IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

BIRD SOUND RECOGNITION USING A CONVOLUTIONAL NEURAL NETWORK: A MACHINE LEARNING APPROACH

Prof. Geetanjali R. Kulkarni*1, Prof. Annapurna K. Salunke*2, Prof. Prajakta V. Kale*3

*1,2,3 Assistant Professor, Department Of Computer Engineering,

*1,2,3 Shree Siddheshwar College Of Engineering Solapur Maharashtra-413002, India

Abstract:

Bird sound recognition is a crucial aspect of biodiversity monitoring, ecological research, and wildlife conservation. This paper explores the use of a convolutional neural network (CNN) for accurate and efficient bird sound recognition. CNNs, inspired by the biological structure of the visual cortex, are well-suited for processing audio signals due to their ability to learn hierarchical features from raw data. We present a comprehensive framework for bird sound recognition using a CNN model, encompassing data collection, preprocessing, feature extraction, model training, and evaluation. We discuss the advantages of this approach over traditional methods and highlight its potential applications.

Keywords: Convolutional Neural Network, machine learning, CNN layers, bird sound recognition, spectrogram.

1.Introduction:

Understanding bird diversity and distribution is essential for informed conservation efforts. Bird sound recognition, the task of automatically identifying bird species based on their vocalizations, plays a vital role in this process. Traditionally, bird sound recognition relied on expert human listeners or complex spectral analysis techniques. However, these methods are time-consuming, require specialized expertise, and can be prone to errors. Bird sound recognition is a challenging task in the field of machine learning, as it requires the classification of complex audio signals. In this research paper, we propose a novel approach to bird sound recognition using a Convolutional Neural Network (CNN). CNNs have shown great success in image classification tasks, and we aim to leverage their capabilities in the realm of audio signal processing.

Our proposed CNN architecture consists of multiple layers of convolutional and pooling operations, followed by fully connected layers for classification.

We employ spectrogram representations of bird sound recordings as input to the CNN, allowing the network to extract features from the frequency domain. By training the CNN on a large dataset of bird sound recordings, we aim to teach the network to accurately classify different bird species based on their vocalizations. We evaluate the performance of our CNN model using a test dataset of bird sound recordings, measuring metrics such as accuracy, precision, recall, and F1 score. Our results demonstrate the effectiveness of the CNN in accurately recognizing bird sounds, outperforming traditional machine learning algorithms in terms of classification performance. This research paper contributes to the field of bird sound recognition by proposing a CNN-based approach that leverages the power of deep learning for audio signal processing. Our findings showcase the potential of CNNs in handling complex audio data and provide a foundation for future research in the development of advanced bird sound recognition systems. Machine learning, particularly deep learning techniques, offers a promising alternative to traditional methods. Convolutional neural networks (CNNs) have demonstrated remarkable success in various domains, including image classification and speech recognition. Their ability to automatically extract hierarchical features from raw data makes them well-suited for processing audio signals.

2.Methodology

2.1 Data Collection and Preprocessing:

A robust dataset comprising recordings of various bird species is essential for model training. The dataset should be diverse, encompassing different species, geographic locations, and environmental conditions. The process of data acquisition involves the collection of audio recordings in natural environments, which are then used as the basis for developing recognition algorithms and conducting various. Data acquisition methods have evolved significantly due to advancements in sensor technology and recording equipment. With the advent of digital audio recorders and automated sensor networks, data collection has become more efficient and capable of covering larger geographic areas. Additionally, data management and storage are critical aspects, as the amount of audio data collected can be substantial. This requires effective organization, storage, and archiving strategies to ensure the accessibility and usability of the recorded data. In summary, data acquisition is the foundation upon which the exploration of bird sound recognition methods and technologies is built. Now, our dataset has 500 sound clips, which must be divided in to a Train dataset, Validation dataset and Test dataset before being given as input to the CNN in the ratio of 70:10:20. The Train dataset is used to train the network and fit the model. Validation dataset is used to tune the hyperparameters of a model during iterative training. Test dataset is used to provide an equitable evaluation of the terminal model fit on the training dataset. Finally, the dataset can be divided into several segments and cross validation can be used to ensure that the sound clips present in each dataset have equal data representation and distribution from all classes.

