree Classifier, and SVM. The authors compared the classification accuracies of these three algorithms and concluded that SVM gives the highest classification accuracy. Therefore, SVM is used as the algorithm for classification in stress recognition (Zhang et al. 2017). Nanni et al. (2020) proposed an automated classification of animal audio by combining dissimilarities spaces formed by a set of siamese neural network (SNN) with various clustering techniques for training support vector machine (SVM).

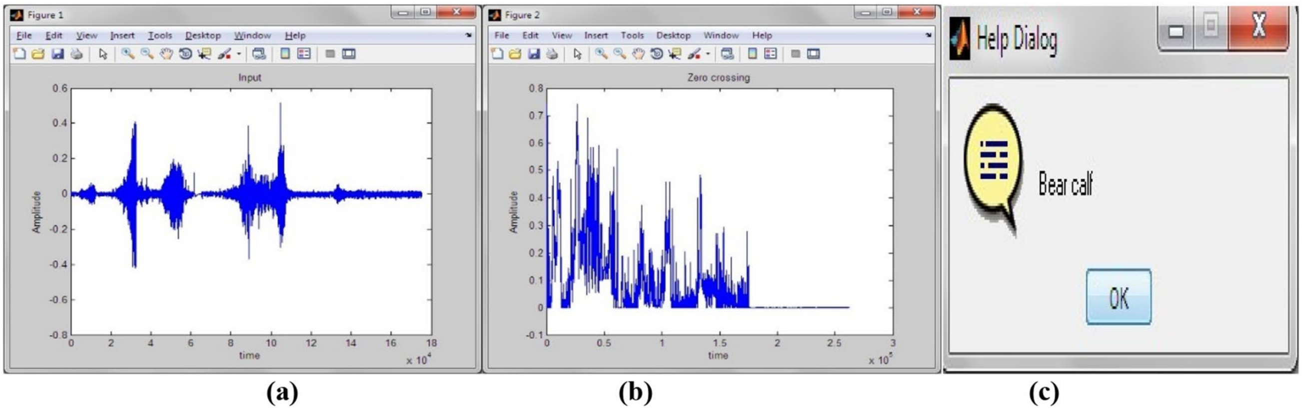


Fig. 7 Results for bear calf detection: **a** input, **b** zero crossing, **c** output

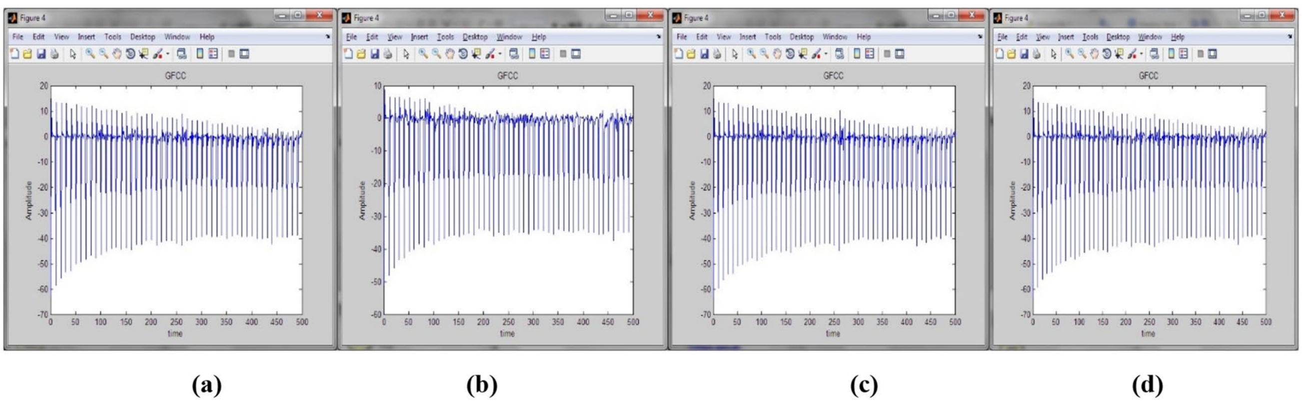


Fig. 8 Software results of GFCC for elephant SOM detection: **a** Roar, **b** Rumble, **c** Trumpet, **d** Cry

The results for the animal identification of target animals were taken in MATLAB 2016 environment (Figs. 3, 4, 5).

