International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Acoustic Signal Detection of Search-Phase Echolocation Bat Calls with CNN

Anala M. R.*[1], Hemavathy R.[2] , Dhanush M.[3]

Submitted: 05/11/2023 **Revised**: 24/12/2023 **Accepted**: 06/01/2024

Abstract: Bats are one of the most diverse species on earth. Not only do they provide highly beneficial services to nature such as pollination and pest control, they provide useful insights on the changes occurring to the ecosystem as a result of anthropogenic change. Therefore, bat tracking helps in the conservation of endangered bat species and also to measure trends in biodiversity. Bats play a vital role in our ecosystem, however, they are least studied and marred with myths. Hence it is necessary to study and monitor their population dynamics to get an insight on roost size, mortality rate, migration pattern, breeding season, etc. Insectivorous bats use "Echolocation" to communicate with each other, to find roosts, to detect prey and obstacles while navigating in flight. This paper proposes a Convolution Neural Network (CNN) based pipeline for automatically detecting search-phase calls produced by echolocating bats in noisy, real-world recordings. Audio files are first converted into spectrogram or time-frequency representation and then denoised. The performance results of this model were compared with other existing models on different evaluation metrics like precision, recall, Receiver Operating Characteristic (ROC) curves, and Precision Recall (PR) curves. The model performed better than the existing systems on three different acoustic datasets. Around 500 more bat calls were detected across all 3 datasets compared to that of the existing systems, with a significant increase in recall of the proposed model, as high as 11%. The proposed detection system proved to be capable of detecting echolocation bat calls reliably.

Keywords: *Bioindicator, Convolutional Neural Networks, Fourier Transform, STFT, Non-Max Suppression, Bat*

1. Introduction

Bat populations are endangered across the globe. There are about 1300 species of bats in the world [1], making them one of the most diverse mammals on earth. India is home to over 119 bat species [2]. Out of which fourteen are mega or fruit bats, and the remaining are micro or insectivorous bats. The study on the diversity of bats in the western ghats of India [3] revealed that nearly 35% of the bat population is threatened. A few potential threats to bats worldwide include pesticide use, hunting, and climatic changes [4]. Bats provide crucial ecological and economic services to nature in terms of dispersing of seeds, pollination, and pest control [5]. They are said to be one of the most useful non-domesticated mammals in terms of the economic value they provide [6].

Monitoring of bat fauna has to be conducted not just to mitigate the decline of bat populations across the world, but to help us understand the effect of man-made activities on the ecosystem due to the bioindication characteristics of bats. Bioindicators are those species that tend to react to the changes occurring in their surrounding environment. Bats are one such species that provide essential information on various alterations taking place in the environment due to their sensitivity to anthropogenic activities [7]. Monitoring of bats or any other bioindicator species for that matter helps us in taking up measures to protect nature.

[1] *RV College of Engineering, Bengaluru, India*
ORCID ID : 0000-0001-7538-8300
[2] *RV College of Engineering, Bengaluru, India*
ORCID ID : 0000-0002-7586-7827
[3] *RV College of Engineering, Bengaluru, India*
** Corresponding Author Email: analamr@rvce.edu.in*

A few negative bio-indications provided by bats include river quality, climate change, forest management, farming practices, and heavy metal pollution [8]. All these factors contribute to the decline in bat colonies beyond the level of replacement as the reproductive rates of bats are quite slow [9] making them vulnerable even to small man-made activities against the ecosystem. Rapid urbanization has also contributed to this decline [10]. This is one of the reasons why they are good indicators of environmental health. Therefore, there is a critical need to effectively monitor and take necessary actions to conserve these diverse species of mammals and eventually move towards a sustainable future. This monitoring of bat species must be treated as a serious concern not merely as an academic exercise as bias during bat assessment can lead to bad management decisions which can have a serious impact on bat fauna [7].

This paper proposes a Convolutional Neural Network based pipeline for effectively automatically detecting search-phase echolocation bat calls.