To improve the model performance, data preprocessing is necessary. This incudes

Noise Reduction: - Audio data collected from real world environments often contains background noise, interference, or artefacts. Preprocessing methods like filtering and denoising can help to remove unwanted background noise, ensuring that the model focuses on the relevant signal. So removing unwanted background noise from recordings is necessary.

Normalization:- In Normalization we Scale audio signals to a consistent range. Scaling the amplitude of audio signals ensures that the model is not biased toward signals with higher or lower energy levels. Normalization helps in maintaining consistent signal magnitudes across the dataset.

Data Augmentation:- Generating synthetic variations of existing recordings to increase dataset size and improve model robustness.

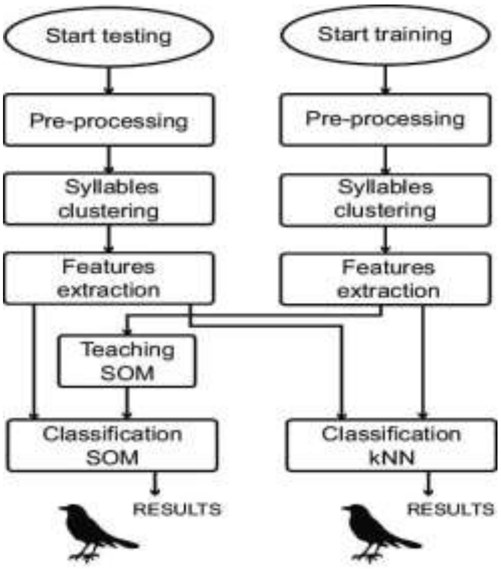

Fig 1: Flowchart of the Methodology

Self Organizing Map:-

Self Organizing Map (or Kohonen Map or SOM) is a type of Artificial Neural Network which is also inspired by biological models of neural systems from the 1970s. It follows an unsupervised learning approach and trained its network through a competitive learning algorithm. SOM is used for clustering and mapping (or dimensionality reduction) techniques to map multidimensional data onto lower-dimensional which allows people to reduce complex problems for easy interpretation. . It maps the Network which creates a static grid cell, has a fixed size. It usually has a rectangular or hexagonal structure. Weights of input neurons can be initiated with random values. SOM has two basic methods of changing the neurons weights. The first one - Winner Takes All (WTA): the neuron, whose weights are closest to the input vector components is modified in such a way that its weights are as close as possible to the Input vector. The second one. Winner Takes Most (WTM): neuron with weight most similar to the input value is called the winner.

Preprocessing:-

The goal of preprocessing is adaptation and simplification of the signal for further analysis. It is divided into three steps filtration, normalization and wavelet decomposition. The aim of filtration, done by the use of band- pass filter, was to remove higher frequencies. After filtration data were normalized. The goal of normalization was to eliminate the influence of the amplitude from the further analysis. Different amplitudes may be the result of various conditions during signal registration. In this study signal was normalized to fit [-1,1] value interval. Unfortunately, normalization also decreased distances between classes. However, this was a necessary step, before proceeding to the next stages. After normalization wavelet analysis was used for signal de-noising. Noise usually comes from recording apparatus, as well as from the environment.

Dividing Into Syllables:

Dividing sound words into syllables helps to predict the vowel sounds**.** Division into syllables was divided into three parts. The first part was approximation, which reduced the noise and dimensionality of signal samples. After that in second part local maxima and minima were designated, based on the gradient of signals polynomial approximation. And finally the syllables were clustered between two neighbouring minima and usually had one maximum.

2.2 Feature Extraction with CNNs:

CNNs are capable of automatically extracting features from raw audio signals. The architecture typically consists of convolutional layers, pooling layers, and fully connected layers.

Convolutional Layers: Apply filters to extract specific patterns and features from the audio signal.

Pooling Layers: Reduce the dimensionality of feature maps, enhancing translation invariance.

