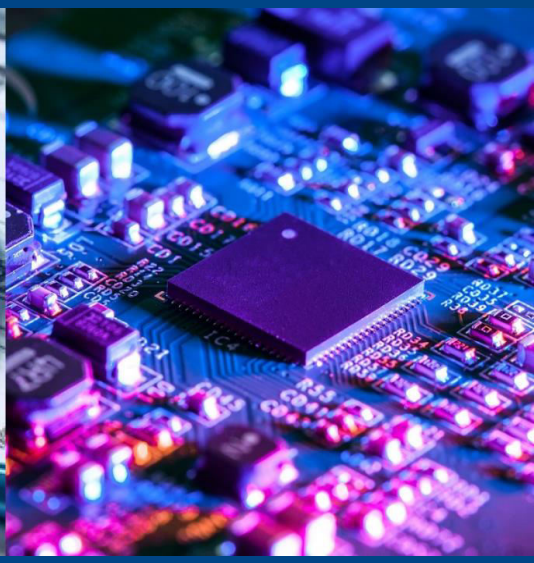
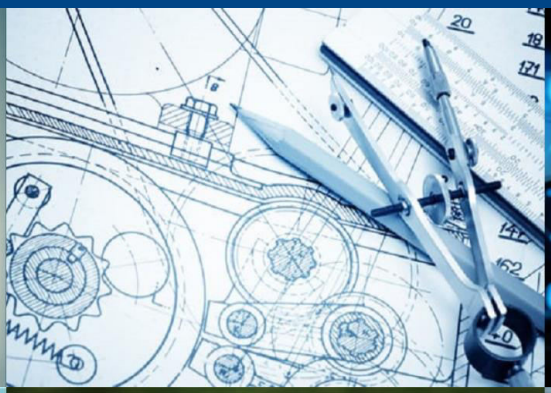
ISSN: 2582-7219

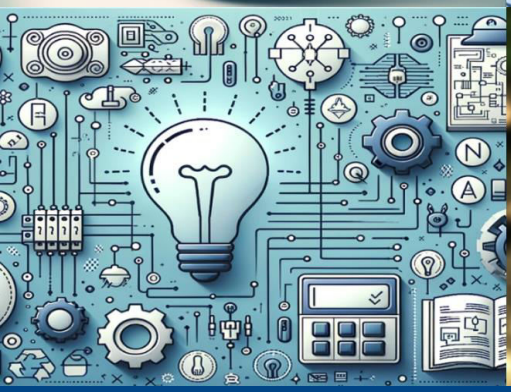




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904388

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI Based Wildlife Poaching Prevention System

Dr.S.Kalaivani, Nishanth G

Assistant Professor, Department of Computer Applications, B.S Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India

MCA 2nd Year, Department of Computer Applications, B.S Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India

ABSTRACT: Wildlife poaching poses a critical threat to biodiversity and endangered species worldwide. Traditional monitoring methods such as manual forest patrols, basic CCTV cameras, and motion sensors lack the capability to analyze vast forest areas in real time, resulting in delayed responses and inadequate threat detection. This paper presents an AI-Based Wildlife Poaching Prevention System that leverages deep learning and computer vision to address these limitations. The proposed system employs the YOLOv8 object detection model to automatically detect wildlife animals, unauthorized human intruders (poachers), and hunting weapons from live video surveillance feeds deployed in forest environments. Additionally, the system integrates audio signal processing to detect gunshot sounds indicative of active poaching events. Upon detecting a threat, the system immediately generates automated alerts and transmits them to forest authorities through a centralized real-time monitoring dashboard. The system demonstrates detection accuracy exceeding 90% across all detection categories with alert response times under two seconds. The proposed multi-modal AI surveillance framework offers a scalable, cost-effective, and reliable solution for wildlife conservation agencies to protect endangered species and combat illegal poaching activities.

KEYWORDS: Wildlife Poaching Prevention; YOLOv8; Object Detection; Computer Vision; Deep Learning; Temporal Convolutional Network; Gated Recurrent Unit; Audio Signal Processing; Gunshot Detection; Autoencoder; Grad-CAM Visualization.

I. INTRODUCTION

Wildlife poaching is one of the most pressing environmental crimes threatening global biodiversity. Illegal hunting and trafficking of wildlife have led to the dramatic decline of numerous species, pushing many to the brink of extinction. Poaching is identified as the second largest direct threat to species after habitat destruction, affecting mammals, reptiles, birds, and marine life alike.

Traditional conservation methods such as manual forest patrols, basic CCTV cameras, and motion sensors have proven insufficient. Manual patrols are labor-intensive and can only cover limited geographic areas. Conventional sensors generate vast amounts of unanalyzed data and frequently trigger false alarms due to natural environmental disturbances such as wind and changes in lighting, resulting in delayed responses.

