ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

BIRD SPECIES IDENTIFICATION USING ANN

K. Haribabu[1], Dr. V. Uma Rani[2]
[1]Student, Department of IT, harisam284@gmail.com
[2]Professor, Department of IT
University College Of Engineering Science And Technology Hyderabad
Jawaharlal Nehru Technological University Hyderabad

ABSTRACT

These days, it is crucial to keep an eye on how human activity is affecting the environment before it causes irreversible damage. One approach to keep an eye on these effects is to monitor population dynamics, biodiversity, and animal breeding behavior. Because they often respond the best to environmental changes, like deforestation or forest fires, birds are among the best animals to study. An estimated 1,370 species, or 13% of all bird species, are threatened with extinction. Even though they are widely distributed, many bird species are difficult for humans to tame. Professionals have been manually tracking the birds up until now, which takes time and is not a practical method. We propose a deep learning method to identify the bird species from audio recordings, which should help ecologists handle this challenge. The aim is to identify bird species automatically using the most recent Artificial Neural Networks (ANN) model using MLP as a classification model, which uses audio inputs to achieve this purpose. Our goal in this study was to improve the categorization accuracy of the current bird species classifier. This shows that the accuracy was 97% for validation and 100% for training. Thus, we can conclude that ANN is capable of successfully identifying the bird species and outwitting the current implementation model.
Keywords: ANN, MLP, Audio inputs.

1. INTRODUCTION

The behavior, size, and shape of the bird ecosystem is incredibly diverse. However, because of human encroachment into their habitats and complete habitat destruction, along with environmental disasters like forest fires and global warming, this biological diversity may be in jeopardy. Due to their tiny and shrinking ranges, 1,481 bird species, or 13.5 percent of all species for whom data is sufficient, are threatened with extinction globally as of 2020.

One of the main goals of monitoring birds is to control and evaluate the surrounding environment. Certain bird species suffer from air and water pollution. Thus, it is possible to identify and prevent environmental issues by identifying bird species. Given their rapid environmental response, birds can also help us detect different living forms in the surrounding area. However, it is not affordable to gather and compile data about bird species because it requires a large quantity of human labor and is more expensive. A reliable system in this case will be a valuable resource for scientists and government officials, offering a multitude of data about birds. Monitoring biodiversity can provide information on population status, migration routes, and environmental health for conservation planning and management. It is possible to identify someone by their voice, image, or video. An audio processing system that records the auditory signals of birds can be used to identify them. It used to be required to touch the bird directly in order to identify or classify it.

OBJECTIVE

An audio processing system that records the auditory signals of birds can be used to identify them. Direct contact with the bird in question was once required for identification or classification. On the other hand, it is now simple to identify birds thanks to the development of automated classification systems.

This procedure entails the following steps:

1) Recording bird music outside
2) Since these recordings are frequently made in loud settings, audio pre-processing techniques are used to enhance signal quality.
3) Extracting features from the audio input.
4) Building an accurate artificial neural network (ANN) model to identify bird species.

PROBLEM STATEMENT

Starting with sources such as the Xeno-canto database or the Macaulay Library of the Cornell Lab of Ornithology, a broad dataset of labeled bird song recordings is gathered to start the process. For consistency, the audio data must be segmented, noise reduced, and normalized during the preprocessing stages. Spectrophotograms, Mel-Frequency Cepstral Coefficients (MFCCs), and other pertinent features are then retrieved. To train an Artificial Neural Network (ANN), the dataset is divided into training, validation, and test sets. To achieve the best results, the selected model is trained and its hyperparameters adjusted.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

Eventually, in order to verify that the model is successful in correctly recognizing bird species based on their songs, its performance is assessed using metrics including accuracy, precision, and cross entropy loss.

