

AI-Powered Sound-Based Wildlife Monitoring System for Real-Time Detection and IoT Alerting

Kamaleshware P¹, Shantha Kumar S², Dr. S Mohana³, Ms. P Rohini⁴

Department of Computer Science and Engineering¹⁻²

M.E., Ph.D., Professor, Department of Computer Science and Engineering³

M.E., Assistant Professor, Department of Computer Science and Engineering⁴

Saranathan College of Engineering, Panjappur, Tamil Nadu, India¹⁻²

Department of Artificial Intelligence and Data Science³⁻⁴

K. Ramakrishnan College of Engineering, Trichy, India³⁻⁴

Abstract: This paper presents an AI-powered wildlife monitoring system designed for early detection of wildlife activity using acoustic sensing, convolutional neural network (CNN)-based intelligence, GPS localization, and Internet of Things (IoT) communication. A high-sensitivity LM393-based sound sensor connected to a NodeMCU ESP8266 monitors ambient events and provides the trigger for intelligent analysis. The project architecture also documents a camera-to-PC CNN inference path, allowing visual verification and emergency assessment in addition to the sound-based detection concept. When a positive wildlife event is obtained, the system acquires geographic coordinates using a GPS module, displays status locally through a 16x2 LCD and I2C interface, transmits event information to an IoT platform such as Blynk or ThingSpeak, and sends an SMS alert to the forest department. Reported experiments achieved more than 92% classification accuracy in a controlled environment and approximately 85% in semi-outdoor conditions. GPS localization was reported within 5-10 m, while alerts were received within 5-10 s of detection. The system operated for several hours of continuous testing and across multiple days of field simulation, although weak Wi-Fi caused occasional cloud transmission failures. The results indicate that a low-cost, non-invasive, and automated sensing architecture can support proactive wildlife monitoring while highlighting the need for larger real-world datasets, robust noise handling, and alternative communication paths for deployment in connectivity-constrained regions.

Keywords: wildlife monitoring, bioacoustics, sound classification, convolutional neural network, NodeMCU ESP8266, GPS, IoT, emergency alert

I. INTRODUCTION

Human-wildlife conflict is a persistent challenge in regions where settlements and agricultural lands border forests and wildlife reserves. Encounters between people and wild animals can lead to crop destruction, property damage, livestock loss, injuries, and retaliatory actions that threaten animal populations. The project report identifies a need for reliable and automated early-warning mechanisms that can detect wildlife presence before conflict escalates. Traditional approaches such as manual patrolling, camera traps, and electric fences can be labor-intensive, costly, maintenance-heavy, or reactive rather than proactive.

Sound provides a useful sensing modality because wildlife activity can generate vocalizations and movement-related acoustic events that may be detected without direct visual contact. The proposed project combines acoustic sensing with artificial intelligence and IoT connectivity so that a detected event can be classified and communicated to remote users. The NodeMCU ESP8266 acts as the central embedded controller; a high-sensitivity sound sensor monitors the environment, a CNN-based intelligence layer interprets detected events, GPS supplies location, and a cloud-connected



interface supports remote monitoring and alerting. This configuration is intended for forest edges, agricultural areas near wildlife corridors, conservation sites, and other locations where timely notification is valuable.

The principal contribution of this work is an integrated, low-cost monitoring pipeline that combines event sensing, AI-assisted recognition, location tagging, on-site display, and remote notification. The source report further documents a camera and PC-based CNN inference path and an object-detection implementation in its appendix. Rather than treating this as a separate project, the present paper interprets it as a complementary visual verification path within the broader wildlife-monitoring architecture. This interpretation is directly supported by the system block diagram, implementation code, and output snapshot contained in the report.

II. RELATED WORK

Research on environmental and animal sound classification has established CNNs as an effective approach for learning discriminative patterns from acoustic representations. The project report cites work on sound classification using CNN and tensor deep-stacking approaches, spectrogram-based animal sound recognition, and environmental sound classification with data augmentation [1]-[3]. End-to-end CNN approaches for environmental sounds are also reported in the cited literature [6], while classical and deep-learning approaches to environmental sound analysis provide additional context for robust acoustic monitoring [14], [19], [20].

