iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38953

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Detecting Bird and Frog Species Using Tropical Soundscape

Anuranjan Pandey[1], Kartik Mane[2], Pooja Thakkar[3]
[1]*Department of Computer Engineering, Army Institute of Technology, Pune, India*
[2, 3]*Department of Information Technology, Shri Govindram Seksaria Institute of Technology and Science, Indore, India*

Abstract: In the tropical jungle, hearing a species is considerably simpler than seeing it. The sounds of many birds and frogs may be heard if we are in the woods, but the bird cannot be seen. It is difficult in this these circumstances for the expert in identifying the many types of insects and harmful species that may be found in the wild. An audio-input model has been developed in this study. Intelligent signal processing is used to extract patterns and characteristics from the audio signal, and the output is used to identify the species. Sound of the birds and frogs vary according to their species in the tropical environment. In this research we have developed a deep learning model, this model enhances the process of recognizing the bird and frog species based on the audio features. The model achieved a high level of accuracy in recognizing the birds and the frog species. The Resnet model which includes block of simple and convolution neural network is effective in recognizing the birds and frog species using the sound of the animal. Above 90 percent of accuracy is achieved for this classification task.
Keywords: Bird Frog Detection, Neural Network, Resnet, CNN.

I. INTRODUCTION

More than 500 million years ago, the Atlantic Ocean had spread sufficiently to bring a warm, wet environment to the Amazon basin, resulting in the formation of the Rainforest. The rainforest was first discovered in the Eocene period. While many of these people work as farmers, they don't do it on a huge scale, and instead they trade in high-value items like honey, skins, and feathers with others who don't live exclusively in the Rainforest. One of the most diverse habitats on the planet, rain forests are home to half of all known species [1]. Reptiles, birds, and other invertebrates species may all be found in rain forests. Reptiles include snakes, turtles, and chameleons, some of which may be harmful and result in a life-threatening scenario if they attack. Most biotic species can only be found in rainforests, according to experts. In addition, scientists think that there are many species still to be found, and they are constantly searching for them [2]. To become an expert on all sorts of species, individuals have previously attempted to become forest rangers by completing the training programmer for forest rangers. Even forest rangers lack this kind of training since many species found exclusively in tropical rainforests are unknown to most people. In order to become an expert, it takes years of real-world experience, but machine learning can assist in automating this process and allowing us to avoid having to know everything about every species.

The existence of rainforest species may serve as a visual indicator of climate change and habitat degradation. It's vital to employ auditory technology that can be used worldwide in order to recognize these species visually. Real-time information and early detection of human influences on the environment may be provided through machine learning approaches, such as neural networks. This information might help conservationists make better judgments [3]. As a means of overcoming this problem, this study utilizes a library named Librosa. [3] A preprocessing strategy based on amplitude may be useful for identifying different species, despite the fact that it doesn't seem to make sense to use the loudness of a sound to distinguish between them.

II. RELATED WORKS

SpecAugment was the term given to a method of data augmentation proposed in [6]. The inputs from the network's features were mixed with it. As an alternative to more conventional methods of augmenting, the author advocated the blocked masks for time steps and the frequency term channels. It was possible to improve performance on some of the most common audio datasets thanks to these augmentation strategies. In this research, [7] authors provided a method for eliminating noise from audio signals that includes a mix of two frequency and time domain processes (Amplitude). There was only a limited amount of success with the deep neural network, but the network's predictions helped to isolate noise from the audio signal. Since semi-supervised learning has seen such rapid growth in recent years, the authors in [8] exploited the Libri-Light dataset's unlabeled audio samples to get better outcomes utilising the technique. In order to teach students with noisy data utilising augmentation approaches, they employed pre-trained gigantic Conformer models that were pre-trained with weights from w2vec.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

As in the case of [9], the authors initially trained a Convolutional neural network for weak supervised learning using data from the samples of audio. After the instruction, they showed their methods in a typical fashion. semi-supervised learning is affected by incorrectly tagged data. The team also carried out a feasibility investigation and calculated how much it would cost to get the poorly labelled data, so they would not have to provide it themselves

The evolution of artificial intelligence relies heavily on the capacity to learn things sequentially. Connectionist models have long assumed that catastrophic forgetting is an unavoidable part of neural networks, although this has not always been the case. You can train networks that are capable of maintaining skill on activities they haven't done in a long time, according to our findings. Slowing down the learning of weights relevant to a given activity is an effective way of remembering previous tasks. Using MNIST handwritten digit datasets and Atari 2600 games, we demonstrate that our technique is scalable and successful in performing a range of classification problems, In this research [9], the research have shown feature of overcoming catastrophoc forgetting in the neural networks.

