### Citation:

Mehyadin, Aska E., et al. “Birds Sound Classification Based on Machine Learning Algorithms.” Asian Journal of Research in Computer Science, June 2021, pp. 1–11. journalajrcos.com, <https://doi.org/10.9734/ajrcos/2021/v9i430227>.

### Dataset Used:

The study used the "Female Feature MFCC" dataset from Kaggle, designed initially for predicting emotions from MFCC values, showing a 94% accuracy for emotion prediction. This dataset was repurposed to classify bird sounds.

### Algorithms and Classification Techniques:

- \*\*Naïve Bayes\*\*

- \*\*J4.8\*\* (a decision tree algorithm)

-\*\*Multilayer Perceptron (MLP)\*\*

### Methodology:

- The methodology included \*\*data collection\*\*, \*\*pre-processing\*\* (noise reduction and filtering), \*\*feature extraction\*\* using MFCC, and \*\*classification\*\* using the mentioned algorithms. The study highlighted the importance of preprocessing to enhance the quality of sound recordings and the effectiveness of MFCC in capturing essential features for classification.

### Accuracy Achieved:

- Accuracy for the algorithms was reported as follows:

- \*\*Naïve Bayes\*\*: 47.45%

- \*\*J4.8\*\*: 78.40%

- \*\*Multilayer Perceptron (MLP)\*\*: 74.68%

- J4.8 not only had the highest accuracy but also a reasonable computational cost with a processing time of \*\*39.4 seconds\*\*.

### ML Models and Performance Evaluation:

The performance of the models was evaluated based on metrics like True Positive Rate, False Positive Rate, Precision, Recall, and F-Measure for various classes (emotions) which were repurposed for the study of bird sound classification.

### Conclusion:

The study concluded that machine learning algorithms could effectively classify and identify bird sounds with notable accuracy. J4.8 was identified as the most effective algorithm for this task due to its high accuracy and moderate computational demand.

### Citation:

Jadhav, Yogesh, et al. “Machine Learning Approach to Classify Birds on the Basis of Their Sound.” 2020 International Conference on Inventive Computation Technologies (ICICT), IEEE, 2020, pp. 69–73. DOI.org (Crossref), <https://doi.org/10.1109/ICICT48043.2020.9112506>.

### Dataset Used:

The bird audio data was collected from the Macaulay Library, eBird, and xeno-canto. The dataset was refined manually using Audacity to cancel noise and create relevant audio clips of 2-3 seconds each. A total of 100 clips for each bird class were generated for further processing.

### Algorithms and Classification Techniques:

The study evaluated different algorithms, including:

-\*\*K-Nearest Neighbors (KNN) \*\*

- \*\*Random Forest \*\*,

- \*\*Multilayer Perceptron\*\*

- \*\*Bayes\*\*

The primary classification model tested various algorithms, and the best-performing one was chosen for implementation. Feature selection was critical, and a combination of 34 different audio features was selected for the highest accuracy, including Zero Crossing Rate, Energy, Entropy of Energy, Spectral Centroid, Spectral Spread, Spectral Entropy, Spectral Flux, Spectral Roll-off, MFCCs, Chroma Vector, and Chroma Deviation.

### Methodology:

Audio files were divided into 3-second clips with 1 second of overlapping windows for classification. Each audio file was processed to predict the bird class using a trained model which provided probabilistic results. The average of the probabilistic results from each segment was used to predict the bird species from the entire audio file. Principal Component analysis was performed for dimension reduction process.

### Accuracy Achieved:

- The Support Vector Machine (SVM) with a radial basis function kernel achieved an accuracy of 96.7% for primary classification.

- The sub-classification based on calls or songs of the classified bird species achieved between 96% to 99% accuracy.

- The system produced a confusion matrix illustrating the accuracy of primary classification.

### ML Models:

- After data visualization and analysis, SVM, KNN, and Random Forest algorithms were shortlisted as potential classifiers for the non-linear dataset.

- The SVM with a radial basis function was chosen for model training due to its high accuracy.

- To avoid overfitting, the 'C' parameter of SVM was tuned, with a value of 10 providing the best trade-off between accuracy and generalization.

### Conclusion:

The proposed system was successful in differentiating audio files of distinct bird classes. While the model performed well in identifying distinct classes, it faced challenges in differentiating closely related species like pigeons and doves due to similar feature representations.

###Citation:

Balemarthy, Siddhardha, et al. Our Practice Of Using Machine Learning To Recognize Species By Voice. arXiv:1810.09078, arXiv, 22 Oct. 2018. arXiv.org, https://doi.org/10.48550/arXiv.1810.09078.