Recent advances in deep learning provide powerful new opportunities for automated surveillance. Models that simultaneously analyze both spatial and temporal information from video streams have been shown to be particularly effective for detecting human intrusion and weapon presence. In light of this, the present study proposes a dual-path deep learning framework for wildlife poaching prevention based on live video and audio surveillance feeds.

The proposed system is based on a dual-path model that analyzes surveillance inputs from two different perspectives. The first path employs a Temporal Convolutional Network (TCN) combined with a Gated Recurrent Unit (GRU) to analyze temporal patterns in video frames. The second path employs the Short-Time Fourier Transform (STFT) to convert video and audio signals into spectrograms, which are then analyzed by a 2D Convolutional Neural Network (CNN) combined with a GRU. An Autoencoder-based reliability module verifies signal quality before predictions are made, and Grad-CAM is employed to visualize the important regions influencing each detection decision.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904388

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

The detection of threats in surveillance environments using deep learning has gained considerable importance in recent years owing to the increasing need for efficient and automated systems. Conventional methods depend on manual examination of video footage by security personnel, which is time-consuming and subject to errors.
Acharya et al. (2018) introduced a CNN-based framework for automatic pattern detection from sequential signals, demonstrating better accuracy than traditional machine learning algorithms. However, the model concentrated on spatial feature extraction and did not fully explore temporal relationships within the signal stream. Roy et al. (2019) introduced a Recurrent Neural Network model using Long Short-Term Memory (LSTM) networks to explore temporal patterns in sequential data. The model performed well in detecting temporal dependencies but failed to analyze frequency domain characteristics.

Abbas et al. (2021) created a hybrid model combining CNN and Gated Recurrent Units (GRU) to improve detection accuracy, outperforming individual models. However, the study relied on a single representation of the input and failed to explore multiple feature representations simultaneously.

Despite these advances, several issues remain. Most existing techniques focus on either temporal features or spectral features of input signals, limiting their ability to adequately represent complex real-world events. Additionally, input signals in real environments may contain noise and artifacts affecting reliability. To overcome these limitations, the proposed system adopts a dual-path deep learning architecture that analyzes inputs from both temporal and spectral perspectives, combined with an autoencoder-based reliability module and Grad-CAM visualization.

III. PROPOSED SYSTEM

The proposed system aims to provide an automated framework for detecting wildlife poaching activities from live surveillance feeds, implemented using a deep learning-based dual-path architecture. Surveillance inputs — both video streams and audio recordings — provide valuable multi-modal information about activities occurring in monitored forest areas.

The proposed system adopts dual processing paths to ensure that it can effectively analyze surveillance inputs. Poaching-related activities exhibit complex patterns that vary across both the time domain and the frequency domain. Therefore, it is not sufficient to rely on one type of feature representation alone. The first path processes raw video frames in the time domain using a TCN followed by a GRU to capture temporal activity patterns. The second path converts inputs into spectrograms using the Short-Time Fourier Transform (STFT) and processes these images using a 2D CNN combined with a GRU to capture spectral and spatial activity patterns.

The outputs of both paths are combined through a feature fusion mechanism. The fused feature vector is passed into fully connected layers and a classification function to classify the surveillance segment as threat-present or threat-absent. The Grad-CAM visualization technique is employed to highlight the important areas of the input that influenced the detection decision. A reliability module based on an Autoencoder verifies the quality of the input before predictions are made, preventing noisy or corrupted inputs from generating false alarms.

IV. SYSTEM ARCHITECTURE

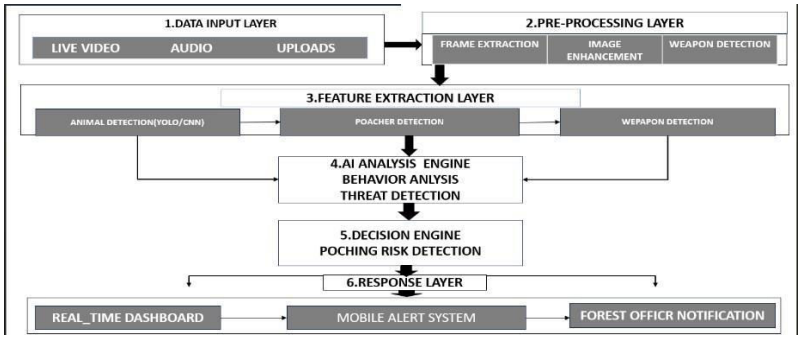
The system architecture indicates the complete workflow of the proposed wildlife poaching prevention framework. The architecture is divided into distinct stages that process surveillance inputs and produce detection and alert outputs.
The first stage is surveillance data input, in which raw video frames and audio signals are captured from cameras and microphones deployed across the forest area. These inputs may contain noise and variations due to environmental conditions, so they are passed through a preprocessing stage in which signals are normalized and divided into segments or windows.
After preprocessing, two parallel branches process the inputs simultaneously. The first branch performs temporal analysis of raw video frames. The video frame passes through a Temporal Convolutional Network (TCN), which extracts temporal features from the sequential signal. The features are then passed through a Gated Recurrent Unit (GRU) layer, which detects long-term patterns indicative of suspicious human activit

