Biological Conservation 309 (2025) 111272

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Living with elephants: Deep learning models performance in examining Asian elephant (*Elephas maximus*) sounds from Sri Lanka and Malaysia with considerations for application

Naufal Rahman Avicena [a,b,c,d,e], Yen Yi Loo [d,e,f], Tomas Maul [g], Noah Thong [d,e], Christopher Chai Thiam Wong [h,l], Shermin de Silva [i,j], Salman Saaban [e,k], Ee Phin Wong [d,e,*]

[a] *Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences & Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China*
[b] *Yunnan International Joint Laboratory of Southeast Asia Biodiversity Conservation & Yunnan Key Laboratory for Conservation of Tropical Rainforests and Asian Elephants, Menglun, Mengla, Yunnan 666303, China*
[c] *Yunnan International Joint Laboratory for the Conservation and Utilization of Tropical Timber Tree Species, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China*
[d] *School of Environmental and Geographical Sciences, University of Nottingham Malaysia, Semenyih 43500, Selangor Darul Ehsan, Malaysia*
[e] *Management and Ecology of Malaysian Elephants (MEME), University of Nottingham Malaysia, Semenyih 43500, Selangor Darul Ehsan, Malaysia*
[f] *Sunway Centre for Planetary Health, Sunway University, Jalan Universiti, Bandar Sunway, 47500 Petaling Jaya, Selangor Darul Ehsan, Malaysia*
[g] *School of Computer Science, University of Nottingham Malaysia, Semenyih 43500, Selangor Darul Ehsan, Malaysia*
[h] *WWF-Malaysia, Petaling Jaya 46150, Selangor Darul Ehsan, Malaysia*
[i] *Trunks and Leaves Inc., Pittsfield, MA 01201, United States of America*
[j] *Department of Ecology Behavior and Evolution, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, United States of America*
[k] *Department of Wildlife and National Parks Peninsular Malaysia, Km. 10 Jalan Cheras, Kuala Lumpur, Malaysia*
[l] *Wildlife Conservation Society—Malaysia Program, Kuching, Sarawak, Malaysia*

ARTICLE INFO

Keywords:
Human-elephant conflict
Human-elephant coexistence
Elephant bioacoustics
Deep learning
Convolutional Neural Network (CNN)
Early warning system

ABSTRACT

Human-elephant conflict (HEC) affects people and wild elephants negatively, and support for harmonious coexistence is needed. With the current human footprint, wildlife is displaced, and people living near wildlife want safe interactions. Conservation interventions are needed to manage human-elephant coexistence in real-time. This research, using deep learning models, provides the fundamental mechanics for acoustic detection of elephants in an automated early-warning system, currently under development. We examine the use of convolutional neural networks (CNNs) for classifying Asian elephant (*Elephas maximus*) sounds and non-elephant sounds. The results demonstrated the ability of CNNs to process bioacoustics data across various sample sizes, with the best-performing model achieving 98.45 % average test accuracy (balanced sample sizes, a k-fold approach with 10 % for testing). But when we infer CNN models built with Sri Lankas elephant vocalizations with unseen Malaysias elephant vocalizations, the performance of the models dropped to an average of 67.93 % accuracy and F1 score between 0.67 and 0.81, regardless of the initial training dataset size. We used Principal Component Analysis to compare 15 sound parameters extracted from spectrograms of elephant calls from Sri Lanka and Malaysia, and found that the sound characteristics between the two subspecies largely overlapped but with some differences. We conclude that the CNN models can detect elephant sounds but perform best with local data. The use of bioacoustic monitoring and automated detection can potentially support harmonious coexistence between humans and elephants, but for endangered species targeted by poachers, safeguards are needed. Additionally, we need discourse on research ethics and local communitys rights.

* Corresponding author.
E-mail address: EePhin.Wong@nottingham.edu.my (E.P. Wong).

https://doi.org/10.1016/j.biocon.2025.111272
Received 1 November 2023; Received in revised form 19 May 2025; Accepted 26 May 2025
Available online 4 June 2025

1. Introduction

Biodiversity loss, alongside climate change and environmental pollution, is now at a rate that is traversing alarmingly past the safety boundaries of what makes our planet livable for humankind (Rockström et al., 2009). The global assessment by the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services (IPBES) highlighted the anthropogenic degradation of nature, with 75 % of terrestrial ecosystems and 66 % of marine environments heavily altered by man, and this is expected to worsen with the event of climate change (IPBES, 2019). The Target 4 of the Convention on Biological Diversity's Kunming-Montreal Global Biodiversity Framework highlighted the need for urgent management actions to effectively manage human-wildlife conflict, to transform conflict to coexistence (IUCN, 2023).

Human-wildlife conflict definition promoted by the IUCN SSC Human-Wildlife Conflict and Coexistence Specialist Group, encompass not only visible impact of conflict on injuries and lives of both human and wildlife, as well as peoples livelihood, but also hidden or perceived negative impacts from the wildlife and from disagreements that arises between different groups of stakeholders (IUCN, 2023). Hence, it is important to consider how different groups of stakeholders can work together as part of conflict resolution. In particular, conflict with wild elephants has always ranked high, provoking strong sentiments from the communities who are affected, members of the public, and the media (Barua, 2010; Kansky and Knight, 2014; Tan et al., 2020; Vasudev and Goswami, 2020).

The Asian elephant (*Elephas maximus*) lives in fragmented populations across 13 countries in Asia, and is listed as 'Endangered' on the IUCN Red List of Threatened Species (Williams et al., 2020). There are four extant subspecies of Asian elephants, including *E.m.indicus*, *E.m. maximus*, *E.m.borneensis,* and *E.m.sumatranus* (Williams et al., 2020). Wild elephants have large home ranges reaching hundreds of square kilometers and their populations often live and roam in areas outside of protected areas (Calabrese et al., 2017; Fernando and Pastorini, 2011). As natural habitats are increasingly lost and fragmented by anthropogenic activities (Sukumar, 2003; Williams et al., 2020), wild elephants frequently come into contact with humans in agricultural land (de la Torre et al., 2021). Human-elephant conflict (HEC) often involves crop damage and property damage, and with occasional cases of injuries or death to humans and elephants (Williams et al., 2020), resulting in negative perceptions towards conflict.

Management of conflict situations is important to safeguard both people and wildlife. The setting up of early warning systems for HEC, like in Annamalai Hills, India, has helped to reduce human injuries and death in an agricultural landscape with human settlements (Kumar and Raghunathan, 2014; persn. Comm Ananda Kumar). This study seeks to complement such efforts with the development of artificial intelligence (AI), here defined as computer systems that has the ability to perform tasks that typically require human intelligence (Fang et al., 2019), and advances in machine learning, to carry out automated detection of elephants, especially where direct sightings of elephants are difficult. This effort is part of a project that is trying to foster a harmonious coexistence between the agriculture communities and elephants in movement corridors and shared landscapes (www.ace-coalition.com). Previous approaches to automatic detection of elephants have been largely dependent on the combination of passive acoustic monitoring (PAM) and supervised learning (Wrege et al., 2017; Swider et al., 2022), which heavily require human interventions in the training process such as feature extraction (Keen et al., 2017). Additionally, such studies have been mainly concentrated in African elephants, with limited research done in Asian elephants.

Elephants are highly social mammals and vocal learners that use different call types to communicate (S. de Silva, 2010; Stoeger and de Silva, 2014). Their vocalization ranges from infrasound (<20 Hz) (Stoeger and de Silva, 2014) to approximately 6 kHz (Nair et al., 2009). The usage of sound recordings collected via Autonomous Recording Units (ARU) can help increase the detection of vocal taxa, as fixed-angle sensors like cameras can only detect when the animal is in view. However, continuous acoustic monitoring generates large amounts of data that is challenging to save and transfer, and require manpower to manually process, making real time detection a challenge. Advances in AI and computer technology can address this bottleneck and carry out automated detection of targeted species using bioacoustics data. Deep learning is a type of machine learning model inspired by the structure and function of the brain, consisting of networks of interconnected nodes used to learn complex relationships between variables for a broad diversity of applications i.e. identifying similar objects in different images or to study the relationship between words and emotions (Emmert-Streib et al., 2020; Mehrish et al., 2023).