### Datasets Used:

The study involves audio datasets containing vocalizations of various species, specifically focused on birds and animals. These datasets include both pre-recorded sounds in controlled environments and field recordings that might contain significant background noise. The datasets are segmented into short clips, each containing a single vocalization event to prevent overlapping and ensure clarity in the sound samples used for training and testing the models.

**###** Methodology and Algorithms:

The core methodology revolves around training machine learning models on audio datasets containing vocalizations of different species, with and without background noise. The goal is to accurately identify species based on their unique vocal signatures. Mel-frequency cepstral coefficients (MFCCs) are extracted from the audio datasets as features that capture the power spectrum of sound. The Hidden Markov Model (HMM) is implemented using TensorFlow to analyze continuous acoustic signals and extract sound events, defining the structure of the audio event detection (AED) system. The system leverages both spectrogram analysis and signal preprocessing (bandpass filtering, noise filtering, and silent region removal) to enhance the quality of the input audio data, aiming to isolate the species-specific sounds from background noise.

### Accuracy Achieved:

The document reports an overall accuracy of 66.7% in detecting species' sounds, a significant achievement considering the complexity of audio recognition tasks, especially in noisy and uncontrolled environments. The improvement in accuracy is attributed to careful preprocessing of audio signals, the use of robust machine learning algorithms, and the iterative refinement of models based on testing and validation against known datasets.

### Conclusion:

The research showcases a sophisticated approach to the problem of species recognition through vocalizations, leveraging a combination of advanced signal processing techniques and machine learning algorithms. The use of MFCCs and HMM, alongside specialized preprocessing steps, demonstrates the effectiveness of their methodology in achieving high accuracy.

### Citation:

Yang, Yiyuan, et al. SSL-Net: A Synergistic Spectral and Learning-Based Network for Efficient Bird Sound Classification. arXiv:2309.08072, arXiv, 23 Dec. 2023. arXiv.org, https://doi.org/10.48550/arXiv.2309.08072.

### Dataset Used:

The study utilized a field-collected bird audio dataset, emphasizing the significance of accurate bird sound classification for ecological studies and habitat protection. The dataset was instrumental in testing the SSL-Net model across 20 classes of bird species, highlighting the need for effective audio analysis in biodiversity conservation efforts.

### Algorithms and Classification Techniques:

SSL-Net introduces a novel dual-branch structure for bird sound classification:

-\*\*Learned Branch\*\*: Utilizes audio-pretrained models, such as BEATs, coupled with a pretrained model-based encoder (ResNet18), to generate semantic feature maps directly from raw audio data.

-\*\*Spectral Branch\*\*: Employs traditional acoustic feature extraction techniques, including MelSpectrogram, STFT, and MFCC, combined with a pretrained encoder (ResNet18), to create spectral feature maps.

### Methodology:

- \*\*Feature Fusion Module\*\*: Three strategies—fixed fusion, shared fusion, and sampling fusion—are proposed for integrating semantic and spectral feature maps, thereby addressing the distribution bias in learned features and enhancing classification robustness.

- \*\*Final Acoustic Classifier\*\*: A lightweight MLP classifier is used for final classification, trained with a cross-entropy loss function.

### Single Branch Methods:

- The study compared baseline and single-branch methods on the WMWB dataset, with the combination of all spectral features (MEL, STFT, MFCC) achieving the highest accuracy of 84.02%.

### Fusion Strategies:

- Different fusion strategies were assessed, with the sampling fusion strategy (combining spectral features and BEATs) leading to the highest accuracy (85.70%) and F1-score (88.79%), suggesting the effectiveness of integrating spectral and learned features.

### Visualization and Performance Insights:

- t-SNE visualizations demonstrated enhanced class separability post-fusion, particularly with the sampling fusion method. This indicates that fused features are more discriminative and conducive to effective classification.

### Impact of Labeled Samples on Performance:

- The F1-score for all methods improved with the increase in the number of labeled samples. Fusion strategies, in general, showed superior performance over single-branch methods, with significant benefits obtained from larger datasets.

### ML Models:

Pretrained models like ResNet18 and audio-specific pretrained models (BEATs, LEAF) were employed as efficient feature extractors, demonstrating the capability of these models to be fine-tuned with minimal labeled data for bird sound classification tasks.

### Conclusion:

SSL-Net addresses the complex challenge of bird sound classification by effectively combining spectral and learned features. It mitigates distribution bias and enhances classification performance with minimal labeling effort. The framework's versatility allows for selecting the most suitable fusion strategies and models based on the resources available. The study's extensive evaluations and experiments with real-world bird audio data validate the SSL-Net's efficiency and robustness, making it a state-of-the-art tool for researchers and engineers in biodiversity conservation and related fields.