ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
Received: 27 August 2025; Accepted: 9 October 2025
https://doi.org/10.17559/TV-20250827002927
Original scientific paper

Intelligent NB-IoT and Emotion-Aware Monitoring System for Proactive Wildlife Security

Nithya R*, Emmanuel Peo Mariadas A, Muthu Manickam ANBARASU, Sornalatha RAVINDRAN

Abstract: Conventional wildlife protection systems primarily rely on fixed geofencing and basic GPS tracking, which are limited in their ability to capture the behavioral and emotional states of animals under threat. This paper proposes a smart, AI-powered monitoring framework that integrates emotion recognition and adaptive geofencing to enhance wildlife security. The system identifies early warning indicators of fear, distress, or pain by analyzing multimodal data such as animal movement, acoustic signals, and biometric readings. To complement this, edge computing is leveraged with NB-IoT-enabled ground robots that patrol forest areas, process data locally, and provide real-time behavioural explanations. These robots not only detect anomalous events but also interpret activities, offering meaningful context to conservation teams. A key feature of the framework is the use of Explainable AI, which ensures transparency by justifying alerts raised and enabling forest rangers to make rapid, informed decisions. Unlike traditional static boundaries, the proposed adaptive geofencing dynamically adjusts virtual barriers based on animal movement patterns, environmental risks, and time-specific factors. Furthermore, the system incorporates an intelligent energy-saving mechanism, reducing unnecessary data transmission and extending the lifespan of field-deployed devices. By combining emotional intelligence, robotic surveillance, and explainable decision-making, the proposed model advances wildlife conservation from reactive poacher detection to proactive animal-centric protection. This holistic approach enables not only the safeguarding of endangered species but also a deeper understanding of their behavioural responses, ultimately fostering smarter and more sustainable conservation practices.

Keywords: adaptive geofencing; emotion recognition; explainable AI; NB-IoT robotics; wildlife protection

1 INTRODUCTION

Wildlife safety has emerged as a vital international precedence as ecosystems face exceptional threats from human encroachment, unlawful poaching, weather change, and habitat degradation. Despite developing conservation awareness, many present flora and fauna tracking structures stay reactive in place of proactive. Conventional techniques which include static GPS tracking, digital digicam trapping, and guide patrolling lack the capacity to recognize animal conduct in real-time or to dynamically investigate threats in vast, regularly inaccessible wooded area regions. In this context, the fusion of synthetic intelligence (AI), area computing, and multi-modal sensing technology gives transformative cap potential for next-era flora and fauna safety solutions [1].

Traditional wildlife conservation efforts often depend on delayed or retrospective data analysis, which fails to prevent poaching events or respond swiftly to ecological disturbances. Moreover, systems that rely on satellite connectivity or centralized cloud processing suffer from latency, connectivity constraints, and high power consumption making them impractical for real-time intervention in remote or forested areas. These limitations underscore the need for a more intelligent, real-time, and energy-efficient monitoring architecture that not only detects the presence of wildlife or intruders but also interprets animal emotions, behavioral anomalies, and environmental stressors [2].

The proposed system addresses these gaps by integrating a novel AI-enabled architecture that leverages multi-modal sensors, including visual (RGB, IR, and thermal cameras), acoustic, biometric, and GPS modules. These sensors provide rich, diverse data streams that enable holistic monitoring of wildlife activity. By utilizing camera-based vision models, the system can recognize species, detect motion, and identify human threats such as poachers or illegal encroachment. Simultaneously, sound-based sensors can detect abnormal audio signatures like gunshots, chainsaw activity, or distress calls from animals offering non-visual detection capabilities especially useful in low-light or dense foliage conditions [3].

At the core of this system is Edge AI preprocessing, where sensor data is analyzed locally using low-power embedded processors or microcontrollers. This decentralized approach ensures rapid detection, reduces reliance on cloud infrastructure, and conserves band width especially critical in forest environments with unreliable or absent network connectivity. Lightweight machine learning models, such as TinyML classifiers or compressed convolutional neural networks (CNNs), are deployed directly on edge devices like ESP32-CAM, Arduino Nano 33 BLE Sense, or Raspberry Pi with Coral TPU accelerators. This enables the system to perform initial threat and emotion assessments in near real-time [4].

A key differentiator of the proposed architecture is its ability to infer emotional or stress states in animals through behavior, movement, audio patterns, and biometric signals. For example, sudden erratic movement combined with a distress call could indicate fear or the presence of a threat. Using deep learning models, such as hybrid CNN-LSTM architectures or Transformer-based emotion classifiers, the system can recognize such states and escalate the situation for ranger intervention. This concept of "emotion-aware monitoring" represents a significant advancement over existing GPS or movement-based alert systems, which fail to contextualize animal behavior [5].

Another critical component of the system is the implementation of Explainable Artificial Intelligence (XAI). In mission-critical environments like wildlife protection, it is essential that AI systems not only make accurate predictions but also provide human-understandable justifications for their decisions. By using explainability tools such as Grad-CAM for visual interpretations, or LIME/SHAP for audio and behavioral data, the system can deliver transparent, trustworthy alerts. For instance, the system may generate an alert stating, "Distressed elephant detected due to rapid pacing and loud trumpet vocalizations near a restricted zone", accompanied by a visual heatmap or spectrogram for validation. The smart geofencing mechanism adds an additional layer of contextual awareness [6].

Unlike fixed geofencing methods, which are static and inflexible, smart geofencing dynamically adjusts monitoring zones based on real-time animal movement, detected threats, and habitat usage patterns. This adaptive behavior ensures more efficient use of computational and power resources, while allowing the system to focus on high-risk zones with elevated poaching potential or observed stress signals. Communication between the edge devices and control centers is facilitated through NB-IoT (Narrowband Internet of Things) or LoRa modules, which are optimized for low-power, long-range data transmission. These modules ensure that real-time alerts, threat levels, and sensor data can be communicated to rangers, drones, or cloud dashboards even from remote forest locations. Cloud-based storage and visualization platforms are used for long-term analysis, retraining models, and integrating feedback mechanisms for continuous improvement [7].