Convolutional neural networks (CNNs) are a type of deep learning architecture inspired by the hierarchical and local connectivity patterns found in the mammalian visual system, for example a cat (Jogin et al., 2018), which is used for processing high-dimensional data such as images, videos, or sounds (Christin et al., 2019; Mehrish et al., 2023; Park et al., 2020). Using prior algorithms, such as multi-layer perceptrons (MLPs) and support vector machines (SVMs), to process data such as sounds generally requires the laborious task of manually designing the extraction process of summary features such as peak frequency and syllable duration for input (Stowell, 2022). CNNs effectively remove this step by having the ability to automatically identify and extract summary features from lightly-preprocessed data, keeping richer information for processing, thereby effectively outperforming prior algorithms by huge margins (Goodfellow et al., 2016; Stowell, 2022). Earlier works of applying CNNs to animal calls came from amphibians and birds, due to both being vocal species (Colonna et al., 2016; Goëau et al., 2016; Stowell, 2022). The application of CNNs to elephants shortly followed suit (Bjorck et al., 2019; Zeppelzauer et al., 2015), and it has proved promising for a given classification task, potentially paving the way for human-elephant conflict management through the development of sound-based early warning systems (Loo et al., 2024; Ramasubramanian et al., 2022; Thomas Leonid and Jayaparvathy, 2022).

The major challenge in using CNNs and other deep neural networks is to increase the performance of automated detection, which depends on the training data (Ribeiro et al., 2020). Training a CNN model requires a sufficient dataset with diverse features, and training labels to achieve better performance (Alzubaidi et al., 2021; Karimi et al., 2020). Insufficient data and variation in training data may lead to poor generalization and affect the overall performance of the model (Alzubaidi et al., 2021). This problem can be surmounted to some degree by creating artificial variations or noise in the training data set through data augmentation techniques (Mehrish et al., 2023; Nolasco et al., 2023; Park et al., 2020). In addition, there is currently a lack of evidence on whether different subspecies of Asian elephants vary in their vocalizations, which is an important gap to fill to provide scalability of CNN models trained on a given population of elephants.

Here, our objective is to apply CNNs to classify elephant and non-elephant sounds to pave the way for automated early warning systems (EWS), which are currently under development. We reviewed existing literature on bioacoustics and deep learning, and assessed the use of a CNN architecture (Dubey and Jain, 2019; Fukushima, 1969) to train and test the feasibility of the algorithm for elephant sounds. Then, we investigated the correlation between dataset sizes and the performance of the model. Finally, we tested the trained models from *E.m.maximus* (Sri Lanka) to classify *E.m.indicus* (Malaysia) sounds. We ran a Principal Component Analysis (PCA) to better understand the characteristics of recorded vocalizations between the two subspecies. This study will be useful as a case study for future applications of AI-bioacoustics methods as the basis of ecological studies and early warning systems for vocal taxa.

2. Materials and methods

2.1. The use of deep learning and CNNs to process bioacoustics data

The Web of Science (Clarivate) search engine was used to examine the trend of using deep learning models for studying wildlife bioacoustics. The search includes scientific articles, proceeding papers and early access papers, but excludes reviews, book chapters and others. Keywords such as "deep learning", "wildlife AND biodiversity", "bioacoustics OR sound OR vocal OR calls", "elephant", and "Convolutional Neural Network OR CNN" were used in combination. This review was revised on the 14th of October 2023.

2.2. Data collection

2.2.1. Data collection for E.m.indicus vocalizations in Malaysia

Elephant sounds for *E.m.indicus* were collected between October 2021 and March 2022 (6 months), from two salt licks, Sira Gajah and Sira Tersau, located in Belum-Temengor Forest Complex, Perak, Malaysia (Appendix A1, Fig. A1.1). At each site, we deployed an ARU approximately two to three meters above ground (Frontier's Lab BAR-LT [www.frontierlabs.com.au] at one site and Wildlife Acoustics Song Meter 4 [www.wildlifeacoustics.com] in the other site) together with three camera traps (Reconyx HyperFire 2). The ARUs were set to a 44,100 Hz sampling rate (Browning et al., 2017) for dual purpose of studying elephants and soundscapes, and 24db gain on both channels (stereo). Both channels were used during CNN data preprocessing. Recordings were saved in Waveform Audio File (.wav) format. Both ARUs were fitted with omnidirectional microphones. The ARUs recorded the soundscape for 24 h a day, which yielded ~300 MB per day. Elephant signals were present at various times of the day. The camera traps, deployed at approximately 1.5 m above ground, were used to validate the presence of elephants during the ARU recording period and aid manual annotation of sounds.

The collected sound data were manually annotated using Raven Sound Analysis Software (Raven Pro ver. 1.6.1, K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology) using spectrogram window with frequency grid spacing = 62.5 Hz and DFT size =512 samples. Each call was annotated by drawing a selection box around the call approximately 2 milliseconds before and after the call of interest to ensure all the features were included. Elephant signals were detected between 1 pm to 1 am on 31/10/2021, 01–02/11/2021, and 13/02/2022. The annotation exercise yielded 3 call types: roars, rumbles, and chirps.

2.2.2. Data collection for E.m.maximus vocalizations in Sri Lanka

Trunks & Leaves™ Inc., an elephant conservation non-profit organization, provided the data for *E.m.maximus*. The data were collected by SdS from 2006 to 2007, in Uda Walawe National Park, Sri Lanka, during direct observations of elephants from 6 am to 6.30 pm on a land vehicle (de Silva, 2010). The vocalizations were recorded using an Earthworks QTC50 microphone shock-mounted inside a Rycote Zeppelin windshield through a Fostex FR-2 field recorder (sampling rate: 48000 Hz) connected to a 12 V lead acid battery (de Silva, 2010). A total of 2960 elephant calls were in the dataset, all annotated by SdS. This dataset contains Bark-Rumbles, Barks, Chirp-Rumbles, and Croak-Rumbles, Growls, and other sounds, but excludes trumpets, squeaks and squeals. Detailed description of the Sri Lankan elephant sounds dataset can be found in de Silva (2010).

2.2.3. Non-elephant class sound data

For the non-elephant class, we used two open-source datasets from the DCASE 2018 Challenge: warblr10k and freefield1010 datasets (Stowell et al., 2019). The Warblr10k dataset consists of smartphone recordings of non-specific sounds such as weather noise, traffic noise, and human sounds, whilst the freefield1010 dataset contains excerpts from field recordings of bird sounds around the world (Stowell et al., 2019). We did not collect negative classes for the dataset from the same study sites; however, it is a recommended and established procedure to do so. Since the application for the detection of elephants will be in other sites with more anthropogenic influence, we opted to use open-source datasets for this study. A follow-up test using one of the models from this study was conducted by Loo et al. (2024) by incorporating sound clips from tropical rainforest without elephants as negative classes, which showed that the model had indeed learnt useful features for the detection of elephant calls in the target domain.

2.3. Training and testing CNN models

2.3.1. Datasets preprocessing for model training

The elephant vocal data and non-elephant sound data were organized into three datasets: TL_DS (2960 *E.m.maximus* sounds +2960 non-elephant sounds), MEME_DS (520 *E.m.indicus* sounds +520 non-elephant sounds), and TL-MEME_DS which is the combination of the first and second datasets from above. The non-elephant sounds were randomly picked from the warblr10k dataset for TL_DS and the freefield1010 dataset for MEME_DS using a stratified sampling method in Python's *NumPy* module (Harris et al., 2020).

