Bird Species Classification using Deep Learning

Yogesh J Prajapati
U V Patel College of Engineering,
Ganpat University, Mehsana – 384001, India
Department of Computer Engineering
yjp01.ganpatuniverisy.ac.in

Rahul Vinayak Nimje
U V Patel College of Engineering,
Ganpat University, Mehsana – 384001, India
Department of Computer Engineering
rahul.nimje8399@gmail.com

Manan D Thakkar
U V Patel College of Engineering,
Ganpat University, Mehsana – 384001, India
Department of Computer Engineering
mdt01@ganpatuniversity.ac.in

***Abstract*— In an environment, people are finding bird species in a specific area for their biodiversity. Birds can be automatically classified based on the songs (sounds) they hear. By this study we can improve accuracy to find out bird species by using deep learning algorithms. The raw audio are recorded by researchers for recognizing audio technique by various fields of DL and ML. This study may bring challenges to understand bird audio and recognizing audio to overcome limitations from previous techniques. This study will make effortless to bird watcher to classify bird species.**

Keywords *– Convolutional Neural Network (CNN), Spectrogram, Mel Frequency Cepstrum Coefficient (MFCC), Audio Preprocessing*

I. INTRODUCTION

There are more than 10000+ bird species are available in world with the different types of voice, shape, color and functions. Simply, the study goal is to create method which we can easily classify birds species by their audible sound. Bird's population in specific area becomes more challenging now a days. In an environment, bird species identification/classification became an important role to identify from images. By using frequency images of bird species is means to predict the bird species and this prediction is made from bird image, bird song, bird video, etc. An audio pre-processing technique can make possible to identify bird species by capturing raw audio. Now the most challenging part is that raw audio data may consist other environmental noise and it's become more complicate. Let solve this complexity by this study.

Fig. 1. Bird Species[23]

A. Problem Formulation in bird species classification

The challenge shows that it can be split into two parts. Signal Classification & Feature Extraction. Maintaining the Integrity of the Specifications.

1) Signal classification: This study describes time-series signal processing using neural networks, with a particular focus on two types of signal classification tasks: tone classification and tennis swing motion classification. There are many ways to define the problem of bird species classification. By considering the bird presence/absence in an audio recording with SIBL, SISL, and SIML problem. The problem in Signal Instance Binary Label (SIBL), that is for each audio recording there is two possible output, either bird presence in audio or bird absence in audio. This allows the system to detect the presence of birds. If we are interested in actual bird species of singing bird, than number of possible increases to observe bird species making Single Instance Single Label (SISL) problem. If we are interested in species recording, need ability to classify multiple bird species w.r.t recoding, making it Single Instance Multiple Label (SIML).

$$w = \arg\min_{w} \sum_{\bar{x} \in X} loss(f(\bar{x}), f_w(\bar{x})).$$

Fig. 2. Signal Classification-Audio-Formula[24]

2) Feature Extraction: Another problem in training neural networks is determining which function to use as input to the network. In particular, how to extract abstract features, which contain as much information about the original input as possible, but with a lower dimensionality to enable efficient training. The design of a feature extraction method can be seen as a trade-off between reparability and contraction. We want a contractor, or feature extractor, _(x) which reduces the dimensions of data point x without sacrificing reparability, that is, we want _(x) 6= _(x0) if f(x) 6= f(x0), where f is a true classifier. If this property holds we say that _ separates f.

II. LITERATURE REVIEW

1) Bird Audio Classification [1]

Publication Year: 2019

Author: Hadar Abukrat and Indu N. Korambath

Journal Name: IEEE Signal Processing Society

a. Summary: Identifying birds from voice recordings is a difficult area to study. Many researchers prefer to classify birds by sound rather than by sight because bird sounds are often detectable, but capturing images is not always easy. Acoustic noise can be a problem, but audio signals can be classified into songs, calls and noises, providing more information about the bird. Bird species identification is also much more difficult with this additional classifier. The frequency range and sound level of a single bird can be very dynamic. Bird watching and recording is popular with hobbyists and scientists alike, especially for conservation status issues. By improving the vocal classification of birds, it helps researchers better understand certain species and the environments in which they live. In this project, the input to the algorithm is a bird sound file & # 40; .mp3 & # 41; Then convert to spectral image, mel spectrum, MFCC (mel frequency cepstrum coefficient). Take the model one step further and average it frame by frame. Use a random forest (RF) classifier and multiple CNN models to make predictions for each input file. Predictions are the birds contained in each file.