Deep-learning methods have also been applied to wildlife surveillance and animal identification [4], [7], [9]-[13], [17], [18]. The report highlights that passive acoustic monitoring can generate large volumes of data, making automated analysis important for conservation. Data augmentation has been specifically explored for animal-audio classification [15], [16]. In the context of human-wildlife conflict, an IoT embedded approach for detection and discrimination of animals has also been reported [8]. These studies support the use of AI-assisted acoustic or visual recognition together with connected sensing, while also motivating the need for robust performance in real-world noise.

The proposed system differs in its emphasis on a practical edge-to-cloud monitoring loop: a low-cost NodeMCU controller interfaces with an acoustic sensor, optional visual inference, GPS, local LCD output, and IoT alerting. The design therefore focuses not only on recognition accuracy but also on response time, localization, portability, and proactive communication.

III. PROPOSED SYSTEM

A. SYSTEM ARCHITECTURE

The system is organized around the NodeMCU ESP8266, which coordinates sensor input, communication, and local display. The project block diagram connects the sound sensor and IoT application to the NodeMCU, while GPS supplies geographical information and an I2C connector drives the LCD display. The same diagram also shows a camera feeding a PC-based CNN/deep-learning stage through a USB-to-TTL interface. This architecture makes the embedded controller the integration point between detection, inference, localization, and alerting.

B. EVENT DETECTION AND AI PROCESSING

During normal operation, the sound sensor continuously monitors environmental activity. A predefined threshold is used to identify events that require further processing. The NodeMCU forwards the detected event through the documented serial/USB-to-TTL path to the CNN-based intelligence stage. The report describes the deep-learning component as responsible for distinguishing meaningful wildlife events from irrelevant environmental sounds and for supporting the decision on whether an event requires emergency action.

The source material does not specify the training-set size, train/test split, optimizer, number of epochs, exact CNN layer configuration, or class-wise precision/recall values for the implemented model. Accordingly, this paper reports only the aggregate performance values actually documented in the project report and does not introduce unverified model hyperparameters or additional metrics.



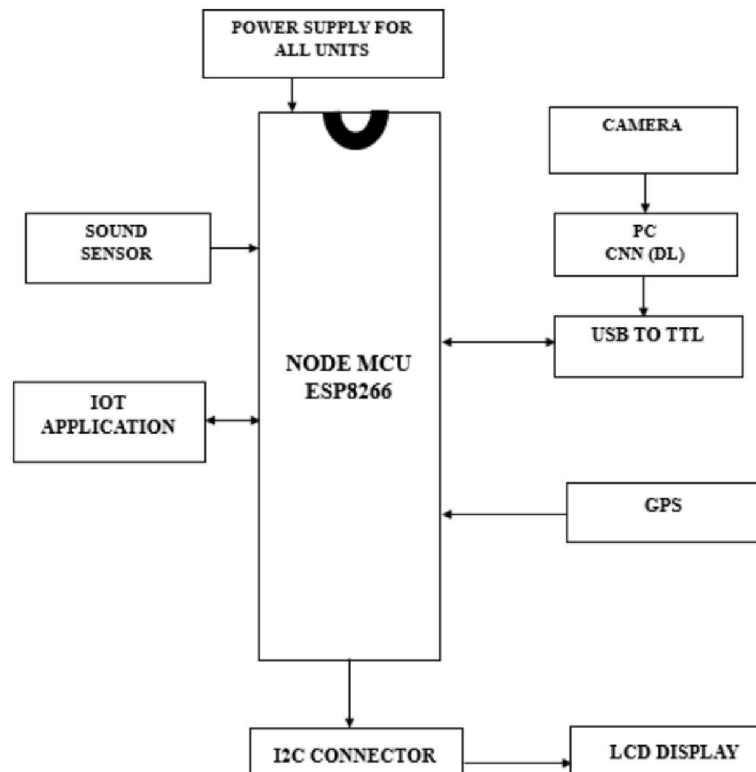


Fig. 1. System block diagram documented in the project report

C. GPS LOCALIZATION, DISPLAY, AND IOT ALERTING

When a positive wildlife event is obtained, the GPS module provides latitude and longitude for the detection. The report identifies a NEO-6M-type GPS module and a serial TX/RX interface with the NodeMCU. The event status and location are displayed locally through a 16x2 LCD connected using an I2C interface, reducing wiring requirements to the SDA and SCL lines. For remote monitoring, the event information can be transmitted to a cloud IoT platform such as Blynk or ThingSpeak. The proposed workflow also includes SMS notification to the forest department so that preventive action can be initiated.