2. Background

2.1. What makes monitoring bats so difficult

One of the biggest reasons why monitoring bats is such a challenging task is there is little to no change in call patterns among bats of different species, and there's a lot of variation among bats of the same species. Though there are many challenges in monitoring bat fauna, this paper embarks on a fortuitous event i.e. bats emit calls in the frequency of 9kHz to 215kHz [11]. From one perspective this frequency range is not used by other fauna, which makes bat monitoring a less tedious task. Still, since bat calls have low intensity, a very high operating frequency range, along a high directional property, recording devices might not be capable enough to capture these calls unless placed very close. The possibility of manual human detection has to be eliminated as the calls are usually above the human audible range of 20Hz-20kHZ. Another reason that makes bat monitoring difficult is the capacity of bats to change their call frequencies when there's interference to their echoes from other bats [12]. Care must be taken while recording bat audio, as human interference can lead bats to abandon sites, so passive acoustic monitoring must be conducted. These are a few challenges that bring out the need for effective acoustic monitoring techniques.

2.2. Legal perspective on monitoring bats

Earlier in this paper the various ecological benefits of monitoring bats were listed. They are great bioindicators that reflect changes occurring in our environment. In this section, the legal perspective of monitoring bats will be discussed. There is a significant decrease in bat fauna in regions where turbines have been installed. Therefore, many legislations have been passed by governments of many nations making it mandatory to conduct bat assessments near existing wind turbines or before installing new turbines. Also, countries like the United Kingdom have come up with a "bat index" to measure trends in biodiversity.

2.3. Types of Bat Calls

Bat calls can mainly be classified into two categories: Echolocation calls and social calls. Echolocation is a characteristic possessed by bats through which they navigate and locate prey. Calls are emitted by bats and through the echoes of their calls, bats perceive their surrounding environments. Fig. 1. shows a spectrogram containing echolocation calls. Echolocation bats pass through 3 different phases: search-phase, approach-phase, and feeding buzzes, from navigating the prey to getting hold of it. The calls get quicker when they move from the search phase to the feeding buzz [11]. The calls emitted by bats to communicate with other bats are known as Social calls. Researchers use echolocation calls to locate and identify different bat species.

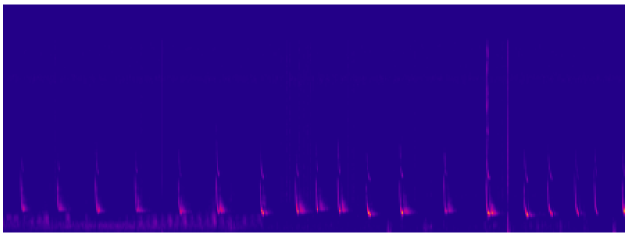

Fig. 1. Echolocation calls of bats

3. Literature Review

Four different machine learning and deep learning techniques to monitor bats are used [13]. This is one of the few papers that has focused on the task of localizing search-phase echolocation bat calls. The dataset was collected by a group of citizen scientists as part of a road-transect acoustic monitoring program in over 14 countries from 2005-2011. This kind of road-transect surveying has been found very helpful in deriving data for various species in Ireland [14]. Initially, audio files are converted into Spectrograms using the Fast-Fourier transform, and these spectrograms are denoised by removing the mean amplitude in each frequency band. Before denoising the Spectrogram, they are cropped to the frequencies of 5kHz to 115kHz as this is usually the frequency range where bats emit calls. The first technique used was the Segmentation method and this method is more like an amplitude thresholding approach, in which a threshold amplitude is chosen, and the Spectrogram image is segregated into two categories: object and background. The points having an amplitude greater than the chosen threshold are classified as objects, and the remaining portion is classified as background. Compared to the remaining methods used, the segmentation method performed poorly. The second approach used was feature engineering-based. Various audio features were extracted and fed into a Random Forest ensemble model. Finally, a Convolutional Neural network-based model was implemented in two ways, CNN-Full for better performance, and CNN-Fast for faster computation. The CNN-Full architecture was made up of 3 layers, whereas the CNN-Fast architecture had 2 layers. Finally, the performance was compared with other commercial systems such as Kaleidoscope, SonoBat, and SCAN'R, and both CNN models stood out against other bat call detectors.

