IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 7, July 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

Automated Bird Species Recognization Using Audio InputSignal Processing

Suresh M R[1], Divya S[2], Navya Prabhu K P[3], Yashaswini K V[4]

Department of Information Science and Engineering, PES College of Engineering, Mandya, India

ABSTRACT: Birds are warm-blooded vertebrates adhering to the Aves class. There are almost 10,000 extant species of birds in the world, each with a unique set of traits and appearances. People frequently find bird watching to be an intriguing hobby. The ability to identify a species of bird accurately requires extensive knowledge in the science ofornithology, which is beyond the scope of human knowledge. In this study, an automated approach that uses deep neural networks to identify a bird's species is presented. Convolutional neural networks are competent machinelearning toolkits that have demonstrated effectiveness in the realms of image processing and sound identification. We have used a prominent image recognition model called Inception v3 that has demonstrated higher than 93%accuracy on the ImageNet dataset. The dataset includes 20 South Indian bird species.

KEYWORDS: Neural Networks, MFCC, Mel-Spectrograms, CNN, Xception, Inception, ResNet50

I. INTRODUCTION

Recognizing bird species based on their sounds is a difficult task due to the challenges in pinpointing distinctive features. Traditional methods that rely on human observers visually inspecting the sounds are time- consuming and subjective. To automate this process, researchers have explored different approaches, including extracting features and using classification algorithms. The choice of feature extraction techniques and classification algorithms significantly impacts the performance of bird recognition systems. In recent years, deep learning models have garnered attention for their ability to classify bird sounds automatically, offering a promising solution for building automated bird sound classification systems.

Ornithologists study bird vocalizations to gain insights into bird distribution and languages, and the introduction of machine learning techniques opens up new possibilities for further research in this intriguing field. Our methodology employs machine learning to assess threatened bird species, not only differentiating between species but also providing information about their taxonomic classification, including their kingdom. The primary objective is to evaluate the effectiveness of specific public or private initiatives aimed at extending the timeline and preventing birds, both threatened and future species, from becoming a concern. Users will benefit by gaining a deeper understanding of the importance of various species and the essential conservation efforts that can be undertaken to preserve biodiversity.

Fig. 1 Harmonizing 20 avian visuals for audio classification

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

II. BACKGROUND AND RELATED WORK

Researchers in the US have created a framework [1] to teach a computer to recognise different bird species. In India, there are about 1500 different bird species, and when you take into account the subspecies, there are even more. We gathered bird audio data from the Macaulay Library, eBird, and xeno-canto. The open-source programme Audacity was used to manually refine thedata. The training model and detection make up the majority of the system. Due to a lack of research and inconsistencies in the online dataset, information on these bird classes has been scarce. With the trained model, we were able to estimate the bird for each audio file, and it provides probabilistic results for each bird. The actual bird from the entire audio stream was predicted by averaging each small probability result. The accuracy of the Support Vector Machine with Radial Basis Function during the classification of the main classes was 96.7%, and during the sub-classification based on the call or song of the classified bird species, it was between 96% and 99%. For audio of 5 to 7 seconds, the server needed approximately 10 to 20 milliseconds to process the file and return the response.

Three sets of classifiers and two feature sets are chosen to categorise the different bird species in this paper [2]. They offered Mel Frequency Cepstral Coefficient performance as a feature for classifying bird species. The highest accuracy with both feature sets is provided by CNN, which has 39 feature coefficients. A 93.37% accuracy rate was attained. By choosing the number of features in the feature set, the accuracy can be increased. Suppressing background noise in bird sound recordings can improve the model's performance even more.

In their thesis[3], the researchers developed an automated system for identifying bird species using audio samples and deep learning. They trained the system on a dataset of 8,000 sound samples, achieving high accuracy using a CNN Net model. The goal was to help bird watchers accurately identify bird species. While some challenges were encountered during training, the researchers discovered that certain labels created by intelligent birds improved accuracy. Overall, the thesis contributes to automated bird species identification and its potential benefits for bird watchers and conservation.