The results for the age estimation of target animals were taken in MATLAB 2016 environment (Figs. 6, 7).

The results for state of mind (SOM) detection of an adult elephant were taken in MATLAB 2016 environment (Fig. 8).

5 Experimental hardware setup

In this section, the experimental hardware verification is discussed in detail. In this work, the processed input audio signal is fed to the ARM 7 microcontroller through Asynchronous Receiver Transmitter (URAT 0) from the camera to personal computer (PC). Here the received audio signal is stored in an external ROM connected to the ARM processor. This is temporarily stored for a certain period (30 days) and it will be erased by cache memory in the ARM processor. If we need these audio signals to store permanently for future use the audio signals can be retrieved and stored in an external storage device such as read-only memory (ROM) or PC. GPS is connected to the microcontroller through interrupt mode (URAT 1). GPS data, such as Latitude, Longitude, Greenwich mean time (GMT) is recorded as a frame in receiver interrupt (URAT 1) for every 15 to 20 s continuously, from this frame we separate latitude, longitude location details that are stored as an array. The stored value is sent via the GSM modem (SIM 900A). Once an animal is detected a text message (as shown in Fig. 9a) is sent to the control unit (Forest Ranger, Radio Station, Local Communities) with the help of GSM, along with the location details of animal movement using GPS. This system will be placed in the animal passageway that is given as Initial Boundary. Once if it reaches the Initial Boundary it will start to detect. The early detection of an animal gives enough time to the forest authorities to chase the animal back inside the forest.

Initial Boundary: Inside the forest, the Animal detection and discrimination unit covering the region.

Actual Boundary: The distance between human settlements to Initial Boundary.

With the internet of things (IoT) the received information from the components like GSM, GPS will be updated

