VIDYASHIP UNIVERSITY

North American Bird Species Identification Using Bird Voice Recordings

PresentedBy:-Vismaya MP (2023UG000169)
Thanushree (2023UG000165)
Shravanthi (2023UG000139)
Mohit (2023UG000167)

Problem Statement

- North America hosts diverse bird species with similar-sounding vocalizations.
- Manual identification from audio is slow, expert-dependent, and error-prone.
- Automated audio sensors generate thousands of recordings, impossible to classify manually.
- Researchers, conservationists, and monitoring agencies need a fast, accurate, automated system to classify birds using sound.

- **Our Goal**
  - Use machine learning to identify bird species using recorded vocalizations.

- **Our Approach**

  a. Collect and preprocess bird audio recordings

  b. Extract acoustic features (MFCC, spectral features)

  c. Reduce dimensionality using PCA

  d. Train & compare ML models: SVM, Random Forest, Gradient Boosting

  e. Select the best classifier for species prediction (SVM)

Data Source and Description

- **Dataset:** Zenodo “North American Bird Species” dataset
- **Size:** 679 records, 10 main columns after cleaning
- **Format :** 44.1 kHz, 16-bit WAV
- **Classes:** 11 labeled species + 1 “unknown” class (12 total)
- Real-world noise present (wind, insects, distance calls)

Conclusion: A high-quality dataset suitable for supervised ML in bioacoustics.

Feature Extraction

- 3101 real bird-call recordings collected from 12 North American species.
- Audio contains natural background noise such as wind, insects, and human activity.
- MFCC features extracted to capture frequency patterns important for distinguishing bird calls.
- Mel-spectrogram statistics used to represent time–frequency energy distribution of each call.
- Spectral centroid, bandwidth & rolloff describe brightness, spread, and high-frequency decay of audio.
- Each recording is represented by a 184-dimensional feature vector combining all extracted descriptors.

Data Cleaning and Feature Engineering

- **Data cleaning**
  - No missing values
  - No invalid/outlier values
  - All 3074 samples retained

- **Feature engineering and encoding**
  - Extracted 183 features per audio recording
  - (MFCCs, spectral centroid, rolloff, bandwidth, ZCR, contrast, mel-spectrogram stats)
  - StandardScaler applied
  - LabelEncoder for species classes

Final: clean numerical feature matrix: 3074 × 183

Models Evaluted

1. **Support Vector Machine (SVM)**

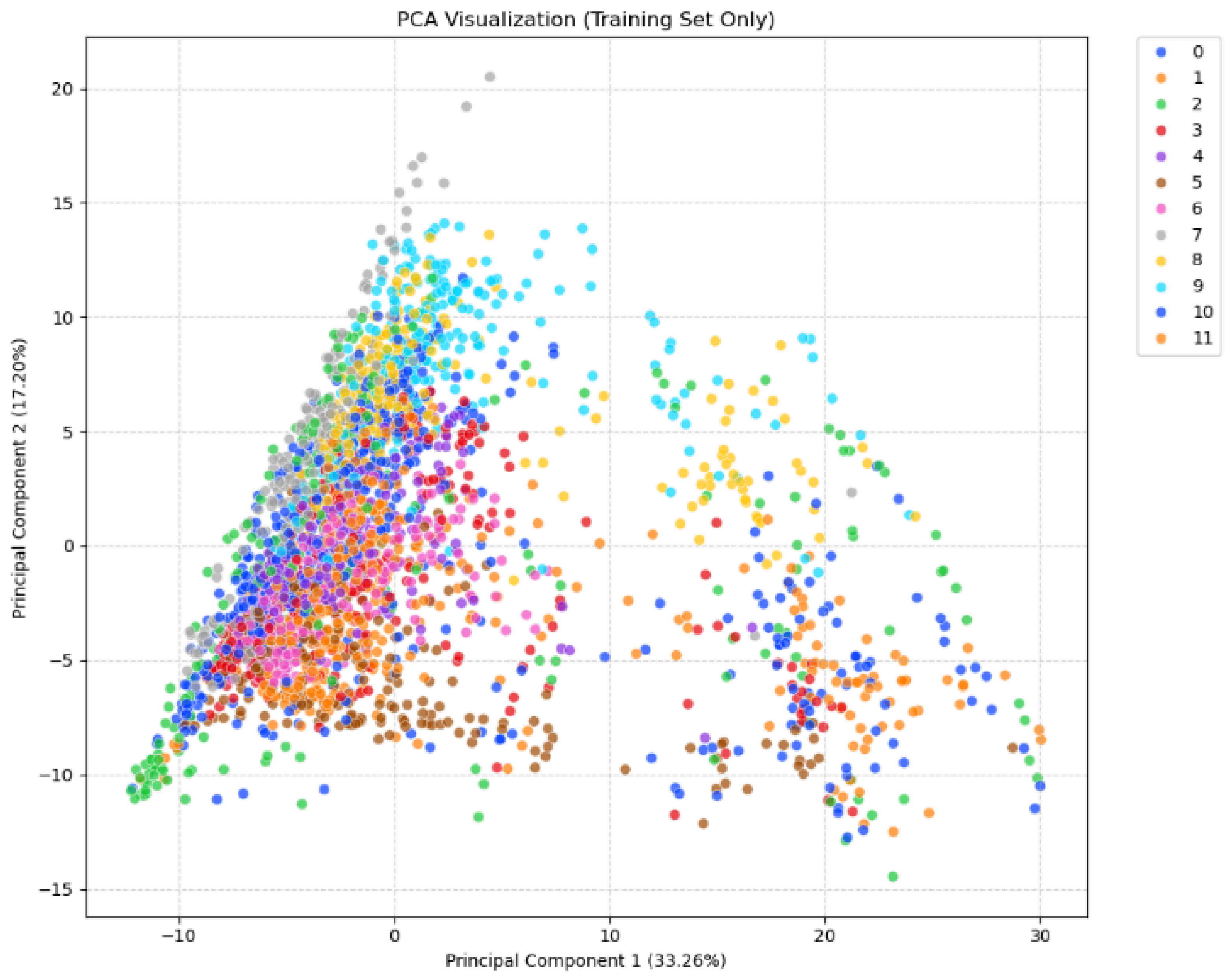
- Learns non-linear boundaries (RBF kernel)
- Best for high-dimensional audio features

2. **Random Forest**

- Combines many trees for stable predictions
- Handles noisy and correlated features

3. **Gradient Boosting**

- Learns patterns by correcting previous errors
- Captures subtle acoustic differences



- **Common Training Steps**
  - Standardization of features
  - PCA reduced to 5 components
  - Same train/test split for all models

Which Model Was Best?

SVM is best Model

- Achieved highest accuracy and F1-score among all models
- Effectively handles non-linear decision boundaries
- Works best with PCA-reduced high-dimensional features
- Shows lowest misclassifications in the confusion matrix
- Most stable and consistent across all 12 species

MODEL	Accuracy	Precision	Recall	F1-Score
Random Forest	86.86%	0.872	0.8763	0.8725
SVM	**90.67%**	**0.9106**	**0.9151**	**0.9104**
Gradient Boosting	84.89%	0.8561	0.852	0.8525

Conclusion

- PCA + SVM achieved the highest accuracy (90.68%)
- ML can effectively automate bird species identification
- Valuable for conservation and eco-acoustic monitoring
- Approach can extend to mobile, IoT, and other sound recognition tasks
- Shows that traditional ML can perform strongly on acoustic data.
- Provides a solid base for future deep learning or larger datasets.

Implications and Limitations

- **Potential implications:**
  - ML can automatically identify bird species from audio, reducing need for experts.
  - Useful for biodiversity monitoring, conservation, and ecological research.
  - Can be deployed in real-time systems like mobile apps, IoT sensors, and forest monitors.
  - Method can extend to other audio tasks such as animal call recognition or environmental sound detection.
- **Limitations:**
  - Background noise in recordings can reduce accuracy.
  - Some species have overlapping acoustic patterns, limiting perfect separation.
  - Dataset includes only 12 species, reducing generalization.
  - Performance depends on feature extraction quality (MFCC, spectral stats).
  - Slight class imbalance affects rare species predictions.

THANK YOU