This study aimed to develop computerized methods[4] for classifying birds based on their distinct sound signatures. By analyzing chirping patterns, researchers achieved an accuracy of 95% and 86% using Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) algorithms, respectively. This approach is helpful for ornithologists and opens up possibilities for investigating audio recordings in noisy environments. The study emphasized the potential application of these methods to a wide range of bird species and suggested that studying additional animal species could contribute to our understanding of complex ecosystems.

A method [5] uses sound waves to identify bird species. It employs Gaussian Mixture Model (GMM) and Self-Organizing Neural Network (SNN) for classification.Bird sounds, including songs and calls, are analyzed by breaking them down into phrases, syllables, and components. Pattern recognition is used to distinguish specific data. The system has a training model that generates wavelet features from input sounds and saves them in a database. The identification model calculates Wavelet Transform characteristics and uses SNN and GMM with the database features for classification. The system produces spectrograms of bird calls using Short- Time Fourier Transform (STFT) with specific parameters.

A recent publication in computer science presents a method [6] for recognizing bird species using audio recordings. The researchers tackled the challenge of audio files. Xeno Canto allows users to search for bird recordings based on species, location, and recording quality, among other parameters. Once the appropriate audio files are located, they can be downloaded in various identifying multiple species from overlapping vocalizations in audio clips. They employed Convolutional Neural Networks (CNNs) and combined multiple sigmoid outputs to make judgments. Their visualization-based system outperformed the MFCCDNN technique, achieving an average F1 score of 0.65 on a test dataset. This work contributes to the field of bird species identification from audio recordings and shows promisefor future advancements.

Priyadarshani et al. [7] conducted a study to assess the state of bird sound identification. They developed a straightforward categorization system to identify Borneo's native bird species based on their sounds. Using a nearest centroid classifier, they achieved a prediction accuracy level of 96%, outperforming more complex methods like support vector machines and K-nearest neighbors. The research demonstrates the effectiveness of the proposed method, providing insights for environmental monitoring and conservation in the Borneo region.

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

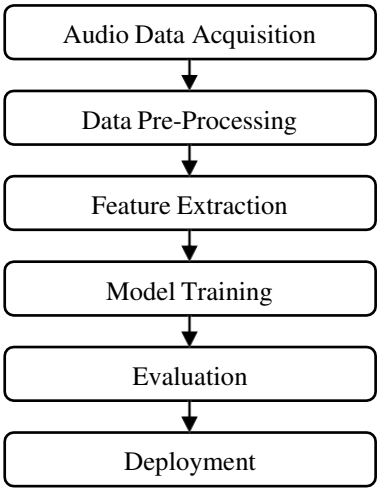
|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

The study paper [8] focuses on utilizing key methodologies such as Human Factor Cepstral Coefficients, Voice Activity Detection System, Hidden Markov Models, k-Nearest Neighbors (kNN) Algorithm, and Support Vector Machine (SVM) Algorithm. The signal is represented by a set of coefficients known as human factor cepstral coefficients. The Voice Activity Detection System analyzes bird vocalization segments and generates a likelihood rate. Hidden Markov models are trained using data specific to each bird species. The k- Nearest Neighbors algorithm is employed to categorize bird species, while the Support Vector Machine algorithm assesses distances. Results indicate an interspecific success rate of 81.2% for species recognition and a success rate of 90.45% for categorizing species into families, demonstrating the effectiveness of the Automatic Bird Species Recognition Based on a Bird Vocalization method.