In this research [10] used a 700,000-video dataset to train a Convolutional neural network with more than 30k labels based on video level classifications. Among the tools they employed were FCDNNs, ResNet, VGG, AlexNet, and Inception. It was determined that a particular model built for an image-related job performed extremely well on the categorization of audio signals, and a larger dataset for training helped to improve the findings.

A time-based signal recognition system called the MCLNN (Masked Conditional Neural Network) and the CLNN (Conditional Neural Network) was introduced in [11]. According to the time-based audio signal, a conditional neural network used a masked architecture with masking to maintain the most important characteristics, and automated the process of combing these features together to discover new combinations.

The experiments mentioned above have shown the potential of neural networks to do categorization on various forms of audio data while also achieving high levels of performance from the model in question. A neural network model for audio classification of bird and frog species is created in this study utilising pretrained weights imagenet/cornell construction and environmental voice collection to prepare for audio classification of bird and frog species.

III. AUGMENTATION METHODS

A. Audio Augumentation

Audio augumentation is the process of increasing the audio dataset by processing the audio by adding noise, time shifting the audio data, increasing decreasing the volume, transformation of audio. Techniques used to enhance the quantity of data in data analysis include adding slightly changed copies of previously existing data or creating new synthetic datasets from existing datasets, which are then utilised in the analysis of the dataset. When training a machine learning model, it functions as a regularizer and reduces overfitting. Oversampling in data analysis has a lot to do with it.

1) *Guassion Noise:* Gaussion noise is noise generation technique in which noise can be added to different data like image, sine audio wave, images, etc. In our research the noise is added to the audio wave. As the name suggests, the statistical noise known as "Gaussian" comes from Carl Friedrich Gauss and has the same probability density function (PDF) as the normal distribution. [12] [13] In other words, the noise has a Gaussian distribution of values. [14] White Gaussian noise is a specific instance in which the values at any two points in time are statistically independent and identically distributed (and hence uncorrelated). Gaussian noise is utilised as additive white noise to create additive white Gaussian noise in communication channel testing and modelling. Digitized photos are susceptible to Gaussian noise because of low light or high temperature on the sensors, as well as the electrical circuits used to transmit them. [1] Although Gaussian noise reduction may be achieved using a spatial filter, the blurring of fine-scaled edges and features may occur when smoothing a picture since they also correlate to blocked high frequencies. Mean (convolution) filtering, median filtering, and Gaussian smoothing are examples of spatial filtering techniques for noise reduction.
2) *Pink Noise:* Pink noise is another type of noise that can be added to the audio wave data. There are two types of 1f noise: pink noise and 1f noise, both of which are signals or processes whose frequency spectrums have a power spectrum density that is inversely proportional to frequency. In pink noise, the quantity of noise energy is proportional to the frequency octave change. In biological systems, pink noise is one of the most prevalent signals. [15] The term is derived from the colour of the visible light produced by this spectrum of power. Unlike white noise, where the strength of each frequency interval is equal, this noise has varying intensity.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

3) *Time Shift:* Time shift is a technique for altering the timeline of an audio wave in order to process the audio timing and produce new types of audio from the original sample. The technique of speeding up or slowing down an audio stream without altering its pitch is known as time stretching. To the contrary: pitch scaling is the process of adjusting pitch without changing the pace. Pitch shift is a effect that uses pitch scaling to alter the pitch of a audio data. This technique can also be used in audio augmentation
4) *Volume Control:* It is possible to convert volume frequency to volume type using this volume control processing approach, which is described below. It makes use of uniform, fade, and cosine processing to generate numerous different forms of audio from a single sample of sound.

IV. BACKGROUND

A. Data Set

The dataset contains sound file of numerous species the dataset consists of features which are described below: recordings are identified by their unique IDs, and species are identified by their unique IDs it's the song type id that's different. Annotated signal start time t min Lowest annotated signal frequency, denoted by f min. time max **-** time at which the annotated signal reaches its maximum value. A signal's maximum frequency is f max. An indicator of whether the label is from the TrainTp (1) or TrainFp(0) file is provided by the is_tp**-** [tfrecords only].

train/ **-** the training audio files

test/ **-** the test audio files; the task is to predict the species found in each audio file

B. Training and Testing Data

Train/Test is a technique for determining the model's correctness. As the name suggests, you divide your data set into two sets: one for training and the other to test. The data in the training set serves as the basis for the algorithm's learning process. Observations in supervised learning issues typically include an observed outcome variable and one or more observed input variables. Test data is data used to assess how well a model performs based on an evaluation measure. As a consequence of this, the test set should not contain observations from the training set. If the test set includes examples from the training set, it will be difficult to distinguish if the algorithm has learned to generalize from the training set or has just recalled it. The dataset is divided into training and testing set as described:

Train_tp.csv : data for training species labels that are _truly positive_ and include time localization.