2. SYSTEM ANALYSIS

2.1 Existing System

Some bird species are becoming increasingly rare, and even when they are discovered, it can be challenging to anticipate their categorization. From a human perspective, birds in different settings naturally appear in varying sizes, shapes, colors, and angles. Furthermore, visual cues are more useful in identifying the species of bird than auditory cues.

Furthermore, it makes more sense for humans to be able to identify the birds from the pictures. For training and testing purposes, this approach makes use of the Caltech-UCSD Birds 200 [CUB-200-2011] dataset. An image is transformed into greyscale format using the convolutional neural network algorithm, and then numerous nodes of comparison are made by utilizing Tensor Flow to generate an autograph.

The four bird species that Chandu B employed were the cuckoo, sparrow, crow, and laughing dove, each of which had 100 input spaces. The dataset comprised 400 samples of bird sound recordings overall. Xeno-canto.com, a website dedicated to the exchange of bird sounds from around the globe, was the source of the bird sound recordings. To prevent overfitting and maintain diversity, every clip, which ranges in length from 5 to 20 seconds, is converted to a fixed sampling frequency of 44100Hz or 48000Hz. Google Recording and Libri Speech ASR datasets provided the data for these examples.

2.2 TECHNIQUE USED

Convolutional Neural Network (CNN).

A temporary database storage of the image occurs whenever a user uploads an input file on the website. CNN is then fed this input file and given a trained dataset, after which it is fed back into the system. Convolutional layers are the building blocks of a CNN. To achieve optimal accuracy in categorization, multiple alignments/features including the head, body, color, beak, form, and overall image of the bird are taken into account. Every alignment undergoes a deep convocational network in order to extract features from several network layers. The classification of the image is then done using an unsupervised approach known as deep learning utilizing CNN.

In addition, the image is classified pixel by pixel using a grey scale technique. After that, these characteristics are combined and sent to the classifier. In this case, potential outcomes will be produced by comparing the input to the learned dataset.

Disadvantage of Existing System

- There is a time limit on the current system.
- Typically, the current system needs a lot more data.
- The current technique requires a lot more training data and is computationally more expensive.
- They struggle to categorize images with diverse positions.
- The quantity of computer power required is likewise higher.

2.3 Proposed System

Many species that we regularly see in our daily lives are also difficult for most people to recognize. These days, bird watching is a hobby, thus such systems have potential business applications. The goal of this research is to create an automated system that uses audio recordings of birds to identify them. It can be challenging to detect and identify birds using auditory signals since noises like rain or traffic sometimes overlap with bird sounds, making the process of bird recognition much more complex.

Because it requires a large number of professionals and is unreliable, manual spectrogram inspection is prone to error and highlights the need for automated systems. Artificial Neural Network is the tool used to do this. Data collection from a dataset that can be found on Kaggle is the initial phase of implementation. This resource includes the bird recordings in.wav format for listening. Bird sounds recorded in.wav format are included in this collection. Users can upload recordings of themselves to Kaggle, an open dataset platform. As a result of the dataset's numerous stated attributes, classes—such as genus and species—are formed and birds are categorized based on these classes.

An ANN classification algorithm is a widely used technique for bioacoustics signal analysis and recognition. We use the multilayer perceptron (MLP) as a classification model. To identify each bird species, the MLP uses a set of predefined attributes as input and generates a unique result. The two steps in this identification process are training and testing. The multilayer perceptron was trained using syllables of specific bird noises during the training phase, which caused the correct MLP output to be activated. The training procedure involves feeding the network with known sounds frequently and then iteratively modifying the network's weighting. . Reducing the overall error between the given and expected results until a predetermined error requirement is met is the aim of this training.

The user can analyze the bird species using the output by using the GUI, or Graphical User Interface. Users can submit audio files and receive results with the GUI.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

ADVANTAGES OF PROPOSED SYSTEM:

- Not only does the suggested system analyze bird sounds faster.
- But it also requires fewer training data to function.
- In contrast to the current system models, the proposed system does not come with a higher price tag.
- As opposed to the current system model, the suggested system may train the dataset entirely from scratch in a matter of hours.
- It also uses less computational power and exhibits strong fault tolerance.