III. METHODOLOGY

1. Audio data acquisition: Audio data acquisition: Collection of recordings of bird noises from a variety of sources, including bird calls, nature recordings, and more. Acquiring audio data from Xeno Canto involves accessing the website's extensive collection of bird recordings and downloading the desired audio files. Xeno Canto allows users to search for bird recordings based on species, location, and recording quality, among other parameters. Once the appropriate audio files are located, they can be downloaded in various file formats, such as WAV orMP3, and used for bird classification or other related projects. Xeno Canto is a valuable resource for researchers, birders, and anyone interested in bird vocalizations.



|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

Fig.2 Methodology

2. Pre-processing: Pre-processing plays a vital role in the bird sound classification pipeline as it involves de-noising, filtering, and converting the audio signal into a suitableformat for analysis. Its main objectives are to remove unwanted noise, improve the quality of the signal, and prepare the audio data for feature extraction. Several common pre-processing algorithms used in bird sound classification are as follows:

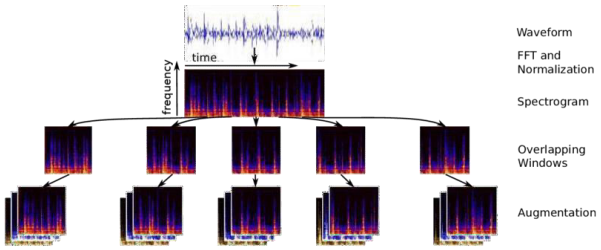


Fig. 3 Pre-Processing Pipeline

a. **Noise reduction**: a. Noise reduction: The objective of noise reduction is to enhance the signal-to-noise ratio by eliminating distracting background noise from the audio stream, such as wind, traffic, and other types of noise. Various algorithms, including Wiener filtering and spectral subtraction, can be employed to achieve this. Spectral Subtraction: This method involves estimating the background noise spectrum by utilizing a segment of the noisy audio signal. The estimated noise spectrum is then subtracted from the overall audio signal. While this technique is efficient and straightforward, there is a possibility of over-subtraction and the introduction of musical noise artifacts.

b. **Filtering:** Using a digital filter to enhance the bird sounds, reduce high or low-frequency noise, and remove extraneous data from the audio channel. Low-pass, high- pass, and band-pass filters are frequently utilized in pre-processing.

c. **Normalization:** To ensure that the data have comparable magnitudes and to avoid any numerical overrun during the processing stage, the audio signal is scaled to a certain range. Several techniques, including min-max normalization, mean normalisation, and z-score normalization, can be used to accomplish this.

d. **Resampling:** matching the sample rate of the labelled data or lowering the processing's computing cost by changing the sample rate of the audio output. Algorithms

3. Feature extraction: identifying and extracting spectral, temporal, and harmonic elements that are pertinent to the description of the bird calls from the audio stream. Featureextraction, which entails turning the raw audio signal into a collection of pertinent and meaningful features that define the bird sounds, is a crucial stage in the classification of bird sounds. The primary steps in feature extraction are as follows:

a. **Representation:** A wavelet transform, spectrogram, or other appropriate representation for feature extraction is created from the audio signal. The audio signal is visualized using this format, and pertinent features that describe the bird calls are extracted.