Train_fp : False positives species designations, with their related temporal locations, as training data.

Train and test folder: The data has train and test set which contains the training audio files and test audio files.

```
class RFCDataset:
    def __init__(self, tp, fp=None, config=None,
                 mode='train', inv_counts=None):
        self.tp = tp
        self.fp = pd.read_csv("../input/rfcxextras/cornell-train.csv")
        self.fp = self.fp[self.fp.ebird_code<'c'].reset_index(drop=True)
        self.fp_root = "../input/birdsong-resampled-train-audio-00/"
        self.inv_counts = inv_counts
        self.config = config
        self.sr = self.config.sr
        self.total_duration = self.config.total_duration
        self.duration = self.config.duration
        self.data_root = self.config.TRAIN_AUDIO_ROOT
        self.nmels = self.config.nmels
        self.fmin, self.fmax = 84, self.sr//2
        self.mode = mode
        self.num_classes = self.config.num_classes
        self.resampler = torchaudio.transforms.Resample(
            orig_freq=48_000, new_freq=self.sr)
        self.mel = torchaudio.transforms.MelSpectrogram(sample_rate=self.sr, n_mels=self.nmels,
                                                        f_min=self.fmin, f_max=self.fmax,
                                                        n_fft=2048)

        self.transform = Compose([
            OneOf([
                GaussianNoiseSNR(min_snr=10),
                PinkNoiseSNR(min_snr=10)
            ]),
            TimeShift(sr=self.sr),
            VolumeControl(p=0.5)
        ])
        self.img_transform = A.Compose([
            A.OneOf([
                A.Cutout(max_h_size=5, max_w_size=20),
                A.CoarseDropout(max_holes=4),
                A.RandomBrightness(p=0.25),
            ], p=0.5)])
        self.num_splits = self.config.total_duration//self.duration
        assert self.config.total_duration == self.duration * \
            self.num_splits, "not a multiple"

    def __len__(self):
        return len(self.tp)

    def __getitem__(self, idx):
        labels = np.zeros((self.num_classes,), dtype=np.float32)
```

Figure 1: Dataset preprocessing

This figure illustrates how preprocessing is carried out on a dataset. The data is preprocessed by applying the above-mentioned procedures such as Gaussian noise, time shift, and volume control on the data set in question.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

```python
class AudioTransform:
    def __init__(self, always_apply=False, p=0.5):
        self.always_apply = always_apply
        self.p = p

    def __call__(self, y: np.ndarray):
        if self.always_apply:
            return self.apply(y)
        else:
            if np.random.rand() < self.p:
                return self.apply(y)
            else:
                return y

    def apply(self, y: np.ndarray):
        raise NotImplementedError
```

Figure 2: Audio Transform

One of the processes in the process of audio augmentation and enhancement is the conversion and preprocessing of audio, which is accomplished via the usage of the audio transform class.

```python
class GaussianNoiseSNR(AudioTransform):
    def __init__(self, always_apply=False, p=0.5, min_snr=5.0, max_snr=20.0, **kwargs):
        super().__init__(always_apply, p)

        self.min_snr = min_snr
        self.max_snr = max_snr

    def apply(self, y: np.ndarray, **params):
        snr = np.random.uniform(self.min_snr, self.max_snr)
        a_signal = np.sqrt(y ** 2).max()
        a_noise = a_signal / (10 ** (snr / 20))

        white_noise = np.random.randn(len(y))
        a_white = np.sqrt(white_noise ** 2).max()
        augmented = (y + white_noise * 1 / a_white * a_noise).astype(y.dtype)
        return augmented
```

Figure 3: Gaussian Noise

Gaussian noise is another audio augmentation and preprocessing procedure that may be used. Another approach is the addition of pink noise to the audio data in order to preprocess and execute audio augmentation.