2) Bird Classification using deep learning [2]

Publication Year: 2020

Author: Piyush Bidwai, Vaibhav Mahalle, Eshan Gandhi, Sharda Dhavale

Journal Name: IRJET Journal

a. Summary: Bird behavior and populations are important issues in today's scenarios. Birds help recognize other organisms in the environment. Basically, identifying the song of a bird is an important and difficult problem. There are also various methods that can be used to track birds. Because many birds move in response to changes in the environment, identifying birds in an automated way is an effective way to assess the number and diversity of birds present in the area. .. Artificial intelligence and machine learning sound like science fiction prophecies when it comes to technical characteristics. Image recognition is one of the most accessible applications. Machine learning built into consumer websites and apps is changing the way data is visually organized and processed. Very efficient image recognition and recognition using deep learning algorithms. This is a machine learning technique that models how the human brain works. This teaches the computer to recognize the visual elements in the image. This algorithm recognizes new patterns and uses a large database to understand images and generate relevant categories and tags. Bird classifications are usually based on catalog classes. Categorizing bird poses adds additional challenges to the category because the classes are similar. Therefore, determining the species of a bird plays an important role in determining the species to which the image of a particular bird species belongs. Bird species detection uses images to predict the species to which a bird belongs.

3) Bird Sound Classification Using Convolutional Neural Network [3]

Publication Year: 2019

Author: ChihYuan Koh1, JawYuan Chang1, ChiangLin Tai1, DaYo Huang1, HanHsing Hsieh2, YiWen Liu1

Journal Name: IRJET Journal

a. Summary: In recent years, community awareness of environmental protection and sustainable development has increased. Due to the important role of birds in the ecosystem, the requirements for automatic classification of bird calls are also increasing. Compared to video surveillance, sound has the advantage of being able to travel long distances undisturbed between the source (in this case a bird) and the recording device. Therefore, a robust system for identifying bird calls may help monitor biodiversity in fixed locations and detect migratory birds along the route. Recognizing the importance of this mission, the LifeCLEF Institute has been holding a contest called BirdCLEF every year since 2014. The purpose of the contest is to identify the birds in the picture. At this year's competition, participants needed to recognize bird calls from scenes recorded in the XenoCanto3 database every 5 seconds and identify species when there was a bird bark. Previous attempts to use machine learning approaches to detect bird callers included decision trees, integrated neural networks (CNNs), and recurrent neural networks (RNNs). For example, a random decision tree was applied and the input consisted of features taken from the histogram statistics. By ranking the importance of the features returned by the decision tree, you can find relevant segments to identify each audio class. Due to the computational cost of obtaining statistics from the spectrum, the decision tree method may not be the best solution to the current Vogel CLEF challenge. Its ability to manage over 600 species is also interesting. The RNN-based model was used in last year's Bird CLEF Challenge. In particular, it employs a short-term long-term memory (LSTM) architecture. He uses continuous information on bird calls. However, it is difficult to converge the model due to vanishing and gradient inflation issues associated with the sigmoid gate function. Moreover, due to the nature of RNNs, it is difficult to perform pre-processing and upscaling. Therefore, CNN-based models appear to be becoming the most common approach in detecting bird callers. Generally, the bird noise spectrum is used as the input, and the model treats the bird noise detection task as an image classification problem. This is intuitive because the trained human eye can recognize bird-specific features such as pitch and timbre in the spectrum.

4) Bird Sound Recognition Using Convolutional Neural Network [4]

Publication Year: 2018

Author: A´gnes Incze 、 Henrietta Bernadett Jancso´ 、 Zolta´n Szila´gyiy 、 Attila Farkasy 、 Csaba Sulyok