3. METHODOLOGY

3.1 MODULES

- Dataset
- Importing the required library collections
- Exploratory Data Analysis of Audio data
- Imbalance Dataset check
- Data Preprocessing
- Splitting the dataset
- Audio Classification Model Creation
- Compile the Model
- Train the Model
- Check the Test Accuracy
- Saving the Trained Model

MODULES DESCSRIPTION:

- Dataset:

We created the system to obtain the input dataset for training and testing purposes in the first module.

Data set link:
https://www.kaggle.com/datasets/vinayshanbhag/bird-song-data-set
Total Number of Files: 2
File Format:
Metadata: CSV
Audio: WAV
Number of Subjects: Bird species (5) {Northern Mockingbird,American Robin,Bewick's wren, Northern Cardinal, Song Sparrow}
Other Details: Metadata includes filenames and bird species names.
The dataset consists of 5,422 Bird Sounds

- Importing the required library collections:

Librosa is a fantastic and highly significant library that facilitates audio and music analysis. To install the library, just use the Pip command. It offers the fundamental components needed to design a model for information retrieval from music. TensorFlow is an excellent package that we will utilize for deep learning modeling.

-Exploratory Data Analysis of Audio data

Under the urban dataset folder, there are five distinct folders. We will attempt to comprehend how to load audio files and how to view them as a waveform before performing any preprocessing. You can use the IPython library and provide it the audio file path directly if you want to load and play the audio file. We have selected the first audio recording from the dog bark category found in the fold 1 folder. Let's load audio data using Librosa now. Librosa thus provides us with two options when we load any audio file. Two-dimensional array is one, and sample rate is the other. Now that Librosa has the audio file above loaded, let's use it to plot the waveform.

The number of samples recorded per second is known as the sampling rate. The reading rate of 22050 is the default for librosa when reading a file. Your selection of library will affect the sample rate.

2-D array: Amplitude samples that have been recorded are shown on the first axis. The number of channels is shown on the second axis. Stereo (which has two channels) and monophonic (which has one channel) are the two types of channels. Librosa loads the data, normalizes it all, and attempts to provide it at a single sample rate.

Let's visualize the wave audio data now. The ability to normalize data printed from Librosa is a crucial concept to grasp. Librosa is becoming more and more popular for audio signal processing for the reasons listed below. It attempts to bring the signal into mono (one channel). A regular pattern is seen because it can represent the audio signal in normalized form between -1 and +1.

It can also observe the sample rate; by default, it transforms it to 22 kHz, but for other libraries, we see it according to a different number.

- Imbalance Dataset check:

We now understand what audio files are and how to see them in an audio format. We will load the CSV data file that is supplied for each audio file and go from format to data exploration to determine the number of entries for each class. Let's examine the file as all we know is the filename and its location. Now examine each class's records using the value counts method. When you look at the output, the majority of the classes have roughly the same number of records, and the data is not skewed.

- Data Preprocessing:

Certain audio recordings are made at a different f requency, such as 44 KHz or 22 KHz. The data will be at 22 KHz when using librosa, and we will then be able to view it in a normalized pattern. We now have the essential duty of extracting some information and storing our data as dependent features (class labels) and independent features (extracted features from the audio stream). Mel Frequency Cepstral coefficients will be utilized to extract separate characteristics from audio

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

streams. MFCCs: An MFCC provides a summary of the frequency distribution over a given window size. Therefore, an analysis of the sound's frequency and temporal characteristics is conceivable. We will be able to recognize characteristics for categorization with this audio representation. As a result, it will attempt to transform audio into features that will aid in classification by using temporal and frequency characteristics. In order to show how MFCC is used in practice, let's start by applying it to a single audio file that we currently use.