We preprocessed the data by applying batch normalization and data augmentation. Batch normalization ensures all audio files have the same dimensions to make the CNN process faster and more stable through rescaling and recentering by adjusting the inputs to each layer, which reduces the internal covariate shift of the network (Ioffe and Szegedy, 2015). We use stereo clips during analysis because the model only accepts inputs in the same dimensions. We then standardized the sampling rate by resampling all the audio data to the same sampling rate of 44,100 Hz and resized all the audio files to the same length (10 s). Silence was added on both sides of the file (beginning and end), to center the elephant sound (Appendix A3, Fig. A3.1).

We then augmented the normalized data using a time-shift technique by randomly shifting the audio in the temporal domain (Doshi et al., 2021; Piczak, 2015). We converted the first augmented data into mel spectrogram, a graphical representation of the sound (with frequency, time and decibels), before applying the SpecAugment technique (Park et al., 2020) to mask random frequencies by adding horizontal bars on top of the spectrogram and mask random time steps by using vertical bars (Appendix A3, Fig. A3.2). The augmentation process resulted in a total of 5920 images ready to be fed to the model for training.

2.4. Convolutional Neural Networks (CNNs)

2.4.1. CNNs architecture

Convolutional neural networks (CNNs) are a deep learning architecture specialized in data types that exhibit grid-like structures (Aggarwal, 2018; Premarathna et al., 2020). In a typical image-based application, the convolutional layer is an essential part of a CNN architecture that carries out feature extraction (Jogin et al., 2018). We used four blocks of convolutional layers followed by ReLU activation in each layer (Dubey and Jain, 2019; Fukushima, 1969). Then, we applied batch normalization to stabilize and speed up the training process by adjusting the inputs to each layer. After the images were processed by the CNN layers, the outputs were downsampled in the adaptive layer to reduce the overall computational cost while retaining the important characteristics of the extracted features. Then, they were fed into a final linear layer to predict the target classes, using cross-entropy loss together with the softmax function to calculate the probabilities. Fig. 1 shows the methods framework and schematic diagram of the CNN architecture. The model summary, including output dimensions of each layer, can be found in Supplementary Materials (Table A5.1).We used an audio classification architecture and framework on GitHub (Doshi et al., 2021) to develop our model using a specialized neural network development framework in Python called *PyTorch* (Paszke et al., 2019). The

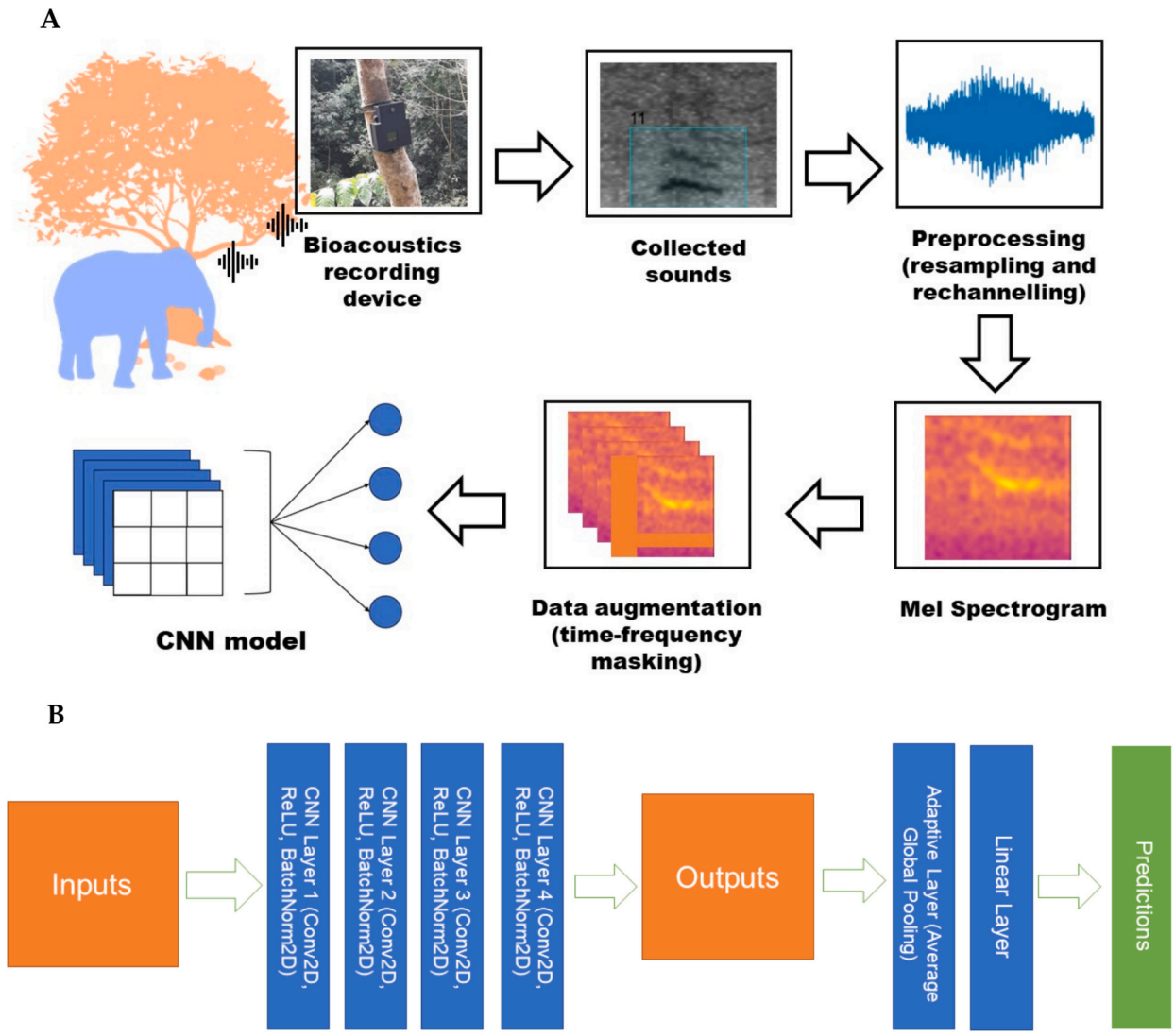

Fig. 1. (A) The methods framework and (B) The CNN architecture with four convolutional layers, one adaptive layer, and a linear layer for prediction. Code info in Appendix A2.

code used in this paper is provided (Avicena, 2020) and can be found in Appendix A2.

2.4.2. *Model training*

We utilized NVIDIA Tesla M60 GPU (6 cores GPU, 56GB RAM, and 360GB disk) through Microsoft Azure Machine Learning Studio to train the CNN models. We aimed to achieve at least 95 % average accuracy and <0.2 average loss for our final model.

The TL_DS dataset was divided into various sizes of equal distribution between positive and negative classes (CNN1 to CNN10, Table 1) to

Table 1
CNN models with datasets, sizes and performance.

Model	Dataset	Dataset sizes (elephant sound + non-elephant sound)	Average training loss	Average test loss	Average training accuracy	Average test accuracy
CNN1	TL_DS	54 files (27 + 27)	1.771	1.830	67.07 %	59.53 %
CNN2	TL_DS	74 files (37 + 37)	1.772	1.871	65.73 %	55.75 %
CNN3	TL_DS	150 files (75 + 75)	1.407	1.313	76.76 %	74.80 %
CNN4	TL_DS	300 files (150 + 150)	1.189	1.131	80.22 %	81.50 %
CNN5	TL_DS	600 files (300 + 300)	0.898	0.856	85.93 %	87.18 %
CNN6	TL_DS	1200 files (600 + 600)	0.597	0.519	91.42 %	93.05 %
CNN7	TL_DS	2400 files (1200 + 1200)	0.364	0.275	94.47 %	96.28 %
CNN8	TL_DS	3600 files (1800 + 1800)	0.240	0.158	96.47 %	97.71 %
CNN9	TL_DS	4800 files (2400 + 2400)	0.199	0.130	96.88 %	97.94 %
CNN10	TL_DS	5920 files (2960 + 2960)	0.167	0.095	97.37 %	98.45 %
Additional models						
CNN11	MEME_DS	1040 files (520 + 520)	0.839	0.747	86.57 %	89.97 %
CNN12	TL-MEME_DS	6960 files (3480 + 3480)	0.226	0.158	95.63 %	96.78 %

train the CNN algorithm and make comparisons between. MEME_DS was used for inference on models trained with the TL_DS dataset and we trained an additional model with it to gauge its performance (CNN11). Lastly, the overall combined dataset TL-MEME_DS was used to train another model (CNN12) to examine its performance.