Journal Name: Research Gate

a. Summary: In recent years, there has been a steady increase in interest in noise detection using algorithms. The

popularity of deep learning and various types of neural networks provides new unexplored mechanisms for solving these classification problems. Among the many niche sound categories that are useful for this type of classification, current research focuses on bird sounds. We define nature lovers and ornithologists as target groups. Birds, which are often difficult to find in the wild, often have difficulty finding songbirds, so they can benefit from practical methods of distinguishing bird species based solely on audible clues. Current research uses transmission learning to optimize existing neural networks to detect bird calls. Many related networks are trained to recognize common features of images. But since the sound is one-dimensional and the image is a two-dimensional signal, do we need a general transformation for compatibility? To do this, use a spectrogram that visually represents the magnitude returned by the Short-Time Fourier Transform (STFT). STFT is a version of the Discrete Fourier Transform (DFT) that not only performs a DFT on a long signal, but also divides the signal into partially overlapping parts, each with a sliding window. Perform a DFT. It provides a two-dimensional spectral representation of sound fragments whose axes are time and frequency. The spectrogram can use the colormap to display the STFT output as images and feeds. it to the associated image-based network. Use a compact, performance-focused mobile net as a starting point. Our experiments provide a comparative study of the number of classes used in the spectrogram and related system configurations such as colormaps. Extensive research has been conducted in recent years to shed light on possible solutions to the problems presented. The growing interest may be due to the biodiversity data assessment campaign, which is an annual challenge to raise awareness of bird clefs. The Bird CLEF 2017 training dataset contains over 36,000 audio files of 1500 types collected by Xenocanto, and the class does not necessarily contain the same number of audio samples. This task focuses on identifying individual audible species and separating multiple overlapping sounds in field recordings.

5) Bird Species Identification Using Deep Learning [5]

Publication Year: 2019

Author: Prof. Pralhad Gavali, Ms. Prachi Abhijeet Mhetre, Ms. Neha Chandrakhant Patil, Ms. Nikita Suresh Bamane, Ms. Harshal Dipak Buva

Journal Name: IRJET Journal

a. Summary: Currently, some bird species are rare and it is difficult to predict when a taxon of bird species is discovered. Of course, birds that appear in different scenarios come in different sizes, shapes, colors, and angles from a human perspective. Images are not a voice classification, but a powerful variation for identifying bird species. Also, the human ability to recognize birds in photographs is easier to understand. Therefore, this method uses the Caltech UCSD Birds200 [CUB2002011] dataset for training and testing purposes. It uses the DCNN (Deep Convolutional Neural Network) algorithm to convert the image to grayscale format and uses TensorFlow to generate a sine to generate several comparison nodes. Compare these different nodes to the test dataset to get a scoresheet. After reviewing the scorecard, you can use the highest score to mark the bird species you need. Experimental analysis of the dataset (eg Caltech UCSD Birds 200 [CUB2002011]) shows that the algorithm provides 80% to 90% bird identification accuracy. I am doing experimental research on the Ubuntu 16.04 operating system using the Tensor thread library.

6) Bird Species Identification Using Spectrogram Based on Multi-Channel [6]

Publication Year: 2021

Author: Zhang, F.; Zhang, L.; Chen, H.; Xie, J.

Journal Name: IRJET Journal

a. Summary: Deep Convolutional Neural Networks (DCNN) has made breakthroughs in identifying bird species using bird bark spectrograms. An integrated function discriminative model (SFIM) with residual blocks and a modified weighted cross-entropy function has been proposed to address the imbalances in the new voice dataset. To further improve the discriminant accuracy, we constructed two multichannel aggregation methods using three SFIMs. One of them combined the outputs of the three SFIM feature extraction parts (feature pooling mode), and the other combined the outputs of the three SFIM classifiers (result pooling mode). SFIM was trained in three spectrograms calculated using short-time Fourier transform, choke frequency spectrum transform, and chirp transform, respectively. Transfer learning has been used in multi-channel models to overcome the lack of learnable model parameters. Using our own vocal dataset as a sample set, the resulting fusion mode model outperformed the other proposed models with a maximum average accuracy (MAP) of 0.914. If you select three spectrogram periods of 100 ms, 300 ms, and 500 ms for comparison, the results show that 300 ms is the best time period for your dataset. We suggest that the duration be determined based on the distribution of the duration of the new syllable. The maximum average classification accuracy (Camp) has reached 0.135 for the performance of the BirdCLEF2019 training dataset. This means that the proposed model has some generalization capability.