We now need to construct the data frame and extract features from each audio sample. Thus, we'll write a function that accepts the filename (or file path if it exists). Using librosa to load the file, we obtain two pieces of information. MFCC for the audio data will be found first, and the mean of the transpose of an array will be found to determine scaled features. We now need to run a loop over every row in the data frame in order to extract every feature for every audio file. To monitor the development, we additionally make use of the TQDM Python Library. We'll set up a unique file path for every file inside the loop, then run the method to extract MFCC features, append features, and matching labels within a freshly created data frame.

- Splitting the dataset:

Make two separate datasets: test and train. 20% of the data are test, and 80% are train.

- Audio Classification Model Creation:

From the audio sample and splitter in the train and test sets, we have retrieved characteristics. We will now use Keras sequential API to create an ANN model. Our output shape, or the number of classes, is 10. We will build an artificial neural network (ANN) with three dense layers; the architecture is described below. First layer: one hundred neurons. The number of features with an activation function of Relu indicates that the input shape is 40. To prevent overfitting, the Dropout layer will be used at a rate of 0.5. There are 200 neurons in the second layer that have a 0.5 dropout rate and an activation function similar to Relu. There are 100 neurons in the third layer once more, activation as Relu, and dropout at a rate of 0.5.

- Compile the Model

Our approach requires the definition of three key components: an optimizer named Adam, an accuracy metric called accuracy score, and a loss function called categorical cross-entropy.

- Train the Model

The model will be trained, and it will be saved in HDF5 format. A 32-piece batch with 250 epochs will be used to train the model. To find out how long it took to train over data, we'll utilize callback, a checkpoint.

- Check the Test Accuracy

Using test data, we will now assess the model. With the training dataset, we achieved an accuracy of about 97%, and with the test data, 100%.

- Saving the Trained Model:

First, use a library such as pickle to save your trained and tested model as a.h5 or.pkl file before putting it into a production-ready environment. Verify if pickles are installed in your surroundings. The model is then dumped into a.h5 file once we import the module.

3.2 ALGORITHM USED IN PROJECT

➢ **Artificial Neural Networks model (ANN model)**

Analyzing and identifying bioacoustics signals is commonly achieved through the use of Artificial Neural Network (ANN) classification algorithms. MLPs, or multilayer perceptrons, are utilized as classification models. Every bird species can be identified by using the MLP, which generates a distinct result based on a predefined set of criteria.

The two phases in this identification process are training and testing. During the training phase, the multilayer perceptron was trained using the syllables of specific bird sounds, which caused the correct MLP output to be activated. The network is trained by continually supplying it with sounds that it is familiar with, and then iteratively modifying the weighting of the network. This training aims to reduce the overall error between the given and expected results until a predetermined error requirement is met.

MLP (Multi-Layer Perceptron):

One kind of Artificial Neural Network (ANN) is the Multi-Layer Perceptron (MLP), which is made up of several layers of neurons, each of which is fully connected to the layer above it.

STRUCTURE OF MLP

1. Input Layer:
   - the network layer in which input data is first received. The input layer's neurons are each associated with a single feature of the input data.
2. Hidden Layers:
   one or more levels of intermediate information that sit between the output and input layers. By using activation functions and weighted connections, the neurons in these layers subject the input data to a number of changes. Complex pattern learning ability of the network can be impacted by variations in the number of neurons and hidden layers per layer.
3. Output Layer:
   The last layer that generates the network's output. The number of anticipated classes and the number of neurons in the output layer match.

KEY COMPONENTS

1. Neurons:

- Fundamental components of the network that calculate input weight sums and transfer the result through an activation function.

2. Weights:

- Characteristics that dictate how strongly neurons connect to one another.
- To reduce prediction error, these weights are modified throughout training.

3. Bias:

- An extra parameter to the weighted sum that is applied before the activation function is applied to improve the model's ability to fit the data.