A feature engineering approach [15] to detect bat calls is used. Initially, the Voice Activity Detector (VAD) from MATLAB's toolbox named Voicebox is applied to the audio files. VAD is used to find the presence of sound. Feature Extraction is applied on each of the detection points obtained through VAD. Three main categories of features are extracted the voice activity detector likelihood, i.e. how likely is there a presence of sound, secondly, the Call sonogram features such as frequency interval and time length, and Image shape features such as Eccentricity, Centroid, and Orientation among others are fed into the Random Forest algorithm for training. One of the key suggestions of this paper is to implement deep neural networks along with VAD. The performance of this work instead of being evaluated by ROC curves like it is usually done for classification algorithms is evaluated by Cost and F1-measure curves. The performance results of this VAD+RF detector are comparable to that of Bat Detective's CNN.

Being a predecessor of the previous paper, the bat species classification task with the same VAD and feature engineering-based approach is implemented [16] with 44 different call features extracted in this work at each of the detection points. Five different classifiers are used i.e. Discriminant Function Analysis, KNN, SVM, Classification and Regression Trees (CRT), and Random Forest. The Random Forest algorithm was best performing with 92% accuracy followed by SVM with 89% accuracy. One of the plus points of this work is that detection was not restricted to search-phase echolocation calls, even approach and terminal phase calls are taken into account while classifying species. As discussed earlier, this is a good approach as in some cases there is little to no change in call patterns among bats of different species, and in some cases, there's a lot of variation among bats of the same species [11].

Two popular CNN architectures - Inception-v1 and ResNet50, along with two CNNs of 6 and 10 convolutional layers each are implemented [17]. Similar to the above work, a spectrogram is extracted from audio, then filtered to be in the frequency range of 15kHZ to 120kHZ, and finally fed into the CNN model. Two CNNs were used, first to detect whether the audio contained a bat call or an environmental noise, and another deep CNN to classify the bat call to its respective species. Of the four architectures used, Resnet50 performed comparatively well.

A CNN model is built to efficiently classify bat species from their calls [18]. Here the CNN model is trained by extracting three types of features from audio files namely Short-time Fourier transform (STFT), Mel-Spectrogram Filter Banks (MSFB), and Mel Frequency Cepstral Coefficients (MFCC). The model trained with MSFB showed positive results with a classification accuracy of 97%. This CNN model contains 4 convolutional layers with a ReLU activation function. One max pooling layer and three average pooling layers are used. Finally, a softmax layer with 8 units is used to predict which of the 8 bat species the call belongs to. Even if good results have been achieved through this work, the motivation behind was in a negative sense i.e. monitoring bats to prevent the spread of the virus. The biggest limitation of this work is that the bat calls had to be manually isolated.

A transfer learning-based approach is implemented where pre-training is performed using the ImageNet dataset (an image database containing over 1.2 million annotated images) [19]. Later, the network was fine-tuned to adapt and recognize bat calls. A novel CNN model named BatNet with 22 layers was proposed by the authors which showed promising results. This BatNet architecture was compared with three traditional CNN architectures ResNet18, ResNet50, and VGG. Out of the four models, the ResNet50 architecture showed an accuracy of 95.6%, followed by the BatNet architecture which achieved an accuracy of 95.4%.

Three machine learning techniques namely Random Forest, SVM, and ANN are used to identify bat species [20]. An audio library containing around 662 search-phase calls of ten bat species found in Uruguay was used to train the models. If a particular bat call didn't match any of the species present in the training dataset, it was classified as an "unknown class". A probability threshold was set up for this purpose. If the probability of the bat call belonging to a particular species was less than the threshold, then the particular call was added to the unknown category. Out of the 3 machine learning models the Random Forest algorithm achieved the best results.

4. Methodology

4.1 Data Source

The model is trained with the iBats dataset, and tested against three datasets namely: iBats (E. Europe), NBP (Norfolk), and iBats (UK). These datasets were collected around the highways of Europe, using road transect acoustic monitoring [21]. The dataset is in the form of .npy files and consists of three columns: Audio file name, Call Positions (in seconds), and file Duration. File duration is 3.84 seconds for all files.

