|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S. No.** | **Name** | **Year** | **Authors** | **Model** | **Proposed Method** | **Dataset used** | **Advantages** | **Disadvantages** | **Results** |
| 1 | Learnable Acoustic Frontends In Bird Activity Detection | 2022 | * Mark Anderson * Naomi Harte | EfficientNet B0 | Learnable:   * Per-Channel Energy Normalisation (PCEN) * Time-Domain Filterbanks (TD) * Spectro-Temporal Filters (STRF) * Learnable Audio Frontend (LEAF)   Non-learnable frontends:   * Mel spectrograms * Linear spectrograms * Log compressed mel spectograms | * DCASE2018 BAD Challenge * BirdVox-DCASE-20k * freefield1010 * warblrb10k | * Good efficiency of computational resources and accuracy. * Possible to deploy on edge-based systems too. | * Work only on clean data (not noisy)Space constraint for learning filterbacks in species agnostic. * Poor performance for interpretable learnable filterbacks. | Per-Channel Energy Normalization (PCEN) is best with an overall accuracy of 89.9%.  Can conclude that compressing the learnable spectrogram magnitudes gives the most increase in performance |
| 2 | Few-shot Long-Tailed Bird Audio Recognition | 2022 | * Marcos V. Conde * Ui-Jin Choi | tf\_efficientnet\_b0\_ns | * 9 backbones, 22 models * CNN * tf\_efficientnet\_b0\_ns | * BirdCLEF 2022 * BirdVox-DCASE-20k * freefield1010 * train\_soundscapes | * Small dataset is enough * Weak labels can also be used to train * Fine-grained vocals can be classified very fast | * Overfitting of certain classes | Achieved a local validation (CV) of 0.8745 |
| 3 | A Survey On Bird Species Identification Using  Audio Signal Processing And Neural Network | 2022 | * Prof. Pooja Wale * Abhishek Mankar * Pratik Padale * Sanket Gawade * Prasanna Ghogare | CNN | Using the image of the spectrogram for classification using a CNN.  Detects parts and extract features from various convolution layers to be aggregated and then put into a classifier. | NA | Predicts getting a good accuracy | Model not fine-tuned on a dataset | NA |
| 4 | Weakly-Supervised Classification and Detection of  Bird Sounds in the Wild. A BirdCLEF 2021 Solution | 2021 | * Marcos V. Conde * Kumar Shubham * Prateek Agnihotri * Nitin D. Movva * Szilard Bessenyei Marcos V. Conde * Kumar Shubham * Prateek Agnihotri * Nitin D. Movva   Szilard Bessenyei | * PANN DenseNet-121 * ResNeSt-50 * EfficientNet-B0 | * For pre-processing 6 Augmentations are done. * 13 (CNN) models are taken into the ensemble and a Support Vector Classifier is used in the second stage. | BirdCLEF 2021 Challenge | Birds can also be identified from Noisy Soundscapes. | * Hardware requirements of mobile not satisfied. * High inference time. | * Ensemble of models worked better * Smaller architectures gave better results. * Final model gave 0.7801 as the F1 score. |
| 5 | Audio-based Bird Species Identification with Deep  Convolutional Neural Networks | 2020 | Mario Lasseck | * DCNN * Pre-trained:   + InceptionV3   + ResNet152   + DualPathNet92   + InceptionV4   + DensNet   + InceptionResNetV   + Xception   + NasNet | Data Loading:   * Audio extracted from the file of fixed duration(5 seconds) * Fourier transform applied * Normalization * Conversion of power and linear spectrogram using logarithm * Discarding frequencies that are high and low * resize spectrogram * Change grayscale to RGB   Data Augmentation:   * pitch shift * frequency stretch * time interval dropout * time stretch * frequency * interpolation filters * color jitter   The first run used a single model for both subtasks  Second run, an ensemble of the first two is done.  Third run ensemble with a new model.  Fourth run, same model with more snapshots of same data. | * LifeCLEF 2018 * BirdCLEF 2016 | * Extracted audio from randomized positions * Good train validation split * A variety of data augmentations applied to increase the dataset and prevent overfitting | * Since piece-wise time stretching is applied the impact of duration jitter is not found. * This is because of augmentation. * Certain augmentation techniques can be avoided like degradation of quality. | InceptionV3 was the best model. |