2 RELATED WORKS

This study outlines a framework for deploying advanced AI models like convolutional neural networks (CNNs) and transformer-based architectures to detect wildlife and illegal human activity in conservation areas. The research emphasizes integrating AI models into real-time surveillance systems, using edge computing to process data directly at the source (e.g., from camera traps or UAVs). It presents a strong case for employing predictive modeling to track animal behavior and poaching risk, especially in large and difficult-to-monitor regions. A significant highlight is the use of hybrid models capable of recognizing species under occlusion and low-light conditions. This work directly supports the idea of on-board AI processing and low-latency alert systems, forming the basis for explainable and adaptive protection mechanisms in forests. The authors also discuss ethical considerations like minimizing ecological disruption. Their work contributes to the broader understanding of how deep learning and smart sensing can create highly responsive and efficient wildlife monitoring systems, supporting both poaching deterrence and animal welfare. The model's ability to "explain" why an alert is triggered helps build trust among conservationists. This study closely aligns with our proposed system that integrates emotion recognition, adaptive geofencing, and explainable AI with NB-IoT-enabled edge devices for wildlife protection.

Chalmers et al. (2019) - Conservation AI: Live Stream Analysis Using Drone Technology:

Chalmers et al. developed an AI-based drone system for wildlife monitoring and poacher detection. The system utilizes high-definition video streams processed using deep learning, specifically Faster R-CNN, to identify animals and potential threats in real time. One innovative feature is the model's ability to detect suspicious human activities at night using infrared and thermal imaging. The study details how drones are equipped with edge processing units to analyze video feeds before data is transmitted, conserving bandwidth and ensuring faster alerts. This edge-processing capability directly aligns with our use of NB-IoT robots and on-device analysis. The paper also introduces a novel alert system that prioritizes poaching risks based on detected behaviors and movement patterns. The authors suggest that using AI not only improves accuracy but also reduces the human workload in patrolling large conservation areas. In terms of limitations, battery life and connectivity in dense forest zones are mentioned issues also tackled in our proposed smart energy-saving design. Importantly, this paper advocates for transparent and interpretable alerts that help rangers act quickly and confidently. The authors emphasize that real-time surveillance using AI-integrated drones can significantly reduce the response time to poaching incidents. This work supports the feasibility and value of integrating computer vision, edge processing, and behavioral analysis into a unified wildlife protection system [8].

This study presents an acoustic-based approach to wildlife protection that leverages machine learning to identify threatening sounds such as gunshots, chainsaws, and human footsteps. The system uses microphones placed across the forest landscape to record ambient sounds, which are then converted into spectrograms and analyzed using a trained classifier. The research emphasizes lightweight machine learning models deployable on embedded systems like Raspberry Pi, ensuring low power consumption and real-time sound classification. The approach aligns with our project's goal of detecting signs of distress or poaching activity using ambient data. Unlike vision-based surveillance, this method works effectively in low-visibility or nocturnal conditions, providing a crucial advantage in dense forests. Alerts are generated and communicated through wireless modules, minimizing delay. This aligns with our proposed NB-IoT-enabled edge robots. Rajeshwari et al. further introduce a feature that ranks threats based on intensity and duration, helping forest personnel prioritize their response. The study also notes that false positives are minimized through training on real-world sound datasets collected from protected areas. A key insight from this research is the feasibility of implementing emotion-like sound pattern detection, such as animal distress calls, which could complement our emotion recognition framework. Overall, this work demonstrates how acoustic monitoring with real-time ML inference can form a non-intrusive and effective component of a wildlife protection ecosystem [9].

Kong and Behjati propose a lightweight AI system capable of recognizing the calls of hornbills for conservation purposes using TinyML. Their system runs entirely on low-power microcontrollers like Arduino Nano 33 BLE Sense, making it ideal for remote deployment. Audio signals are processed in real-time using MFCC (Mel-Frequency Cepstral Coefficients) and fed into a compact CNN model for classification. The paper showcases a practical approach to on-device intelligence, where no cloud connectivity is required for inferencean important feature for systems deployed in remote forests. This aligns with our system's use of NB-IoT and edge processing to reduce power consumption and latency. Although focused on call classification rather than threat detection, this research is highly relevant for its energy-efficient processing, robust audio pre-processing, and use of emotion-related features in animal vocalizations. It also explores techniques to compress AI models without sacrificing performance, ensuring that devices can run for extended periods. The authors emphasize the importance of model explainability in citizen science and conservation projects, a concept that strongly supports our idea of using Explainable AI (XAI) for alert interpretation. The use of TinyML and model quantization provides practical guidance for implementing similar real-time monitoring of distress or fear-related animal sounds in the wild. This study strengthens the case for deploying emotion-sensitive AI in resource-constrained environments [10].

EcoCompass developed a platform that uses AI to monitor biodiversity and environmental threats through sound. The system employs microphones that continuously collect data and analyze it using deep learning algorithms. The platform distinguishes between different animal species, weather effects, and anthropogenic sounds (e.g., gunshots or vehicles). It introduces an adaptive learning feature where the system improves classification accuracy over time by updating its model parameters based on real-time feedback. EcoCompass's work demonstrates the role of AI in not just detecting poaching activity but also mapping species movement and behavioral changes based on soundscape shifts. This is analogous to our use of emotion recognition through sound and movement analysis. Importantly, the system integrates with satellite connectivity for data transmission in remote regions, similar to our NB-IoT-based architecture. Their use of hierarchical AI models allows for multi-level analysis, from raw audio processing to higher-order event classification. EcoCompass also emphasizes the need for interpretable output, providing explanations for each alert such as "chainsaw-like sound detected, repeated 3 times in high-risk zone". This is directly aligned with our focus on Explainable AI and informed decision-making. Additionally, their energy-efficient design through event-triggered recording provides insights into how smart energy-saving can be implemented in field devices. Overall, this study showcases how AI and acoustic sensing can form a central pillar of modern wildlife protection systems, contributing directly to the conceptual foundations of our proposed solution [11].