III. WHY IT IS IMPORTANT?

Birdwatching by bird sounds is important for many environmental and scientific purposes. Like, [8]

- Reduce the need for volunteers in such biological projects
- observed easily by experienced bird watchers
- identify and count birds in a specific area
- to estimate long-term population trends

A. Bird Classification

The modern Linnaean taxonomy classifies species according to their degree of relationship. Species of the same genus share a more recent common ancestor than species of different genera. The same applies to the grouping of genera into a family by family and neck. Studies comparing the

DNA of new species shed new light on (and continue to) the relationship between birds, but nevertheless many groupings based on the original morphological grounds It turned out to be true. Birds are better known than any other animal group, but new species are discovered almost every year. Approximately 10,000 newly identified species of birds are known and are classified into more than 200 families, more than 2200 genera and 29 eyes. About two-thirds (about 6,000) of known species belong to the order Passerine. The remaining 4000 or so belong to 28 orders of nearly 100 families and about 1000 genera. The following summary of how birds today are classified into different eyes, families, and genera gives an idea of bird species diversity. [21]

B. Birds Evolution

The oldest chook but discovered, is thought from numerous fossils recovered from satisfactory slate deposits in Germany. This magpie-sized animal lived in a tropical environment about 150 million years ago. Archaeopteryx had flight feathers and a fused furcular (i.e., wishbone) reminiscent of the ones of present day birds; those capabilities are, so far, particular in Jurassic animals. Its strong bones and shortage of a keel at the breastbone confined its powers of flight. This made it extra a glider than a powered flyer and it is able to were succesful handiest of gliding from perch to perch. Archeopteryx had many traits discovered additionally in theropod dinosaurs, and plenty of scientists now agree with that birds advanced from feathered dinosaurs. Most fossil birds of the Cretaceous Period (14.265 million years ago) belong to extinct organizations consisting of the enantiornithes, or "contrary birds," characterised via way of means of capabilities of scapula and leg bones which can be contrary to the ones of present day birds. Birds that survived the past due Cretaceous mass extinction various into species which can be present day in look and might frequently be labeled as present day chook organizations.[21]

IV. PROCESS & TECHNIQUES

Techniques which can used in audio signals:[8]

- For low noise single bird sounds, the energy-based time domain approach is reliable.
- For multiple bird calls in a noisy environment, or multiple bird calls in a noisy environment, segmentation uses a 2D time frequency approach.
- Spectrogram Segmentation.[6]

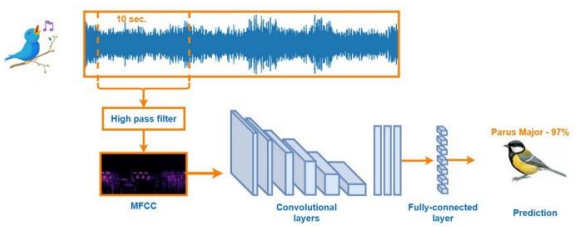


Fig. 3. Solution Overview-Audio Data Preproc essing and Neural Network Models

V. LITERATURE SURVEY

After doing literature survey, it is seen that many researches are using CNN, DCNN and multiple CNN in bird classification. It is find that after using spectrogram extraction, researchers getting more accuracy to classify/predict of bird species from raw data. It is found that many organizations like Caltech-UCSD Birds 200, Bird CLEF etc. are organizing bird species classification with large dataset. Any researcher has not implement this project in village areas.

A. Advantage

- Today's youth are showing their interest in the field like bird species classification.
- Experienced bird watching helps identify birds.
- Models can also implement in specific areas.
- Villager can also user this model for protecting their crops from harmful birds.
- Search for knowledge in various fields such as normalization, image recognition by data expansion, and deep CNN.
- CNN was developed for image data and is possibly the most efficient and flexible model for image classification problems.

B. Dis-advantage

- It is difficult to identify bird species from loud noises.
- During the monsoon season, the sensor may not work due to heavy rain.
- Need to explore more places for better and large data of birds.

C. Problem Statement

- It is notify that many researchers has researched and create model for bird species classification with different methods, but no model are implemented in village.
- So the villages can grow crops in their farms without hesitating of any harmful bird.
- By implementing this project in village, that villager will always have information of bird's presents in their farm.