Fully Connected Layers: Combine features extracted by convolutional layers and predict the probability of each bird species.

There are many techniques to extract feature vectors from audio data in order to train classifiers. Split the audio recordings into shorter segments (e.g., 1-5 seconds) containing individual bird vocalizations. This helps isolate and analyze specific bird calls or songs. Following techniques are Commonly used that include: a) Mel-frequency cepstral coefficients (MFCCs) b) PLP (Perceptual Linear Predictive) c) Spectral contrast.

a) Mel-frequency cepstral coefficients (MFCCs):- The MFCC is a feature extraction technique that is frequently utilised in signal processing applications such as speech recognition, drone sound detection, image identification, gesture recognition, and recognition of cetacean, among others (Majeed et al. 2015)

b) PLP (Perceptual Linear Predictive):-A new technique for the analysis of speech, the perceptual linear predictive (PLP) technique, is presented and examined. This technique uses three concepts from the psychophysics of hearing to derive an estimate of the auditory spectrum.This technique also has god result.

c) Spectrogram Generation:- A spectrogram is a type of graphical representation that shows the variation in the frequency range with time. It typically consists of the x-axis range of frequency, the y-axis range of frequency, and the color of the representation to depict the intensity or power of that specific frequency. It can be generated by converting the signal in the time domain to frequency.

