

Fusion-Based Hybrid Ensemble Learning for Intelligent Bioacoustics Species Classification in Ecological Monitoring

B. Rajesh Reddy^{1*}, Tallapally Akshay Kumar², C. Bheemulu², Yarva Ashok²

¹Assistant Professor, ²UG Student, ^{1,2}Department of Computer Science and Engineering

^{1,2}Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: B. Rajesh Reddy (rajeshreddy.reddy54@gmail.com)

ABSTRACT

The increasing focus on biodiversity preservation and ecological monitoring has created a demand for intelligent, non-intrusive systems capable of accurately recognizing animal species. Bioacoustic analysis, which relies on animal vocalizations, serves as an efficient alternative to conventional visual monitoring methods, especially in environments with dense vegetation or limited visibility. In this regard, this study presents an advanced machine learning framework for automatic animal species classification using bioacoustic data. The system processes raw audio inputs and extracts a diverse range of acoustic features, including Mel-Frequency Cepstral Coefficients (MFCC), chroma features, Mel spectrogram, spectral contrast, zero-crossing rate, and root mean square energy. These features effectively represent the temporal, spectral, and harmonic characteristics of animal sounds, allowing reliable differentiation between species. Several classification algorithms, such as Decision Tree Classifier (DT), Gradient Boosting Classifier (GB), and Nearest Centroid Classifier (NCC), are employed to assess baseline performance. To improve classification results and address the shortcomings of individual models, a hybrid soft voting technique is proposed by integrating Support Vector Machine (SVM) and Light Gradient Boosting Machine (LGBM) through a probabilistic voting mechanism. This approach combines the output probabilities of both models to enhance prediction reliability and generalization capability. Additionally, the system includes a Graphical User Interface (GUI) that supports dataset handling, feature extraction, model training, and real-time prediction. Experimental findings indicate that the proposed hybrid model delivers superior performance in terms of accuracy, precision, recall, and F1-score (Harmonic Mean of Precision and Recall). Overall, the study demonstrates the effectiveness of bioacoustic intelligence as a scalable and efficient solution for wildlife monitoring applications.

Keywords: Machine learning, Acoustic feature extraction, Mel-Frequency Cepstral Coefficients (MFCC), Gradient Boosting, Support Vector Machine (SVM), Light Gradient Boosting Machine (LGBM).

1. INTRODUCTION

Automatic sound classification, also known as acoustic or audio classification, is a technique used to identify and categorize different types of sounds through signal processing methods and machine learning algorithms. This method is an important component of modern audio informatics, where the efficient management, retrieval, and automated analysis of audio data are crucial [1]. An important area within this field is bioacoustics, which involves the study of sounds generated by living organisms and their interaction with the surrounding environment. Bioacoustic analysis has become increasingly valuable in ecological monitoring, especially for observing biodiversity in situations where visual monitoring is difficult or not feasible [2].

Animal vocalizations contain unique acoustic signatures that encode species-specific characteristics as well as behavioural and environmental information. Bioacoustics research can broadly be divided into three primary areas. The first area deals with the detection and classification of animal sounds for

species identification across varying behavioural conditions. The second focuses on understanding how these sounds influence communication, behaviour, and emotional responses among organisms, as shown in Figure 1. The third area investigates the underlying mechanisms and physiological processes involved in sound production [3].

Despite notable progress in automated bioacoustic classification, there remains a gap between bioacoustic and general acoustic research. General acoustic studies typically emphasize the development and evaluation of machine learning algorithms for diverse sound classification tasks, whereas bioacoustic research is predominantly application-driven, with a focus on species recognition. A comprehensive review of studies conducted over the past two decades indicates that techniques such as Short-Time Fourier Transform (STFT) and feature extraction methods are widely utilized; however, several advanced learning approaches successfully applied in general acoustics have not yet been fully explored in bioacoustic applications [4, 5].

2. LITERATURE SURVEY

Mutanu et al. [6] conducted a comprehensive survey that examined bioacoustics classification in comparison with general acoustic studies to provide a broad understanding of the research landscape. The study reviewed a total of 124 research papers spanning multiple years, offering a representative overview of advancements in the field. It identified major application domains within bioacoustics, such as species identification and environmental monitoring, while also analyzing commonly used audio transformation techniques and feature extraction methods. Furthermore, the survey evaluated the classification algorithms employed in bioacoustic systems, highlighting both traditional and modern approaches. The findings emphasized the growing importance of machine learning in automating acoustic analysis while also pointing out gaps between bioacoustic and general acoustic research. Karaaslan et al. [7] proposed an automated framework for detecting and classifying animal vocalizations, aiming to improve efficiency in behavioral analysis. The methodology includes a preprocessing stage where relevant sound segments are isolated from continuous audio recordings to eliminate noise and irrelevant data. Feature extraction is performed using techniques such as Short-Time Fourier Transform (STFT), Mel-Frequency Cepstral Coefficients (MFCCs), and Linear-Frequency Cepstral Coefficients (LFCCs), which capture important acoustic characteristics. These features are then fed into various convolutional neural network (CNN) architectures, including AlexNet, DenseNet, EfficientNet, ResNet50, and ResNet152, to evaluate classification performance. Experimental results demonstrate that deep learning models significantly enhance classification accuracy, with different architectures showing varying levels of effectiveness depending on the feature representation used.

Aliane et al. [8] presented a detailed literature review focusing on the integration of drone technology and artificial intelligence for wildlife monitoring. The study explored two major dimensions: the methodologies and algorithms used in AI-driven wildlife monitoring systems, and the challenges and opportunities associated with their implementation. It analyzed different operational frameworks and real-world applications, emphasizing how AI techniques are applied in aerial data collection and analysis. Additionally, the review highlighted several limitations, including technical challenges, data acquisition constraints, and regulatory issues. Despite these challenges, the study identified significant opportunities for improving wildlife conservation through the integration of drones and AI technologies. Oswald et al. [9] contributed by reviewing key research works that demonstrate best practices in various bioacoustic subfields. Instead of presenting a traditional guideline-based approach, the study focused on case-based analysis to highlight successful applications of bioacoustics. It examined how bioacoustic techniques can be used to determine the identity, location, and population size of animal species within ecosystems. The study discussed practical examples where bioacoustic

methods were applied effectively for species identification, population analysis, and biodiversity assessment. These case studies provided valuable insights into the real-world applicability of bioacoustic technologies. Sharma et al. [10] analyzed the increasing adoption of artificial intelligence techniques in wildlife acoustic monitoring. Their review indicated a significant rise in the use of AI-based approaches, particularly for bird species classification, followed by mammals. The study identified Convolutional Neural Networks (CNNs) as the most widely used algorithm due to their superior performance compared to traditional methods. The findings demonstrated that AI-driven approaches provide higher accuracy and efficiency in processing complex acoustic data. This highlights the potential of AI technologies in advancing ecological research and improving the understanding of wildlife behavior and population dynamics.

Chao et al. [11] proposed an approach for frog species classification based on the symmetry properties of bioacoustic spectral features. The study employed machine learning models such as Support Vector Machine (SVM) and Feedforward Neural Network (FNNA) for classification tasks. It emphasized the importance of symmetry concepts in simplifying complex data representations and improving learning efficiency. Feature extraction was performed using Mel-Frequency Cepstral Coefficients (MFCC), which effectively capture spectral characteristics of frog vocalizations. The integration of symmetry-based analysis with machine learning techniques enabled improved classification performance. Manikandan et al. [12] explored the application of edge computing and TinyML frameworks for bioacoustic analysis in commercial poultry environments. The study addressed challenges related to scalability, computational limitations, and real-time processing in complex acoustic environments. It demonstrated the potential of deploying lightweight models for applications such as emotion recognition, disease detection, and behavioral analysis. Additionally, the research identified key limitations, including the lack of standardized datasets, inconsistencies in evaluation protocols, and challenges in model interpretability. These findings highlight the need for more robust and standardized approaches in bioacoustic research.

Ratnayake et al. [13] conducted a systematic review on the application of machine learning techniques for bee species identification. The study analyzed 26 research publications selected from major scientific databases using a keyword-based search approach. It compared shallow learning methods with deep learning approaches, revealing a growing preference for deep learning models, particularly after 2020. The findings indicated that deep learning techniques are more effective in handling complex and high-dimensional acoustic data, leading to improved classification accuracy. This demonstrates the evolving trend toward advanced AI techniques in bioacoustic research. Trapanotto et al. [14] evaluated the performance of pretrained convolutional neural networks on a lion roar dataset consisting of approximately 150 samples from five individual lions. The study utilized multiple feature representations, including MFCC, spectrogram, Mel spectrogram, VGGish features, Stockwell transform, and LM spectrogram. Models such as VGG16, ResNet50, and AlexNet were retrained and evaluated individually as well as in ensemble configurations. The results showed that ensemble models achieved over 98% accuracy, outperforming individual models, which achieved over 95% accuracy. This highlights the effectiveness of combining multiple deep learning models for improved classification performance. Dewmini et al. [15] focused on elephant sound classification using raw audio signals without converting them into spectrograms. The study evaluated lightweight models such as MobileNet, YAMNET, and RawNet, and introduced a novel model named ElephantCallerNet. The proposed model achieved an accuracy of 89%, demonstrating the feasibility of direct audio-based classification. Bayesian optimization techniques were used to fine-tune hyperparameters such as learning rate, dropout rate, and kernel size, leading to improved model performance. This work highlights the potential of lightweight models for real-time and resource-constrained applications.

3. PROPOSED METHODOLOGY

The proposed study introduces a well-defined analytical framework for recognizing animal species using bioacoustic signals with the help of advanced machine learning methods. The process starts with collecting and organizing animal audio recordings, followed by systematic preprocessing and extraction of relevant features. These extracted acoustic features represent key properties such as frequency patterns, temporal changes, and harmonic structures found in animal sounds. The generated feature set is then utilized by various machine learning models to perform species classification. To enhance accuracy and reliability, a hybrid soft voting strategy is implemented, which combines multiple learning algorithms through a probabilistic fusion approach. The system is equipped with an interactive graphical interface that allows smooth execution of essential tasks, including dataset selection, feature extraction, model training, performance evaluation, and real-time prediction as shown in Fig. 2. A lightweight storage system is incorporated to handle trained models and user authentication data efficiently, ensuring secure and optimized data management.

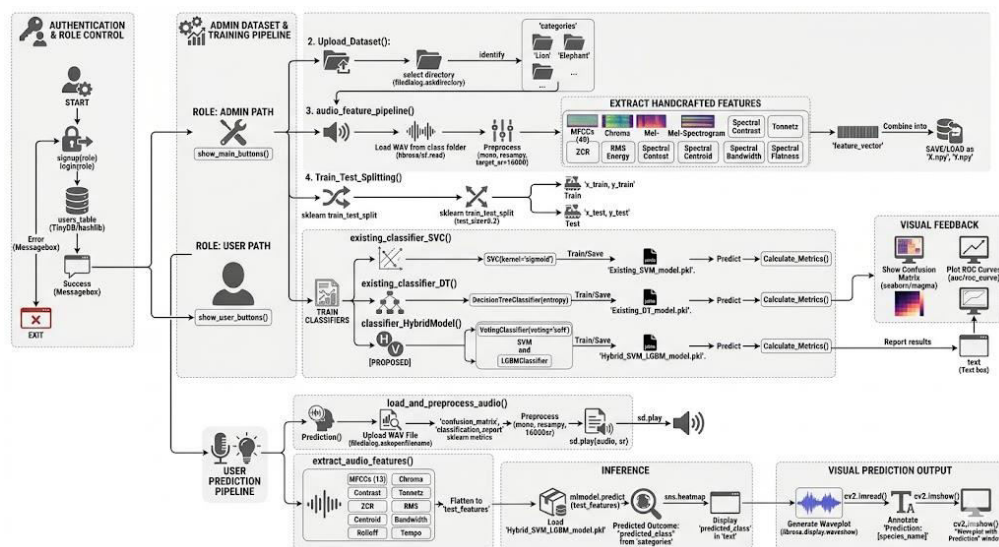


Fig. 2: Proposed system architecture

The system is integrated with an interactive graphical interface that enables seamless execution of key operations such as dataset selection, feature extraction, model training, performance analysis, and real-time prediction. A lightweight storage mechanism is used to manage trained models and user authentication data, ensuring efficient data handling and secure access. Additionally, the framework supports audio-based prediction workflows, allowing users to input new sound recordings and obtain classification results instantly. Continuous evaluation and retraining mechanisms are also included to enhance system adaptability and maintain performance as new bioacoustic data becomes available.

1. User Interface (Client Application)

The framework provides a localized desktop graphical interface developed to ensure a seamless and accessible user experience.

- **Operational Control:** Offers dedicated modules for secure login, dataset management, model training, and performance visualization.
- **Audio Interaction:** Allows users to browse local storage, select specific animal sound files (WAV/MP3), and submit them for automated species identification.
- **Interface Flow:** All user-initiated actions are captured by the GUI and routed to the central processing module for execution.

2. Application Processing Module

This layer serves as the core orchestration engine, managing the lifecycle of an audio signal from raw input to final classification.

- **Task Coordination:** Synchronizes the transitions between noise filtering, feature extraction, and the execution of the machine learning pipeline.
- **Data Flow Management:** Ensures that audio buffers are processed efficiently, supporting the real-time identification of incoming acoustic signals.
- **Efficiency:** Optimized for low-latency performance to generate instant results upon audio submission.

3. Authentication Database (User Storage)

A lightweight storage perimeter is integrated to manage secure system access and maintain data integrity.

- **Security Management:** Maintains records of usernames and encrypted passwords to prevent unauthorized access to the analytical tools.
- **Access Control:** Verifies login credentials and manages user roles, ensuring that sensitive training and evaluation modules are accessed only by authorized personnel.

4. Audio Dataset (Bioacoustic Collection)

The primary intelligence source consists of a diverse collection of animal vocalizations organized into specialized categories.

- **Categorized Records:** Audio samples are organized into folders by species, providing the labeled "ground truth" for supervised learning.
- **Acoustic Diversity:** Captures a wide range of vocalizations representing unique acoustic patterns, essential for building a model that generalizes well across different environments and individuals.

5. Audio Preprocessing and Feature Extraction

Raw bioacoustic signals are refined and converted into a multi-dimensional numerical format that captures the essence of the animal's call.

- **Signal Refinement:** Involves noise reduction, resampling to a standard frequency, and amplitude normalization to ensure consistency across the dataset.
- **Feature Suite:** Extracts a comprehensive set of features, including:
 - **MFCC & Mel Spectrogram:** Captures the spectral envelope and frequency distribution.
 - **Chroma & Spectral Contrast:** Identifies harmonic content and the "brightness" of the sound.
 - **Zero-Crossing Rate & RMS Energy:** Measures the temporal characteristics and loudness of the vocalization.

6. Machine Learning Classification Models

The framework evaluates the extracted feature vectors through multiple algorithmic paradigms to ensure a robust comparative analysis.

- **SVM:** Learns optimal high-dimensional hyperplanes to separate species based on their acoustic features.
- **LDA:** Focuses on maximizing class separability through dimensionality reduction.
- **DT:** Employs rule-based logic for interpretable classification of acoustic patterns.
- **Hybrid Soft Voting:** The flagship model that integrates SVM and LGBM to leverage both geometric and gradient-boosting strengths.

7. Hybrid Ensemble Learning Module

A critical component of the framework that utilizes a fusion-based strategy to maximize predictive accuracy.

- **Soft Voting Mechanism:** Integrates predictions from both SVM and LGBM.
- **Probabilistic Aggregation:** Instead of a simple "majority win," the system aggregates the probability scores from both models to produce a final, highly confident prediction.
- **Robustness:** This ensemble approach significantly improves generalization, reducing the likelihood of misclassification caused by environmental noise or overlapping vocalizations.

8. Prediction and Output Generation

The system translates complex acoustic modelling into human-readable species identification.

- **Visual Reporting:** Displays the predicted species name directly on the user interface along with confidence scores.
- **Waveform Visualization:** Generates a real-time waveform plot of the input audio, providing a visual confirmation of the signal being analysed.
- **Interpretation Support:** Provides additional prediction details to help researchers understand the model's reasoning.

9. Audio Prediction Workflow

The architecture is designed for a highly responsive, real-time operational flow.

- **Flexible Ingestion:** Supports the instant upload and classification of new, unseen audio files.
- **Automated Pipeline:** Once a file is selected, it automatically flows through the noise-handling, feature extraction, and ensemble classification stages without user intervention.

10. Model Evaluation and Retraining

The framework includes a diagnostic and adaptive layer to ensure the identification accuracy remains high over time.

- **Quantitative Metrics:** Measures performance using Accuracy, Precision, Recall, and the F1-score.
- **Visual Diagnostics:** Generates confusion matrices and ROC curves to visualize species-specific accuracy and identify potential overlaps in acoustic signatures.
- **Adaptive Learning:** Supports retraining the framework with new bioacoustic data, allowing the system to evolve and identify new species or regional vocal variations.

4. RESULTS DISCUSSION

The system implementation integrates deep learning with machine learning algorithms in a graphical Python interface using Tkinter. It facilitates end-to-end functionality, starting from dataset upload and feature extraction to classification and prediction. The overall structure is modular, and each module is represented by a dedicated function triggered through the GUI buttons. The backend leverages pre-trained deep learning embeddings from YAMNet and multiple traditional ML classifiers for accurate and interpretable classification. The implementation of this project is structured in a modular and interactive Python program, powered by the Tkinter library for the GUI interface. The primary goal of the system is to allow users to classify animal species from .wav audio recordings by extracting features using deep learning and passing them through traditional and ensemble machine learning classifiers.

Figure 3 presents a comprehensive exploratory analysis of a bear vocalization signal through multiple audio feature representations. The waveform shows variations in amplitude over time, indicating segments of low and high acoustic activity corresponding to different vocal bursts. The MFCC (Mel-Frequency Cepstral Coefficients) plot captures the timbral texture of the sound, reflecting consistent spectral patterns that characterize the bear's vocal signature. The chroma features highlight pitch class intensities, revealing harmonic distributions and tonal variations across time. The Mel spectrogram provides a time-frequency representation, clearly showing energy concentration in lower and mid-frequency bands typical of animal calls. The spectral contrast emphasizes the difference between harmonic and noise components, indicating structured acoustic patterns. The zero-crossing rate (ZCR) illustrates signal noisiness, with peaks corresponding to rapid waveform changes, while the RMS energy plot reflects the loudness dynamics, identifying high-energy vocal segments. Together, these features collectively describe the acoustic, spectral, and temporal characteristics essential for accurate species classification.

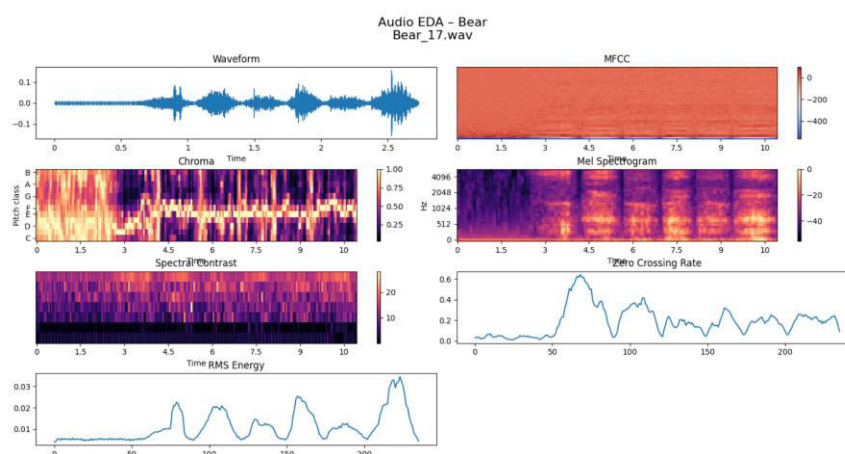


Figure 3: Bioacoustic Feature Visualization of Bear Vocalization

Figure 4 illustrates a detailed exploratory analysis of a dolphin vocalization using multiple acoustic feature representations. The waveform shows a dense and continuous signal with relatively consistent amplitude, indicating sustained vocal emissions typical of dolphin clicks and whistles. The MFCC plot captures the spectral envelope and reveals stable timbral characteristics with subtle temporal variations. The chroma features highlight pitch class distributions, showing distinct tonal patterns that reflect harmonic structures in dolphin sounds. The Mel spectrogram displays strong energy concentration across a wide frequency range, particularly in higher bands, which is characteristic of dolphin acoustic behavior. The spectral contrast emphasizes variations between harmonic and noise components, indicating structured yet complex sound patterns. The zero-crossing rate (ZCR) shows moderate

fluctuations, representing rapid signal changes associated with high-frequency clicks, while the RMS energy plot indicates a relatively stable energy level throughout the signal with minor variations. These features collectively capture the temporal, spectral, and harmonic richness of dolphin vocalizations, enabling effective species-level discrimination.

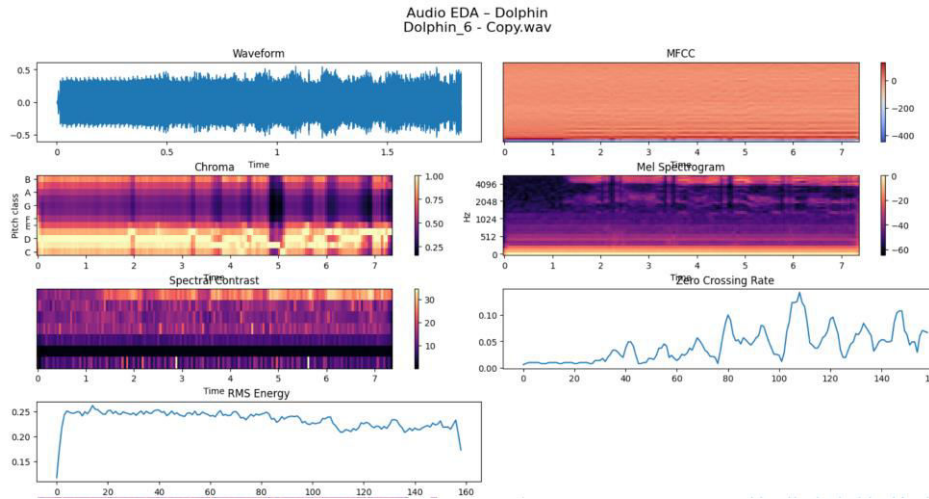


Figure 4: Bioacoustic Feature Visualization of dolphin Vocalization

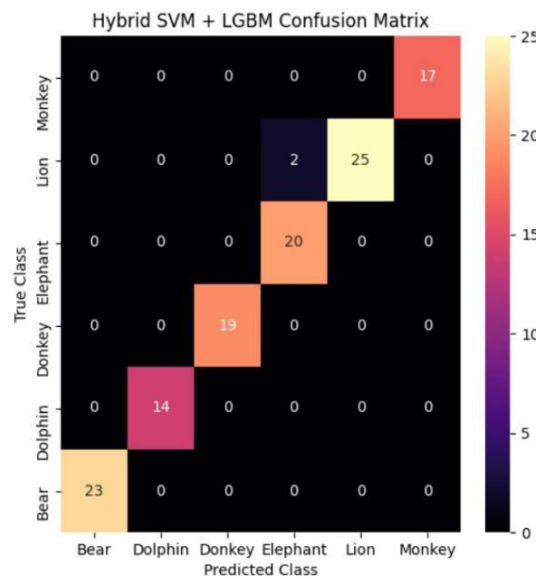


Figure 5: Confusion Matrix obtained using Hybrid soft voting model

In Figure 5 Hybrid soft voting model Confusion Matrix this figure presents the confusion matrix of the proposed Hybrid Soft Voting Model. Strong diagonal dominance indicates highly accurate classification across all animal species. Very few misclassifications are observed, demonstrating effective learning of acoustic patterns. The hybrid model successfully separates overlapping feature distributions. Ensemble learning improves robustness and reduces bias seen in individual classifiers. This result confirms the superiority of the proposed approach. Balanced predictions across classes indicate strong generalization. Thus, the hybrid model significantly outperforms existing methods.

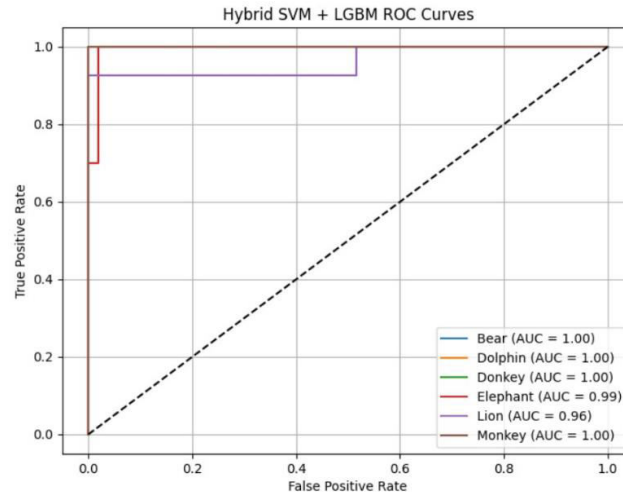


Figure 6: ROC curve obtained using hybrid soft voting model.

Figure 6 shows the ROC curves for the proposed Hybrid soft voting model. The curves approach the top-left corner, indicating excellent classification performance. AUC values close to 1.0 confirm near-perfect discrimination for most animal species. This reflects the effectiveness of soft voting and gradient boosting. The hybrid model captures both global and local feature patterns. ROC analysis validates robustness across all classes. These results confirm high reliability for real-world applications. Hence, the proposed model achieves superior performance.

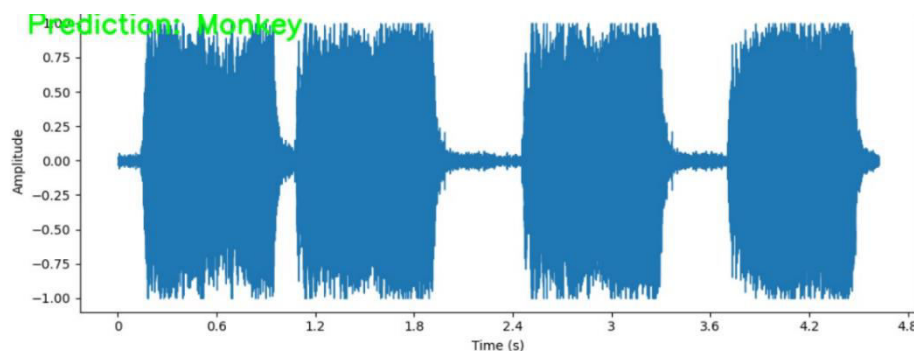


Figure 7: Audio Waveform Visualization with Predicted Animal Species.

Figure 7 shows Audio Prediction waveform visualization generated during the prediction phase of the system. The uploaded test audio signal is displayed in the time–amplitude domain. The predicted animal species label is overlaid on the waveform for clear interpretation. This visualization allows users to correlate acoustic patterns with classification results. The system also plays the audio during prediction for auditory verification. The waveform confirms that the audio signal was processed successfully. Such visual feedback enhances system transparency and user confidence. This figure demonstrates real-time prediction and result visualization capability.

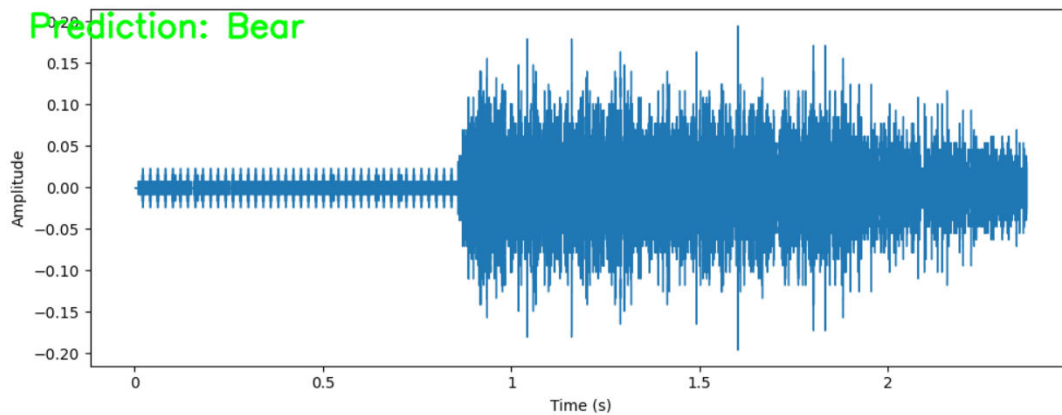


Figure 8: Audio Waveform Visualization with Predicted Animal Species.

Figure 8 shows Audio Prediction waveform visualization generated during the prediction phase of the system. The uploaded test audio signal is displayed in the time–amplitude domain. The predicted animal species label is overlaid on the waveform for clear interpretation. This visualization allows users to correlate acoustic patterns with classification results. The system also plays the audio during prediction for auditory verification. The waveform confirms that the audio signal was processed successfully. Such visual feedback enhances system transparency and user confidence. This figure demonstrates real-time prediction and result visualization capability

Comparative Analysis

The comparative analysis focuses on evaluating the performance of different machine learning models for animal species classification using bioacoustic features. It aims to examine how effectively each model can learn and distinguish complex acoustic patterns present in animal vocalizations. By comparing models such as SVM, Decision Tree, Gradient Boosting, and the proposed Soft Voting approach, the study highlights variations in accuracy, precision, recall, and F1-score. This analysis helps in identifying the strengths and limitations of individual classifiers as well as the advantages of ensemble techniques. It also provides insights into model stability, generalization ability, and prediction consistency. The comparison plays a crucial role in determining the most suitable approach for accurate and reliable bioacoustic classification.

Table 1: Performance Comparison of Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM model	11.66%	1.94%	16.66%	3.48%
DT model	91.66%	91.84%	92.53%	91.62%
GB model	96.66%	97.09%	97.09%	97.09%
Soft Voting model	98.33%	98.48%	98.76%	98.56%

The performance comparison clearly shows significant variation among the implemented models for bioacoustic classification. The SVM model achieved a very low accuracy of 11.66%, with poor precision (1.94%), recall (16.66%), and F1-score (3.48%), indicating its inability to effectively learn complex acoustic patterns. The DT model demonstrated a strong improvement, achieving an accuracy

of 91.66%, along with balanced precision (91.84%) and recall (92.53%), showing reliable classification performance. The GB model further enhanced the results, reaching an accuracy of 96.66% and maintaining high precision, recall, and F1-score values (97.09%), reflecting its capability to capture intricate feature relationships. The proposed Soft Voting model achieved the highest performance with an accuracy of 98.33%, precision of 98.48%, recall of 98.76%, and F1-score of 98.56%. This improvement highlights the effectiveness of combining multiple models to leverage their strengths. Overall, the results confirm that ensemble learning significantly enhances classification accuracy and robustness compared to individual models.

5. CONCLUSION

This study introduces an intelligent bioacoustic analysis framework for precise identification of animal species using machine learning methods. By extracting a variety of acoustic features such as MFCC, chroma, spectral contrast, and energy-based attributes, the system successfully captures the distinct patterns present in animal vocalizations. Several classification models were analyzed, and the Soft Voting approach combining SVM and LGBM achieved better performance compared to individual models. The use of soft voting improved prediction consistency by utilizing the strengths of both algorithms. Enhanced results were observed in terms of accuracy, precision, recall, and F1-score, along with improved generalization on unseen data. The system also maintains computational efficiency through feature optimization and reuse of trained models, thereby reducing processing overhead. Additionally, visualization methods such as confusion matrices and ROC curves confirmed the reliability of the classification outcomes. The inclusion of an interactive interface further improves usability and supports real-time prediction. Overall, this work contributes to scalable and automated solutions for wildlife monitoring and ecological studies, effectively addressing challenges related to high-dimensional and complex bioacoustic data.

REFERENCES

- [1] Penar, W.; Magiera, A.; Klocek, C. Applications of bioacoustics in animal ecology. *Ecol. Complex.* 2020, 43, 100847.
- [2] Choi, Y.K.; Kim, K.M.; Jung, J.W.; Chun, S.Y.; Park, K.S. Acoustic intruder detection system for home security. *IEEE Trans. Consum. Electron.* 2005, 51, 130–138
- [3] Shah, S.K.; Tariq, Z.; Lee, Y. Iot based urban noise monitoring in deep learning using historical reports. In *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 9–12 December 2019; pp. 4179–4184.
- [4] Vacher, M.; Serignat, J.F.; Chaillol, S.; Istrate, D.; Popescu, V. Speech and sound use in a remote monitoring system for health care. In *International Conference on Text, Speech and Dialogue*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 711–718.
- [5] Olivieri, M.; Malvermi, R.; Pezzoli, M.; Zanoni, M.; Gonzalez, S.; Antonacci, F.; Sarti, A. Audio information retrieval and musical acoustics. *IEEE Instrum. Meas. Mag.* 2021, 24, 10–20.
- [6] Mutanu, L.; Gohil, J.; Gupta, K.; Wagio, P.; Kotonya, G. A Review of Automated Bioacoustics and General Acoustics Classification Research. *Sensors* 2022, 22, 8361. <https://doi.org/10.3390/s22218361>.
- [7] Karaaslan, M.; Turkoglu, B.; Kaya, E.; Asuroglu, T. Voice Analysis in Dogs with Deep Learning: Development of a Fully Automatic Voice Analysis System for Bioacoustics Studies. *Sensors* 2024, 24, 7978. <https://doi.org/10.3390/s24247978>.
- [8] Aliane, N. Drones and AI-Driven Solutions for Wildlife Monitoring. *Drones* 2025, 9, 455. <https://doi.org/10.3390/drones9070455>.

- [9] Oswald, J.N.; Van Cise, A.M.; Dassow, A.; Elliott, T.; Johnson, M.T.; Ravnani, A.; Podos, J. A Collection of Best Practices for the Collection and Analysis of Bioacoustic Data. *Appl. Sci.* 2022, 12, 12046. <https://doi.org/10.3390/app122312046>.
- [10] Sharma, S.; Sato, K.; Gautam, B.P. A Methodological Literature Review of Acoustic Wildlife Monitoring Using Artificial Intelligence Tools and Techniques. *Sustainability* 2023, 15, 7128. <https://doi.org/10.3390/su15097128>.
- [11] Chao, K.-W.; Hu, N.-Z.; Chao, Y.-C.; Su, C.-K.; Chiu, W.-H. Implementation of Artificial Intelligence for Classification of Frogs in Bioacoustics. *Symmetry* 2019, 11, 1454. <https://doi.org/10.3390/sym11121454>.
- [12] Manikandan, V.; Neethirajan, S. AI-Powered Vocalization Analysis in Poultry: Systematic Review of Health, Behavior, and Welfare Monitoring. *Sensors* 2025, 25, 4058. <https://doi.org/10.3390/s25134058>.
- [13] Ratnayake, A.M.B.; Yasin, H.M.; Naim, A.G.; Abas, P.E. Buzzing through Data: Advancing Bee Species Identification with Machine Learning. *Appl. Syst. Innov.* 2024, 7, 62. <https://doi.org/10.3390/asi7040062>.
- [14] Trapanotto, M.; Nanni, L.; Brahnam, S.; Guo, X. Convolutional Neural Networks for the Identification of African Lions from Individual Vocalizations. *J. Imaging* 2022, 8, 96. <https://doi.org/10.3390/jimaging8040096>.
- [15] Dewmini, H.; Meedeniya, D.; Perera, C. Elephant Sound Classification Using Deep Learning Optimization. *Sensors* 2025, 25, 352. <https://doi.org/10.3390/s25020352>.