4.2 Pre-processing

Audio signals are initially in time domain representation. Since CNN works only on images, audio signals need to be converted from the Time domain to the Time-frequency domain. The Fourier transform is used to split audio signals into individual frequencies. Each audio signal is composed of multiple single frequencies, which can be decomposed through Fourier transform [22]. Fourier transform works by choosing a frequency and optimizing the phase for that particular frequency. The magnitude would be the maximum at the optimal phase. The whole point of this is to measure how much amount of a particular frequency is present in the audio signal. The optimal phase can be calculated with (1).

$$\varphi_f = argmax_{\varphi \epsilon [0,1)}(\int s(t) \times sin(2\Pi(ft - \varphi))dt \quad (1)$$

where, φ_f = optimized phase

s(t) = original signal

And the magnitude corresponds to the amount of a particular frequency being present in the signal, which would be obtained at optimal phase. The magnitude is expressed as shown in (2).

$$d_f = max_{\varphi \epsilon [0,1)} (\int s(t) \times sin(2\Pi(ft - \varphi))dt \quad (2)$$

where s(t) = original signal

Fourier transform in terms of complex numbers can be expressed with (3)

$$\hat{g}(f) = \int g(t)\, e^{-i2\Pi f t} dt \quad (3)$$

where, g(t) = original signal

Audio signals are continuous, but due to hardware constraints continuous audio signals cannot be processed. For this purpose, the analog signal is converted into a digital signal employing sampling or quantization. The sampling rate means the number of samples considered in one second of the audio file. Usually, a sampling rate of 44100 is used. This is because the sampling rate must be twice as large as the largest frequency expected to be captured according to the Nyquist-Sampling theorem. As the maximum frequency humans can perceive is 20kHZ, 44100 (44100 / 2 = 22050Hz > 20kHz) as the sampling rate can be used. Fourier transform for discrete signals is expressed as shown in (4)

$$\hat{x}(k) = \sum_{n=0}^{N-1} x(n)\, e^{-i2\Pi n \frac{k}{N}}$$

(k = 0 : N - 1) (4)

The above formula is suitable to convert audio from the time domain to the frequency domain. The drawback of this approach is that it doesn't provide information about time. The information about which sound occurred before and which sound isn't present. To solve this problem, Discrete Fourier Transform is applied to small overlapping windows. This technique is known as Short-Time Fourier Transform (STFT). The STFT is expressed as shown in (5)

$$S(m,k) = \sum_{n=0}^{N-1} x(n+mH) \cdot w(n) \cdot e^{-i2\Pi n \frac{k}{N}}$$

(5)

where, m = 1, 2, 3.. (represents window number)

H = hop size

W(n) = Hann Window

To generate a window, the original audio signal is multiplied with the Hann window. This selects only a portion of the audio, by suppressing the rest of the signal. The Hann window is represented as shown in (6)

$$W(n) = 0.5 \times \left(1 - cos\left(\frac{2\Pi n}{K-1}\right)\right) \quad (6)$$

(n = 1…K)

Finally, (7) helps to visualize the spectrogram,

$$Y(m,k) = |S(m,k)|^2 \quad . \quad (7)$$

The computed spectrogram is denoised by subtracting the mean amplitude across each frequency band. An audio file in time-frequency representation is shown in Fig. 2.

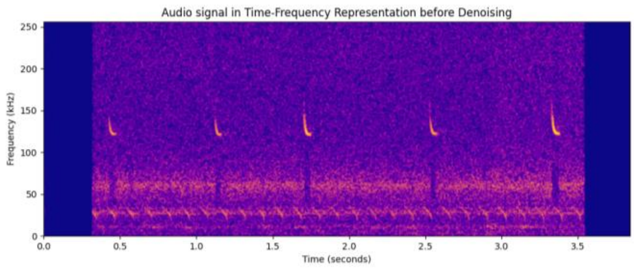

Fig. 2. Audio Signal in Time-Frequency domain before denoising

4.3 Median Clipping

Once the spectrograms of the audio files are obtained, they need to be denoised. The technique used to perform denoising is Median clipping. This is performed by removing the median amplitude in each frequency band. An audio file in time-frequency representation after denoising is shown in Fig. 3.