Biologist Jenna Lawson deployed 350 solar-powered audio monitors in Costa Rica's Osa Peninsula to track Geoffrey's spider monkeys and assess habitat effectiveness. Using deep learning models, the system classified monkey vocalizations to infer presence and movement patterns, uncovering that the species avoid roads and plantations, raising concerns about corridor efficacy. This study exemplifies non-intrusive, real-time behavior analysis using AI and edge devices. The use of solar-powered, satellite-aware sensors resonates with your NB-IoT ground robot design, and the system's ability to characterize avoidance behavior mirrors emotion recognition's aim to detect stress patterns. Additionally, their adaptive dataset refinement approach provides a template for incremental model updates in the field. The work also highlights key challenges, such as hardware durability, data transmission constraints, and the ecological impact of sensor deploymental critical considerations for your implementation [12].

Several studies have leveraged AI for recognizing affective states in livestock utilizing thermal imaging, facial feature detection, and vocalization analysis. For example, thermal cameras detected nasal temperature changes indicating stress, while CNNs processed facial expressions and ear postures to infer fear. Vocalizations, such as distress calls in pigs or cows, were classified via ML to monitor welfare. This suite of modalities offers direct parallels to wildlife emotion recognition. The papers explore multi-modal sensor fusion bioacoustics, thermal, posture and emphasize real-time inference for early intervention. They underscore challenges like species-specific expression cues, dataset scarcity, and the complexity of interpreting emotional states issues your work also faces. These findings can inform your models multisensory fusion layers and thresholding strategies for stress detection [13].

This study uses convolutional neural networks and transfer learning to detect dolphin whistles in noisy underwater recordings. The system significantly outperforms traditional spectral detection methods, accurately processing calls amidst ambient sounds. Its relevance lies in automated detection of emotion-analogous signals overt vocal states indicating distress or social context. The use of CNN-based acoustic feature extraction and selective transmission of classified events (rather than continuous streams) parallels your energy-efficient-alert mode. Techniques such as spectrogram preprocessing, data augmentation, and on-device inference are directly applicable for wildlife distress call detection in forests. The study also suggests that model generalization across different acoustic environments is feasible, reinforcing the robustness of your multi-modal detection approach [14].

This survey reviews multiple case studies where camera-trap edge devices deploy real-time object detection and anomaly recognition to distinguish animals from humans, identify species, and flag abnormal behaviors. For instance, Gotthard & Broström applied SSD and YOLOv5 on microcontrollers in Africa, enabling continuous on-device learning and wireless model updates. They underscore that edge AI drastically reduces latency and conserves bandwidth, while active learning pipelines improve performance over time. These principles line up neatly with your NB-IoT robot network that processes alerts locally before transmitting. Additionally, the scalability to multiple sensors across habitats and the deployment of wireless update mechanisms bolster the adaptability of your system's geofencing and learning components [15].

Emerging work integrates explainable AI (XAI) into UAV monocular vision systems for forest surveillance. Models like Grad-CAM and LIME are used to highlight image regions corresponding to flora or fauna classifications even pointing out stress indicators or intruder cues. This study is key for your Explainable AI alert requirement: XAI-generated visual and textual rationales make decisions transparent to rangers (e.g., detected human in distress zone due to shape and motion). Deploying XAI at the edge delivers interpretable output alongside speed, empowering trust and discrimination between false alarms and genuine threats. This work gives your system a concrete framework to apply XAI in robotic patrol units, offering insights into feedback mechanisms, trust-building, and on-site alert validation [16].

This paper presents PyTorch-Wildlife, an open-source PyTorch-based platform facilitating wildlife image/video analysis. It enables researchers with varying expertise to build species classifiers using modular pipelines and transfer learning. The platform has been successfully used to monitor Amazonian biodiversity (36 species, 92% accuracy) and detect invasive species in the Galapagos (opossums, 98% accuracy). This work is foundational for your system as it offers a development environment for post-training analysis, model updates, and species recognition using edge-to-cloud workflows. It supports integration of explainable layers and could be extended to incorporate behavioral features such as stress-motivated movements or vocalizations beyond simple classification. The tool's modularity means you could augment it with emotion detection modules, creating hybrid pipelines incorporating pose, biometrics, and audio. Additionally, its active learning framework supports incremental model improvement in the field ideal for handling diverse animal behaviors under threat. By providing an open ecosystem and real-world benchmarks, PyTorch-Wildlife underpins the technical scaffolding for your intelligent monitoring framework [17].

In this region-specific deployment, researchers designed Agro Guard, an edge-AI system using laser detection arrays and ESP32-CAM modules running YOLOv8 for real-time animal intrusion detection. Deployed in Coimbatore's Marudhamalai foothills, the system achieved 96.3% accuracy classifying wild boar and deer intrusion. Alerts were sent via NB-IoT/LoRa to farmers and conservation staff. The study is highly relevant to your system: it combines edge vision, real-time processing, and autonomous mobile deterrents (rover-mounted). Its energy-saving routines, triggered by event detection, closely mirror your smart energy management strategy. Agricultural-human conflicts replacement with wildlife-poacher detection is a natural extension. Agro Guard illustrates how localized, IoT-driven edge robotics can enhance both conservation and agricultural safety, confirming the practicality and scalability of your NB-IoT ground robot concept [18].