C. Model Evaluation Metrics

Typically, deep learning algorithms made the loss numbers available to the public. An inaccurate prediction will result in a loss, which is a technical term used to refer to the penalty that will be assessed as a result of the error. Further, if the model projection is correct, the loss value would be zero. More specifically, due to this, the objective is to lower the loss values by using biases and weights in the model. In addition to the loss measure, which deep learning systems utilize, accuracy is a metric that is commonly used by academics to evaluate prediction performance. The accuracy of forecasts is defined as the percentage of correct predictions.

The accuracy formula is as follows:

$$\text{Accuracy} = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})}$$

D. Loss Improvement

It is possible that the catastrophic forgetting of a neural network will have a detrimental influence on the performance of a model if the data is not balanced. It is possible that the model's performance may be adversely affected as a consequence of this problem for this reason, the BCE Loss optimizer is modified by altering the pos_weight parameter in the Loss function, which helps to overcome the problem. When comparing the goal and the output, BCELoss is a function that quantifies the Binary Cross Entropy between the two. For the purpose of preventing other species of data from having an impact on the leader board score, the BCE Loss is optimised to the fullest degree feasible before being applied.

V. ALGORITHMS AND METHODS

A. Convolutional Neural Network

It is projected that the CNN, a sort of artificial neural networks that has proven useful in many computer vision applications, will continue to receive attention across a wide range of disciplines, including forecasting, in the coming years. CNNs are meant to acquire spatial hierarchies of features in a dynamic and seamless manner through backpropagation, and they do so by employing a range of building blocks, including layers such as convolution, pooling, and full connection, in order to accomplish this goal. Providing a high-level review of the underlying concepts of CNN and their application to this report is the goal of this part.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Local connectivity was a driving force for the development of convolutional neural networks. This connection exhibit a high is referred to the field of responsiveness of a node. weighted sums from the weighted sums table are substituted with the weighted sums from the weighted sums table to establish local connectivity. A type of neural network is convolutional neural networks. Layers of the convolutional neural network convolves the input.To make a feature map, use the weight matrix. To put it differently, the weight matrix. Using the input vector and the weight matrix, the dot product is computed by sliding across the input. In contrast to typical neural networks, the output feature map's values all have the same weights. This indicates that every node in the output detects the same pattern. CNNs lower the total amount of learnable parameters due to their local connection and shared weights, As a result, training is more efficient. A convolutional neural network works by constructing a weight matrix in each layer that can extract the needed data from the input

The weight, height, and no of channels are all factors to consider are commonly considered three-dimensional inputs to a convolutional layer. To construct the feature output map, The first layer applies a set of 3D filters to all of the input channel to convolve the data in other words. Consider the following example: $x = (x_j)^{N-1}_{j=0}$, a one-dimensional input of size N with no zero padding. Convoluting each filters w1h for h = 1,..., M1 with the following input yields the output feature map of the first layer:

$a^1(i, k) = (w^1{}_k * x)(i) = \Sigma\, (w^1{}_k\,(j)x(i - j))$ where j=-∞ to ∞.

Features may be identified in a time invariant manner, but number of trainable parameters is decreased, because all of the elements in feature map have the same weightage. The network has output after convolutional, whose count and size depends on the size of filter used and final layer uses the number of filter. To minimise errors in network output, the model's weights are trained depending on what our model needs to learn.

B. Resnet Neural Network

A deep neural network with additional layers has been used in every successful CNN-based design since AlexNet's victory in ImageNet 2012 to lower error rates. A typical issue in deep learning, known as Vanishing/Exploding gradient, occurs as the number of layers in a neural network increases. This results in a gradient that is either zero or too huge. Thus, as the number of layers is increased, so does the training and testing error rate. This design introduces the Residual Network idea to address the vanishing/exploding gradient issue. We employ a method called "skip connections" in this network. Skipping many stages of training and connecting straight to output is possible with the skip connection.

This type of artificial neural network is based on pyramidal cells in the cerebral cortex, and it is known as a residual neural network (ResNet). Skip connections or shortcuts are used in residual neural networks to go around some layers. Nonlinearities (ReLU) and batch normalisation are common in typical ResNet models, which are implemented in double or triple layer skips. [16] When discussing residual neural networks, a non-residual network is referred to as a simple network.

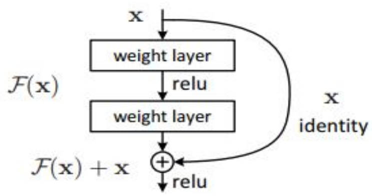

Figure 4: Resnet Model
Source – [17] *Deep Residual Learning for Image Recognition*

Since the introduction of Residual blocks, it has become possible to train incredibly deep networks, and the ResNet model is built using these blocks. The ResNet model is built of Residual blocks, which solves the problem of training extremely deep networks.