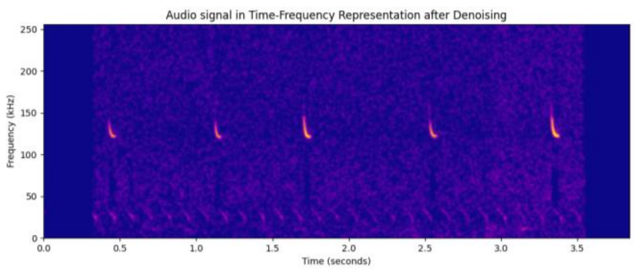

Fig. 3. Audio Signal in Time-Frequency domain after denoising

4.4 Dataset preparation

Since a sliding window CNN is to be implemented, windows of fixed size have to be generated which will then be fed to the CNN model. Images must be of the same size when being trained through a CNN model. Therefore, windows of size 130 x 20 from the audio files in time-frequency representation are extracted. Extraction of windows is followed for both training and testing files.

4.5 Data Augmentation

Having only positive instances of training data would make the training model to be highly biased. To avoid this equal number of negative instances needs to be included. The number of generated windows would still be less even after adding negative instances. Therefore, to increase the amount of training dataset, windows of 15ms before and after a true call can also be considered as a positive window, and the corresponding number of negative calls which are randomly generated is also augmented to the training dataset.

4.6 Classification through CNN

Convolutional Neural Networks (CNN) is a type of neural network widely used for the task of image localization and classification. A novel CNN architecture is proposed in this work to efficiently detect bat calls. Unlike machine learning algorithms where features have to be manually given, CNN is capable of learning directly from the time-frequency representation of audio files. CNN works only on images, therefore audio files are converted into spectrograms. In this work, 4 convolutional layers are used. The configuration of the CNN architecture used is as follows

- 48 Convolutional Filters (7x7)
- ○ Max Pooling layer (2x2)
- 64 Convolutional 64 Filters (5x5)
- ○ Average Pooling layer (2x2)
- 48 Convolutional 48 Filters (3x3)
- ○ Max Pooling layer (2x2)
- 64 Convolutional Filters (3x3)
- ○ Max Pooling layer (2x2)
- Fully Connected Layer
- Softmax Layer

The CNN architecture is depicted through Fig. 4.

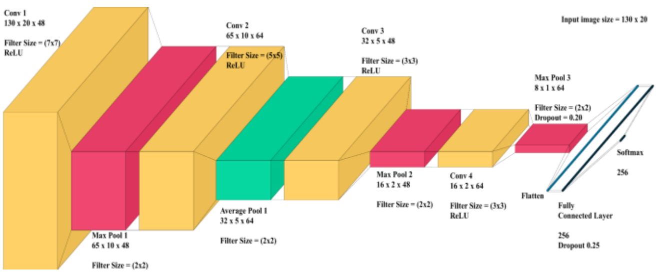

Fig. 4. CNN architecture of the proposed model

4.7 Non-Max Suppression (NMS)

NMS is used when the same object is wrapped around different bounding boxes. Here the most optimal bounding box is used and others are discarded. This happens by taking the maximum probability and suppressing the other boxes through non-max probabilities. NMS is used in this work in the detection stage. The signals taken as input are denoised, and sliding window CNN can be applied to the spectrogram. The model gives out probabilities of the start position of each bat call which can be seen in Fig. 5, and through Non Max Suppression technique the optimal bounding box is plotted.

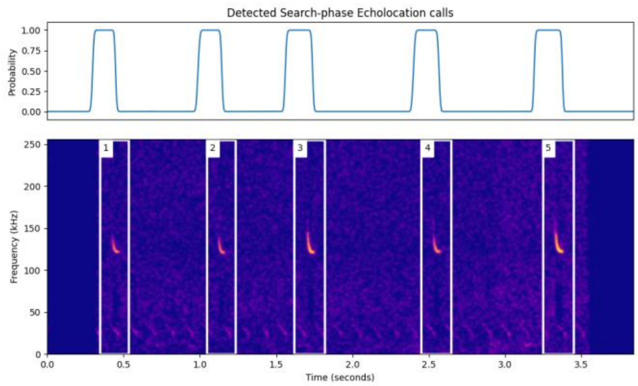

Fig. 5. Detected search-phase echolocation bat calls

4.8 Regularization

Training A large number of parameters can lead to overfitting, and the model is said to have high variance. This happens when the model performs too well on training data by trying to fit all the points instead of generalizing it to newer test data. Dropout regularization technique is used to reduce overfitting. This technique works by disabling a certain number of neurons. Dropouts varying from 0.20 to 0.30 are used after each pooling layer.