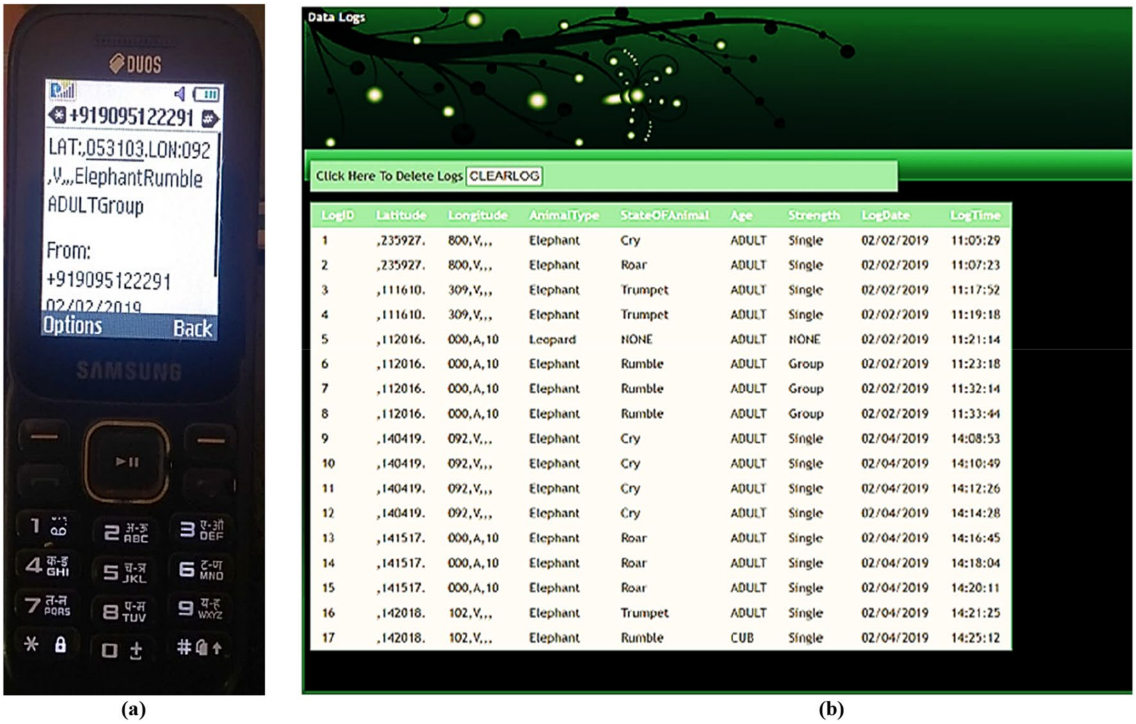


Fig. 9 Experimental hardware test setup for animal detection and discrimination system: **a** SMS received, **b** web page shows the information received by the animal detection and discrimination system using internet of things (IoT)

in the web page created and this acts as a database and it will be very essential for forest officials for further reference.

IoT will bridge the gap between devices and will also enable to monitor animals from anywhere. The information received from the device (GSM, GPS) is live and continuous and it is updated automatically in a separate webpage created with a login ID and Password (as shown in Fig. 9b). It is ensured that the system is operating efficiently. If the component connected with the processor is not working properly then, we can easily identify it by checking this web page. The software results were taken in the MATLAB environment. Overall, the proposed system acts as an Early Warning and Monitoring Module, which highlighted the efficiency in the detection of animals based on the acoustic content of target animals.

6 Evaluation

The dataset for our work was collected from the forest department and used online resources for gathering animal audio dataset. We have trained with 70–100 audio samples for each animal (Elephant, Leopard, Bear), for the testing phase we have taken 40–60 samples for each animal. Performance is measured based on the total cross-validation,

$$Accuracy = \frac{True\ Positive(TP)}{(True\ Positive(TP) + False\ Negative(FN))} \tag{18}$$

$$Precision = \frac{True\ Positive(TP)}{(True\ Positive(TP) + False\ Positive(FP))} \tag{19}$$

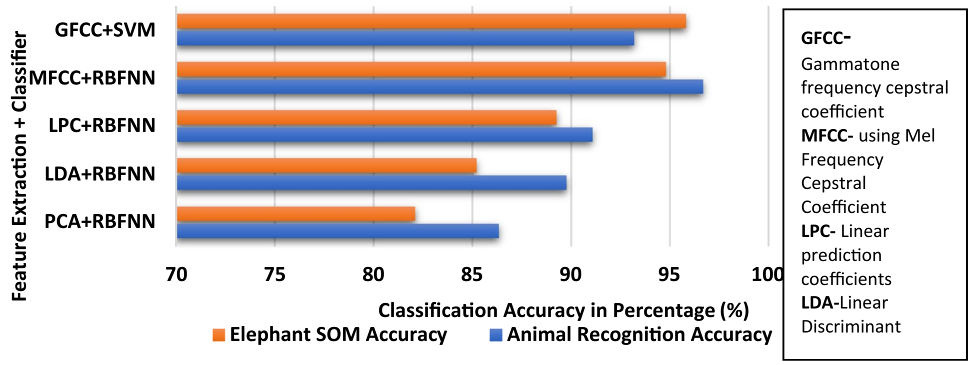
True Positive (TP)—Correctly detected
False Positive (FP)—Wrongly detected
False Negative (FN)—Failed to detect

Table 3 Performance analysis of radial basis function neural network (RBFNN) classifier for animal recognition

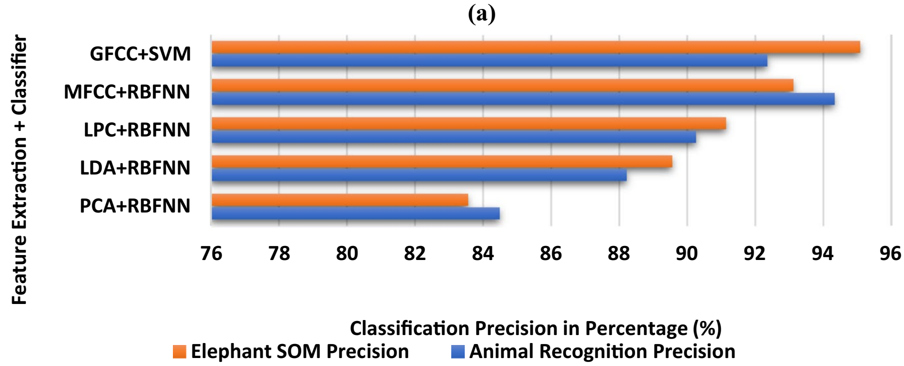
Samples	Training samples	Testing samples	TP	FP	FN	Accuracy (%)	Precision (%)
Elephant	100	40	39	2	1	97.50	95.12
Leopard	70	40	38	2	2	95.00	95.00
Bear	65	40	39	3	1	97.50	92.86