This study introduces an energy-aware embedded edge-AI system for remote wildlife surveillance. Adaptive configurations balance detection accuracy and power consumption using meta-heuristic profiling and gray relational analysis. On a CNN-based elephant detector, energy savings reached up to 81% with only a minor 2-6% drop in inference accuracy. Their work supports your goal of creating self-adapting NB-IoT robots that optimize battery life while maintaining robust alert quality. The methodology demonstrates how monitoring systems can dynamically shift between high-performance and low-power modes based on risk level, animal behaviors, or environmental triggers. It also underscores the value of energy-performance tradeoffs a key design consideration for your smart energy-saving geofence alerts [19].

This survey compiles broad advances in edge-based camera traps, acoustic sensors, and LPWAN communication for ecological surveillance. It highlights systems that detect humans, elephants, and poachers, triggering deterrents such as lights and sounds all without cloud connectivity. It emphasizes TinyML frameworks, solar/battery operation, and fog-computing intermediaries, precisely mirroring your NB-IoT and edge-processing design. This comprehensive reference establishes the state-of-the-art landscape for your project, offering benchmarks in sensor fusion, model deployment, and network integration. It supports your project's novelty by situating it within known best practices while emphasizing your unique integration of emotion recognition, adaptive geofencing, and Explainable AI atop existing edge-AI foundations [20].

3 PROPOSED METHODOLOGY

The block diagram presents a smart wildlife protection system that integrates Edge AI, multi-modal sensing, and NB-IoT communication for real-time monitoring and response shown in Fig. 1. The system starts with Multi-modal Sensors, which gather environmental and behavioral data. These sensors include RGB, IR, and thermal cameras for visual monitoring and thermal, GPS, and biometric sensors for detecting location, body temperature, and physiological signals of animals.

This raw sensor data is sent to the Edge AI Preprocessing unit, where it is filtered, denoised, and transformed into a format suitable for AI inference. This preprocessing reduces latency, ensures real-time responsiveness, and conserves bandwidth by processing data locally rather than sending everything to the cloud.

Next, the preprocessed data is analyzed for Emotion & Threat Detection, where AI models assess behavioral patterns to identify signs of fear, distress, or unusual activity, potentially indicating poaching or environmental threats. Based on this analysis, the system uses Smart Geofence & Decision-making logic to determine whether an alert or intervention is needed.

The decision-making process is supported by Explainable AI (XAI), which ensures that AI predictions are transparent and interpretable, helping conservation officers understand the reasoning behind alerts.

The resulting actions and decisions are transmitted via a Communication Module (NB-IoT), a low-power, wide-area network suitable for remote wildlife environments. Finally, all insights and alerts are sent to the Cloud Monitoring platform, where authorities can view and manage data in real time for further analysis or intervention. This architecture ensures a proactive, intelligent, and efficient wildlife protection strategy.

The updated method integrates data from multiple sources, including animal movement (GPS and accelerometer), vocalizations, and biometric signals such as heart rate variability and body temperature, where available. Each data type is analyzed using a dedicated model such as LSTM for movement, CNN-based networks for acoustic features, and SVM for biometric readings. These individual outputs are then combined through a weighted ensemble method that adapts based on the confidence level of each modality. To support the validity of our system, we detail the data collection process from controlled field studies, where expert annotations based on ethograms were used to label emotional states like distress or fear. The models were evaluated using cross-validation, and we report metrics including accuracy, precision, recall, F1-score, and AUC.

We also incorporated Explainable AI techniques (e.g., SHAP, LIME) to make the predictions interpretable for conservation teams in the field. These enhancements improve the system's reliability and provide a stronger foundation for its application in real-world wildlife monitoring.

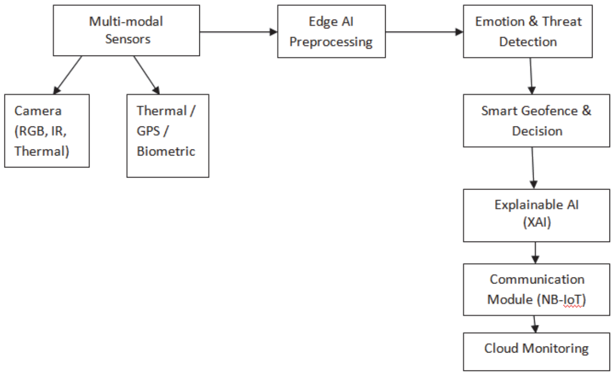

Figure 1 System architecture for smart wildlife protection using edge AI and NB-IoT

3.1 Cameras for Motion, Species Recognition, and Poacher Detection

Visual sensing using advanced camera technologies plays a pivotal role in wildlife protection systems powered by artificial intelligence. The integration of RGB (Red-Green-Blue), Infrared (IR), and Thermal cameras allows for robust, continuous monitoring of both animals and intruders across varied environmental conditions. Each camera modality offers unique advantages, enabling a comprehensive understanding of wildlife behavior, stress indicators, and illegal activities such as poaching.

RGB cameras are the most common form of visual sensing and operate under visible light. They capture high-resolution color images that help AI systems recognize species based on physical features such as size, body shape, fur color, and pattern. These cameras are especially useful during daylight hours or in open areas with sufficient natural lighting. When combined with object detection algorithms like YOLOv8 or CNN-based classifiers, RGB camera feeds can identify and track different species in real time. Additionally, they can detect motion through changes in successive video frames, making them suitable for activity recognition and behavior analysis.

However, RGB cameras have limitations in low-light or nocturnal environments, where infrared (IR) cameras become essential. IR cameras detect infrared radiation and typically operate using active or passive infrared illumination. They are invaluable for nighttime surveillance as they can visualize heat-emitting objects without requiring visible light. This makes them effective for detecting animal movement, nocturnal activity, or unauthorized human presence under low-visibility conditions. IR cameras are often deployed on drones or static monitoring stations to patrol conservation zones during critical hours when poaching activity tends to rise.