D. COMPLEMENTARY VISUAL VERIFICATION

Although the project title emphasizes sound-based monitoring, the documented implementation includes a camera-to-PC inference path and an object-detection program based on TensorFlow. The appendix loads a frozen inference graph and label map, processes frames from a camera stream, and applies a detection-score threshold above 0.70 for an injury-related output condition. A representative output image shows a detected wildlife injury region. This evidence supports interpreting the camera/CNN stage as a complementary visual verification capability rather than as evidence that the complete system is exclusively audio-only.



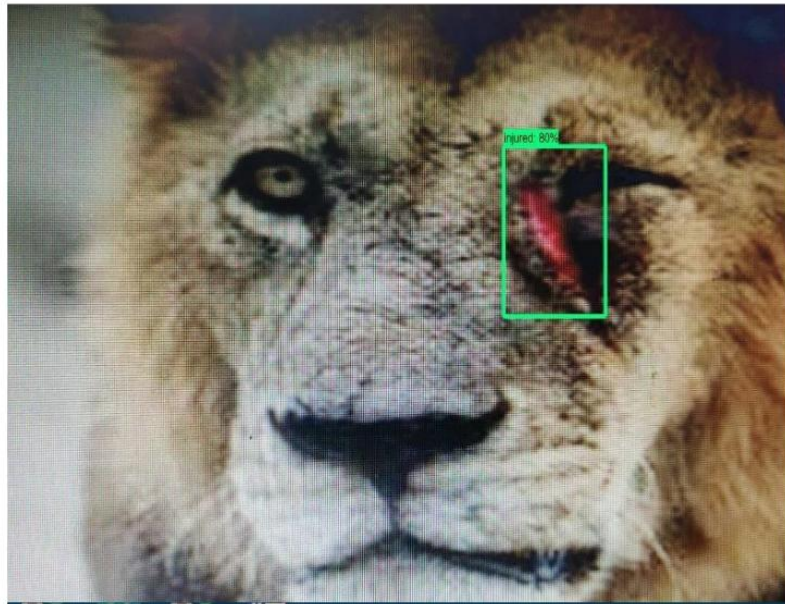


Fig. 2. Representative CNN-based visual detection output documented in the implementation

E. HARDWARE AND SOFTWARE COMPONENTS

TABLE I: MAJOR COMPONENTS AND THEIR FUNCTIONS

Component / Tool	Role in the Proposed System
NodeMCU ESP8266	Central controller for sensor interfacing, communication, and local display management.
LM393-based sound sensor	Threshold-based detection of environmental acoustic activity.
Camera and PC CNN path	Complementary visual inference path documented in the block diagram and appendix.
NEO-6M-type GPS	Provides latitude and longitude for detected events.
16x2 LCD + I2C	Displays detection status and location locally with reduced wiring.
USB-to-TTL interface	Provides the documented serial communication path between NodeMCU and the PC/CNN stage.
Blynk / ThingSpeak	Remote IoT monitoring, data logging, and dashboard access.
Arduino IDE, Python, TensorFlow, Keras, NumPy, SciPy, OpenCV	Software environment and libraries listed in the project report.

IV. IMPLEMENTATION AND EXPERIMENTAL SETUP

A. OPERATIONAL WORKFLOW

The operational sequence follows an event-driven loop. First, the sound sensor monitors the environment and the NodeMCU checks the signal against a predefined threshold. When the threshold is exceeded, the event is forwarded to the AI processing stage. If the AI stage identifies a wildlife-related event, the controller retrieves GPS coordinates, updates the LCD, transfers the event to the IoT platform, and issues an SMS alert. The system then returns to



continuous monitoring. The documented camera path can additionally provide visual evidence for the event, particularly for the injury-detection functionality shown in the implementation.

B. TEST CONDITIONS

The project report evaluates the system in controlled and semi-natural or semi-outdoor conditions. Pre-recorded and real-time animal sounds were used as inputs, with examples including elephant trumpets, tiger growls, monkey calls, and dog barks. The evaluation considered classification accuracy, location accuracy, alert responsiveness, continuous operation, power behavior, and robustness to environmental noise.