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

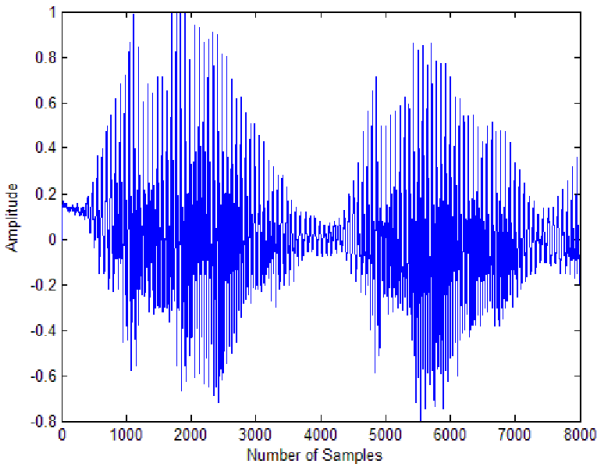


Fig. 4 Graphical Representation of Audio Signal

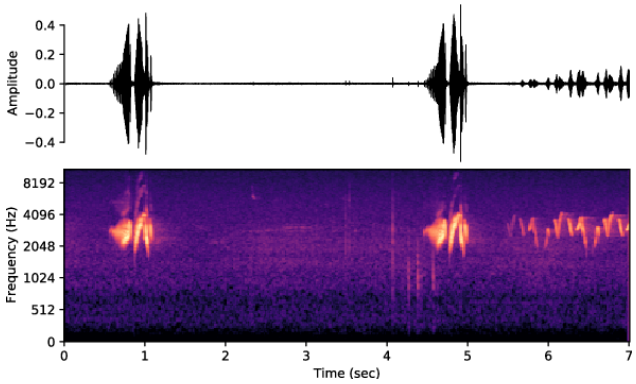


Fig. 5 Audio Spectrogram Representation

b. **Feature selection:** Based on how well they can distinguish between various bird sounds, a subset of the most important elements is chosen from the audiorepresentation. Different procedures, including linear discriminant analysis (LDA), mutual information and principal component analysis (PCA) can be used to do this.

c. **Feature extraction:** In order to create a feature vector, which serves as the input for the machine learning model, the chosen features are retrieved from the audio representation. Several techniques, including cepstral features, frequency-domain features, and time-domain features, can be used to do this.

4. Model training: To classify bird sounds using audio, collected features and labeled data can be used to train a CNN neural network. CNNs, a type of deep learning model, are commonly employed in image and audio classification. They learn the mapping between audio signals and bird species by utilizing convolutional layers, activation functions, pooling layers, and fully connected layers. This enables effective classification of bird sounds based on their audio characteristics.

a. ResNet
b. Inception
c. Xception

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

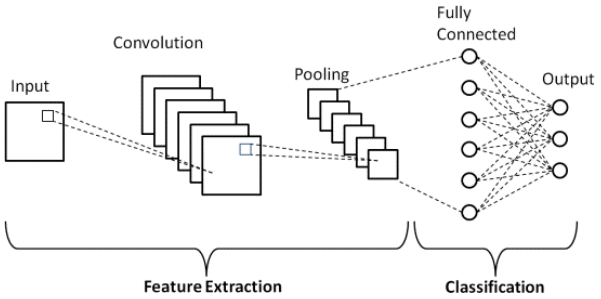


Fig. 6 CNN Model

5. Evaluation: Evaluating the performance of the trained model in terms of accuracy, precision, recall, and F1 score.

6. Deployment: Utilizing the trained model to classify bird noises on a massive scale or in real-time in a mobile application.a single vector by averaging each channel. The fully connected layer takes this vector as input and maps it to a set of logits, which are then transformed into a probability distribution over the bird species (or other classes) using the softmax function.

Convolutional layer:

$$Output\ size = ((input\ size - kernel\ size + 2 * padding) / stride) + 1$$

where:

- input size: the size of the input feature map
- kernel size: the size of the convolutional kernel
- padding: the amount of zero padding applied tothe input feature map
- stride: the stride of the convolution operationBatch normalization layer:

$$Output = (Input - mean) / sqrt(variance + epsilon) * gamma + beta$$

where:

- Input: the input to the batch normalization layer
- mean: the mean of the input
- variance: the variance of the input
- epsilon: a small constant added to the variance toavoid division by zero
- gamma: a learnable scale parameter
- beta: a learnable shift parameterReLU activation:

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

3.1 RESNET50

where:
Output = max(Input, 0)
ResNet50 is a variant of the ResNet architecture that was introduced by Microsoft in 2016. It is a deep neural network with 50 layers, sitting between the shallower ResNet-18 and ResNet-34, and the deeper ResNet-101 and ResNet-152. ResNet50 has gained significant popularity in computer vision tasks such as imageclassification, object detection, and segmentation.

One of the key features of ResNet50 is its use of residual connections to address the issue of vanishing gradients during training. By employing residual blocks, the model allows the input signal to bypass certain layers and be added to the output of the block. This approach enables ResNet50 to learn more complex features while mitigating the problem of diminishing gradients.