4.9 Optimization

The optimization function used is Nesterov's variant of Stochastic Gradient Descent. The task of an optimization algorithm is to speed up the training process of a neural network. With each epoch, the weights and biases of the neural network have to be updated such that the model performs better than it did in the current epoch. This is achieved by tracing down the Loss Curve’s minima. Gradient descent is one such optimization algorithm. A few limitations of Gradient descent are that it can take quite a lot of resources to update the weights since its progress is slow. By using another parameter named momentum, the learning rate of the neural network can be accelerated. The Nesterov’s variant of gradient descent is used in this work. This optimizer uses momentum to speed up the process of training. The hyperparameters used were

- Learning Rate: 0.01
- Momentum: 0.9

The loss function used was Binary Cross Entropy. This loss function is used when the output is either 0 or 1. Like other types of loss functions it calculates how far the predicted output is from the expected output. This loss function is mathematically represented by equation (8).

$$Binary\ cross\ entropy = -\frac{1}{N}\sum_{i=1}^{N} \left(y_i \cdot log(p(y_i)) + (1-y_i)\cdot log(1-p(y_i))\right) \quad \ldots(8)$$

5. Results

Table 1 suggests the proposed model performed better than existing systems. Around 500 more bat calls were detected by the proposed model as compared to the existing systems as listed in Table 2. Though the increase in precision is negligible, there is a significant increase in recall of the proposed model, as high as 11% compared to CNN-Fast, and 9% compared to the CNN-Full model on the iBats (UK) dataset.

Table 1. Performance comparison with existing systems

	Proposed Model	**Bat Detective's CNN-Full**	**Bat Detective's CNN-Fast**
	Precision		
iBats (UK)	0.9767	0.9661	0.9624
iBats (E. Europe)	0.9682	0.9558	0.9528
NBP (Norfolk)	0.9747	0.9639	0.9579
	Recall		
iBats (UK)	0.8301	0.7458	0.7268
iBats (E. Europe)	0.9108	0.8717	0.8659
NBP (Norfolk)	0.8111	0.7975	0.7799

Table 2. Number of calls detected on comparison with existing systems

	Proposed Model	**Bat Detective's CNN-Full**	**Bat Detective's CNN-Fast**
	Number of Calls Detected		
iBats (UK)	2097	1888	1793
iBats(E. Europe)	4383	4196	4147
NBP (Norfolk)	3273	3165	3166
Total	9753	9249	9106

The ROC curve plots sensitivity (or True Positive Rate) against specificity (1 – False Positive Rate). Classifiers having higher AUC ROC are better at the task of classification compared to classifiers with lower AUC ROC. A random classifier is expected to have an ROC curve passing through the origin where the True Positive Rate is equal to the False Positive Rate and its AUC is 0.5. The area under an ROC curve for an optimal classifier is 1. The proposed model was again compared with the existing systems using ROC curves and achieved higher AUC on all three datasets. The ROC Curve on the iBats (UK) dataset is shown in Fig. 6.