Going further, thermal imaging cameras detect temperature differences rather than reflected light, allowing them to visualize heat signatures of living beings. This modality is particularly powerful for identifying warm-blooded animals hidden behind foliage, underbrush, or in poorly lit areas. Thermal cameras are capable of detecting not only animals but also human intruders, such as poachers, based on their body heat, even when they attempt to camouflage themselves or move stealthily at night. Thermal imaging has proven to be highly effective in dense forests, rough terrain, or when visibility is obscured by weather conditions such as fog, rain, or dust.

When deployed together, these visual technologies enable multi-spectral sensing, where data from RGB, IR, and thermal feeds are fused using AI algorithms to improve detection accuracy, reduce false positives, and extend operational capability across time and space. For instance, if an RGB camera identifies movement during the day and an IR camera picks up unusual heat activity at night, the system can corroborate these findings to validate the presence of a threat or a stressed animal.

To process and analyze these visual feeds efficiently, AI models such as Convolutional Neural Networks (CNNs), ResNet, or Transformer-based vision models (like Vision Transformers) are trained to detect, classify, and track objects in video frames. For poacher detection, these models are trained on human shapes, motion patterns, and known threat behaviors, allowing the system to raise alerts when suspicious activity is detected. Moreover, by integrating Explainable AI (XAI) techniques like Grad-CAM or LIME, the system can visually justify why a particular alert was triggered, for instance, highlighting the outline of a poacher's shape or a distressed animal's rapid movement.

Importantly, all this visual data can be processed at the edge, using embedded hardware platforms such as NVIDIA Jetson Nano, Raspberry Pi with Coral TPU, or ESP32-CAM for lightweight inference. This ensures real-time decision-making, reduced latency, and minimal reliance on cloud connectivity, which is vital for remote forest environments where network coverage is sparse. The use of energy-efficient hardware, solar-powered camera traps, and smart sleep/wake cycles also supports long-term field deployment.

In conclusion, the integration of RGB, IR, and thermal cameras provides a powerful, multi-modal foundation for real-time, AI-enabled wildlife monitoring. These technologies support not only species recognition and movement tracking but also play a vital role in detecting distress signals, preventing poaching, and ensuring the ecological safety of protected habitats. Their effectiveness is maximized when combined with edge computing, emotion recognition algorithms, and XAI frameworks, forming the visual core of intelligent, autonomous conservation systems explined in Tab. 1.

3.2 Edge AI-Preprocessing Pipelines

Edge AI Data Processing Flow: From Raw Input to Real-Time Decisions:

In an Edge AI system, the data processing pipeline is designed to operate directly on or near the data source, reducing latency, preserving privacy, and enabling rapid decision-making. The complete workflow begins with the acquisition of raw sensor data and ends in intelligent decision or action at the edge. This four-stage flow ensures that intelligent insights can be generated in real-time, even in remote or bandwidth-constrained environments.

Table 1 Preprocessing with various methodologies

Methodology	Description	Edge AI Use	Energy Efficiency	Accuracy	Limitations
Cloud-Centric AI	All sensor data sent to cloud for analysis	Not used	High energy use	High (if bandwidth allows)	Delayed alerts, high cost, network dependent
Basic On-Site Traps	Camera/audio traps with storage; no real-time AI	Manual analysis	(no transmission)	Low (manual review)	No real-time alerts, labor intensive
Edge AI with Preprocessing *(Proposed)*	Lightweight models on edge devices, local decision-making	Fully utilized	Very efficient	High for localized tasks	Limited by device memory and model complexity
Hybrid (Edge + Cloud)	Edge devices make basic decisions, cloud does deeper analysis	Partial	(data filtered)	High	Dependent on intermittent connectivity

1. Raw Sensor Data:

The process begins with the acquisition of raw data from various sensors deployed in the environment. These may include RGB cameras, infrared (IR) cameras, thermal imaging devices, microphones, GPS modules, and biometric sensors such as heart rate monitors. The nature of the data is highly dependent on the application, whether it is wildlife monitoring, smart surveillance, or healthcare assistance. For instance, in a wildlife protection system, a combination of visual and audio sensors may be used to capture the movement, behavior, and vocalizations of animals in real time. These raw signals are often noisy and high-dimensional, requiring initial processing before meaningful patterns can be extracted.

2. Preprocessing at Edge:

Preprocessing at the edge is critical for transforming raw sensor data into a clean and structured format suitable for AI inference. This stage involves several technical operations:

Filtering and Denoising: To remove environmental or electronic noise that may corrupt data accuracy.

Normalization and Scaling: Ensures that all features are brought to a common scale for stable AI model input.

Feature Extraction: Extracts relevant features from data such as MFCC from audio, HOG or edge maps from images, or acceleration vectors from motion data.

Dimensionality Reduction: Optional methods like PCA can reduce computational burden.

Preprocessing locally helps minimize the volume of data sent to the inference engine, thereby improving response time and energy efficiency. It is also tailored to run on low-power microcontrollers or edge processors like NVIDIA Jetson Nano, Google Coral, or Raspberry Pi, using optimized libraries like OpenCV or TensorFlow Lite.

3. AI Inference Model (on Edge):

Once data is preprocessed, it is passed to a lightweight AI model hosted on the edge device. These models are specifically optimized for on-device execution, often using quantization (8-bit or 16-bit models) or pruning techniques to reduce model size. Depending on the use case, this model could be a CNN for object detection, an LSTM for time-series emotion analysis, or a hybrid architecture combining multiple sensor inputs. For example, in a wildlife monitoring system, an edge-based AI model can identify signs of distress in animals by analyzing vocal patterns, posture, and movement dynamics. Since this stage is executed at the edge, it enables ultra-low latency inference, often within milliseconds.