| 6 | Audio Based Bird Species Identification using Deep Learning Techniques | 2018 | * Elias Sprengel * Martin Jaggi * Yannic Kilcher * Thomas Hofmann | CNN | Data Augmentation:   * Time Shift * Pitch shift * Combining same class files * Noise   Model:   * 6 layers – 5 convolutional, 1 dense. * All convolution layers have rectified activation function. * Convolution layer is followed by max pooling. | BirdCLEF2016 | Foreground species were identified perfectly. | * Background species not detected well * Training with data augmentation creates a loss for the detection of background species | The highest recorded accuracy score in 2018 was 0.58 and the Mean Average Precision (MAP) of 0.69. |
| 7 | Audio-only Bird Species Automated Identification Method with  Limited Training Data Based on Multi-Channel Deep  Convolutional Neural Networks | 2018 | * XIE Jiang-jian * DING Chang-qing * LI Wen-bin * CAI Cheng-hao | VGG-16 pretrained | * Data accumulated * Preprocessing of the signal using Hamming window. * Spectogram calculation using Mel-frequency Cepstral Transform (MFCT), Short Time Fourier Transform (STFT), and Chirplet Transform (CT). * DCNN model (VGG16) pretrained on ImageNet is used to extract features and sent to 2 convolution layers followed by softmax to identify 18birds. * Accuracy was improved by using 3 models parallelly fusing outputs of features from the VGG16 model. | Own dataset made from wild birds recorded in Beijing Song-Shan National Nature Reserve | * Lesser trained parameters * Can be used with small samples as makes use of transfer learning. * Ch spectrogram works better than Mel. | Denoising was not performed thus can’t identify background birds. | Highest Mean Average Precision (MAP) of 0.9998 |
| 8 | Automated Bird Species Identification using Audio  Signal Processing and Neural Networks | 2020 | * Chandu B * Akash Munikoti * Karthik S Murthy * Ganesh Murthy V * Chaitra Nagaraj | AlexNet - Convolutional Neural Network (CNN) | * Pre Emphasis on higher frequency * Division into frames and removal of silence * Reconstruction by joining frames * Generating spectrograms through Fourier Transform * Pre trained AlexNet * Model trained on bird data and validated then applied to recognize a human | * Bird sounds in xeno-canto.com * Human voice in Google Audio-set * Human voice in LibriSpeech ASR Corpus | * First paper to work on removing silence in the audio during frame generation * Has minimum hardware requirements can be run on mobile | * Input collected in an ideal environment without noise, doesn’t work well in the presence of noise | 91% accuracy in the real environment |
| 9 | Bird Identification from Timestamped,  Geotagged Audio Recordings | 2019 | Jan Schlüter | Ensemble CNN and Multi-Layer Perceptron | * Convolution Neural Network (CNN) for pre-processing and Multi-Layer Perceptron (MLP) for prediction * Mel spectrogram used * Utilizes global pooling * Find Global predictions from local predictions considering it as a multiple-instance learning problem | BirdCLEF 2018 | * Can use arbitrary-length recordings * Doesn’t use averaging done by choosing manually or a majority voting. * Made use of audio and metadata for predictions. | Can either be used to target background or foreground species. Map value decreases if done to fit both.  Adding/presence of noise causes loss in the result. | Maximum foreground MAP 0.768 and background MAP of 0.698. |
| 10 | Bird species identification using spectrogram and dissimilarity approach | 2018 | * Rafael H.D. Zottesso * Yandre M.G. Costaa * Diego Bertolinia * Luiz E.S. Oliveira | Support Vector Machine (SVM) | * Zones are selected (vertical and horizontal) in the spectrogram to identify information-rich regions. * Features are extracted using Local Binary Pattern (LBP), Local phase quantization (LPQ), and Robust local binary pattern (RLBP). * Solution found based on dissimilarity * 8 subsets created from the dataset | LifeCLEF bird identification task 2015 | No Need to train the model if a new class introduced | * Only classes 23 to 915v identified in the dataset. * Cant compare results with other models as first time use of subsets | 71% identification rate in 915 classes |