We applied the k-fold cross-validation method during training for model performance estimation (Fushiki, 2011). Datasets were split into 10 folds of equal size, whereby nine folds were used for training the model and one-fold was used for testing, and the order was swapped until all data were used for training and testing. For each fold, we set each batch size to include 128 images to run through 10 epochs. In our case, we chose 10 epochs in our experiment following multiple tests with various epoch numbers, where we found that the model stopped improving after 10 epochs.

We used accuracy and loss metrics to evaluate the models with respect to training and validation sets (Aggarwal, 2018). The cross-entropy loss function (range: 0–1) was used to calculate the loss metrics (Appendix A4), whereby 0 is closest to the truth (Aggarwal, 2018). We compared the performance of the model after undergoing 10 epochs. We used R statistical software version 4.2.2 (R Core Team, 2022) to carry out Pearson's correlation for dataset sizes with accuracy and loss respectively.

2.4.3. Model validation

We used four evaluation metrics: accuracy, precision, recall, and F1 scores (Appendix A4) (Aggarwal, 2018). Accuracy is the percentage of correct predictions when compared to the total number of predictions. When used with a balanced dataset, with equal number of positive and negative detections, it is less prone to bias, and the accuracy score reflects well the performance of the model. Precision is the ratio between true positives and the sum of true positives and false positives, while recall is the ratio between true positives and the sum of true positives and false negatives. The F1 score is the harmonic mean of precision and recall.

2.5. Comparison of E.m.indicus acoustic parameters with E.m.maximus

We compared the acoustic parameters of *E.m. indicus* and *E.m. maximus* using Principal Component Analysis (PCA) with a subset of 44 sounds from *E.m.indicus* (Malaysia) and 70 sounds from *E.m.maximus* (Sri Lanka) of different call types. Only elephant vocalizations that did not overlap with other non-target signals were selected for analysis. There were 15 acoustic parameters measured (Appendix A5, Table A5.1) using Raven Pro (Ver. 2.0, K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology). Using individual PCA loadings (PC1 to PC4 in turn) as dependent variables, comparison was made between *E.m. indicus* and *E.m. maximus* acoustics using linear models. The PCA were carried out using the *factoextra* and linear model analysis using *stats* packages respectively in R statistical software version 4.2.2 (R Core Team, 2022).

3. Results

3.1. The use of deep learning and CNNs for processing bioacoustics data

An initial search on the keyword "deep learning" on Web of Science revealed 281,753 scientific and proceeding papers. When filtered with keywords "wildlife OR biodiversity", the hits dropped to 890, indicating the publications involving deep learning models for wildlife and biodiversity research is about 0.32 % in comparison to other fields of study such as computer science, medical, engineering, telecommunications, and others. When the results were further filtered using keywords "bioacoustics OR sound OR vocal OR calls", the number of search results dropped to 83, with one paper each on Asian elephants and African elephants. From the 83 papers, 28 utilized the CNN approach to study the bioacoustics of birds, frogs, soundscapes (environmental sounds), bats, whales, dolphins, fish, gibbons, and other wildlife.

The overall trend for publications on deep learning shows 83.50 % of the papers were published in the last five years. When we examined across all fields of research, how many "deep learning" scientific and proceeding papers were on "bioacoustics OR sound OR vocal OR calls", this resulted in 3904 hits, whereby 1107 used the CNN approach.

3.2. Performance of CNN models on training and evaluation sets

Our *E.m.maximus* CNN models' performance increased steadily with dataset sizes (Table 1, Fig. 2). From **CNN8** (1800 *E.m.maximus* sounds +1800 non-elephant sounds) and onwards, the model met the benchmark set in this study (**CNN8**: train accuracy: 96.47 %, test accuracy: 97.71 %, test loss: 0.158; **CNN9** (2400 + 2400): train accuracy: 96.88 %, test accuracy: 97.94 %, test loss: 0.130; **CNN10** (2960 + 2960): train accuracy: 97.37 %, test accuracy: 98.45 %, test loss: 0.095). Across 10 epochs of training, the learning curves (accuracy and loss) show various degrees of convergence (Appendix A6, Fig. A6.1 and Fig. A6.2). From Pearson's correlation, there are significant positive relationships between dataset sizes and train-test accuracy (**train accuracy**: $r = 0.81$, 95 % C.I. [0.37,0.95], df = 8, p-value = 0.004; **test accuracy**: $r = 0.75$, 95 % C.I. [0.22,0.94], df = 8, p-value = 0.013); and significant negative relationships between dataset sizes and train-test loss (**train loss**: $r = -0.86$, 95 % C.I. [−0.97, −0.49], df = 8, p-value = 0.002; **test loss**: $r = -0.84$, 95 % C.I. [−0.96, −0.44], df = 8, p-value = 0.002).

The performance of CNN11 (MEME_DS, 520 + 520), trained on *E.m. indicus* sounds collected in Malaysia, rank predictably between CNN5 (300 + 300) and CNN6 (600 + 600), with a train accuracy of 86.57 %, test accuracy of 89.97 % and test loss of 0.747. Whilst the overall model CNN12 (TL_MEME_DS (3480 + 3480), trained on both *E.m.indicus* and *E.m.maximus* sound files, performed lower than CNN10 despite having bigger data size (Table 1).

3.3. Inference on MEME_DS

The dataset sizes for models (CNN1 to CNN10) trained with *E.m. maximus* (TL_DS) vocalization had little effect on the model performances when performing inference on unseen data from MEME_DS (520 *E.m.indicus* sounds +520 non-elephant sounds), as models achieved accuracy ranging from 62 % to 78 %., precision (0.66–0.83), recall value (0.62–0.78), and F1 scores (0.64–0.81) (Table 2). The confusion matrices for the models are illustrated in the Appendix, Fig. A6.3.

3.4. Comparison of acoustic parameters between E.m.indicus and E.m. maximus

The principal component analysis (PCA) comparison of 15 acoustic parameters (Appendix A5, Table A5.1 & Table A5.2) from *E.m.indicus* and *E.m.maximus,* found four principal components (PC1 to PC4) that contributed to 86.99 % of the variation explained, with the acoustic space of *E.m.indicus* and *E.m.maximus* generally overlapping (Appendix A5, Table A5.3). When using linear models to test the PC scores of the sound samples, we found that *E.m.indicus* and *E.m.maximus* elephant sounds were significantly different in **PC1** ($F_{(1,112)} = 14.23$, p-value<0.001; $\text{Intercept}_{\text{E.m.indicus}} = 1.186$, 95 % C.I.[0.391, 1.982], $t = 2.96$, p-value = 0.004; $ß_{\text{E.m.maximus}} = -1.932$, 95 % C.I.[−2.947, −0.917], $t = -3.77$, p-value<0.001) and **PC4** ($F_{(1,112)} = 11.40$, p-value = 0.001; $\text{Intercept}_{\text{E.m.indicus}} = 0.392$, 95 % C.I.[0.098, 0.686], $t = 2.65$, p-value = 0.009; $ß_{\text{E.m.maximus}} = -0.639$, 95 % C.I.[−1.013, −0.264], $t = -3.38$, p-value = 0.001), but not in **PC2** and **PC3.** The variables contributing to **PC1** are Center Frequency (Hz), Frequency 5 % (Hz), Frequency 25 % (Hz), Frequency 75 % (Hz), Frequency 95 % (Hz), Max Frequency (Hz), and Peak Frequency (Hz); while the variables contributing to **PC4** are Center Time (s), -Delta FrEq. (Hz), and Peak Time Relative (Appendix A5, Fig. A5.1).