4. Activation Functions:

- Functions that are applied to every neuron's output to give the model non-linearity so that it can recognize intricate patterns.
- Rectified Linear Unit, or ReLU, is a common activation function.

TRAINING AN MLP

1. Forward Propagation:

- The method of obtaining the outcome by sending input data over a network. After applying the activation function and computing the weighted sum of its inputs, each neuron sends the output to the layer below.

2. Loss Function:

- A function that calculates the discrepancy between the output that was expected and what was obtained.
- Standard loss function The classification method is based on cross-entropy loss.

3. Backpropagation:

An algorithm that modifies the network's weights in response to errors The process entails determining the gradient of the loss function concerning every weight and subsequently modifying the weights in a manner that minimizes the loss.

4. Optimization Algorithm:

We have updated the weights throughout training using optimizers such as Adam.

FUNCTIONAL REQUIREMENTS:

-Load Audio Data: Gathering relevant metadata and loading audio files.

-Feature extraction: taking audio recordings and extracting the Mel Frequency Cepstral Coefficients (MFCC) features.

-Data Preprocessing: Divide data into training and testing sets and encode class labels.

-Model Training: Utilizing the features that were extracted, train a neural network model.

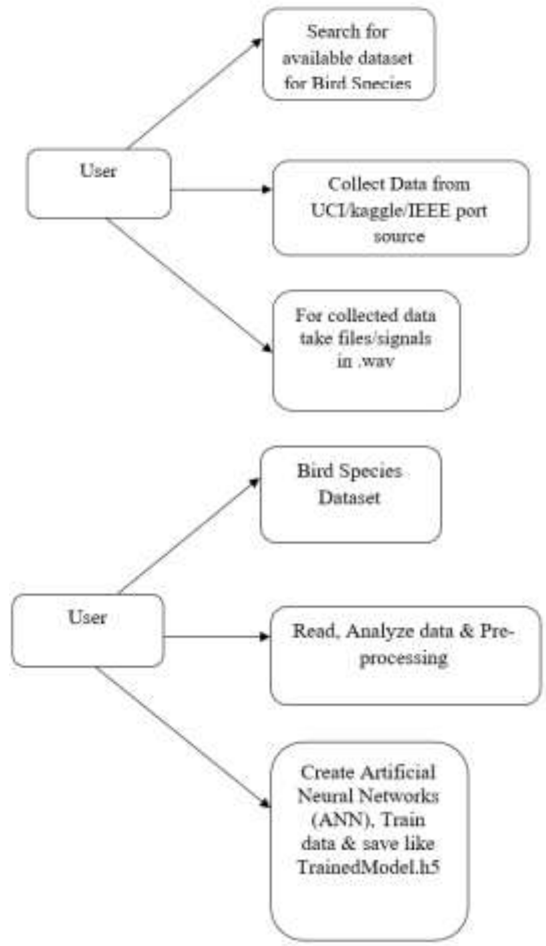
-Model Assessment: Assess how well the model performs on the test set.

-Model Saving: Save the trained model for future use.

NON-FUNCTIONAL REQUIREMENTS:

- Utilization
- dependability
- performance

4. DATA FLOW DIAGRAM



INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

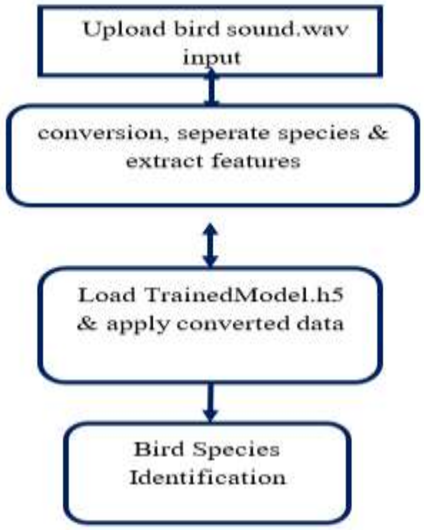


Fig: Flow Diagram

5. SYSTEM ARCHITECTURE

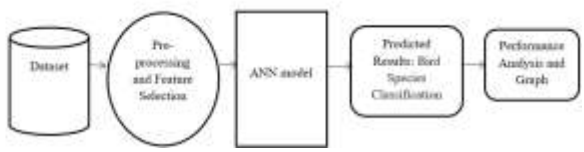
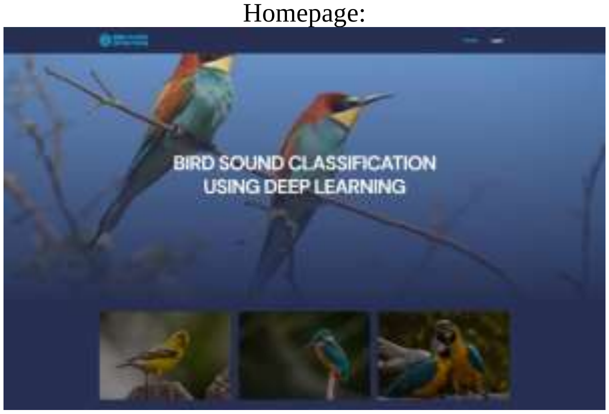


Fig: System Architecture

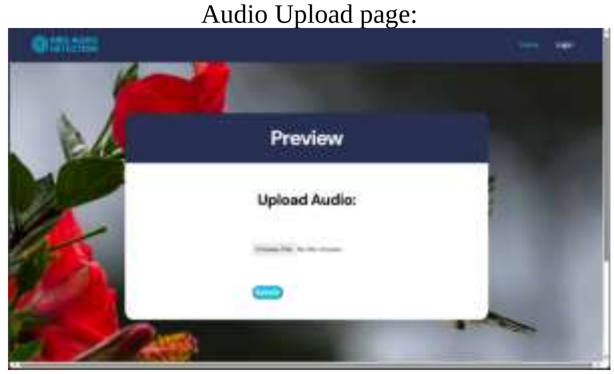
6. RESULTS

Homepage:

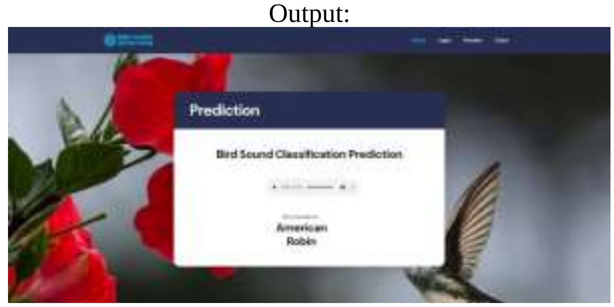


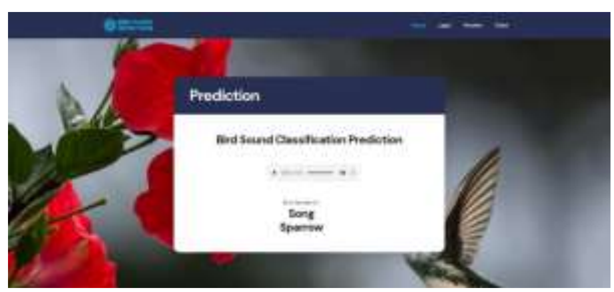
Login page:

Audio Upload page:



Output:





7. FUTURE ENHANCEMENT

When identifying and classifying bird species, this technique makes it possible to work on a larger number of classes, which leads to more accurate discoveries. When used well, this program can be a very beneficial tool for tracking a variety of other species, identifying natural habitats, and calculating the size of bird populations. Users who enjoy wildlife and the

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

environment may find a user-friendly programme useful as well. We can expand the range of birds it can identify in the future.