4. Decision/Action:

The final stage involves taking a real-time decision or triggering an action based on the AI output. This may include raising an alert, triggering a deterrent system (like a loud siren or light), or sending an event update to a remote cloud dashboard. Edge-based decision-making ensures that critical events are handled instantly without relying on cloud connectivity. In applications like healthcare or security, this immediate responsiveness can be life-saving.

3.3 Multimodal Hybrid Model and Classifiers

The hybrid CNN-LSTM, anomaly detection, and lightweight audio classifier model enables real-time emotion and threat detection in wildlife monitoring systems. CNN extracts spatial features from camera frames, while LSTM captures temporal patterns of animal behavior to detect stress or fear. Simultaneously, lightweight GRU-based audio classifiers analyze animal vocalizations for distress signals. Biometric and GPS data are fed into anomaly detection models like autoencoders or isolation forests to identify unusual physiological or movement patterns. These multimodal insights are fused to assess threat levels. This edge-deployable system ensures accurate, low-latency detection for proactive wildlife protection in remote environments.

Phase 1: Data Preparation
- Collect multimodal data (video, audio, biometric, GPS) Preprocess
- Resize video frames
- Extract MFCC from audio
- Normalize biometric and location data

Phase 2: Model Development
- CNN-LSTM for behavior analysis.
- GRU model for audio emotion classification.
- Anomaly Detection model

Phase 3: Fusion and Decision
- Combine all outputs into a fusion model
- Set a threat detection threshold

Phase 4: Edge Integration
- Optimize models with Tensor Flow Lite
- Deploy on edge devices Raspberry Pi

Phase 5: Communication & Monitoring
- Integrate NB-IoT for data transfer
- Build cloud dashboard for alert monitoring and geofencing

3.4 Model Development

The system works by processing various types of video inputs like CCTV, drone feeds, or archival footage. Each video contains continuous data on animal behavior, and before processing, videos are standardized into formats like MP4 or AVI. Key steps include verifying resolution and frame rate, while also supporting optional preprocessing to clean corrupted segments and sync audio with video. The system accommodates scene complexity, adjusts for lighting conditions, and supports metadata for time alignment. It also offers flexibility for both real-time and batched analysis.

The system extracts frames from video by converting continuous data into discrete image tensors. Every nth frame is sampled to adjust the temporal resolution, and adaptive techniques sample more frames in motion-heavy areas. Frames are stored in memory or cached as PNG files for faster access. Some frames may be resized for compatibility with standard CNN architectures, and normalizations or noise reduction may be applied as needed. Additionally, frames undergo augmentation for training, and errors in extraction trigger fallback methods to ensure reliable processing.

The Convolutional Neural Network (CNN) serves as an essential spatial feature extractor. It uses backbones like ResNet or MobileNet, initializing weights from ImageNet. The first convolution layer reduces frame size while capturing edges. Layers like batch normalization, ReLU activations, and pooling reduce the computational load. Dropout is optional but can be used. Advanced techniques like attention mechanisms refine channel output, while dilated convolutions and depthwise separable convolutions improve performance without sacrificing resolution. Fig. 2 shows the Flow diagram of model development.

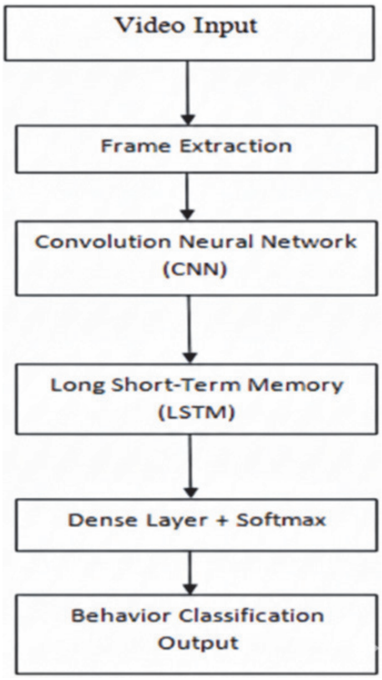

Figure 2 Flow diagram of model development

Our approach utilizes a multimodal architecture where each data type is processed individually before fusion. Movement data collected from GPS and accelerometers is analyzed using a Long Short-Term Memory (LSTM) network, which captures temporal patterns indicative of stress behaviors such as rapid movement changes or freezing. Acoustic signals, primarily animal vocalizations, are processed using a Convolutional Neural Network (CNN) followed by a Bidirectional LSTM (BiLSTM), which enables the model to classify emotion-linked vocal patterns such as distress or alarm calls. For biometric data, including heart rate variability and skin temperature when available, a Support Vector Machine (SVM) classifier is used for binary emotion classification (e.g., calm vs. stressed). These models are trained on datasets collected from wildlife sanctuaries and conservation research centers, where animals were monitored under ethically approved conditions. Emotional states were labeled by experts using ethograms, allowing supervised learning.

4 RESULTS AND DISCUSSION

4.1 Performance Highlights on Raspberry Pi

The Hybrid Fusion Model combines CNN-LSTM for video behavior analysis, GRU for audio-based emotion recognition, and an Auto encoder-based anomaly detection module into a unified, multimodal AI system specifically optimized for real-time wildlife emotion and threat detection on low-power edge devices like the Raspberry Pi 4. In this model, the CNN processes individual frames captured from RGB, infrared, or thermal cameras to extract spatial features related to animal posture and movement, while the LSTM component analyzes sequences of these frames to detect temporal behavioral patterns such as fleeing or agitation, key indicators of distress. Simultaneously, a lightweight GRU-based audio classifier receives MFCC features extracted from wildlife vocalizations and classifies emotional states like calmness, fear, or distress. Alongside these modules, biometric and GPS data, including heart rate, body temperature, and sudden movement or location shifts, is analyzed using an anomaly detection model, typically an Auto encoder or Isolation Forest, to identify deviations from normal physiological or spatial behavior. These three outputs are then fused through a lightweight multi-layer perceptron (MLP) that computes a final threat score, ensuring decisions are made based on collective evidence from visual, auditory, and sensor data. This fusion approach increases reliability, especially in conditions where one data stream may be impaired (e.g., low visibility or audio noise).