The feature extraction layers of ResNet50 consist of several convolutional layers with different kernel sizes. These layers are followed by batch normalization and ReLU activation. By progressively processing the input spectrogram, starting from low-level features like edges and corners and advancing to more abstract features like shapes and textures, the model learns a hierarchy of increasingly complex representations. The output of these layers is a feature map that captures the most salient features of the spectrogram.
ResNet50's classification layers consist of a global average pooling layer, a fully connected layer, and a softmax activation function. The global average pooling layer reduces the spatial dimensions of the feature map to Input: the input to the ReLU activation functionGlobal average pooling:

$$Output = 1/N * sum(x)$$

where:
- x: the input feature map
- N: the number of elements in the feature mapFully connected layer:

$$Output = Wx + b$$

where:
- x: the input to the fully connected layer
- W: the weight matrix
- b: the bias vectorSoftmax activation:

$$Output = exp(x) / sum(exp(x))$$

where:
- x: the input to the softmax function

3.2 INCEPTIONV3

InceptionV3 is a deep convolutional neural network architecture that was introduced by Google in 2015. It is an extension of the original Inception architecture and is designed to improve the performance of image classification tasks by reducing the computational complexity of the model while maintaining or even improving its accuracy.

The InceptionV3 architecture uses a module called the "Inception module" that consists of parallel convolutional layers with different kernel sizes and pooling operations. These parallel layers are concatenated along the channel dimension and fed as input to the next layer. This approach allows the model to learn both local and global features from the input image, improving its ability to recognize complex patterns.

In addition to the Inception module, InceptionV3 also uses several other techniques to improve its performance, including:

- Factorized 7x7 convolutions: The originalInception architecture used a 7x7 convolutional layer to process

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

the input image. InceptionV3 replaces this layer with two 3x3 convolutions, reducing the number of parameters and computational cost.

- Auxiliary classifiers: InceptionV3 adds two auxiliary classifiers to the network, which aretrained to predict the class labels in the middle of the network. This approach helps to mitigate the problem of vanishing gradients and improves the overall accuracy of the model.
- Batch normalization: InceptionV3 uses batch normalization after each convolutional layer, which helps to improve the convergence rate and generalization ability of the model.
- Regularization: InceptionV3 uses several regularization techniques, including the dropout and L2 weight decay, to prevent overfitting and improve the model's ability to generalize to new data.

3.3 XCEPTION

Xception, introduced by Google in 2016, is a deep convolutional neural network architecture that extends the capabilities of the Inception model. Its primary objective is to enhance the performance of image classification tasks by reducing the computational complexity of the model while maintaining or improving its accuracy.A distinctive feature of the Xception architecture is the replacement of standard convolutional layers commonly used in neuralnetworks with depthwise separable convolutions. These convolutions consist of two stages: a depthwise convolution that applies a single filter to each input channel, followed by a pointwise convolution that combines the outputs of the depthwise convolution using a 1x1 filter. This innovative approach effectively reduces the number of parameters and computational costswhile simultaneously enhancing the model's ability to learn complex features.

In addition to the utilization of depthwise separable convolutions, Xception incorporates several other techniques to further optimize its performance. These include:

- Residual connections, similar to those found in the ResNet architecture, enable the model to learn from the discrepancy between a layer's input and output. This mechanism significantly improves the model's capacity to capture long-range dependencies present in the data.
- Batch normalization is implemented after each convolutional layer, leading to improved convergence rates and enhanced generalization capabilities ofthe model.
- Xception employs various regularization techniques such as dropout and L2 weight decay to effectively prevent overfitting and improve the model'sability to generalize well to new, unseen data.

IV. RESULTS

The initial computation of mel-spectrograms utilized a hop size of 1024 ms and a window size of 10 ms. Subsequently, the CNN model was fed segmented mel- spectrograms in short-duration chunks. By using Mel-frequency bins, each bird sound was transformed into its corresponding mel-spectrogram representation. The training process employed the Adam optimizer. The model was trained for 25 epochs with a batch size of 32. During the experiment, 10% of the dataset was reserved for validation. The performance of the model was evaluated using Precision, Recall, and F1-score metrics.