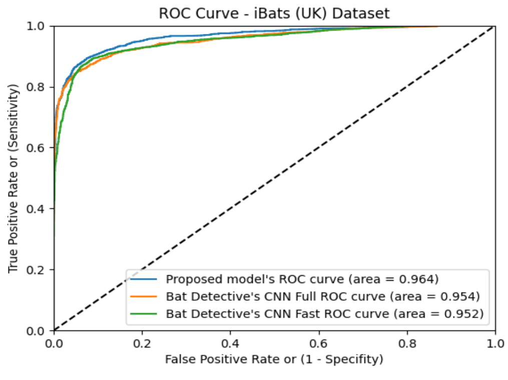

Fig. 6. ROC Curve on iBats (UK) Dataset

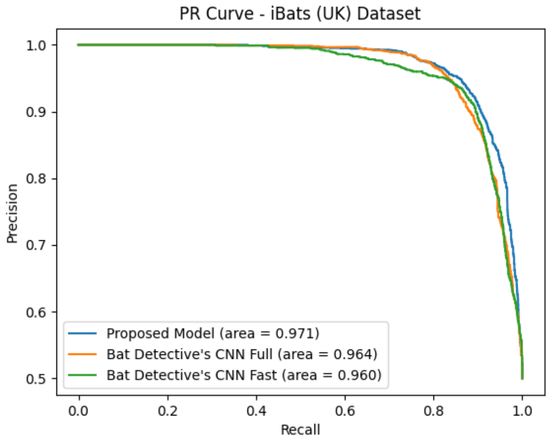

Fig. 7. PR Curve on iBats (UK) Dataset

It is desired that the algorithm should have both high precision and high recall. The proposed classifier is expected to have high recall as well as precision. However, this is not always possible. The PR AUC of an optimal curve is 1. The PR Curve on iBats (UK) dataset is shown in Fig. 7. This figure implies that the proposed model (represented by a blue line) performs this classification task better than the other two classifiers (represented by the orange and green lines). Usually, classifiers with a higher area under the ROC curve will have a higher area under the PR Curves. It can be observed that the proposed model is performing well on all three datasets compared to the existing systems.

6. Conclusion and Future Scope

In this work a CNN model was developed for effectively detecting echolocation calls from noisy audio recordings. Around 500 more bat calls were detected across all 3 datasets compared to that of the existing systems, with a significant increase in recall of the proposed model. Although a highly efficient model using CNN was developed, there is always room for improvement. One way by which the performance of these classifiers can be increased is by considering social calls along with echolocation calls to identify species. Also, other environmental features such as the type of vegetation, temperature, and season can be exploited to get more insights into bat activity. The common problem observed in most of the work is the lack of data. Therefore, both governmental and non-governmental agencies must take up this daunting task which will not only help in protecting endangered bat species but also in forming appropriate policies and taking necessary actions in preserving the environment.

Acknowledgements

This research was supported by Nvidia Corporation as part of the Applied Research Accelerator Program. The authors thank NVIDIA Corporation for funding and donating the hardware used to conduct this project. We thank Mr Rajesh Puttaswamaiah from Bat Conservation India Trust, Bengaluru who provided insight and expertise that greatly assisted the research.

References

[1] Fenton MB, Simmons NB. Bats: A World of Science and Mystery. USA: University of Chicago Press, Chicago.

[2] Bates PJJ, Harrison DL. Bats of the Indian Subcontinent. UK: Harrison Zoological Museum; 1997

[3] Korad V, Yardi K, Raut R. Diversity and distribution of bats in the Western Ghats of India. Zoo's Print Journal. 2007 July 21:22(7):2752-2758. doi: 10.11609/JoTT.ZPJ.1563.2752-8.

[4] Frick WF, Kingston T, Flanders J. A review of the major threats and challenges to global bat conservation. Ann N Y Acad Sci. 2020 Jun;1469(1):5-25. doi: 10.1111/nyas.14045.

[5] Kunz TH, Braun de Torrez E, Bauer D, Lobova T, Fleming TH. Ecosystem services provided by bats. Ann N Y Acad Sci. 2011 Mar;1223:1-38. doi: 10.1111/j.1749-6632.2011.06004.x.

[6] Boyles JG, Sole CL, Cryan PM, McCracken GF. On Estimating the Economic Value of Insectivorous Bats: Prospects and Priorities for Biologists. Bat Evolution, Ecology, and Conservation. Springer, ed. Adams R, Pedersen S, pp. 501–515, New York, 2013.

[7] Danilo R, Christian CV. The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for a sound. Ecological Indicators, 2016 July,66:598-602. doi: 10.1016/j.ecolind.2016.02.036.