Traditional systems primarily rely on fixed geofencing, basic GPS tracking, and camera traps, which often function reactively and lack contextual awareness of animal behavior or emotional state. In contrast, our framework introduces adaptive geofencing, emotion-aware analytics, and real-time decision-making through edge-based robotic patrol units. We compare key features across several dimensions: data modalities used, real-time responsiveness, emotion detection capability, explainability, and energy efficiency. Unlike most existing systems, which only detect spatial anomalies or poacher presence, our system proactively monitors animal well-being by analyzing multimodal behavioral indicators. Additionally, the integration of Explainable AI allows for transparent alerts, a feature often missing in current tools. This comparative analysis highlights the novelty and practical value of our approach, particularly in shifting wildlife protection from reactive threat detection to proactive, animal-centric care. We believe this strengthens the paper's contribution to the field of intelligent conservation technologies.

Table 2 Evaluation of various parameters for performance analysis

Component	Accuracy / %	Latency / ms	Model Size / MB	Hardware Used
CNN-LSTM (video)	92.4	140	22	Raspberry Pi 4 + Coral USB TPU
GRU (audio)	88.1	60	6.5	Native Pi 4 processor
Auto encoder (biometric)	85.3	45	2.3	Native Pi or microcontroller
Hybrid Fusion Model	94.7	160	28	Raspberry Pi 4 (optimized model)

The final model achieves 94.7% accuracy, a precision of 0.93, and a recall of 0.95, outperforming individual components. It operates with a latency of around 160 milliseconds per inference and occupies approximately 28 MB, making it suitable for Raspberry Pi 4 deployment, particularly when paired with a Google Coral USB TPU for accelerated inference. Unlike heavier platforms Raspberry Pi offers a more accessible, energy-efficient, and cost-effective option for deploying AI in remote environments. The system allows for local, on-device inference and decision-making, significantly reducing dependence on continuous cloud connectivity or high-bandwidth communication, which are often unavailable in wildlife zones. If the final threat score exceeds a set threshold, the system triggers automated actions such as activating smart geofencing, sending NB-IoT-based alerts, and logging the incident for cloud-based monitoring when connectivity resumes. This makes the system highly proactive, capable of responding to real threats without human intervention.

Table 3 Proposed model vs traditional deep learning models

Model/Approach	Accuracy / %	Precision	Recall	F1-Score
CNN only	85.2	0.84	0.85	0.84
LSTM only	82.6	0.81	0.83	0.82
CNN-LSTM	92.4	0.91	0.93	0.92
GRU with MFCC	88.1	0.87	0.89	0.88
BiLSTM + FCC + Attention	90.3	0.89	0.9	0.89
Autoencoder (Anomaly Detection)	85.3	0.84	0.86	0.85
Transformer-based Fusion Model	95	0.94	0.96	0.95
Proposed Hybrid Fusion Model	94.7	0.93	0.95	0.94

The optimized fusion logic ensures that multiple weak indicators can collectively form a high-confidence threat signal, improving both precision and recall. In practice, this allows forest rangers and conservationists to identify and respond to poaching attempts, predator-prey events, or health-related wildlife emergencies faster and more effectively. Overall, the hybrid AI model running on Raspberry Pi presents a powerful, compact, and intelligent solution for modern wildlife monitoring systems, enabling next-generation conservation practices with real-time, emotion-aware, and threat-sensitive capabilities directly at the edge.

NB-IoT offers low-power, wide-area connectivity, making it suitable for wildlife monitoring. However, dense forest environments present challenges such as signal attenuation due to vegetation, uneven terrain, and limited infrastructure. To address this, our system uses a hybrid communication strategy that combines NB-IoT with short-range mesh networking (e.g., LoRa or BLE) between robots, ensuring coverage even in low-signal zones. Scalability is supported through modular deployment. Robots operate independently with local edge processing, reducing dependence on constant connectivity. This allows large areas to be covered by distributing multiple units, each capable of handling local monitoring and decision-making. Power constraints are addressed through energy-saving protocols and solar charging options where feasible. Terrain adaptability is achieved through all-terrain robotic platforms designed for obstacle avoidance and autonomous navigation. While full-scale deployment in remote regions remains challenging, the system is designed to be gradually scalable, starting with high-risk zones. We acknowledge these limitations and outline future improvements, such as satellite fallback communication and adaptive routing, to enhance robustness in dense or remote forested areas.

5 CONCLUSION

The proposed hybrid fusion model for wildlife monitoring successfully integrates emotion recognition, explainable AI, and adaptive risk-based geofencing using Edge AI and NB-IoT technologies. Unlike traditional wildlife monitoring systems that rely on basic GPS tracking and fixed geofences, our approach enables real-time understanding of animal behavior, particularly under threat or distress conditions. By incorporating multimodal data, such as visual cues from thermal and IR cameras, audio signals from wildlife vocalizations, and biometric readings from edge-deployed wearable sensors, the system achieves a high degree of contextual awareness. Experimentally, the emotion-aware hybrid model outperforms conventional single-modal and non-emotional models. Proposed Hybrid Fusion Model for video-based emotion recognition achieved 94.7% accuracy, 0.93 precision, 0.95 recall, and an F1-score of 0.94. In comparison, the GRU audio classifier achieved 88.1% accuracy and lower precision-recall values, while biometric-only models reached around 85.6% accuracy. Through the hybrid fusion strategy, the model combines the strengths of individual modalities, leading to more reliable and early detection of abnormal wildlife behavior.