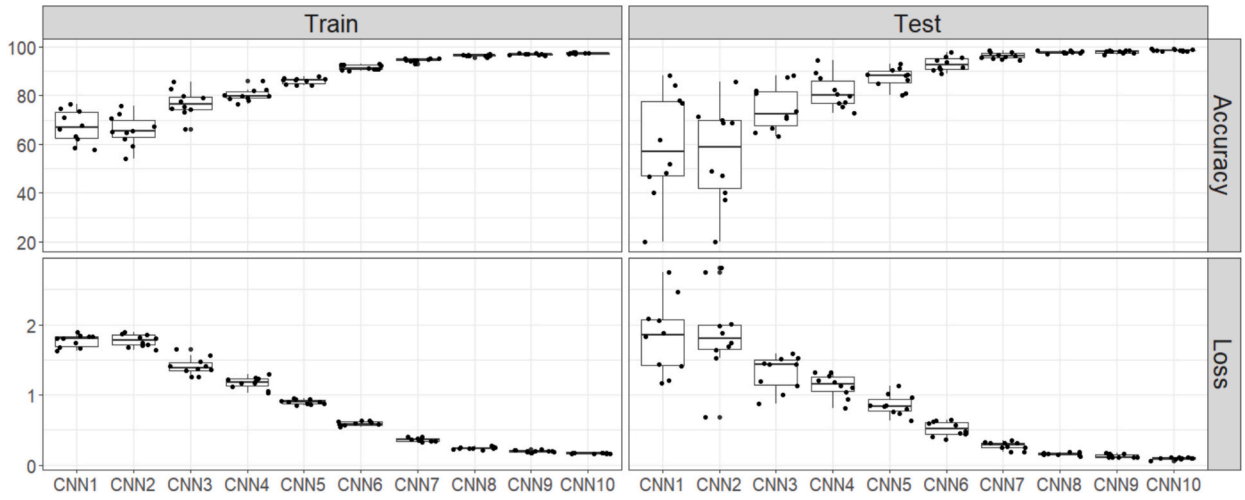

Fig. 2. Test and train accuracy with test and train loss across 10 folds validation for CNN1 to CNN10.

Table 2
Inference results on MEME_DS (*E.m.indicus*) for CNN1-CNN10 models build from TL_DS (*E.m.maximus*) Sri Lanka.

Models from TL_DS *E.m.maximus* sounds	Accuracy	Precision	Recall	(Precision-recall)	F1-Score
CNN1 (27 elephant +27 non-elephant files)	63.46 %	0.72	0.63	0.09	0.67
CNN2 (37 elephant +37 non-elephant files)	78.27 %	0.83	0.78	0.05	0.81
CNN3 (75 elephant +75 non-elephant files)	72.50 %	0.82	0.72	0.10	0.77
CNN4 (150 elephant +150 non-elephant files)	70.58 %	0.76	0.70	0.06	0.73
CNN5 (300 elephant +300 non-elephant files)	65.48 %	0.71	0.65	0.06	0.68
CNN6 (600 elephant +600 non-elephant files)	62.60 %	0.66	0.62	0.04	0.64
CNN7 (1200 elephant +1200 non-elephant files)	64.23 %	0.67	0.64	0.03	0.65
CNN8 (1800 elephant +1800 non-elephant files)	65.67 %	0.67	0.65	0.02	0.66
CNN9 (2400 elephant +2400 non-elephant files)	69.13 %	0.71	0.69	0.02	0.70
CNN10 (2960 elephant +2960 non-elephant files)	67.40 %	0.68	0.67	0.01	0.67

4. Discussion

The use of convolutional neural networks (CNNs) is yet to be fully explored for bioacoustics of wildlife and environment (28 out of 281,753 studies). Although deep learning models have become popular in the last five years, their application for bioacoustics study in the field of wildlife and biodiversity is still limited. In this study, we investigated the correlation between training data sizes with accuracy and loss performance, and subsequently assessed the generalization capability of models trained with *Elephas maximus maximus* in relation to unseen data from *Elephas maximus indicus*. Additionally, we explored the vocal characteristics between the two elephant subspecies.

Our results show that elephant vocalizations can be successfully classified by CNN-based architecture. Models that were trained on smaller datasets achieved poorer performance levels whilst models with larger datasets achieved better and more stable performances (Fig. 2). Four of our models (**CNN8**–1800 elephant sounds, **CNN9**–2400 elephant sounds, **CNN10**–2960 elephant sounds, and **CNN12**–3480 elephant sounds) exceeded our benchmark performance with CNN10 being our best-performing model with 98.45 % test accuracy. Usually, models can be fine-tune by adjusting the hyperparameters such as by increasing the epochs and the convolutional layers.

We found a significant positive relationship between dataset sizes for training and model test performance (Table 1, Fig. 2). Generally, deep learning models will have higher performance levels when trained on larger and more diverse datasets (Ribeiro et al., 2020) and when the data is labelled adequately (Karimi et al., 2020; Ribeiro et al., 2020). For instance, our smallest model surpassing 95 % accuracy (**CNN8**: train accuracy: 96.47 %, test accuracy: 97.71 %), required 1800 labelled elephant sound files. Potentially, transfer learning – a technique involving fine-tuning existing models developed by others (Christin et al., 2019; de Silva et al., 2022; Emmert-Streib et al., 2020; Ghani et al., 2023) – can be explored in future research. Usually, tweaks to the models can be made by fine-tuning the hyperparameters such as by increasing the epochs and the convolutional layers to increase learning. There are existing large state-of-the-art models, such as the VGG16, a deep CNN consisting of 138 million parameters trained on 14 million images belonging to 22,000 classes, that are available for wildlife conservationists to utilize (Emmert-Streib et al., 2020).

Our CNN models built with only *E.m.maximus* vocalizations (CNN1 to CNN10), although achieved high test accuracy, did not perform well regardless of training data size, when inferred with unseen *E.m.indicus* vocalizations. The CNNs have difficulty in differentiating unseen *E.m. indicus* sounds from Malaysia, resulting in higher false negatives and average recall values between 0.62 and 0.78 (Table 2 & Appendix A6, Fig. A6.3). In a follow up study, we found CNN models built with only *E. m.indicus* vocalization collected from several localities in Malaysia, performed better when inferred with unseen *E.m.indicus* vocalization from Malaysia (Loo et al., 2024), highlighting the importance of local data in training and improving CNN models. When the vocal acoustics parameters were compared with PCA, we found that the sound characteristics between Asian elephants from Sri Lanka and Malaysia, overlapped highly (Appendix A5, Table A5.3 & Fig. A5.1), with PC1 and PC4 showing significant differences (p-values <0.05) between the two subspecies, but not so for PC2 and PC3. Due to data limitation, we were unable to control for age groups, male or female, individuals, and morphology (body size) in the comparison, which we hope future research can explore further.

Previously, a direct comparison on the percentage of call types for Asian elephants in Sri Lanka and India had detected differences (de Silva, 2010). A larger comparative study, found differences in call combination, order and frequency of call types between Asian elephants and the two African elephant species, and between different populations of the same elephant species (Pardo et al., 2019). These findings raise the question whether differences in elephant vocalizations indicate differences in syntax or dialects. There are studies describing animal vocal learning abilities of other taxa such as birds (Henry et al., 2015), whales (Ford, 2009), and primates (Zürcher et al., 2019). As dialects are often attributed to developmental learning (ontogeny) instead of genetic origin, akin to human culture, there are possibilities of differences in vocalization behavior arising between populations in distinct localities (Ford, 2009; Henry et al., 2015; Zürcher et al., 2019). It is argued that differences in vocalization behavior between elephant populations are associated with the function of the calls in relation to site-specific conditions including habitat types (i.e. forest or open grasslands) and presences of threats in the surrounding areas (de Silva, 2010; Pardo et al., 2019). Hence, it is entirely possible that the differences in *E.m. maximus* and *E.m.indicus* sounds detected in this study may not be due to phylogenetic or genetic differentiation of these populations, but instead caused by differences in call usage due to habitat (i.e. open grassland in Sri Lanka and forest in Malaysia), differences in the recording conditions between the two countries, or both. Additionally, a study on African elephants found that an elephant can have distinct calls to address other elephants individually, which opens up a world of possibilities for the study of elephant communication (Pardo et al., 2024).