VI. PROPOSED MODEL

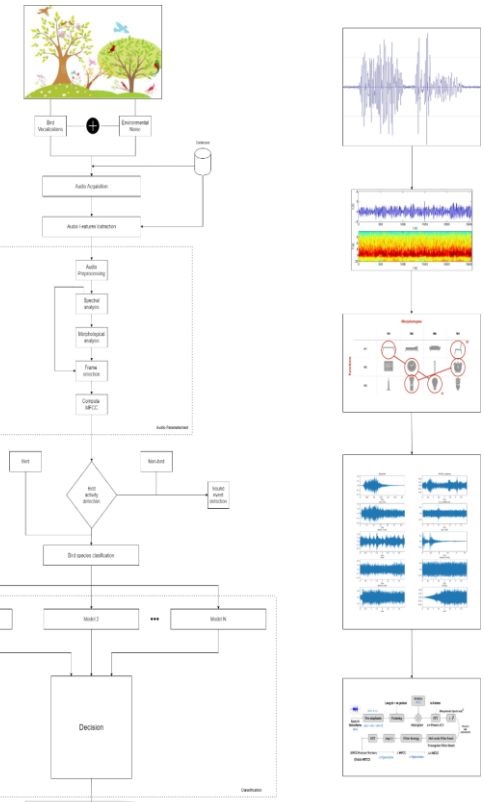
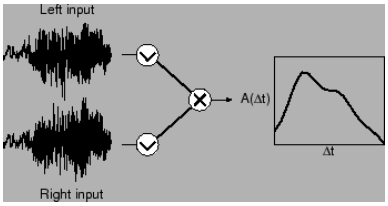


Fig. 4. Proposed Model of Bird species classification

A. Audio Pre-Processing

An audio pre-processing can also be called audio/sound classification. By this study we will perform different comprehensive experiment in audio preprocessing using time-frequency representation, Logarithmic magnitude compression, frequency weighting, & scaling are commonly used in audio preprocessing. So, by below way we can classify the recorded audio into two different classes.



$$A(\Delta t) = \sum_{t} |x_L(t)||x_R(t+\Delta t)|$$

Fig. 5. Audio pre-processing [14]

Fig. 6. Audio pre-processing Formula [14]

B. Spectral Analysis

Spectral analysis analyzes several frequency terms such as energy and eigenvalues. In some areas, spectral analysis may be related to chemistry and physics. The audible frequency indicates another frequency generated by the audio data. In the audio spectrum, tones are a representation of tone samples. Classify the various spectra present in the audio dataset and create a frequency diagram. There are multiple frequencies for all audible frequencies divided by the morphological analysis process.[15]

Critical Function 1	Critical Function 2	Critical Function 3	Critical Function 3
Solution 11	Solution 12	Solution 13	Solution 14
Solution 21	Solution 22	Solution 23	Solution 24
Solution 31	Solution 32	Solution 33	Solution 34

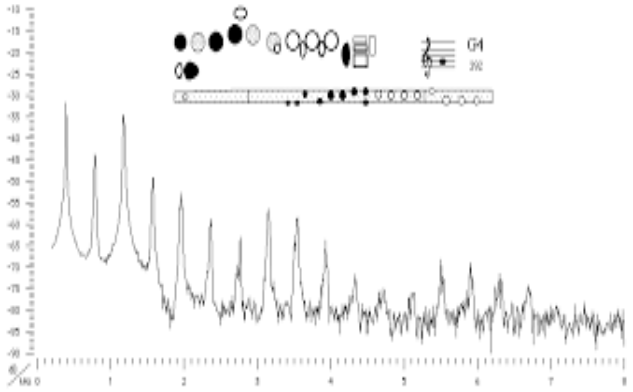


Fig. 7. Spectral Analysis [15]

C. Morphological Analysis

In morphological analysis it analysis the frequencies made from spectral analysis. The morphological analysis is a general analysis for exploring the methods of all possible solution to multiple dimensional and non-quantities problems. There are several frequencies are present in audio all frequencies and divided through morphological analysis process.[16]

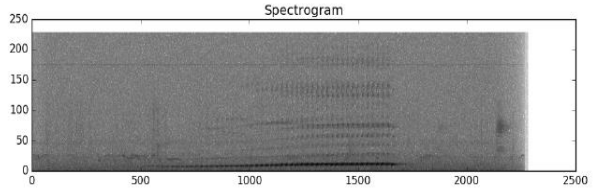


Fig. 8. Morphological analysis example [16]

D. Frame Selection

After Morphological analysis its time to divide different frequencies inti frame with the help of frame selection method. It consist audio frames, sample frames, amplitude information at point of time. Thre are many frams are played in a sequence to create or produce frequencies. In case of audio frame there is around 44000+ frames/sec are generated.[17]

Fig. 9. Frame selection from recording [17]

E. Mel-Frequency cepstral coefficient

Mixed frequency coefficients (MFCC) are the coefficients that together make up the MFC. This frequency distortion allows, for example, better sound representation in audio compression. The MFCC is usually derived as follows: Performs a Fourier transform on the signal (the window part of the signal). [18]

• Frame the signal into short frames.