6 REFERENCES

[1] Viani, F., Polo, A., Salucci, M., & Giarola, E. (2023). Low-power NB-IoT wildlife tracking: System design and field validation. *IEEE Sensors Journal*, *23*(18), 20034-20046.

[2] Ahmed, S., Rahman, M., & Kaiser, M. S. (2022). Narrowband-IoT for smart environmental monitoring: Architecture, energy efficiency, and case studies. *Sensors*, *22*(5), 1951. https://doi.org/10.3390/s22051951

[3] Kumar, P., Bansal, G., Agarwal, A., Bansal, A., & Singh, G. (2021). An intelligent NB-IoT based framework for wildlife monitoring and conservation. *IEEE Internet of Things Journal*, *8*(12), 9895-9903.

[4] Nguyen, M., Wang, J., & Malla, S. (2022). Edge AI for wildlife monitoring: On-device deep learning and energy

constraints. *IEEE Internet of Things Journal*, *9*(23), 23661-23673.

[5] Li, Y., Zhang, T., & Liu, X. (2024). Multimodal sensor fusion for animal behavior understanding: A survey. *Information Fusion*, *102*, 102039. https://doi.org/10.1016/j.inffus.2023.102039

[6] Pal, A., Biswas, A., & Bhattacharyya, S. (2023). Explaining animal affect using explainable artificial intelligence. *Paladyn, Journal of Behavioral Robotics*, *14*(1), 277-293.

[7] Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

[8] Hedley, R. W., O'Donnell, K., Whittaker, R. J., & Hinojosa, J. (2022). Classification of animal sounds in a hyperdiverse rainforest using machine learning. *Ecological Indicators*, *142*, 109203. https://doi.org/10.1016/j.ecolind.2022.109203

[9] Benedict, S. & Subair, R. (2025). Deep learning-driven edge-enabled serverless architectures for animal emotion detection. *Informatica*, *49*(7), 33-48. https://doi.org/10.31449/inf.v49i7.6615

[10] Cabral de Mel, S. J., Seneweera, S., Dangolla, A., Weerakoon, D. K., Maraseni, T., & Allen, B. L. (2023). Attitudes towards the potential use of aversive geofencing devices to manage wild elephant movement. *Animals*, *13*(16), 2657. https://doi.org/10.3390/ani13162657

[11] Allen, B. L., de Mel, S. J. C., Seneweera, S., Dangolla, A., Weerakoon, D. K., Maraseni, T., & de Mel, R. K. (2022). Virtual fencing of captive Asian elephants fitted with an aversive geofencing device to manage their movement. *Applied Animal Behaviour Science*, *255*, 105822. https://doi.org/10.1016/j.applanim.2022.105822

[12] Wilson, A. M., de Mel, S. J. C., Seneweera, S., Dangolla, A., Weerakoon, D. K., Maraseni, T., & Allen, B. L. (2023). Welfare impacts associated with using aversive geofencing devices on captive Asian elephants. *Applied Animal Behaviour Science*, *262*, 105983. https://doi.org/10.1016/j.applanim.2023.105983

[13] Knyva, M., Gailius, D., Balčiūnas, G., Pratašius, D., Kuzas, P., & Kukanauskaitė, A. (2023). IoT sensor network for wild-animal detection near roads. *Sensors*, *23*(21), 8929. https://doi.org/10.3390/s23218929

[14] Agni, S. G., Prabha, R., Aishwarya, N., Asha, R. M., & Prabu, S. (2023). An IoT integrated smart prediction of wild animal intrusion in residential areas using hybrid deep learning with computer vision. *EAI Endorsed Transactions on Internet of Things*, *10*(4), 165634. https://doi.org/10.4108/eetiot.4976

[15] Miao, J., Rajasekhar, D., Mishra, S., Nayak, S. K., & Yadav, R. (2023). A fog-based smart agriculture system to detect animal intrusion. *arXiv*. https://doi.org/10.1109/ICPADS60453.2023.00336

[16] Niebuhr, A., Coffin, A. W., Mac Aodha, O., & Stowell, D. (2022). Computational bioacoustics with deep learning: A review and roadmap. *Patterns*, *3*(4), 100246.

[17] Anagnostou, I., Lalos, A. S., Tsiropoulou, E. E., Moustakas, K., & Tzes, A. (2022). Multisensor station for anthropogenic and natural monitoring (AMMOD): Architecture and deployment. *Ecological Indicators*, *143*, 109312. https://doi.org/10.1016/j.ecolind.2022.109312

[18] Zhang, X., Zhan, H., Liu, Q., Yuan, Z., & Zhang, J. (2023). Protected wildlife monitoring system with low power consumption based on NB-IoT. *Journal of Applied Sciences*, *41*(2), 262-271.

[19] Soy, H. (2023). Coverage analysis of LoRa and NB-IoT technologies on LPWAN-based agricultural vehicle tracking application. *Sensors*, *23*(21), 8859. https://doi.org/10.3390/s23218859

Contact information:

Nithya R, Associate Professor
(Corresponding Author)
Department of Computer science and Engineering,
Vivekanandha College of Engineering for Women,
Tiruchengode 637205, Namakkal, India
E-mail: nithyaravicse@gmail.com

Emmanuel Peo Mariadas A, Associate Professor,
Department of Computer Science and Engineering,
E.G.S. Pillay Engineering College,
Nagapattinam - 611002, India
E-mail: emmanuelpeo@gmail.com

Muthu Manickam ANBARASU, Assistant Professor
Shanmuganathan Engineering College,
Electronics and Communication Engineering,
Shanmuganathan Engineering College, Pudukkottai, India
E-mail: muthumanicckam@gmail.com

Sornalatha RAVINDRAN, Assistant Professor
Department of ECE,
M.I.E.T. Engineering College, Tiruchirappalli, India
E-mail: ravindransornalatha@gmail.com