C. REPORTING SCOPE

The available project documentation provides aggregate experimental observations rather than a complete reproducibility package. In particular, it does not provide the exact number of training samples used for the final implemented model, the train/validation/test partition, class-wise confusion-matrix counts, or detailed network hyperparameters. These omissions do not invalidate the reported prototype measurements, but they limit the extent to which the model results can be independently reproduced from the report alone. The present paper therefore preserves the reported aggregate figures without extrapolating beyond the documented evidence.

V. RESULTS AND DISCUSSION

A. CLASSIFICATION PERFORMANCE

In the controlled environment with minimal background noise, the report states that the deep-learning model achieved more than 92% accuracy for correctly classifying wild-animal sounds. The reported confusion-matrix observation indicates a high true-positive rate for distinct and loud calls, with elephants and tigers identified as examples. Under semi-outdoor conditions containing wind, rustling leaves, birds, and human voices, the reported accuracy decreased to approximately 85%. False positives were associated mainly with mechanical or environmental noises that resembled animal sounds. This difference demonstrates the sensitivity of acoustic recognition to the deployment environment and motivates continued collection of diverse real-world sound data.

TABLE II: REPORTED PROTOTYPE PERFORMANCE

Measure	Reported value	Interpretation
Controlled-environment accuracy	>92%	High aggregate classification accuracy under low-noise conditions.
Semi-outdoor accuracy	~85%	Reduced performance with wind, vegetation, birds, and human voice noise.
GPS localization	5-10 m	Reported location error suitable for the intended early-warning use case.
Alert latency	5-10 s	Time from detection to remote alert receipt in reported field tests.
Continuous operation	Several hours; multi-day simulation	Stable operation reported, with occasional cloud failures under weak Wi-Fi.

B. RESPONSE TIME AND LOCALIZATION

The system obtained GPS coordinates consistently within a reported 5-10 m margin. Once an event was successfully detected and classified, field tests reported alert receipt within 5-10 s. The combination of location tagging, local LCD feedback, cloud data access, and SMS notification allows a user to move from detection to actionable response without waiting for manual review of stored media.



C. POWER, CONNECTIVITY, AND DEPLOYABILITY

The prototype was operated for several hours of continuous testing and across multiple days of field simulation. The report indicates that the low-power NodeMCU architecture can be supported by portable battery or solar arrangements, which is relevant to remote forest boundaries. The main connectivity limitation was weak Wi-Fi: cloud transmission occasionally failed when network coverage was poor. The report suggests temporary local storage and later synchronization as a mitigation, and identifies cellular or LoRa-based communication as potential alternatives for connectivity-challenged locations.

D. ROBUSTNESS AND LIMITATIONS

Environmental noise is the principal observed accuracy limitation. Overlapping human and machine sounds can produce false positives, while quiet or stealthy animals may produce insufficient acoustic evidence for detection. The system is therefore best viewed as an early-warning component rather than a complete substitute for all wildlife-sensing technologies. The non-invasive nature of the approach is a key strength, because it avoids physical tagging and does not require continuous visual contact with animals.

VI. FUTURE ENHANCEMENTS

Future work should prioritize larger and more geographically diverse wildlife-audio datasets, stronger augmentation for background noise, and systematic evaluation with class-wise precision, recall, F1-score, ROC/AUC where appropriate, and a fully specified train/validation/test protocol. The project report also motivates adaptive detection thresholds so that quiet rural areas and noisy settlement boundaries can be treated differently. The acoustic pipeline could be complemented by motion, infrared, or thermal sensing to reduce missed detections for animals that make little sound. At the communication layer, offline buffering followed by synchronization, together with cellular or LoRa links, could improve reliability in remote regions. A multi-node deployment could aggregate detections over a larger forest boundary and support movement-pattern analysis. The literature reviewed in the project report also points toward data augmentation, spectrogram-based representations, ensemble CNNs, and newer transformer-style acoustic models as potential directions for improving robustness when more training data become available [2], [3], [15], [16], [19].

VII. CONCLUSION

This work demonstrates a practical architecture for AI-assisted wildlife monitoring that combines acoustic sensing, CNN-based intelligence, GPS localization, IoT communication, and local display. The reported prototype achieved more than 92% accuracy in controlled testing and approximately 85% in semi-outdoor conditions, with GPS localization within 5-10 m and alert delivery within 5-10 s. The implementation is non-invasive, portable, and suitable for continuous monitoring at forest boundaries and other wildlife-prone areas. At the same time, the observed sensitivity to environmental noise, limited acoustic evidence from quiet animals, and dependence on Wi-Fi show that further field validation is necessary. The documented camera/CNN path provides a useful complementary verification mode, making the overall prototype a promising foundation for multimodal wildlife early-warning systems.