• Calculate periodogram estimates of the power spectrum on a frame-by-frame basis.

• Apply the Mel filter bank to the power spectrum and calculate the total energy of each filter.

• Get the logarithm of all energies in the filter bank.

• Get the energy DCT from the logarithmic filter bank.

• Keep DCT factor 213 and discard the rest. [19]

Mel Frequency Cepstral Coefficients (MFCCs) are the standard choice of audio features used in speech recognition, and most of their success is due to the amount of information contained in such compressed audio signals. is. Mel frequency cepstrum is similar to how the human ear processes sound. The following steps are used to calculate the MFCC.*I)* compute the power magnitude spectrogram of the signal,

II) Join the FFT bin to the Mel frequency bin,

III) Take the logs of these powers, and

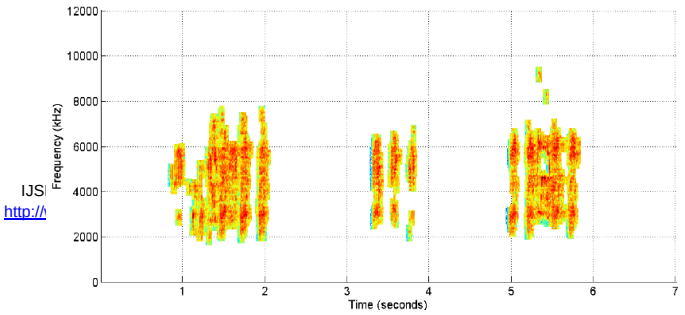
IV) Perform a Discrete Cosine Transform.

MFCC is the amplitude of the resulting spectrum.

F. Data Augmentation

Data expansion is a way to increase the number of training samples in a dataset by expanding the training data. The Bird CLEF dataset is one of the largest bird bark datasets available, but the number of training examples by bird species is fairly small. An average of about 30 samples per sound class.

Fig. 10. Spectrogram Frame [23]



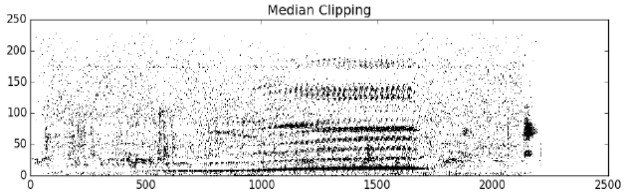


Fig. 11. Median Clipping Frame [23]

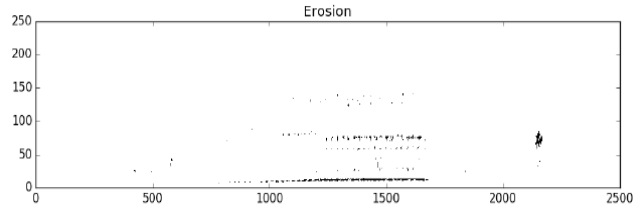


Fig. 12. Erosion Frame [23]

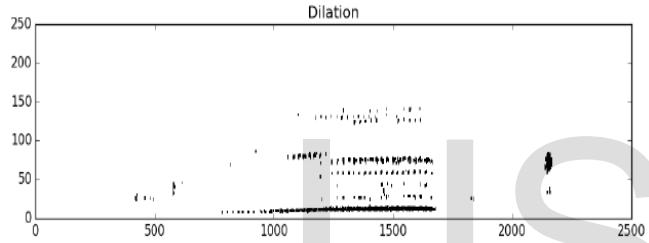


Fig. 13. Dilation Frame [23]

G. Spectogram

The spectrogram of a discrete audio signal ¯x = x1… xn is calculated in two or three steps. First, a short-time Fourier transform (STFT) is applied to the audio signal. The STFT is computed in a standard way by splitting the signal into different overlapping frames, and then compute the Discrete Time Fourier Transform (DTFT) for each frame, which results in a matrix with complex values.

Fig. 14. STFT Formula [23]

Where xn is the input signal at time n, w (n) is a length M = 512 Hann window cantered around n, R = 128 is the hop size between consecutive frames. That is, use a size 512 Hanwindow with a 75% overlap. We use the librosa.stft method of the library librosa to compute the STFT. Secondly, the squared amplitude of the magnitude of the STFT is computed, which we call an amp spectrogram, and thirdly the natural logarithm of the amplitude spectrogram is computed which we refer to as a log spectrogram.