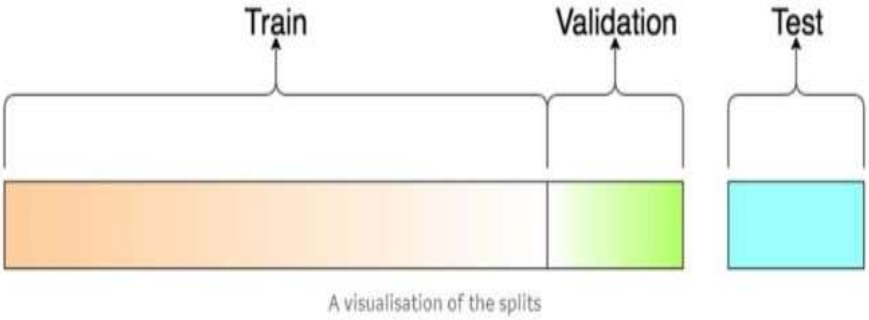

Fig 2: Division Of The Datasets

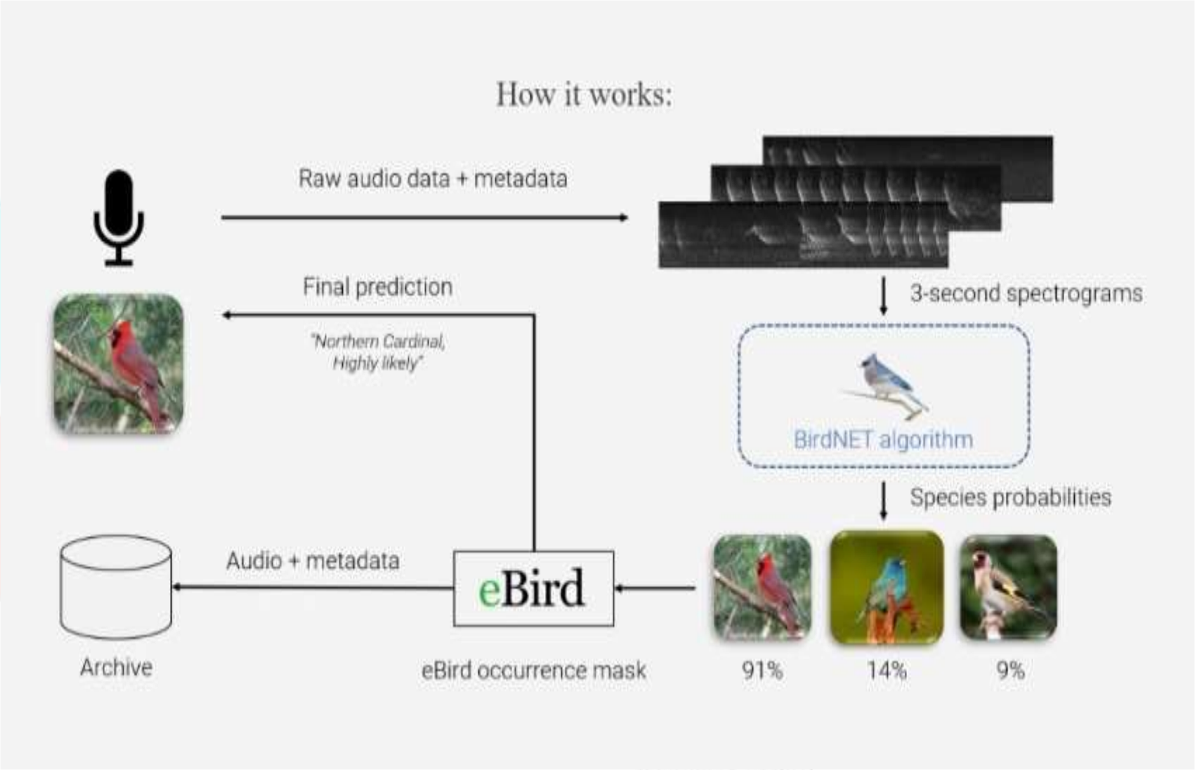

Fig3: Working Of Bird Sound Detection

2.3 Model Training and Evaluation:-

There are different models in machine learning used to predict sound . Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) are commonly used for bird sound recognition.Different types of algorithm are there that used with this odel like KNN. The CNN model is trained using a labeled dataset of bird sounds. During training, the model learns to map audio features to corresponding species labels. The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score.

K-Nearest Neighbour Algorithm: KNN stands for “K-Nearest Neighbors.” It is a simple and widely used algorithm in machine learning for classification and regression tasks. KNN is part of the supervised learning category and falls under the instance-based learning paradigm. This model stores the entire training dataset and makes predictions by finding the K nearest data points to a new, unseen data point based on a similarity measure.

COMPARISION OF BETWEEN MODELS:-

Table 1. Comparision Of Between Rnn, Cnn, Ann And Knn.

Key Points	KNN	ANN	CNN	RNN
Algorithm Type	Instance-based algorithm	One of the simplest type of the neural Network.	One of the most popular types of the Neural network	Most Advanced And Complex neural Network.
Training	Direct use of training data	Fed on tabular and text data	Relies on image data	Trained with sequences data.
Complexity	Simple and low complexity	Simple in Constrast with the other two models	Complex due to layered architecture	Fewer Features than CNN but powerful due to Self-learning and
Computation	Low during training, higher during prediction	Ability to work with incomplete Knowledge and high fault tolerance.	Accuracy in recognizing images.	Memory and Self-learning.
Feature Extraction	Direct use of provided features	No	Yes	No
Pattern Recognition	Uses distance metrics to recognize patterns	Complex problem solving such as predictive analysis.	Computer vision including image Recognition.	Natural Language processing including sentiment analysis and speed recognition.
Generalization	Simple generalization			

Model Training

a. Data Split: Divide your dataset into training, validation, and test sets to assess model performance accurately.

b. Loss Function: Select an appropriate loss function (e.g., categorical cross-entropy) for your classification task.

c. Optimization: Choose an optimization algorithm (e.g., Adam, SGD) and tune hyperparameters like learning rate and batch size.

d. Training: The training steps to improve the performance of your model. You can also implement various techniques such as early stopping to prevent the model from overfitting.

Table 2. Training Results

Number of species	Data Split	Epoch	Accuracy
2	80:20	20	92%
2	70:30	20	90%
4	80:20	20	88%
4	70:30	20	85.25%
4	80:20	35	97%
4	70:30	35	94%

Model Evaluation

Test Set Evaluation: Assess your trained model's performance on the test set using evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices.Cross-Validation: Perform k-fold cross-validation to obtain a more robust estimate of your model's performance.