The first thing we note in the diagram above is that there is a direct link that bypasses several layers of the model. The 'skip connection' lies at the heart of residual blocks. This connection causes the output to be different. A bias term is added to the weights of the layer before the input 'X' is multiplied by the layer's weights. [17] There is always the need to validate our model's correctness to ensure that it has not been overfitted or underfitted, and cross validation is a method for doing this. Only after the model has been trained with data is this validation step carried out. Overfitting is a statistical term that refers to when a model's fit to the data is too tight. In order to accurately anticipate future data, the fitted line must pass through every point on the graph. When a statistical model or machine learning algorithm fails to adequately capture the data's structure, it is known as under fitting. Cross validation is the method employed to keep the model from becoming over fitting and under fitting.

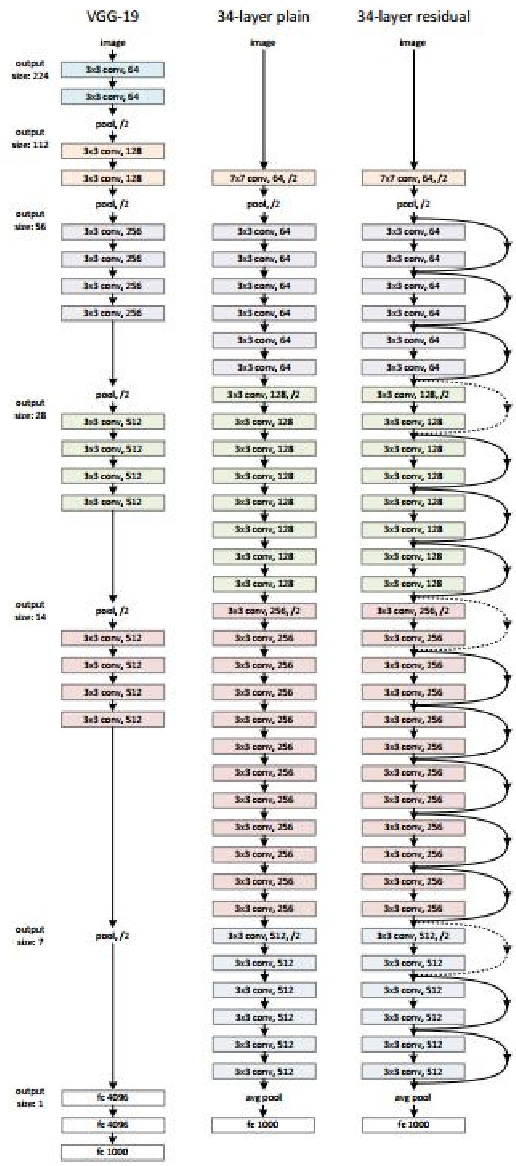

Figure 5: Resnet architecture
Source – [17] *Deep Residual Learning for Image Recognition*

VGG-19-inspired design includes a 34-layer plain network that incorporates shortcut and skip connection methods. As may be seen in the following diagram, the architecture is transformed into a residual network by these "skip connections" or "remaining blocks."

VI. RESULTS

The performance of the model while using the Resnet model and the Loss optimization step is shown in the table below. Although the accuracy of the model is high, it may be made even more accurate by increasing the number of data points used in the model's training and validation.

	s0	s1	s2	s3	s4	s5	s6	s7
count	1992.000000	1992.000000	1992.000000	1992.000000	1992.000000	1992.000000	1992.000000	1992.000000
mean	0.142701	0.246283	0.104052	0.968922	0.064141	0.107820	0.019908	0.338629
std	0.318920	0.382863	0.261135	0.104282	0.180991	0.243327	0.078952	0.416363
min	0.000006	0.000048	0.000001	0.113772	0.000008	0.000013	0.000004	0.000035
25%	0.000276	0.002222	0.000472	0.995523	0.000565	0.001535	0.000324	0.006540
50%	0.001721	0.018099	0.002804	0.999362	0.002965	0.008881	0.002687	0.052465
75%	0.028897	0.353673	0.024783	0.999836	0.021072	0.056754	0.013742	0.874864
max	1.000000	0.999995	1.000000	0.999997	0.999969	0.999971	0.999922	0.999986