Fig. 15. Amp spectrogram and log spectrogram Formula [23]

VII. Research Scope

There is still more scope to improve accuracy in bird species classification in specific areas and villages areas. So how we can achieve mentions this as shown below,

- To use more and large Datasets.
- Noisy audio file which recorded in environment and implementing methods in this kind of files.
- Robust Techniques
- Real time audio recording

VIII. Acknowledgement

Categorizing bird species is a tedious task. Machine learning algorithms can be used to create models that can effectively identify bird species based on bird characteristics. In this study, there were many challenges in capturing audio from specific areas and classifying birds. With the help of deep learning algorithms and other models, this study will be simplified. I have a large dataset that has a lot of memory to store.

References

[1] Hadar Abukrat? and Indu N. Korambath??University of California San Diego, La Jolla, CA 92093-0238. BIRD AUDIO CLASSIFICATION

[2] Piyush Bidwai, Vaibhav Mahalle, Eshan Gandhi, Sharda Dhavale. Bird Classification using DeepLearningVolume: 07 Issue: 05 | May 2020e-ISSN: 2395-0056p-ISSN: 2395-0072 www.irjet.net

[3] Chih-Yuan Koh1?, Jaw-Yuan Chang1?, Chiang-LinTai1, Da-Yo Huang1, Han-Hsing Hsieh2, and Yi-Wen Liu1 hank0924097907.ee04@g2.nctu.edu.tw. Bird Sound Classification using Convolutional Neural Networks.

[4] A´ gnes Incze, Henrietta-Bernadett Jancso´, Zolta´n Szila´gyiy, yCodespring Ltd. Bird Sound Recognition Using a Convolutional Neural Network.

[5] Prof. Pralhad Gavali, Ms. Prachi Abhijeet Mhetre, Ms. Neha Chandrakhant Patil, Ms. Nikita Suresh Bamane, Ms.

$$STFT\{\bar{\mathrm{x}}\}(m,\omega) \equiv X_m(\omega) = \sum_{n=-\infty}^{\infty} x_n w(n-mR) e^{-j\omega n}$$

Harshal Dipak Buva Bird Species Identification using Deep Learning ISSN: 2278-0181 Vol. 8 Issue 04, April-2019.

[6] Zhang, F.; Zhang, L.; Chen, H.; Xie, J. Bird Species Identification Using Spectrogram Based on Multi-Channel Fusion of DCNNs. Entropy 2021, 23, 1507. https://doi.org/10.3390/e2311157 Academic Editor: Ke

$$ampspectrogram\{\bar{\mathrm{x}}\}(\omega) \equiv |\bar{\mathbf{X}}(\omega)|^2$$

$$logspectrogram\{\bar{\mathrm{x}}\}(\omega) \equiv log_e(|\bar{\mathbf{X}}(\omega)|^2)$$

GuReceived: 5 August 2021Accepte.d: 12 November 2021Published: 13 November 2021.

[7]https://ieeexplore.ieee.org/abstract/document/8281775/authors#authors

[8]https://www.slideshare.net/ParthaSarathiKar3/bird-species-classification-based-on-their-sound

[9]https://www.vedantu.com/question-answer/the-soundproducing-organ-in-a-singing-bird-is-a-class-11-biology-cbse-5f4fdfce4ec624396d234cf0

[10] https://en.wikipedia.org/wiki/Village_weaver

[11] https://en.wikipedia.org/wiki/Bird

[12]https://towardsdatascience.com/sound-based-bird-classification-965d0ecacb2b

[13]https://www.ijltet.org/journal/152539753656.%202443.pdf

[14]https://www.usenix.org/legacy/publications/library/proceedings/es99/full_papers/lee/lee_html/node6.html

[15] https://newt.phys.unsw.edu.au/jw/sound.spectrum.html

[16] https://en.wikipedia.org/wiki/Morphological_analysis

[17]https://stackoverflow.com/questions/3957025/what-does-a-audio-frame-contain

[18]http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[19]https://www.thecanadianencyclopedia.ca/en/article/bird-classification-and-evolution.

[20] https://www.ijedr.org/papers/IJEDR2102047.pdf

[21] https://www.trendhunter.com/trends/bird-species