In contrast to the suggested design, the implementation of MFCC-CNN utilized the librosa Python package tocompute the Mel Frequency Cepstral Coefficients (MFCC) in the front end. Our CNN model consisted of hidden layers with 256 perceptrons in each layer. TheReLU activation function and 50% dropout were applied. The output layer contained ten perceptrons, one for each class, followed by the sigmoid activation function. The training process used Adam optimization for 25 epochs with a batch size of 32.Moving on to the outcomes of the deep learning techniques employed, a confusion matrix was used to summarize the performance of the classification models on a set of test data. It provided information about the true positives, true negatives, false positives, and false negatives produced by the models.Specifically, in the context of bird sound classification using Resnet50, InceptionV3, and Xception, the confusion matrix aided in evaluating the performance of each model.In addition to the confusion matrix, a classification report was used to provide a more detailed view of the model's performance. It included precision, recall, and F1 score for each class in the classificationproblem. This allowed for a comparative analysis of the performance of the different models in the bird sound classification task using Resnet50, InceptionV3, andXception

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.12070019 |

V. CONCLUSION

The current focus of existing systems is on classifying a limited number of species. However, this paper specifically concentrates on classifying bird species found in South India. The study involves the development of a mobile application that utilizes a machine-learning model for bird audio classification. This model can automatically classify bird species by analyzing audio recordings. The model's output is typically presented as a prediction for each detected bird species. The specific format of the output depends on the implementation and use case, with common outputs including:

Species class label: The model predicts and identifies the species of the bird present in the audio recording.

Species description: The model provides a description of the bird species detected in the audio recording.

Automated bird species identification using neural networks holds immense potential as a reliable and efficient method. The ability to accurately identify different bird species based on their vocalizations has several applications, including bird monitoring, conservation efforts, and ecological studies. By leveraging deep learning techniques, researchers can develop highly precise models that excel at recognizing bird species.As neural network technology continues to advance, we can anticipate significant developments in the automated identification of bird species. These advancements have the potential to revolutionize our understanding of bird populations and behavior, leading to profound insights in the field.

REFERENCES

1. Jadhav, Y., Patil, V., & Parasar, D. (2020, February). Machine learning approach to classify birds based on their sound. In 2020 International Conference on Inventive Computation Technologies (ICICT) (pp. 69-73). IEEE.
2. Rane, G., Rege, P. P., & Patole, R. (2021). Bird classification based on Bird Sounds. 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN), 1143–1147. https://doi.org/10.1109/spin52536.2021.9566071
3. Jasim, H. A., Ahmed, S. R., Ibrahim, A. A., & Duru,
4. A. D. (2022, June). Classify Bird Species Audio by Augment Convolutional Neural Network. In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-6). IEEE.
5. Supriya, Prakrithi R., Shwetha Bhat, and Santhosh S. Shivani. "Classification of birds based on their sound patterns using GMM and SVM classifiers." Int Res J Eng Technol 5, no. 2004 (2018): 4708-4711.
6. Mohanty, Ricky, Bandi Kumar Mallik, and Sandeep Singh Solanki. "Automatic bird species recognition based on spiking neural network." In International Conference on Nanoelectronics, Circuits and Communication Systems, pp. 343-353. Springer, Singapore, 2020.
7. Rajan, Rajeev, and A. Noumida. "Multi-label Bird Species Classification Using Transfer Learning." In 2021 International Conference on Communication, Control and Information Sciences (ICCISc), vol. 1, pp. 1-5. IEEE, 2021.
8. Ramashini, Murugiaya, Pg Emeroylariffion Abas, Ulmar Grafe, and Liyanage C. De Silva. "Bird sounds classification using linear discriminant analysis." In 2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), pp. 1-6. IEEE, 2019.
9. Stastny, Jiri, Michal Munk, and Lubos Juranek. "Automatic bird species recognition based on birds vocalization." EURASIP Journal on Audio, Speech, and Music Processing 2018, no. 1 (2018): 1-7.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details