Table 4 Performance analysis of support vector machine (SVM) classifier for SOM detection of elephant

Samples	Training samples	Testing samples	TP	FP	FN	Accuracy (%)	Precision (%)
Roar	100	60	58	3	2	96.67	95.08
Rumble	100	60	59	1	1	98.33	98.33
Trumpet	100	60	57	2	3	95.00	96.61
Cry	100	60	56	2	4	93.33	96.55



(a)



(b)

Fig. 10 Performance of different feature extraction and classifier for animal recognition and SOM detection of elephant: **a** accuracy, **b** precision

Performance validation reveals that the radial basis function neural network (RBFNN) classifier gives better accuracy in animal recognition and support vector machine (SVM) classifier achieves greater accuracy in SOM detection of an elephant as shown in Tables 3 and 4. The performance of the different classifiers for animal recognition and elephant SOM detection is shown in Fig. 10a, b.

7 Conclusion

In this paper, we have introduced real-time animal detection and discrimination using the vocal spectral content of animals for avoiding Human–wildlife conflict (HWC). This system is an early warning and monitoring module, which results in animal identification, age estimation, elephant state of mind (SOM), and GPS location details are sent as SMS via GSM to forest ranger, local communities, and fm radio/local channel. The internet of things (IoT) concept is introduced here to update the received information through a web server and stored it in the cloud. A web page is created with a User ID and Password and this will be given to the higher authorities in the forest department. On this Web page, GPS Information like Latitude and Longitude, animal type, state of mind (SOM) in case of elephant, age, strength, log date, log time is provided. Earlier detection of animal roaming near human settlements can help in reducing man–animal conflict. Overall, the proposed system has highlighted the efficiency in the detection and discrimination of animals based on the acoustic content of target animals.

Acknowledgements My heartfelt gratitude to my mentor and my research guide Dr. Chitra Selvi Shokkalingam, Department of Electrical and Electronics Engineering, Anna University-University College of Engineering, Thirukkuvalai for her constant support and valuable guidance, and supervision. I sincerely acknowledge the University Grants Commission, New Delhi for providing me SRF (Senior Research Fellow) Fellowship under National Fellowship Scheme.

Declarations

Conflict of interest The authors have no conflict of interest in submitting the manuscript to this journal.

References

Anni DJS, Sangaiah AK (2015) An early warning system to prevent human–elephant conflict and tracking of elephant using seismic sensors. In: Satapathy S, Govardhan A, Raju K, Mandal J (eds) Emerging ICT for bridging the future—proceedings of the 49th annual convention of the Computer Society of India (CSI) volume 1. Advances in intelligent systems and computing, vol 337. Springer, Cham. https://doi.org/10.1007/978-3-319-13728-5_67

Anni JS, Sangaiah AK (2018) Wireless integrated sensor network: boundary intellect system for elephant detection via cognitive theory and fuzzy cognitive maps. Future Gener Comput Syst 83:522–534 (**ISSN 0167-739X**)

Ayoub B, Jamal K, Arsalane Z (2016) Gammatone frequency cepstral coefficients for speaker identification over VoIP networks. In: 2016 international conference on information technology for organizations development (IT4OD), Fez, pp 1–5. https://doi.org/10.1109/IT4OD.2016.7479293

Bjorck J, Rappazzo BH, Chen D, Bernstein R, Wrege PH, Gomes CP (2019) Automatic detection and compression for passive acoustic monitoring of the african forest elephant. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01):476–484. https://doi.org/10.1609/aaai.v33i01.3301476