Nonetheless, the differences in call frequencies or order should not affect the performance of the deep learning models if the finer acoustics parameters of the calls between populations are similar, and when there are sufficient samples of different sound quality and call types included in training data and converted to mel spectrograms. The differences detected by the PCA on sound acoustics parameters of *E.m.maximus* in Sri Lanka and *E.m.indicus* in Malaysia, together with poor inference by the CNN models across subspecies, suggest the need for further investigation. Furthermore, considering that the physical size of *E.m.maximus* does differ from *E.m.indicus*, there could be underlying mechanism (causal) explanation at play, that could influence differences in vocal acoustics parameters between these two subspecies. All these are interesting possibilities for future investigation. Future studies using standardized methods and equipment would be required to examine this area of research further.

CNN models are fundamental for many automated early warning detection systems for human-elephant conflict mitigation due to their powerful classification ability of lightly preprocessed data, which is evident in their use case in multiple conflict-prone areas (Gunasekara et al., 2021; Premarathna et al., 2020). CNN-based early warning systems have been mainly applied for visually detecting elephants, for example, in prevention of elephant-vehicle collisions (Gunasekara et al., 2021). Applications using bioacoustics, however, are scarce despite their advantage of covering larger areas with fewer devices.

However, building a sound-based early warning system for HEC mitigation comes with its own unique challenges. First, the diversity of elephant vocalization types might introduce signals that have not been included in the model training datasets, potentially causing false negatives. Second, external noises such as airplane or car sound, can potentially lead to false positives because their harmonic structure is similar to elephant rumbles (Zeppelzauer et al., 2015). These two problems can be overcome by having more sound diversity in the training datasets. Local data, in this case, is preferable to increase the performance of the model and help account for potential heterogeneity in the soundscape where the early warning systems will be deployed. Additionally, having some manual verifications from time-to-time and returning feedback to the system will allow the model to perform better in a local context.

Another important challenge to address is to solve potential hardware constraints. Sound data are particularly memory-intensive and processing them on-the-fly for automated detection requires a computationally strong processor. Additionally, this hardware must be energy efficient enough to process such data and ideally connected to a network such as cellular or WIFI signals to issue alarms when elephants are detected nearby, which can be a difficult task since many HEC happens in rural areas where access to networks and electricity might be scarce (Matsuura et al., 2024; Sampson et al., 2021). Memory constraints can be addressed, for example, by temporarily storing the full data only for sound processing, retaining important signals for further research purposes, and deleting redundant data. Moreover, a continuous supply of electricity, such as solar energy, can also be embedded into the hardware, however, this might not entirely be feasible in areas with thick vegetation and scarce sunlight due to clouds and rain. However, there are such systems already available for diverse purposes, although publications in this field are limited, as bioacoustics research is not yet as common as visual-based camera trap studies. Some notable examples include Rainforest Connection by Topher White (www.rfcx.org), Rainforest Listening (www.rainforestlistening.com) by Rainforest Partnership, and Amrita Elephant Watch by AMMACHI Labs (www.amma.org).

Since elephants are endangered species targeted by poachers, the implementation of early warning systems must minimize the risk of accidentally revealing locations of elephants to poachers. The future system can be run independently by a trusted entity or be incorporated into wildlife conservation and enforcement initiatives in protected areas and reserves (Wich and Piel, 2021). Safeguard measures can be taken, for example, the location of the endangered animal can be automatically scrambled over a given radius before being shared with stakeholders on the ground so that the risk of poaching is reduced, while providing sufficient alerts to increase the safety of people. This system can be used to help foster collaboration between communities, governmental and enforcement agencies, plantations, and wildlife conservation organizations.

In community areas, there are social and ethical concerns raised over the privacy of people captured in photos or videos without their consent (Sharma et al., 2020). This evokes the discussion for a code of conduct for researchers, and the use of AI to automatically blur people's faces and still allow data to be used for research (Sharma et al., 2020). Certification of standards and social impact studies involving indigenous communities should be emphasized on free, prior, and informed consent before or at the start of the project, to ensure the communities are aware of the purpose of the project and are given the chance to clarify any concerns. Some of these concerns are also applicable to bioacoustics research.

Overall, our study offers the framework to use a deep learning approach for processing bioacoustics data which have several potential real-world applications, from preventing elephant-vehicle collision to managing human-elephant conflict, and promoting safer human-elephant coexistence. Further research is needed to test the model in real-world scenarios and to create models with a low proportion of false negatives, which can be detrimental if such algorithms were to be used for early warning systems. Additionally, engineering works are needed to design a working low-cost, low-energy prototype for use.

5. Conclusion

The application of sound-based deep learning models can have important implications for human-elephant conflict mitigation. Our study provides the first step in developing an early warning system to detect elephants based on their calls. Further research will be needed on testing the robustness of the model in real-world scenarios and building low-cost efficient prototypes for implementation in conflict-prone areas. Ultimately, the development and effective implementation of such systems will require close collaboration between researchers, government agencies, and the impacted people alike, and we encourage all stakeholders to collaborate to ensure positive outcomes from the development of the model and future systems.

The use of AI in wildlife conservation has the potential to improve or add new dimensions to ecological studies and assist in processing a large amount of data. Technologies using AI are now rapidly influencing and changing human societies in various fields, from mass communication, entertainment, writing, art, autonomous vehicles to medical diagnosis. Advances in AI have the potential to support conservation of endangered species.

CRediT authorship contribution statement

Naufal Rahman Avicena: Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft. **Yen Yi Loo:** Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Writing – review & editing, Writing – original draft. **Tomas Maul:** Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Conceptualization, Writing – review & editing. **Noah Thong:** Project administration, Methodology, Investigation, Formal analysis, Data curation, Writing – review & editing. **Christopher Chai Thiam Wong:** Supervision, Resources, Investigation, Data curation, Writing – review & editing. **Shermin de Silva:** Resources, Investigation, Data curation, Writing – review & editing, Writing – original draft. **Salman Saaban:** Supervision, Resources, Project administration, Investigation, Writing – review & editing. **Ee Phin Wong:** Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft.

Funding information

This work was supported by Sime Darby Foundation [NVHH0007 & NVLO0001]; and Microsoft AI for Earth.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Noah Thong, Yen Yi Loo reports financial support was provided by Sime Darby Foundation. Ee Phin Wong reports a relationship with Sime Darby Foundation that includes: funding grants. Tomas Maul reports a relationship with Microsoft AI for Earth that includes: funding grants. We do not receive any commercial benefit from the publication of this paper, not the authors nor the organizations they are affiliated, or the funders involved. There is no conflict of interest as there are no consultancies nor commercialization involved at this moment. We have tried our best to present our research as neutral and as fair as possible to both elephants and communities involved. We are not editors for any journals linked to Biological Conservation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful for the support given by Sime Darby Foundation and Microsoft AI for Earth. We thank the Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN), Forestry Department of Peninsular Malaysia (Perhutanan), and Forestry Department of Perak (Perhutanan Perak) for their support in our study and for granting us research permits. We are also thankful for the immense help from our interns for sound annotation: Ashraft Yusni and Rehannah Zelda, field assistants: Param bin Pura, Sudin A/L Din, and Muhammad Tauhid bin Tunil, and colleagues especially Lim Jia Cherng, Praveena Chackrapani, Chan Yik Khan, Muhammad Amin Rusli, Cedric Tan Kai Wei, and Amir Aminuddin for their thorough support throughout the studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2025.111272.

