Fusion-Augmented Ensemble Learning for Intelligent Interpretation of Bioacoustic Signals in Wildlife Ecosystems

P Navya[1], Kasturi Kaveri[2], Kommuka Amulya[2], Orunganti Uday[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering

[1,2]Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India.

To Cite this Article
P Navya, Kasturi Kaveri, Kommuka Amulya, Orunganti Uday, "Fusion-Augmented Ensemble Learning for Intelligent Interpretation of Bioacoustic Signals in Wildlife Ecosystems", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 480-489, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp480-489

Submitted: 28-02-2026 Accepted: 02-04-2026 Published: 10-04-2026

ABSTRACT

The rapid advancement of wildlife monitoring and ecological research has increased the demand for intelligent systems capable of automatically identifying animal species using non-invasive techniques. Bioacoustic recordings, which capture animal vocalizations, provide a reliable and scalable alternative to traditional visual monitoring, particularly in dense forests and low-visibility environments. With the growing integration of machine learning and audio signal processing, automated analysis of animal sounds has emerged as a promising approach for biodiversity conservation and environmental surveillance. This research proposes an advanced bioacoustic species classification system using a fusion-based hybrid ensemble learning framework. The system processes raw audio recordings and converts them into meaningful feature representations through techniques such as Mel-Frequency Cepstral Coefficients (MFCC), chroma features, spectral contrast, Mel spectrogram, zero-crossing rate, and root mean square energy. These features effectively capture the temporal, spectral, and harmonic properties of animal sounds, enabling accurate species differentiation. Multiple machine learning models, including Nearest Centroid Classifier (NCC), Decision Tree Classifier (DTC), and Gradient Boosting Classifier (GBC), are implemented and evaluated. To address the limitations of individual models, a hybrid ensemble approach combining Support Vector Machine (SVM) and Light Gradient Boosting Machine (LGBM) using a soft voting mechanism is introduced. This approach aggregates probability outputs to improve prediction accuracy. Experimental results demonstrate that the proposed model significantly outperforms traditional methods in terms of accuracy, precision, recall, and F1-score. The system provides a robust, scalable, and efficient solution for real-time wildlife monitoring and ecological research.

Keywords: Bioacoustics, Animal Species Classification, Machine Learning, Feature Extraction, Ensemble Learning, Support Vector Machine (SVM), Light Gradient Boosting Machine (LGBM), Wildlife Monitoring.

1. INTRODUCTION

Automatic acoustic classification also referred to as audio or sound classification, involves the detection or recognition of sound using audio informatics for storage and retrieval, and machine learning techniques for autonomous classification [1]. Bioacoustics is the branch of acoustics that is concerned with sounds produced by or affecting living organisms. Bioacoustics is often used in acoustic sensing to monitor biodiversity, especially in visually inaccessible areas [2]. Animal acoustic emissions contain

ISSN: 3049-0952
www.jsetms.com

species-specific information that reflects the character and behavior of different living organisms. As shown as figure 1 there are three main application areas of bioacoustics. The first focuses on the classification and analysis of sounds vocalized by different animal species. Its primary aim is to identify sounds that characterize species in different behavioral contexts. The second is concerned with integrating sound signals vocalized by animals with behavioral contexts to understand how the sounds affect the behavior and emotions of the receiver. The third explores the production mechanisms used in sound vocalization processes [3]. The survey presented in this paper explores how current research in automated bioacoustics classification differs from traditional acoustic classification with respect to the techniques used and application areas.

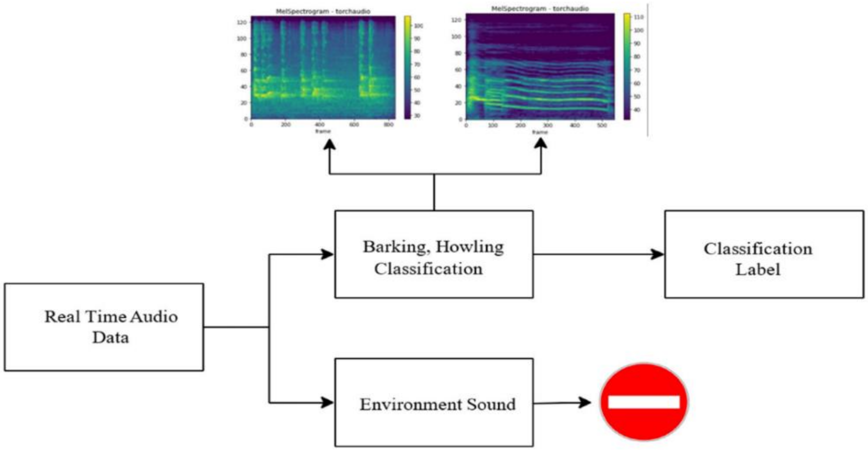

Figure 1: Workflow Diagram of Real-Time Animal Sound Classification System.

They use the term "general acoustic studies" to refer to acoustic research whose primary focus is neither living or non-living organisms.The scope of survey is limited to studies that use machine learning as the primary tool for automating acoustic classification. The survey is intended to be a representative rather than an exhaustive review of the state of the research. The survey reviewed 124 publications, spanning 21 years, from 2000–2021. Only papers published in the English language were reviewed. To the best of knowledge, no recent studies have been undertaken to examine the state of research in this important and fast-growing research area. Survey highlights the advances in automated bioacoustics classification, but also identifies the challenges and opportunities presented.

For example, they note that the automated classification techniques used bioacoustics still lag behind those in general acoustics. A number of machine learning techniques that have been successfully used in general acoustics are yet to be tested in bioacoustics classification [4]. Findings show that current research in bioacoustics is mainly concerned with applications that involve species classification while general acoustic research is primarily concerned with identifying suitable machine-learning algorithms for classifying general sounds. The short-term Fourier transformation (STFT) technique was the most popular audio transformation technique for both bioacoustics and general acoustics studies. Although feature extraction techniques were popular in both bioacoustics and general acoustics research, linear prediction cepstral coefficients (LPCCs) techniques were more popular in general acoustics [5].

2.LITERATURE SURVEY

Mutanu, et al. [6] examined bioacoustics classification alongside general acoustics to provide a representative picture of the research landscape. The survey reviewed 124 studies spanning eight years of research. The survey identifies the key application areas in bioacoustics research and the techniques used in audio transformation and feature extraction. The survey also examines the classification algorithms used in bioacoustics systems. Karaaslan, et al. [7] proposed an automated system for detecting and classifying animal vocalizations, enhancing efficiency in behaviour analysis. The system uses a preprocessing step to segment relevant sound regions from audio recordings, followed by feature extraction using Short-Time Fourier Transform (STFT), Mel-frequency cepstral coefficients (MFCCs), and linear-frequency cepstral coefficients (LFCCs). These features are input into convolutional neural network (CNN) classifiers to evaluate performance. Experimental results demonstrate the effectiveness of different CNN models and feature extraction methods, with AlexNet, DenseNet, EfficientNet, ResNet50, and ResNet152 being evaluated.

Aliane, et al. [8] conducted a comprehensive literature review to examine the recent advancements in drone and AI systems for wildlife monitoring, focusing on two critical dimensions: (1) Methodologies, algorithms, and applications, analyzing the AI techniques employed in wildlife monitoring, including their operational frameworks and real-world implementations. (2) Challenges and opportunities, identifying current limitations, including technical hurdles and regulatory constraints, as well as exploring the untapped potential in drone and AI integration to enhance wildlife monitoring and conservation efforts. Oswald, et al. [9] contributed towards filling this gap. Instead of a classical list of “dos” and “don’ts”, they review some key papers which, they believe, embody best practices in several bioacoustic subfields. In the first three case studies, they discuss how bioacoustics can help identify the ‘who’, ‘where’ and ‘how many’ of animals within a given ecosystem. Specifically, they review cases in which bioacoustic methods have been applied with success to draw inferences regarding species identification, population structure, and biodiversity.

Sharma, et al. [10] reviewed showed a significant rise in the utilization of AI techniques in wildlife acoustic monitoring over this period, with birds (N = 26) gaining the most popularity, followed by mammals (N = 12). The most commonly used AI algorithm in this field was Convolutional Neural Network, which was found to be more accurate and beneficial than previous categorization methods in acoustic wildlife monitoring. This highlights the potential for AI to play a crucial role in advancing their understanding of wildlife populations and ecosystems. Chao, et al. [11] presented the implementation of artificial intelligence (AI) for classification of frogs in symmetry of the bioacoustics spectral by using the feedforward neural network approach (FNNA) and support vector machine (SVM). Recently, the symmetry concept has been applied in physics, and in mathematics to help make mathematical models tractable to achieve the best learning performance. Owing to the symmetry of the bioacoustics spectral, feature extraction can be achieved by integrating the techniques of Mel-scale frequency cepstral coefficient (MFCC) and mentioned machine learning algorithms, such as SVM, neural network, and so on.

Manikandan, et al. [12] revealed compelling evidence for edge computing deployment via TinyML frameworks, addressing critical scalability challenges in commercial poultry environments characterized by acoustic complexity and computational constraints. Advanced applications spanning emotion recognition, disease detection, and behavioral phenotyping demonstrate unprecedented potential for real-time welfare assessment. Through rigorous bibliometric co-occurrence mapping and thematic clustering analysis, this review exposes persistent methodological bottlenecks: dataset standardization deficits, evaluation protocol inconsistencies, and algorithmic interpretability limitations. Ratnayake, et al. [13] reviewed systematically explores the application of machine learning (ML) techniques in bee species determination, shedding light on the transformative potential of ML in entomology. Conducting a keyword-based search in the Scopus and Web of Science databases with manual screening resulted in 26 relevant publications. Focusing on shallow and deep learning studies, their analysis reveals a significant inclination towards deep learning, particularly post-2020, underscoring its ability to handle complex, high-dimensional data for accurate species identification.

Trapanotto, et al. [14] evaluated the performance of three pretrained CNNs (VGG16, ResNet50, and AlexNet) on a small, publicly available lion roar dataset containing approximately 150 samples taken from five male lions. Each of these networks is retrained on eight representations of the samples: MFCCs, spectrogram, and Mel spectrogram, along with several new ones, such as VGGish and stockwell, and those based on the recently proposed LM spectrogram. The performance of these networks, both individually and in ensembles, is analyzed and corroborated using the Equal Error Rate and shown to surpass previous classification attempts on this dataset; the best single network achieved over 95% accuracy and the best ensembles over 98% accuracy. Dewmini, et al. [15] addressed elephant sound classification utilizing raw audio processing. Their focus lies on exploring lightweight models suitable for deployment on resource-constrained edge devices, including Mobile Net, YAMNET, and Raw Net, alongside introducing a novel model termed ElephantCallerNet. Notably, their investigation reveals that the proposed ElephantCallerNet achieves an impressive accuracy of 89% in classifying raw audio directly without converting it to spectrograms. Leveraging Bayesian optimization techniques, they fine-tuned crucial parameters such as learning rate, dropout, and kernel size, thereby enhancing the model's performance.

3. PROPOSED SYSTEM

This research focuses on the intelligent analysis of acoustic signals to identify fault conditions in electric motors using a combination of deep learning–based feature extraction and machine learning classification techniques. Motor-generated sound signals carry valuable information about internal mechanical and electrical anomalies, but their complex, non-stationary nature makes manual interpretation difficult and unreliable. To address this challenge, the study adopts an automated pipeline that transforms raw audio recordings into meaningful representations and applies data-driven learning models for accurate fault classification. By leveraging pretrained deep audio embeddings and ensemble-based learning, the approach enhances fault discrimination capability while reducing dependence on handcrafted features. The inclusion of an interactive graphical interface further supports practical usability, enabling seamless data handling, model training, evaluation, and real-time prediction, thereby aligning the research with modern intelligent maintenance and monitoring systems as shown in figure 2.

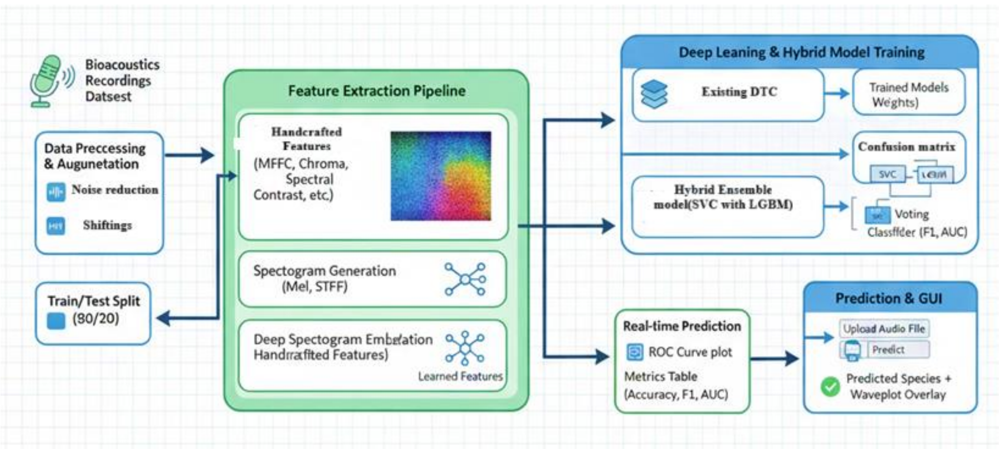

Figure 2: Proposed System Architecture for Animal Species Classification from Audio Inputs.

1. Dataset Acquisition: Acoustic recordings are collected and organized into class-wise folders representing different motor operating or fault conditions.

2. Audio Preprocessing: Input audio signals are converted to mono format, normalized, and resampled to a standard sampling rate to ensure consistency across the dataset.

3. Deep Feature Extraction: High-level acoustic embeddings are extracted from raw waveforms using a pretrained deep neural audio model, capturing both temporal and spectral characteristics.

4. Feature Caching: Extracted feature vectors and corresponding labels are stored to avoid redundant computation during repeated experiments.

5. Train–Test Splitting: The dataset is divided into training and testing subsets to enable unbiased performance evaluation.

6. Model Training: Multiple classifiers, including conventional and ensemble-based models, are trained using the extracted deep audio features.

7. Performance Evaluation: The trained models are evaluated using standard metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC curves.

8. Fault Prediction: New acoustic samples are analysed to predict the corresponding motor condition, supported by waveform visualization for interpretability.

9. User Interaction: A graphical user interface enables easy execution of each stage, making the system suitable for both research experimentation and practical deployment.

4. RESULTS ANALYSIS

The results of the proposed system demonstrate that the hybrid soft voting ensemble model (SVM + LGBM) achieves superior performance compared to individual classifiers such as NCC and Decision Tree. By combining the strengths of both models, the system effectively captures complex patterns in bioacoustic signals, leading to higher accuracy, precision, recall, and F1-score. The model shows strong robustness in handling noisy and high-dimensional audio data. Evaluation metrics and confusion matrices indicate improved classification consistency across multiple animal species. The results validate that the proposed ensemble approach provides a reliable and efficient solution for automated animal sound classification in real-world environments.

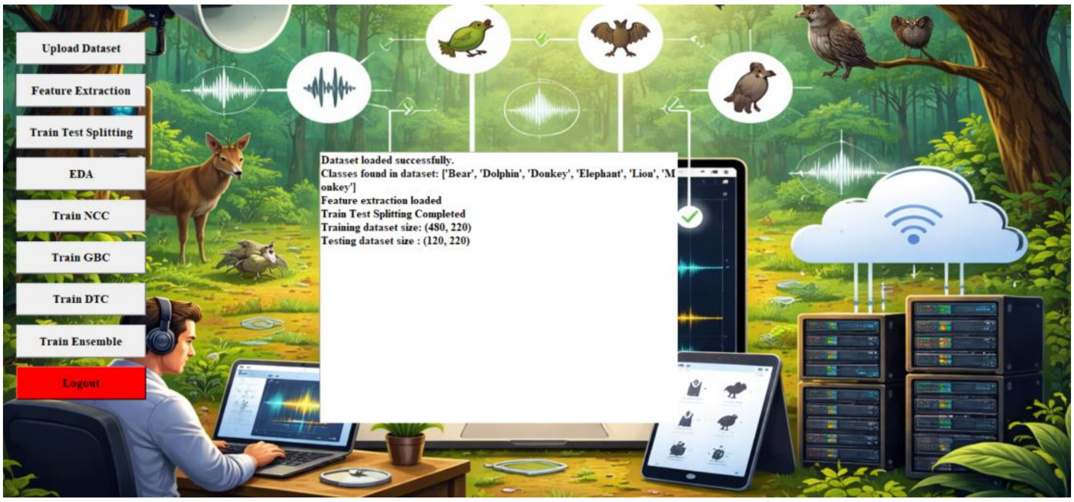

Figure 3: Audio Feature Extraction Status Confirmation Screen.

Figure 3: represents the successful splitting of dataset into training and testing stage in the system. After preprocessing, the system extracts handcrafted bioacoustic features such as MFCCs, chroma, mel-spectrogram, spectral contrast, and energy-based attributes. The extracted features are consolidated into numerical vectors for each audio sample. The confirmation message indicates that features have been successfully saved for future reuse. Storing features prevents redundant computation during repeated training sessions. This step ensures uniform feature representation across all animal classes. Feature

extraction directly influences model learning quality and accuracy. Hence, this figure validates a critical preprocessing milestone.

The 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.

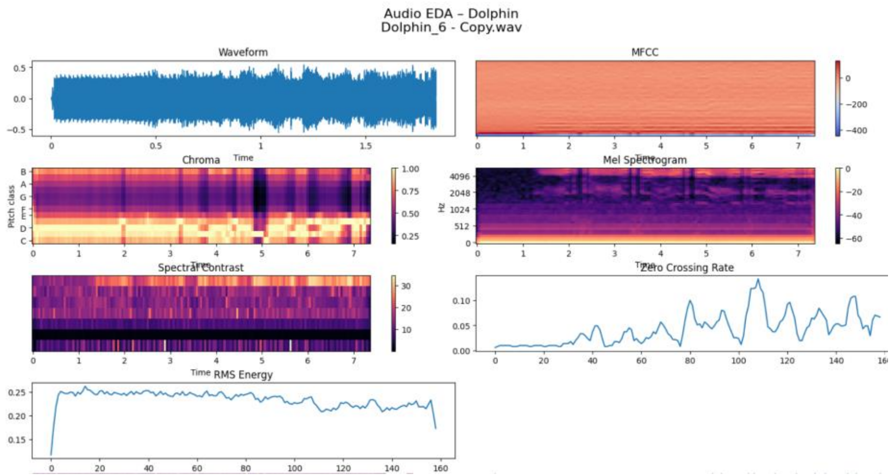

Figure 4: Bioacoustic Feature Visualization of dolphin Vocalization

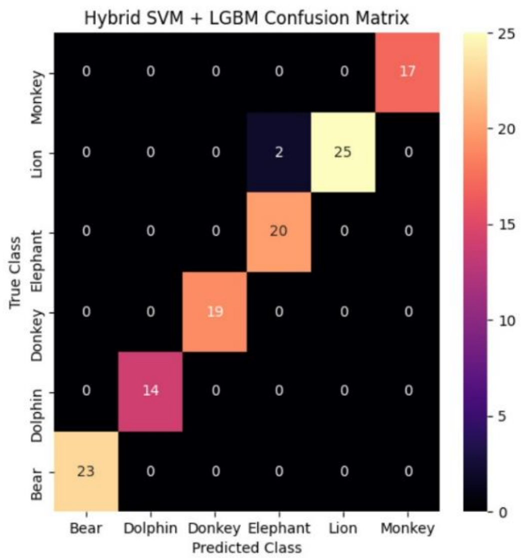

Figure 5: Confusion Matrix obtained using Hybrid soft voting model

 ISSN: 3049-0952
www.jsetms.com

In figure 5 Hybrid soft voting model Confusion Matrix this figure presents the confusion matrix of the proposed Hybrid SOFT VOTING MODEL 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 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.

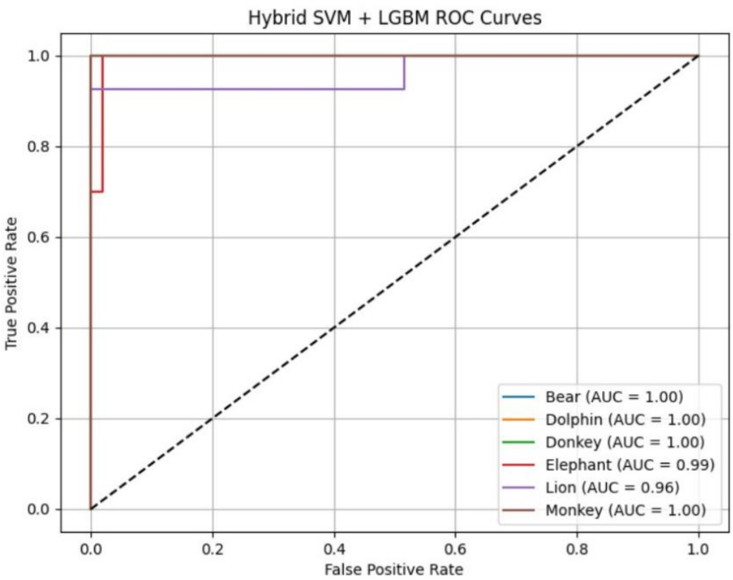

Figure 6: ROC curve obtained using hybrid soft voting model.

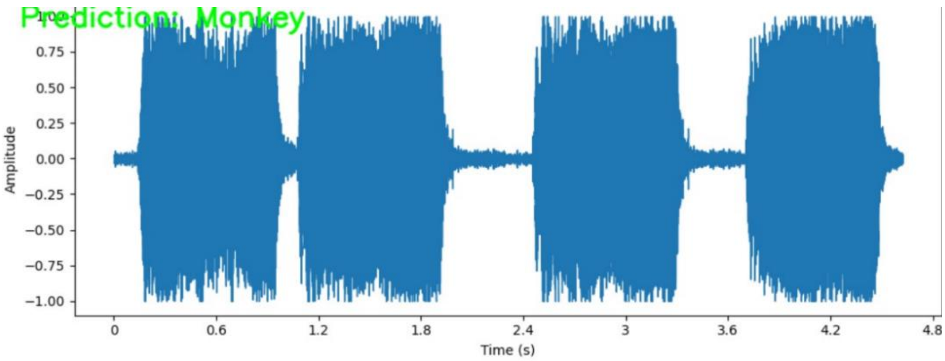

Figure 7: Audio Waveform Visualization with Predicted Animal Species.

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 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.

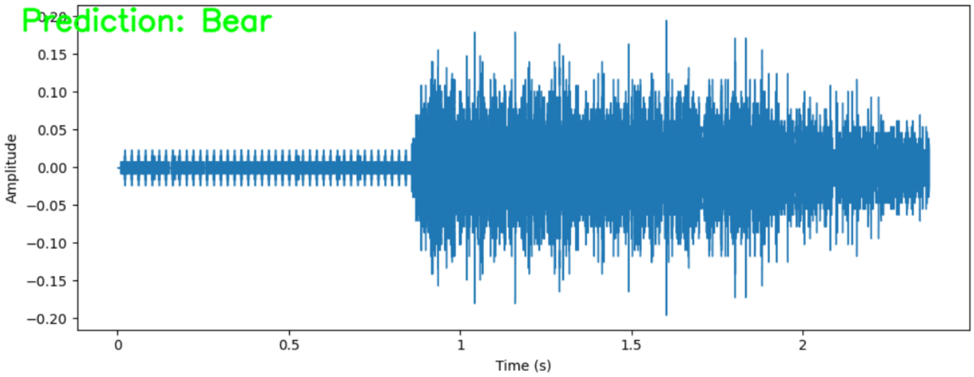

Figure 8: Audio Waveform Visualization with Predicted Animal Species.

4.1 Comparative Analysis

The comparative performance analysis of the implemented models clearly demonstrates a progressive improvement in classification capability from traditional to advanced ensemble techniques. The NCC model shows relatively low performance with an accuracy of 46.66%, indicating its limitation in handling complex, high-dimensional bioacoustic features due to its simple centroid-based decision mechanism, which results in lower precision (41.85%), recall (44.05%), and F1-score (42.62%). The Decision Tree Classifier (DTC) significantly improves performance, achieving 91.66% accuracy, as it effectively captures feature-based decision rules, reflected in balanced precision (91.84%) and recall (92.53%).

Further enhancement is observed with the Gradient Boosting Classifier (GBC), which reaches 96.66% accuracy and consistently high precision, recall, and F1-score (97.09% each), demonstrating its strength in learning complex patterns through iterative boosting. The highest performance is achieved by the proposed Soft Voting model (SOFT VOTING MODEL), which attains 98.33% accuracy, along with superior precision (98.48%), recall (98.76%), and F1-score (98.56%), highlighting the effectiveness of combining complementary models to improve generalization and robustness. The results validate that the hybrid ensemble approach significantly outperforms individual models by leveraging both margin-based and boosting-based learning for accurate bioacoustic species classification.

Table 1: Performance Comparison of Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
NCC model	46.66%	41.85%	44.05%	42.62%
DTC model	91.66%	91.84%	92.53%	91.62%
GBC model	96.66%	97.09%	97.09%	97.09%

Proposed soft voting model	98.33%	98.48%	98.76%	98.56%

5. CONCLUSION

The research successfully developed an Automated Animal Species Identification System using bioacoustics recordings and machine learning techniques. A comprehensive audio processing pipeline was implemented, including preprocessing, handcrafted feature extraction, model training, evaluation, and real-time prediction through a user-friendly Tkinter-based graphical interface. The experimental results clearly demonstrate that traditional classifiers such as SVM and DTC exhibit limited performance when applied independently to complex and overlapping animal vocalization features. The proposed Hybrid SVM with Light GBM (Soft Voting) model significantly outperformed existing approaches, achieving an accuracy of 98.33%, an F1-score of 98.56%, and an ROC–AUC of 99.26%. This substantial improvement is attributed to the ensemble learning strategy, which effectively combines the margin-based decision capability of SVM with the gradient-boosting strength of Light GBM. The hybrid model demonstrated robust generalization across all animal species, reduced misclassification errors, and improved class balance handling. The system proves to be reliable, efficient, and suitable for real-world wildlife monitoring and bioacoustics research applications.

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] Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8.

[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.; Ravignani, 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] Cyril, H. P., & Kumara, S. Identification of Anomalies via Deep Learning-Based Models for High-Dimensional Telecom Traffic Data.

[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.