Borden G, Harris K, Raphael L (1994) Speech science primer: "Physiology, acoustics, and perception of speech (3rd ed.)." Williams and Wilkins, Baltimore

Boussaid L, Hassine M (2018) Arabic isolated word recognition system using hybrid feature extraction techniques and neural network. Int J Speech Technol 21:29

Broersen PMT (2006) Automatic autocorrelation and spectral analysis 1. Spectrum analysis—Statistical methods 2. Signal processing—Statistical methods 3. Autocorrelation (Statistics) 4. Time-series analysis". Springer-Verlag London Limited, I. Title 543.5'0727. ISBN-13: 9781846283284. ISBN-10: 1846283280

Buchan S, Mahu R, Wuth J, Balcazar-Cabrera N, Gutiérrez L, Neira S, Yoma N (2019) An unsupervised hidden Markov model-based system for the detection and classification of blue whale vocalizations off Chile. Bioacoustics 29:1–28

Clemins P, Johnson M (2003) Application of speech recognition to African elephant (*Loxodonta africana*) vocalizations. In: ICASSP, IEEE international conference on acoustics, speech, and signal processing—proceedings, vol 1, p I-484

Darras E, Pütz P, Fahrurrozi KR, Tscharntke T (2016) Measuring sound detection spaces for acoustic animal sampling and monitoring. Biol Conserv 201:0006–3207 (**ISSN 29-37**)

Devi KJ, Thongam K (2019) Automatic speaker recognition with enhanced swallow swarm optimization and ensemble classification model from speech signals. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-019-01414-y

Gragido W, Pirc J, Selby N, Molina D (2013) Chapter 4—signal-to-noise ratio. Blackhatonomics. Syngress, pp 45–55. https://doi.org/10.1016/B978-1-59-749740-4.00004-6

Gutierrez-Galan D, Dominguez-Morales JP, Cerezuela-Escudero E, Rios-Navarro A, Tapiador-Morales R, Rivas-Perez M, Dominguez-Morales M, Jimenez-Fernandez A, Linares-Barranco A (2017) Embedded neural network for real-time animal behavior classification. Neurocomputing. https://doi.org/10.1016/j.neucom.2017.03.090

Hsu C-W, Lin C-J (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13(2):415–425. https://doi.org/10.1109/72.991427

Huang Y, Tian K, Wu A (2019) Feature fusion methods research based on deep belief networks for speech emotion recognition under noise condition. J Ambient Intell Human Comput 10:1787–1798. https://doi.org/10.1007/s12652-017-0644-8

Kaya H, Salah AA, Karpov A, Frolova O, Grigorev A, Lyakso E (2017) Emotion, age, and gender classification in children's speech by humans and machines. Comput Speech Lang 46:0885–2308 (**ISSN 268-283**)

Keen S, Ross JC, Griffiths ET, Lanzone M, Farnsworth A (2014) A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). Eco Inform 21:25–33

Kuchibhotla S, Vankayalapati HD, Anne KR (2016) An optimal two stage feature selection for speech emotion recognition using acoustic features. Int J Speech Technol 19:657

Lee C-H, Chou C-H, Han C-C, Huang R-Z (2006) Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. Pattern Recogn Lett 27:93–101

Lenin J, Sukumar R (2011) Action plan for the mitigation of elephant–human conflict in india. Final report to the US fish and wildlife service. Asian Nature Conservation Foundation, Bangalore

Leonid TT, Jayaparvathy R (2020) Statistical-model based voice activity identification for human–elephant conflict mitigation. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-02005-y

Luque J, Larios DF, Personal E, Barbancho J, León C (2016) "Evaluation of Mpeg-7-based audio descriptors for animal voice recognition over wireless acoustic sensor" networks. Sensors 16(5):717

Matuska S, Hudec R, Kamencay P, Benco M, Zachariasova M (2014) Classification of wild animals based on SVM and local descriptors. AASRI Procedia 9:2212–6716 (**ISSN 25-30**)