s8	s9	...	s14	s15	s16	s17	s18	s19	s20
1992.000000	1992.000000	...	1992.000000	1992.000000	1992.000000	1992.000000	1992.000000	1992.000000	1992.0
0.077564	0.077711	...	0.124676	0.355564	0.088935	0.053946	0.457720	0.027381	0.0456
0.226736	0.195630	...	0.315350	0.411342	0.234976	0.172299	0.472631	0.087098	0.1347
0.000003	0.000032	...	0.000002	0.000184	0.000021	0.000002	0.000014	0.000013	0.0000
0.000315	0.001035	...	0.000200	0.019059	0.000540	0.000213	0.001583	0.000824	0.0003
0.001446	0.006213	...	0.001256	0.099076	0.002391	0.001870	0.102336	0.006180	0.0027
0.009145	0.039391	...	0.012742	0.887977	0.017701	0.018972	0.998323	0.020058	0.0230
0.999969	0.999995	...	1.000000	0.999978	0.999996	0.999998	0.999996	1.000000	0.9999

Figure 6: Statistics of the classification of the classes.

The statistics of the categorization performed by the model are shown in the results table above. S0 to S23 correspond to the categories of birds and the species of frogs.

The findings demonstrate that the model has a high degree of accuracy. The model may be utilized in Internet of Things devices to categories the species of birds and frogs that are present in the environment in real time. 96 percent of the results for species categorization are delivered by the model, which achieves its high accuracy using cross-validation and hyper parameter optimization.

VII. CONCLUSION

Deep learning algorithms, in particular, are gaining prominence among scientists as a result of recent developments in the creation of advanced machine learning-based methods. There are scholars from a wide range of fields participating. The most important question is how accurate and successful these new algorithms are when compared to older methods. These neural network based model can be used in new approaches to solve the problem that takes more time and effort.

There are various methods that may be used in this research to classify animal species based on their audio. Using acoustic signals, the models used to categories tropical rainforest species functioned exceptionally well. The audio data is utilized to train and test the ResNet model. In terms of species classification, the model is delivering 96 percent of the results.

The performance of the models can be improved and a wider number of species can be easily classified if more data is collected on the animal species. IOT devices could be used in the rain forest to implement this approach.

REFERENCES

[1] L. E. A. D. G. C. H. R. F. e. a. Y. Malhi, "Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest," Proc. of the National Academy of Sciences of the United States of America, vol. 106, pp. 20610-20615, 2009.
[2] A. N. E. B. R. C. F. S. M. B. e. a. P. H. S. Brancalion, "Global restoration opportunities in tropical rainforest landscapes," Science Advances , vol. 5, p. 7, 2019.
[3] R. C. S. A. D. Research Prediction Competition, "Automate the detection of bird and frog species in a tropical soundscape".
[4] C. R. D. L. D. P. E. E. B. e. a. B. McFree, "Audio and music signal analysis in python," Proc. of the 14th python in science conf, vol. vol 8, 2015.
[5] D. S. Park, W. Chan, C. C. Y. Zhang and B. Z. e. al., "Specaugment: A simple data augmentation method for automatic speech recognition," Proc. Interspeech, pp. 2613-2617, 2019.
[6] M. Michelashvili and L. Wolf, "Speech denoising by accumulating per-frequency modeling fluctuations," arXiv, pp. 1-5, 2019.
[7] J. Q. D. S. P. W. H. C. C. C. e. a. Y. Zhang, "Pushing the limits of semi-supervised learning for automatic speech recognition,," arXiv, 2020.
[8] A. K. A. G. H. a. B. R. A. Shah, "A closer look at weak label learning for audio events," IEEE, 2018.
[9] R. P. N. R. J. V. G. D. A. A. R. K. M. J. Q. T. R. A. G.-B. James Kirkpatricka, "Overcoming catastrophic forgetting in neural networks".
[10] S. C. D. P. E. J. F. G. A. J. e. a. S. Hershey, "“CNN architectures for large-scale audio classification," IEEE, 2017.
[11] D. C. a. J. R. F. Medhat, "Masked conditional neural networks for audio classification," Springer, 2017.
[12] T. Barbu, "Variational Image Denoising Approach with Diffusion Porous Media Flow," 2013.
[13] B. Truax, "Handbook for Acoustic Ecology," 2017.
[14] B. Truax, " Image Restoration: Introduction to Signal and Image Processing," 2012.
[15] Szendro, "Pink-Noise Behaviour of Biosystems," PMID, 2001.
[16] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," arXiv, 2015.
[17] X. Z. S. R. J. S. Kaiming He, "Deep Residual Learning for Image Recognition," arXiv, 2015.