8. CONCLUSION

In this project, an Artificial Neural Network (ANN) model for automatic bird species recognition is brought forth. In order to aid them in their specific research, some researchers proposed an animal species recognition system in response to the effects of climate change and the number of endangered animals. With the help of artificial neural networks, we have constructed a system in this research to identify bird sounds (ANN). The tones of each bird's calls vary. Python is used to apply ANN in order to categorize and identify the bird noises. The first step in obtaining data for each type of bird is to use all necessary data regarding the power spectral density of birds. Training ANN to recognize different bird species is the next step in the process. A single bird can identify at a time. Last but not least, a graphical user interface (GUI) for identifying bird sounds has been created. To identify different species of birds, the user must provide audio input of bird sounds. This project is completed successfully and can be utilized to identify different types of birds.

REFERENCES:

[1] Aarti Madhavi, R. P. (2018). Deep Learning Based Audio Classifier for Bird Species. IJSDR

[2] C., A. P. (2015). Automatic Recognition of Birds Through Audio Spectral Analysis . Fifth International Conference on Advances in Computing and Communication, 395–398.

[3] Chandu B, A. M. (2020). Automated Bird Species Identificat ion using Audio. 2020 International Conference on Artificial Intelligence and Signal Processing (AISP).

[4] Evangelista, T. L., Priolli, T. M., Jr, C. N., Angelico, B. A., & Kaestner, C. A. (2014). Automatic Segmentat ion of Audio Signals for Bird Species Identification. IEEE International Symposium on Multimedia, 223–228.

[5] Graciarena, M., Delplanche, M., Shriberg, E., & Stolcke, A. (2011). Bird species recognition combining acoustic and sequence modeling. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 341-344.

[6] Guillermo Sarasa, A. G. (2017). An Approach of Algorithmic Clustering Based on String Compression to Identify Bird Songs Species in Xeno-canto Database. 3rd International Conference on Frontiers of Signal Processing.

[7] Incze, A., Jancso, H.-B., Szilagyi, Z., Farkas, A., & Sulyok, C. (2018). Bird Sound Recognition Using Convolutional Neural Network. IEEE 16th International Symposium on Intelligent Systems and Informatics, 000295–000300.

[8] Jancovic, P., & Kokuer, M. (2015). Acoustic recognition of multiple bird species based on penalised maximum likelihood. IEEE Signal Processing Letters, 1-1.

[9] Jancovic, P., Kokuer, M., & Russell, M. (2014). Automatic Detection of Bird Species from Audio Field Recordings using HMM-based Modelling of Frequency Tracks. EUSIPCO, 8252–8256.

[10] Lopes, M. T., Gioppo, L. L., Higushi, T. T., Kaestner, C. A., Silla Jr., C. N., & Koerich, A. L. (2011). Automatic Bird Species Identification for Large Number of Species. IEEE International Symposium on Multimedia, 117-122

[11] M. M. M. Sukri, U. Fadlilah, S. Saon, A. K. Mahamad, M. M. Som and A. Sidek, "Bird Sound Identification based on Artificial Neural," IEEE 2020 IEEE Student Conference on Research and Development (SCOReD), pp. 342-345, 2020.

[12] Manoharan, J. Samuel. "Study of Variants of Extreme Learning Machine (ELM) Brands and its Performance Measure on Classification Algorithm." Journal of Soft Computing Paradigm (JSCP) 3, no. 02 (2021): 83-95

[13] Narasimhan, R., Fern, X. Z., & Raich, R. (2017). Simultaneous Segmentation And Classification Of Bird Song Using Cnn. IEEE Conference, 146-150.

[14] Rai, P., Golchha, V., Srivastava, A., Vyas, G., & Mishra, S. (2016). An automatic classificat ion of bird species using audio feature ext ract ion and support vector machines. Internat ional Conference on Inventive Computation Technologies, 1-5.

[15] Vivekanandam, B. "Speedy Image Crowd Counting by Light Weight Convolutional Neural Network." Journal of Innovative Image Processing 3, no. 3 (2021): 208-222.