Mustafa MB, Yusoof MAM, Don ZM (2018) Speech emotion recognition research: an analysis of research focus. Int J Speech Technol 21:137

Nanni L, Brahnam S, Lumini A, Maguolo G (2020) Animal sound classification using dissimilarity spaces. Appl Sci 10:8578

Ogawa A, Hori T (2017) Error detection and accuracy estimation in automatic speech recognition using deep bidirectional recurrent neural networks. Speech Commun 89:0167–6393 (**ISSN 70-83**)

Oikarinen TP, Srinivasan K, Meisner O, Hyman JB, Parmar S, Desimone R, Landman R, Feng G (2018) Deep convolutional network for animal sound classification and source attribution using dual audio recordings. bioRxiv 437004

Padmanabhan J, Premkumar MJJ (2015) Machine learning in automatic speech recognition: a survey. IETE Tech Rev 32(4):240–251. https://doi.org/10.1080/02564602.2015.1010611

Poshitha D, Suduwella C, Sayakkara A, Sandaruwan D, Keppitiyagama C, De Zoysa K et al (2015) Listening to the giants: using elephant infra-sound to solve the human–elephant conflict. In: Proceedings of the 6th ACM workshop on real-world wireless sensor networks (Seoul). https://doi.org/10.1145/2820990.2821000.

Prabu (2016) An efficient surveillance system to detect elephant intrusion into forest. Int J Adv Eng Technol 7:166–171 (**E-ISSN 0976-3945**)

Stoeger AS, Heilmann G, Zeppelzauer M, Ganswindt A, Hensman S, Charlton BD (2012) "Visualizing sound emission of elephant vocalizations: evidence for two rumble production types" 2012. PLoS ONE 7(11):e48907

Stoeger AS, Zeppelzauer M, Baotic A (2014) Age group estimation in free-ranging African elephants based on acoustic cues of low-frequency rumbles. Bioacoustics 23(3):231–246

Venkataraman AB, Saandeep R, Baskaran N, Roy M, Madhivanan A, Sukumar R (2005) Using satellite telemetry to mitigate elephant–human conflict: an experiment in northern West Bengal, India. Curr Sci 88:1827–1831. Available online at: https://www.jstor.org/stable/24110372

Viljoen JJ, Ganswindt A, Reynecke C, Stoeger AS, Langbauer WR Jr (2015) Vocal stress associated with a translocation of a family herd of African elephants (*Loxodoa africana*) in the Kruger National Park, South Africa. Bioacoustics 24(1):1–12 (**Taylor & Francis**)

Wang H, Zhang C (2019) The application of gammatone frequency cepstral coefficients for forensic voice comparison under noisy conditions. Aust J Forensic Sci. https://doi.org/10.1080/00450618.2019.1584830

Wood JD, O'Connell-Rodwell CE, Klemperer SL (2005) Using seismic sensors to detect elephants and other large mammals: a potential census technique. J Appl Ecol 42:587–594

Wu Y, Wang H, Zhang B, Du K-L (2012) Using radial basis function networks for function approximation and classification. ISRN Appl Math. https://doi.org/10.5402/2012/324194

Yildiz O, Arslan A (2018) Automated auscultative diagnosis system for evaluation of phonocardiogram signals associated with heart murmur diseases. Gazi Univ J Sci 31:112–124

Zeppelzauer M, Stoeger AS (2015) Establishing the fundamentals for an elephant early warning and monitoring system. BMC Res Notes 8:409

Zeppelzauer M, Hensman S, Stoeger AS (2015) Towards an automated acoustic detection system for free ranging elephants. Bioacoustics 24:13–29. https://doi.org/10.1080/09524622.2014.906321

Zhang Bo, Morère Y, Sieler L, Langlet C, Bolmont B, Bourhis G (2017) Reaction time and physiological signals for stress recognition. Biomed Signal Process Control 38:1746–8094 (**ISSN 100-107**)

Zhao Z, Zhang S-H, Xu Z-Y, Bellisario K, Dai N-H, Omrani H, Pijanowski BC (2017) Automated bird acoustic event detection and robust species classification. Ecol Inform 39:99–108

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations