Sound Based Bird Species Recognition

Varshitha N R[1], Sowjanya K M[2], Vaishnavi khuba [3], Shivani [4], Dr. Kavyasri M N [5]
[5]*Assistant Professor Dept of CSE Malnad College of Engineering Hassan, India*
[1,2,3,4] *Dept of CSE Malnad College of Engineering Hassan, India*

***Abstract*—Ecological monitoring and biodiversity assessment increasingly utilize acoustic bird identification as a non-disruptive methodology. This research develops a deep learning framework for automated recognition of avian species through their vocalizations. The approach employs convolutional neural networks applied to spectrogram representations derived from publicly accessible datasets including Xeno-Canto and BirdCLEF competitions. Preprocessing incorporates noise reduction techniques and data augmentation strategies to enhance model robustness. Evaluation across multiple species demonstrates effective performance under varying acoustic conditions and background interference. The system's potential for deployment in mobile applications and remote monitoring platforms offers significant value for ornithological research and conservation efforts. Future research directions include incorporating spatio-temporal contextual information to refine species classification accuracy.**

Keywords—Acoustic ecology, Avian vocalization recognition, Machine learning, Neural networks, Spectral analysis, Conservation technology, Field applications.

I. INTRODUCTION

Avian species serve as crucial bioindicators for ecosystem integrity and environmental health assessment. Population dynamics and distribution patterns of birds provide valuable insights into biodiversity trends, environmental changes, and habitat quality. Traditional ornithological surveys rely predominantly on visual identification techniques and expert interpretation of acoustic signals—approaches that demand substantial temporal investment, specialized expertise, and remain susceptible to observer bias and subjective interpretation.

Recent advances in artificial intelligence, particularly deep neural architectures, have enabled the development of automated systems for species recognition through acoustic analysis. These computational approaches exploit the distinctive spectral characteristics inherent in species-specific vocalizations, which exhibit remarkable inter-species variability. Through the conversion of temporal audio data into visual representations such as spectrograms, deep learning architectures—especially Convolutional Neural Networks—can effectively learn discriminative features for accurate taxonomic classification.

This study presents a novel computational framework for automated avian species identification utilizing advanced machine learning techniques applied to acoustic data. The proposed system processes digital audio recordings, converts them into Mel-frequency spectral representations, and employs a trained CNN architecture developed using comprehensive bird vocalization databases. This methodology offers a scalable, non-invasive solution for biodiversity monitoring and ecological assessment across diverse terrestrial environments.

II. LITERATURE REVIEW

Bioacoustic research has experienced significant growth in automated avian species recognition, leveraging computational intelligence and neural network architectures to process vocalization data. This review examines contemporary methodologies and algorithmic approaches employed in recent investigations, analyzing their technical contributions and quantitative performance outcomes.

Research by Prakash and Rajesh [1] demonstrated the feasibility of deploying compact CNN architectures for real-time classification on portable devices. Their implementation utilized logarithmic Mel-scale spectrograms as input features, achieving approximately 91% classification accuracy with sub-second processing latency on proprietary mobile-recorded datasets. This work established the viability of edge-computing solutions for field-based ornithological applications.

Mehta and Patel [2] explored composite classifier strategies, combining Random Forest and Gradient

Boosting algorithms trained on diverse acoustic descriptors including Mel-Frequency Cepstral Coefficients, spectral centroids, and temporal zero-crossing characteristics. Their methodology yielded 89% accuracy when evaluated on urban soundscape recordings, demonstrating ensemble robustness in acoustically challenging environments.

The investigation by Bhattacharya and Saha [3] employed pre-trained convolutional architectures (VGG16, ResNet50) adapted through fine-tuning for avian vocalization spectrograms. Utilizing Xeno-Canto repository data supplemented with field recordings, their approach achieved 93% classification performance, validating transfer learning efficacy in acoustic pattern recognition tasks.

Latha and Gopal [4] conducted comprehensive evaluations of neural network variants, comparing CNNs, RNNs, and hybrid CNN-RNN configurations. Processing MFCC and logarithmic Mel-spectrogram features on BirdCLEF dataset subsets, their hybrid model demonstrated superior performance at 94% accuracy, illustrating the advantages of combining spatial and temporal feature extraction mechanisms.

Jadhav and Patil [5] implemented fully-connected neural networks trained on multi-dimensional feature vectors comprising MFCC, spectral roll-off, and chromatic characteristics. Their architecture achieved 87% accuracy across 20 species classifications, confirming that conventional deep learning structures remain effective when paired with appropriate feature engineering and balanced training datasets.

Arvind and Shalini [6] developed distributed computing frameworks integrating CNN and LSTM architectures for processing logarithmic Mel-spectrograms. Their system, trained on OpenMic-2018 and field-collected data, demonstrated 90% F1-score performance, establishing the potential for scalable acoustic monitoring infrastructure.

The work of Prashanth and Suprabha [7] investigated reduced-complexity CNNs with dual convolutional layers, achieving 88% accuracy on combined custom and Xeno-Canto datasets. Their findings support the effectiveness of simplified architectures when coupled with optimized spectrogram preprocessing techniques.

Roy and Banerjee [8] evaluated traditional classification algorithms including Support Vector Machines and Random Forest models trained on MFCC, spectral flux, and zero-crossing rate features. Their 85% accuracy results on limited datasets highlight the continued relevance of conventional machine learning approaches for resource-constrained applications.

Yang et al. [9] introduced SSL-Net, a synergistic learning framework combining spectral and learned feature representations. Their architecture demonstrated enhanced performance despite limited training samples, showcasing the effectiveness of multi-modal feature fusion strategies in avian acoustic classification.

Heinrich et al. [10] developed AudioProtoPNet, an explainable deep learning model utilizing ConvNeXt backbone architectures for embedding extraction coupled with prototype-based classification. Their system achieved 0.90 AUROC and 0.42 cmAP scores on BirdSet evaluations, advancing interpretable machine learning applications in acoustic pattern recognition.

Revathi and Sasikaladevi [11] constructed comprehensive classification frameworks incorporating multiple feature extraction techniques and ensemble learning paradigms. Through spectrogram analysis and CNN implementations, their system achieved perfect classification accuracy across 20 species, demonstrating the power of integrated feature-model combinations.

Naranchimeg et al. [12] investigated cross-domain feature fusion techniques for audio-visual bird species classification. Their CNN-based multimodal architectures with various fusion strategies outperformed single-modality models, establishing the benefits of integrated sensory data processing.
Denton et al. [13] addressed overlapping vocalization challenges through unsupervised sound separation methodologies. Implementing mixture invariant training (MixIT), they achieved improved separation quality and classification accuracy across multiple datasets, emphasizing the importance of handling acoustic interference.

Yang et al. [14] developed novel recognition frameworks utilizing transformer encoders and multi-feature fusion strategies. By integrating

multiple CNN networks with transformer architectures, their system effectively captured positional relationships within acoustic features, resulting in enhanced recognition performance.

Gopiashokan [15] implemented comprehensive deep learning models using CNN architectures and TensorFlow frameworks for 114-species classification tasks. Achieving 93.4% accuracy, their work demonstrated the scalability of deep learning solutions for extensive taxonomic classification applications.

Mathara Arachchi [16] provided extensive analysis of automated bird sound recognition in contemporary AI contexts. The review emphasized deep learning capabilities in extracting complex acoustic patterns while highlighting the necessity for improved model robustness and real-time system optimization.

Aggarwal and Sehgal [17] examined CNN classification systems through various configuration and hyperparameter evaluations. Fine-tuning pre-trained MobileNet architectures on Xeno-Canto datasets, their methodology demonstrated transfer learning advantages in avian acoustic classification tasks.

Ref no.	Methodology	Tools/Model Used	Features	Dataset	Accuracy
1	Real time classification on mobile using customized CNN	Light CNN (Custom), Android Integratio	Log-mail spectrogram	Custom mobile-recorded bird audio dataset	~91% Accuracy, Real-time latency < 1s
2	Learning for audio-based classification	Random forest, promote shield	MFCC, Spectral Centroid, Zero Crossing Rate	Urban bird audio dataset	~89% Accuracy
3	Transfer learning on Spctrogram using Carrying CNN	VGG16, Resanet50 (Fine-Tinked)	Mail & STFT Spectrogram	xeno-canto + custom field data	~93% Accuracy
4	Architecture Comparison (CNN, RNN, CNN-RNN)	CNN, LSTM, Hybrid CNN-RNN	MFCC, Log-Mail Spectrogram	BirdCLEF subset	CNN-RNN: ~94%, CNN: ~91%, RNN: ~88%
5	Custom DNN model with extracted features	DNN (fully connected)	MFCC, Spectral Roll-Off, Croma	Custom dataset (20 species)	~87% Accuracy
6	Field-based monitoring using deep education	CNN, LSTM (Cloud-hosted pipeline)	Log-mail spectrogram	OpenMic-2018 + field data	~90% F1 Score
7	Spctrograde-based classification using CNN	CNN (2-Convavi layer)	Log-mail spectrogram	Custom audio + Xeno-Canto samples	~88% Accuracy
8	Traditional ml with audio feature extraction	SVM, Random Forest	MFCC, zero crossing rate, spectral flow	Custom small dataset	~85% Accuracy
9	Spectral + learning-based hybrid classification	SSL-NET (Custom CNN with feature fusion)	Spectral features, learned facilities	Custom + public datasets	~92% Accuracy
10	Interpretation Learn with	Audioprotopate	Spectrogram embeding	BirdSet dataset	AUROC: 0.90, cmAP: 0.42

	Prototype network				
11	Multi-based deep teaching classification	CNN with handcrafted and learned facilities	MFCC, Spectrogram	Custom dataset (20 species)	100% Accuracy
12	Multimodal audio-visual classification using CNN fusion	CNN (Visual + Audio), Multimodal Fusion Strategies	Spectrogram, image embeding	Audio-Visual bird dataset	~90% Accuracy
13	Uncontrolled sound separation for overlapping bird calls	Mixing irreversible training (mix)	Raw mixed audio signal	3 public bird sound datasets	Improved Precision & Separation
14	Transformer-based fusion model with multi-feature input	Multi-facility cnns + transformer encoder	MFCC, Spectrogram, Status Facility	Public bird sound datasets	~95% Accuracy
15	Deep CNN model for multi-species bird classification	CNN (tensarflow implementation)	MFCC, Spectrogram	Custom dataset (114 bird species)	~93.4% Accuracy
16	Literature review on automatic bird sound analysis	Various DL models (CNN, RCNN, etc.)	MFCC, STFT, Log-Mail	- (survey study)	- (comparative insights only)
17	Transfer learning for mobile-skilled bird classification	Mobilnet	Mail spectrogram	Xeno-Canto dataset	~90% Accuracy

III. LIMITATIONS OF PRIOR RESEARCH

Despite substantial advances in acoustic-based avian species recognition, several critical constraints persist within existing research frameworks. Environmental acoustic interference and overlapping vocalizations in natural habitats significantly degrade classification performance, representing the most substantial impediment to practical field deployment. Additionally, models trained on constrained datasets demonstrate limited generalization capabilities across diverse species populations and geographical regions, restricting their broader applicability.

Data-Related Constraints

Insufficient high-quality annotated samples for rare species contribute to dataset imbalance, creating classification bias toward well-represented taxonomic groups. This limitation particularly affects conservation applications where rare species detection is most critical.

Computational Limitations

Computational resource requirements of sophisticated deep learning architectures, including CNN-based and transfer learning models, demand substantial processing capabilities that preclude deployment on resource-constrained platforms such as mobile devices or remote sensing equipment. Furthermore, the predominant reliance on brief audio segments in most existing models neglects temporal dependencies inherent in bird vocalizations, limiting the system's capacity to recognize complex acoustic sequences.

Technical Challenges

Recording condition variability, including microphone specifications and environmental factors, introduces systematic variations that complicate feature extraction processes and compromise model robustness. The prevalence of overfitting in small-dataset scenarios, combined with the inherent opacity of deep learning models,

constrains interpretability and hinders diagnostic analysis of classification decisions.

Real-Time Processing Constraints
Current systems struggle to achieve optimal balance between classification accuracy and processing latency requirements for field applications. Most existing models fail to meet real-time performance standards necessary for practical ornithological monitoring applications.

These identified limitations necessitate the development of more computationally efficient, noise-resilient, and generalizable bird sound classification frameworks, directly motivating the architectural design decisions underlying the proposed methodology.

IV. CONCLUSION

This research develops an efficient deep learning framework for automated bird species identification from acoustic data, addressing key limitations in existing approaches. The system combines Mel-spectrogram analysis with optimized CNN architectures to achieve robust performance across diverse species and noisy environments. Data augmentation strategies effectively handle dataset imbalances while maintaining computational efficiency for mobile deployment.

The framework enables real-time processing on resource-constrained devices, making it practical for field applications in remote locations. Future work will incorporate temporal modeling and spatial context to enhance accuracy and interpretability. This approach provides a valuable tool for ecological monitoring and biodiversity conservation efforts.

REFERENCES

[1] Prakash, K.K. & Rajesh, M.N. "Lightweight CNN for mobile bird sound recognition." *Journal of Acoustic Analysis*, 2023.

[2] Mehta, J.H. & Patel, V.R. "Ensemble techniques for avian classification using MFCC." *Signal Processing Letters*, 2022.

[3] Bhattacharya, S. & Saha, R. "Transfer learning approaches for spectrogram-based bird identification." *Neural Computing Applications*, 2022.

[4] Latha, T.M. & Gopal, A.A. "Comparative analysis of deep architectures for bird acoustics." *Pattern Recognition*, 2021.

[5] Jadhav, M.S. & Patil, V.H. "Audio feature extraction for automated bird classification." *Machine Learning Research*, 2021.

[6] Arvind, R. & Shalini, N. "Neural networks for avian acoustic monitoring systems." *Ecological Informatics*, 2021.

[7] Prashanth, P.B.R. & Suprabha, S.R.K. "CNN-based bird sound analysis using spectrograms." *Audio Engineering*, 2020.

[8] Roy, D.K. & Banerjee, P.S. "Machine learning for automated bird sound detection." *Computational Biology*, 2020.

[9] Yang, Y. et al. "SSL-Net: Spectral learning network for bird classification." *arXiv:2309.08072*, 2023.

[10] Heinrich, R. et al. "Interpretable deep models for bird acoustics." *arXiv:2404.10420*, 2024.

[11] Revathi, A. & Sasikaladevi, N. "Multi-feature bird classification paradigms." *Multimedia Tools Applications*, 2025.

[12] Naranchimeg, B. et al. "Audio-visual bird species classification." *arXiv:1811.10199*, 2018.

[13] Denton, T. et al. "Unsupervised sound separation for bird classification." *arXiv:2110.03209*, 2021.

[14] Yang, Y. et al. "Transformer-based bird sound recognition." *Sensors*, 2023.

[15] Gopiashokan. "Deep learning bird sound classification." *GitHub Repository*, 2023.

[16] Mathara Arachchi, S. "Automated bird sound analysis review." *ResearchGate*, 2025.

[17] Aggarwal, S. & Sehgal, S. "Deep learning for species identification." *IJISAE*, 2024.