[8] Russo D, Salinas-Ramos VB, Cistrone L, Smeraldo S, Bosso L, Ancillotto L. Do We Need to Use Bats as Bioindicators? Biology (Basel). 2021 Jul 21;10(8):693. doi: 10.3390/biology10080693.

[9] Jones G, Jacobs DS, Kunz TH, Willig MR, Racey PA. Carpe noctem: the importance of bats as bioindicators. Endang Species Res. 2009;8:93-115. doi: 10.3354/esr00182.

[10] Santini L, González-Suárez M, Russo D, Gonzalez-Voyer A, von Hardenberg A, Ancillotto L. One strategy does not fit all: determinants of urban adaptation in mammals. Ecology Letters. 2018 Dec 20;22(2):365-376. doi: 10.1111/ele.13199.

[11] Walters CL, Collen A, Lucas T, Mroz K, Sayer CA, Jones KE. Challenges of Using Bioacoustics to Globally Monitor Bats. Bat Evolution, Ecology, and Conservation. Springer, ed. Adams R, Pedersen S, pp.

479–499, New York, 2013. doi: 10.1007/978-1-4614-7397-8_23.

[12] Russo D, Ancillotto L, Jones G. Bats are still not birds in the digital era: echolocation call variation and why it matters for bat species identification. Canadian Journal of Zoology. 2018;96(2):63-78. doi: 10.1139/cjz-2017-0089.

[13] Mac Aodha O, Gibb R, Barlow KE, Browning E, Firman M, Freeman R, Harder B, Kinsey L, Mead GR, Newson SE, Pandourski I, Parsons S, Russ J, Szodoray-Paradi A, Szodoray-Paradi F, Tilova E, Girolami M, Brostow G, Jones KE. Bat detective-Deep learning tools for bat acoustic signal detection. PLoS Comput Biol. 2018 Mar 8;14(3):e1005995. doi: 10.1371/journal.pcbi.1005995.

[14] Roche N, Langton S, Aughney T, Russ JM, Marnell F, Lynn D, Catto C. A car-based monitoring method reveals new information on bat populations and distributions in Ireland. Animal Conservation. 2011 June 21.14(6):642-651. doi: 10.1111/j.1469-1795.2011.00470.x.

[15] Ruiz AT, Equihua J, Martínez S, Robredo E, Palm G, Schwenker F. Detection of Bat Acoustics Signals Using Voice Activity Detection Techniques with Random Forests Classification. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Springer, ed. Vera-Rodriguez R, Fierrez J, Morales A, pp. 253-261, Madrid, 2019.

[16] Ruiz AT, Jung K, Tschapka M, Schwenker F, Palm G. Automated identification method for detection and classification of neotropical bats. In 8th International Conference of Pattern Recognition Systems, pp. 1-6, Madrid, 2017.

[17] Schwab E, Pogrebnoj S, Freund M, Flossmann F, Vogl S, Frommolt KH. Automated bat call classification using deep convolutional neural networks. Bioacoustics. 2022 Mar 20:1-6. doi: 10.1080/09524622.2022.2050816.

[18] Zualkernan I, Judas J, Mahbub T, Bhagwagar A, Chand P. A tiny CNN architecture for identifying bat species from echolocation calls. In 2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G), pp. 81-86, 2020.

[19] Hu J, Huang W, Su Y, Liu Y, Xiao P. BatNet++: A Robust Deep Learning-Based Predicting Models for Calls Recognition. In 2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA), pp. 260-263, 2020.

[20] Nuñez GB, Lemus G, Wolf MM, Rodales AL, González EM, Crisci C. The first artificial intelligence algorithm for identification of bat species in Uruguay. Ecological Informatics. 2018 Jul 1;46:97-102. doi: 10.1016/j.ecoinf.2018.05.005.

[21] Bat Detective - Deep Learning Tools for Bat Acoustic Signal Detection, 2018, [online] Available: http://visual.cs.ucl.ac.uk/pubs/batDetective/

[22] Müller Meinard, Fundamentals of Music Processing: Using python and Jupyter Notebooks, 2nd ed. Springer, 2021.