| 11 | Bird Species Identification using Audio Signal Processing and Neural Networks | 2022 | * Amol Dhakne * Vaishnav M. Kuduvan * Aniket Palhade * Tarun Kanjwani * Rushikesh Kshirsagar | AlexNet - Convolutional Neural Network (CNN) | * For pre-processing silence removal, framing and reconstruction was done. * AlexNet was used for classification of spectrograms created. | Bird song downloaded from xenocanto.com | Can work with noisy data ( with human voice) | Only four species can be identified. | 97% accuracy was achieved in real time environment. |
| 12 | Convolutional Recurrent Neural Networks  For Bird Audio Detection | 2017 | * Emre C¸ akır * Sharath Adavanne * Giambattista Parascandolo * Konstantinos Drossos * Tuomas Virtanen | Convolutional Recurrent Neural Network (CRNN) | * Temporal max pooling for file level estimation unlike frame level. * Acoustic features are mapped to binary estimates of a bird song. * Spectrum is got through Short-Time Fourier Transform and Hamming window. * Convolution layers followed by GRU (Gated Recurrent Unit), max pooling, and finally a feed-forward layer. | Bird Audio Detection challenge | * Can extract features of high level which aren’t changed based on temporal features or spectral because of the CNN (Convolutional Neural Network) part of architecture. * Can learn long temporal context because of RNN (Recurrent Neural Network) part of architecture. * Less complex compared to other merthods | * Data augmentation not done * Final result performance dropped, reason not found. | * In the Bird Audio Detection Challenge was placed second. * For unseen data for evaluation the Area under ROC Curve (AUC) is 88.5%. |
| 13 | Detection of Bird Species Through Sounds | 2021 | * Shirley Cheng * Julie Wang | LeNet inspired CNN (Convolutional Neural Network) | * The CNN is based on LeNet architecture. * ReLu activation function with cross entropy loss function * Multi class classification of spectrogram images * 3 models with increasing layers and complexity with max pooling and batch normalization for all convolution layers * Accuracy improved using L2 and Dropout regularization | xeno-canto public dataset | NA | * Overfitting of the dataset * RAM constraints | Training set accuracy of 98% and testing set accuracy was 67%. |
| 14 | Recognizing Birds from Sound - The 2018 BirdCLEF Baseline System | 2018 | * Stefan Kahl * Thomas Wilhelm-Stein * Holger Klinck * Danny Kowerk * Maximilian Eibl | CNN (Convolutional Neural Network) | * Uses a CNN with 7 weighted layers and only 1 densely connected layer ( final layer) * Uses grouped convolutions * Increase performance by increasing features | * 2018 LifeCLEF bird identification * 2018 BirdCLEF Soundscape task | * Easy to use settings for CNN layout provided | * Long training time * Not using Metadata * More fine tuning when extraction of spectrogram is required. | The Mean Label Ranking Average Precision for background species and foreground species on local validation was 0.564. |
| 15 | Stacked Convolutional and Recurrent Neural Networks for Bird Audio Detection | 2017 | * Sharath Adavanne * Konstantinos Drossos * Emre C¸akır * Tuomas Virtanen | Stacked Convolutional and Bidirectional Recurrent Neural Network (CBRNN) | * Spectograms generated from audio * CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), and FC (Fully Connected) layers are stacked. * Output single node with a range between 0 and 1. 0 is the existence of birds call, 1 absence. * log Mel-Band Energy feature (mbe) and magnitudes (dom-freq) used as a features * Data Augmentation * Domain adaptation using test mixing | Bird audio detection challenge | * Dom-freq and mbe features improve positive detection. * First to have used domain adaptation | * Classifier more exposed to positive cases because of data augmentation. * Data Augmentation not useful | Highest Area Under the ROC curve (AUC) was 95.5% for development data and for evaluation data 88.1% on 5 cross validation |