Data availability

Data will be made available on request.

References

Aggarwal, C.C., 2018. Neural networks and deep learning: a textbook. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-94463-0.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 53. https://doi.org/10.1186/s40537-021-00444-8.

Avicena, N.R., 2020. CNN-based bioacoustics classification of elephant and non-elephant sounds. [online] URL: available on GitHub. https://github.com/aalavicena/Deep-Learning-Elephant-Bioacoustics.

Barua, M., 2010. Whose issue? Representations of human-elephant conflict in Indian and international media. Sci. Commun. 32, 55–75. https://doi.org/10.1177/1075547009353177.

Bjorck, J., Rappazzo, B., Chen, D., Bernstein, R., Wrege, P., Gomes, C., 2019. Automatic detection and compression for passive acoustic monitoring of the African Forest elephant. Proc. AAAI Conf. Artif. Intell. 33, 476–484. https://doi.org/10.1609/aaai.v33i01.3301476.

Browning, E., Gibb, R., Glover-Kapfer, P., Jones, K.E., 2017. Passive acoustic monitoring in ecology and conservation. (Report). WWF-UK. https://doi.org/10.25607/OBP-876.

Calabrese, A., Calabrese, J.M., Songer, M., Wegmann, M., Hedges, S., Rose, R., Leimgruber, P., 2017. Conservation status of Asian elephants: the influence of habitat and governance. Biodivers. Conserv. 26, 2067–2081. https://doi.org/10.1007/S10531-017-1345-5/METRICS.

Christin, S., Hervet, É., Lecomte, N., 2019. Applications for deep learning in ecology. Methods Ecol. Evol. 10, 1632–1644. https://doi.org/10.1111/2041-210X.13256.

Colonna, J., Peet, T., Ferreira, C.A., Jorge, A.M., Gomes, E.F., Gama, J., 2016. Automatic Classification of Anuran Sounds Using Convolutional Neural Networks, in: Proceedings of the Ninth International C* Conference on Computer Science & Software Engineering, C3S2E '16. Association for Computing Machinery, New York, NY, USA, pp. 73–78. https://doi.org/10.1145/2948992.2949016.

de la Torre, J.A., Wong, E.P., Lechner, A.M., Zulaikha, N., Zawawi, A., Abdul-Patah, P., Saaban, S., Goossens, B., Campos-Arceiz, A., 2021. There will be conflict – agricultural landscapes are prime, rather than marginal, habitats for Asian elephants. Anim. Conserv. 24, 720–732. https://doi.org/10.1111/ACV.12668.

de Silva, E.M.K., Kumarasinghe, P., Indrajith, K.K.D.A.K., Pushpakumara, T.V., Vimukthi, R.D.Y., de Zoysa, K., Gunawardana, K., de Silva, S., 2022. Feasibility of using convolutional neural networks for individual-identification of wild Asian elephants. Mamm. Biol. 102, 909–919. https://doi.org/10.1007/s42991-021-00206-2.

de Silva, S., 2010. Acoustic communication in the Asian elephant, Elephas maximus maximus. Behaviour 147, 825–852.

Doshi, R., Chen, Y., Jiang, L., Zhang, X., Biadsy, F., Ramabhadran, B., Chu, F., Rosenberg, A., Moreno, P.J., 2021. Extending Parrotron: An End-to-End, Speech Conversion and Speech Recognition Model for Atypical Speech, in: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6988–6992. https://doi.org/10.1109/ICASSP39728.2021.9414644.

Dubey, A.K., Jain, V. 2019. Comparative study of convolution neural network's Relu and leaky-Relu activation functions, in: Mishra, S., Sood, Y.R., Tomar, A. (Eds.), Applications of computing, automation and wireless systems in electrical engineering, Lecture Notes in Electrical Engineering. Springer, Singapore, pp. 873–880. https://doi.org/10.1007/978-981-13-6772-4_76.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., Dehmer, M., 2020. An introductory review of deep learning for prediction models with big data. Front. Artif. Intell. 3.

Fang, F., Tambe, M., Dilkina, B., Plumptre, A.J., 2019. Artificial intelligence and conservation. Cambridge University Press.

Fernando, P., Pastorini, J., 2011. Range-wide status of Asian elephants. Gajah 35, 15–20.

Ford, J.K.B. 2009. Dialects, in: Perrin, W.F., Würsig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of marine mammals (Second Edition). Academic Press, London, pp. 310–311. https://doi.org/10.1016/B978-0-12-373553-9.00075-4.

Fukushima, K., 1969. Visual feature extraction by a multilayered network of analog threshold elements. IEEE Trans. Syst. Sci. Cybernet. 5, 322–333. https://doi.org/10.1109/TSSC.1969.300225.

Fushiki, T., 2011. Estimation of prediction error by using K-fold cross-validation. Stat. Comput. 21, 137–146. https://doi.org/10.1007/s11222-009-9153-8.

Ghani, B., Denton, T., Kahl, S., Klinck, H., 2023. Global birdsong embeddings enable superior transfer learning for bioacoustics classification. Sci. Rep. 13, 22876 (2023). https://doi.org/10.1038/s41598-023-49989-z.

Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., Joly, A., 2016. LifeCLEF bird identification task 2016: The arrival of deep learning. In: CLEF: Conference and Labs of the Evaluation Forum. Évora, Portugal, pp. 440–449.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press.
Gunasekara, S., Jayasuriya, M., Harischandra, N., Samaranayake, L., Dissanayake, G., 2021. A convolutional neural network based early warning system to prevent elephant-train collisions, in: 2021 IEEE 16th international conference on industrial and information systems (ICIIS). In: Presented at the 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), pp. 271–276. https://doi.org/10.1109/ICIIS53135.2021.9660651.
Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2.
Henry, L., Barbu, S., Lemasson, A., Hausberger, M., 2015. Dialects in animals: evidence, development and potential functions. AB&C 2, 132–155. https://doi.org/10.12966/abc.05.03.2015.
Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd international conference on machine learning. In: Presented at the International Conference on Machine Learning, PMLR, pp. 448–456.
IPBES, 2019. Global Assessment Report on Biodiversity and Ecosystem Services. IPBES.
IUCN, 2023. IUCN SSC Guidelines on Human-Wildlife Conflict and Coexistence. IUCN SSC Human-Wildlife Conflict & Coexistence Specialist Group, Gland, Switzerland.
Jogin, M., Mohana M.S., Madhulika, Divya, G.D., Meghana, R.K., Apoorva, S., 2018. Feature extraction using convolution neural networks (CNN) and deep learning, in: 2018 3rd IEEE international conference on recent trends in electronics, Information & Communication Technology (RTEICT). In: Presented at the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 2319–2323. https://doi.org/10.1109/RTEICT42901.2018.9012507.
Kansky, R., Knight, A.T., 2014. Key factors driving attitudes towards large mammals in conflict with humans. Biol. Conserv. 179, 93–105. https://doi.org/10.1016/j.biocon.2014.09.008.
Karimi, D., Peters, J.M., Ouaalam, A., Prabhu, S.P., Sahin, M., Krueger, D.A., Kolevzon, A., Eng, C., Warfield, S.K., Gholipour, A., 2020. Learning to Detect Brain Lesions from Noisy Annotations, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Presented at the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1910–1914. https://doi.org/10.1109/ISBI45749.2020.9098599.
Keen, S.C., Shiu, Y., Wrege, P.H., Rowland, E.D., 2017. Automated detection of low-frequency rumbles of forest elephants: a critical tool for their conservation. J. Acoust. Soc. Am. 141 (4), 2715–2716. https://doi.org/10.1121/1.4979476.
Kumar, Ananda, Raghunathan, G., 2014. Fostering human-elephant coexistence in the Valparai landscape, Annamalai Tiger Reserve, Tamil Nadu., in: Human-wildlife conflict in the mountains of SAARC region: compilation of successful management strategies and practices. South Asian Association for Regional Cooperation (SAARC) Forestry Centre, Thimphu, Bhutan.
Loo, Y.Y., Avicena, N.R., Thong, N., Marghoobul Haque, A., Nhlabatsi, Y.T., Yousif Abdalla Abakar, S., Ng, K.H., Wong, E.P., 2024. WildTechAlert: Deep learning models for real-time detection of elephant presences using bioacoustics in an early warning system to support human-elephant coexistence, in 2023 13th International Conference on Brain Inspired Cognitive Systems proceedings. https://doi.org/10.1007/978-981-97-1417-9_36.
Matsuura, N., Nomoto, M., Terada, S., Yobo, C.M., Memiaghe, H.R., Moussavou, G.-M., 2024. Human-elephant conflict in the African rainforest landscape: crop-raiding situations and damage mitigation strategies in rural Gabon. Front. Conserv. Sci. 5. https://doi.org/10.3389/fcosc.2024.1356174.
Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., Poria, S., 2023. A review of deep learning techniques for speech processing. Inform. Fusion 99, 101869. https://doi.org/10.1016/j.inffus.2023.101869.
Nair, S., Balakrishnan, R., Seelamantula, C.S., Sukumar, R., 2009. Vocalizations of wild Asian elephants (Elephas maximus): structural classification and social context. J. Acoust. Soc. Am. 126, 2768–2778. https://doi.org/10.1121/1.3224717.
Nolasco, I., Singh, S., Morfi, V., Lostanlen, V., Strandburg-Peshkin, A., Vidaña-Vila, E., Gill, L., Pamula, H., Whitehead, H., Kiskin, I., Jensen, F.H., Morford, J., Emmerson, M.G., Versace, E., Grout, E., Liu, H., Ghani, B., Stowell, D., 2023. Learning to detect an animal sound from five examples. Eco. Inform. 77, 102258. https://doi.org/10.1016/j.ecoinf.2023.102258.
Pardo, M.A., Poole, J.H., Stoeger, A.S., Wrege, P.H., O'Connell-Rodwell, C.E., Padmalal, U.K., de Silva, S., 2019. Differences in combinatorial calls among the 3 elephant species cannot be explained by phylogeny. Behav. Ecol. 30, 809–820. https://doi.org/10.1093/beheco/arz018.
Pardo, M.A., Fristrup, K., Lolchuragi, D.S., Poole, J.H., Granli, P., Moss, C., Douglas-Hamilton, I., Wittemyer, G., 2024. African elephants address one another with individually specific name-like calls. Nat. Ecol. Evol. 1–12. https://doi.org/10.1038/s41559-024-02420-w.
Park, D.S., Zhang, Y., Chiu, C.-C., Chen, Y., Li, B., Chan, W., Le, Q.V., Wu, Y., 2020. Specaugment on Large Scale Datasets, in: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6879–6883. https://doi.org/10.1109/ICASSP40776.2020.9053205.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.
Piczak, K.J., 2015. Environmental sound classification with convolutional neural networks, in: 2015 IEEE 25th international Workshop on Machine Learning for Signal Processing (MLSP). Presented at the 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6. https://doi.org/10.1109/MLSP.2015.7324337.
Premarathna, K.S.P., Rathnayaka, R.M.K.T., Charles, J., 2020. An elephant detection system to prevent human-elephant conflict and tracking of elephant using deep learning, in: 2020 5th international conference on information technology research (ICITR). In: Presented at the 2020 5th International Conference on Information Technology Research (ICITR), pp. 1–6. https://doi.org/10.1109/ICITR51448.2020.9310798.
R Core Team, 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Ramasubramanian, C., Lokiah, S., Viswanath, Y., Jamthe, S., 2022. Averting human-elephant conflict using IoT and machine learning of elephant vocalizations. In: In: 2022 IEEE 8th World Forum on Internet of Things (WF-IoT). Presented at the 2022 IEEE 8th World Forum on Internet of Things (WF-IoT). IEEE, Yokohama, Japan, pp. 1–6. https://doi.org/10.1109/WF-IoT54382.2022.10152220.
Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P.S., Andersson, C.R., Macfarlane, P.W., Meira Jr., W., Schön, T.B., Ribeiro, A.L.P., 2020. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat. Commun. 11, 1760. https://doi.org/10.1038/s41467-020-15432-4.
Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J., 2009. Planetary boundaries: exploring the safe operating space for humanity. Ecol. Soc. 14.
Sampson, C., Rodriguez, S.L., Leimgruber, P., Huang, Q., Tonkyn, D., 2021. A quantitative assessment of the indirect impacts of human-elephant conflict. PLoS One 16, e0253784. https://doi.org/10.1371/journal.pone.0253784.
Sharma, K., Fiechter, M., George, T., Young, J., Alexander, J.S., Bijoor, A., Suryawanshi, K., Mishra, C., 2020. Conservation and people: towards an ethical code of conduct for the use of camera traps in wildlife research. Ecol. Solut. Evid. 1, e12033. https://doi.org/10.1002/2688-8319.12033.
Stoeger, A.S., de Silva, S., 2014. African and Asian elephant vocal communication: a cross-species comparison. In: Witzany, G. (Ed.), Biocommunication of animals. Springer Netherlands, Dordrecht, pp. 21–39. https://doi.org/10.1007/978-94-007-7414-8_3.
Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap - PMC. PeerJ 10. https://doi.org/10.7717/peerj.13152.
Stowell, D., Wood, M.D., Pamuła, H., Stylianou, Y., Glotin, H., 2019. Automatic acoustic detection of birds through deep learning: the first bird audio detection challenge. Methods Ecol. Evol. 10, 368–380. https://doi.org/10.1111/2041-210X.13103.
Sukumar, R., 2003. The living elephants: evolutionary ecology, behavior, and conservation. Oxford University Press, New York.
Swider, C.R., Gemelli, C.F., Wrege, P.H., Parks, S.E., 2022. Passive acoustic monitoring reveals behavioural response of African forest elephants to gunfire events. Afr. J. Ecol. 60 (4), 882–894. https://doi.org/10.1111/aje.13070.
Tan, A.S.L., de la Torre, J.A., Wong, E.P., Thuppil, V., Campos-Arceiz, A., 2020. Factors affecting urban and rural tolerance towards conflict-prone endangered megafauna in peninsular Malaysia. Glob. Ecol. Conserv. 23, e01179. https://doi.org/10.1016/j.gecco.2020.e01179.
Thomas Leonid, T., Jayaparvathy, R., 2022. Classification of elephant sounds using parallel convolutional neural network. Intel. Auto. Soft Comput. 32, 1415–1426. https://doi.org/10.32604/iasc.2022.021939.
Vasudev, D., Goswami, V.R., 2020. A Bayesian hierarchical approach to quantifying stakeholder attitudes toward conservation in the presence of reporting error. Conserv. Biol. 34, 515–526. https://doi.org/10.1111/cobi.13392.
Wich, S.A., Piel, A.K., 2021. Conservation technology. Oxford University Press.
Williams, C., Tiwari, S.K., Goswami, V.R., de Silva, S., Kumar, A., Baskaran, N., Yoganand, K., Menon, V., 2020. *Elephas maximus*. The IUCN red list of threatened species.
Wrege, P.H., Rowland, E.D., Keen, S., Shiu, Y., 2017. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. Methods in Ecology and Evolution 8, 1292–1301. https://doi.org/10.1111/2041-210X.12730.
Zeppelzauer, M., Hensman, S., Stoeger, A.S., 2015. Towards an automated acoustic detection system for free-ranging elephants. Bioacoustics 24, 13–29. https://doi.org/10.1080/09524622.2014.906321.
Zürcher, Y., Willems, E.P., Burkart, J.M., 2019. Are dialects socially learned in marmoset monkeys? Evidence from translocation experiments. PLOS ONE 14, e0222486. https://doi.org/10.1371/journal.pone.0222486.