ACKNOWLEDGMENT

The author acknowledges the project work and documentation that provided the system architecture, implementation details, and experimental observations used to prepare this manuscript.

REFERENCES

- [1] A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey and P. Tiwari, "Sound classification using convolutional neural network and tensor deep stacking network," *IEEE Access*, vol. 7, pp. 7717-7727, 2019.
- [2] Y. Li, H. Huang and Z. Wu, "Animal sound recognition based on double feature of spectrogram," *Chinese Journal of Electronics*, vol. 28, no. 4, pp. 667-673, 2019.



- [3] Zohaib Mushtaq, Shun-Feng Su and Quoc-Viet Tran, "Spectral images based environmental sound classification using CNN with meaningful data augmentation," *Applied Acoustics*, vol. 172, pp. 107581, 2021.
- [4] Ruilong Chen, Ruth Little, Lyudmila Mihaylova, Richard Delahay and Ruth Cox, "Wildlife surveillance using deep learning methods," *Ecology and Evolution*, vol. 9, no. 17, pp. 9453-9466, 2019.
- [5] Zachary J. Ruff, Damon B. Lesmeister, Cara L. Appel and Christopher M. Sullivan, "Workflow and convolutional neural network for automated identification of animal sounds," *Ecological Indicators*, vol. 124, pp. 107419, 2021.
- [6] Yuji Tokozume and Tatsuya Harada, "Learning environmental sounds with end-to-end convolutional neural network," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.
- [7] Gyanendra K. Verma and Pragya Gupta, "Wild animal detection from highly cluttered images using deep convolutional neural network," *International Journal of Computational Intelligence and Applications*, vol. 17, no. 04, pp. 1850021, 2018.
- [8] Surya Thangavel and Chitra Selvi Shokkalingam, "The IoT based embedded system for the detection and discrimination of animals to avoid human-wildlife conflict," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 6, pp. 3065-3081, 2022.
- [9] Mohammad Sadegh Norouzzadeh et al., "Automatically identifying, counting and describing wild animals in camera-trap images with deep learning," *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. E5716-E5725, 2018.
- [10] Anamika Dhillon and Gyanendra K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85-112, 2020.
- [11] M. Clapham, E. Miller, M. Nguyen and C. T. Darimont, "Automated facial recognition for wildlife that lack unique markings: A deep learning approach for brown bears," *Ecology and Evolution*, vol. 10, no. 23, pp. 12883-12892, 2020.
- [12] J. Peng, D. Wang, X. Liao, Q. Shao, Z. Sun, H. Yue et al., "Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 364-376, 2020.
- [13] Frank Schindler and Volker Steinhage, "Identification of animals and recognition their actions in wildlife videos using deep learning techniques," *Ecological Informatics*, vol. 61, p. 101215, 2021.
- [14] Karol J. Piczak, "Environmental sound classification with convolutional neural networks," 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), 2015.
- [15] Loris Nanni et al., "Ensemble of convolutional neural networks to improve animal audio classification," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, pp. 1-14, 2020.
- [16] Loris Nanni, Gianluca Maguolo and Michelangelo Paci, "Data augmentation approaches for improving animal audio classification," *Ecological Informatics*, vol. 57, p. 101084, 2020.
- [17] Marco Willi et al., "Identifying animal species in camera trap images using deep learning and citizen science," *Methods in Ecology and Evolution*, vol. 10, no. 1, pp. 80-91, 2019.
- [18] Mohammad Sadegh Norouzzadeh et al., "A deep active learning system for species identification and counting in camera trap images," *Methods in Ecology and Evolution*, vol. 12, no. 1, pp. 150-161, 2021.
- [19] Justin Salamon and Juan Pablo Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279-283, 2017.
- [20] S. Abdoli, P. Cardinal and A. Lameiras Koerich, "End-to-end environmental sound classification using a 1D convolutional neural network," *Expert Systems with Applications*, vol. 136, pp. 252-263, 2019.