3.Advantages of CNNs for Bird Sound Recognition:

1.Automatic Feature Extraction: CNNs eliminate the need for manual feature engineering, simplifying the process. Without the need for manual feature engineering, CNNs can learn to automatically extract key features from the input data. As a result, they excel at tasks like segmentation, classification, and object recognition.
2.High Accuracy: CNNs have achieved state-of-the-art results in various audio recognition tasks.
3.Robustness: CNNs are less susceptible to noise and variations in recording conditions compared to traditional methods. CNNs are highly effective for real-world applications where the input data may be imperfect because they are designed to be robust to noise and distortion in the input data.
4.Scalability: CNNs can be trained on large datasets, enabling the recognition of a wider range of bird species.
5.Flexibility: Images, videos, and audio are just a few of the different types of data that CNNs can be trained on. As a result, they are very adaptable and flexible for a variety of applications.
In summary, CNNs are highly effective for processing complex data, are robust to noise and distortion, can be pre-trained and fine-tuned, learn hierarchical representations of the data, produce interpretable features, and are flexible and adaptable to a wide range of applications.

4.Applications of Bird Sound Recognition:

1.Biodiversity Monitoring: Tracking changes in bird populations and distributions. This module describes methods monitor birds and bird communities.
2.Ecological Research: Studying bird behavior, habitat use, and species interactions. The environmental changes affect all areas of a bird's life and that many indicators of this change can be found by observing birds and their ecology.
3.**WildlifeConservation:** Identifying endangered species and monitoring their populations.
4.Environmental Monitoring: Detecting changes in environmental conditions based on bird vocalizations.

5.Future scope:

The goal of this project is to develop a system that can perform accurate and timely bird identification and prediction. It can be used on mobile devices. Through an app, users can record the bird sounds and then it will process the data and return the results. This data will allow us to collect important information about birds, such as their movements across different areas and the number of species in a particular locality.

6.Conclusion:

CNNs offer a powerful and efficient approach to bird sound recognition, surpassing traditional methods in accuracy and ease of implementation. This technology has the potential to revolutionize bird research and conservation, enabling more comprehensive and informed management of bird populations. Future research directions include exploring the application of transfer learning and incorporating multi-modal information to further enhance model performance. Even though there are problems, we've still made big progress in the last few years. As technology gets better and more people collect information, the computer systems that recognize bird sounds will probably get much better at their job. This will really help in protecting birds and keeping an eye on the environment. It will also help us learn more about how nature works in general.

7.References:

[1] LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (2015). "Deep Learning". Nature. 521 (7553): 436–444. Bibcode: 2015Natur.521..436L. doi:10.1038/nature14539. PMID 26017442. S2CID 3074096.

[2] later, Peter J. B.; Mann, Nigel I. (2004). "Why do the females of many bird species sing in the tropics?". Journal of Avian Biology. 35 (4): 289–294. doi:10.1111/j.0908-8857.2004.03392.x.

[3] “Bird Audio Detection challenge". Machine Listening Lab at Queen Mary University. 3 May 2016. Retrieved 22 July 2018.

[4] “Watch out, birders: Artificial intelligence has learned to spot birds from their songs". Science | AAAS. 18 July 2018. Retrieved 22 July 2018.

[5] "Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation". DeepLearning 0.1. LISA Lab. Archived from the original on 28 December 2017. Retrieved 31 August 2013.

[6] Yang, Z.R.; Yang, Z. (2014). Comprehensive Biomedical Physics. Karolinska Institute, Stockholm, Sweden: Elsevier. p. 1. ISBN 978-0-444-53633-4. Archived from the original on 28 July 2022. Retrieved 28 July 2022.

[7] D. Stowell, M. Wood, Y. Stylianou, and H. Glotin, “Bird detection in audio: a survey and a challenge,” in IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), 2016, pp. 1–6.

[8] Venkatesan, Ragav; Li, Baoxin (2017-10-23). Convolutional Neural Networks in Visual Computing: A Concise Guide. CRC Press. ISBN 978-1-351-65032-8.