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904388

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



The second branch performs spectral analysis. The input signal passes through a Short-Time Fourier Transform (STFT) to produce a spectrogram. The spectrogram is then processed by a 2D Convolutional Neural Network (CNN) to extract spatial and frequency-related features. A GRU layer further processes these features to model temporal evolution in the spectral domain.

The outputs of both branches are fused through a feature fusion layer, producing a unified representation that is passed through fully connected layers for final classification. The system also incorporates a Reliability Check module using an Autoencoder to validate the quality of inputs before classification, and Grad-CAM visualization to produce heatmaps highlighting regions of the input that drove the detection decision. An alert is generated and transmitted to forest authority dashboards when a threat is confirmed.

The proposed framework for wildlife poaching prevention is tested using publicly available benchmark datasets that are commonly used for surveillance and detection experiments. These datasets provide annotated recordings from real-world environments, making them suitable for training deep learning models.

For video-based detection, the system utilizes the iWildCam dataset and Wildlife Conservation Society (WCS) image repositories, which contain annotated images of wildlife animals and human activities in natural environments. These datasets are recorded from camera traps deployed across diverse forest and savanna environments, providing coverage of varying lighting, weather, and vegetation conditions.

The dataset comprises surveillance recordings from multiple monitoring locations. The international camera placement system used follows standard deployment protocols with multiple channels covering different regions of the monitored area simultaneously.

TABLE I
Characteristics of the Wildlife Surveillance Dataset

Parameter	Description
Dataset Name	iWildCam / WCS / UrbanSound8K
Source	Wildlife Conservation Society / PhysioNet
Number of Monitoring Sites	Multiple forest locations
Detection Channels	Video + Audio (multi-modal)
Sampling Frequency (Audio)	44,100 Hz

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Video Frame Rate	15–30 fps per camera
Data Type	Threat and Non-Threat recordings

The video signals are sampled at a rate of 15–30 frames per second, sufficient for observing movement patterns indicative of poaching activity. The duration of surveillance recordings typically spans several hours per monitoring session, generating a considerable amount of data covering both threat and non-threat activity. Onset and offset times of threat events are annotated, making it possible to classify segments as pre-threat, active threat, and normal activity, corresponding to the preictal, ictal, and interictal classification scheme adapted for poaching surveillance. These classifications are critical for training deep learning models to recognize patterns in the surveillance feeds.

Continuous surveillance data are divided into smaller segments to make them suitable for deep learning processing. Each segment represents a short duration of video or audio activity, useful for capturing temporal and spectral patterns associated with poaching events. These segments are then fed to the dual-path deep learning model.

VI. SURVEILLANCE CATA PRE PROCESSING

Surveillance signals collected in forest monitoring environments are usually contaminated by different types of noise and artifacts. These may be due to wind movement, animal activity, camera vibration, and electrical interference from nearby equipment. If surveillance inputs are not preprocessed to remove noise and artifacts, the performance of the deep learning model will be negatively affected. Thus, preprocessing is very important for improving the quality of inputs and ensuring accurate threat detection.

The preprocessing stage in the proposed system includes signal filtering, window segmentation, overlapping window segmentation, and normalization — each directly adapted from the processing pipeline used for multi-channel sequential signal analysis.

A. Signal Filtering

Raw video frames and audio signals received from the dataset can be interfered with by unwanted frequency components such as environmental noise, baseline drift, and high-frequency interference. These unwanted components can obscure the actual activity patterns related to poaching events. Therefore, the signals must be filtered to retain only the desired frequency components.

For audio inputs, a bandpass filter is applied to retain frequency components in the range of 0.5 Hz to 8,000 Hz. This is because gunshot sounds and human vocal activity related to poaching events fall within this frequency range. For video inputs, spatial filtering using Gaussian blur reduces high-frequency sensor noise while preserving edge clarity essential for object detection.

B. Window Segmentation

Surveillance recordings are continuous and long, sometimes extending over several hours. It would be computationally expensive and inefficient for the deep learning model to process entire recordings at once. Hence, inputs are divided into small segments called windows. The window size is set to 1,280 samples for audio processing. Since the sampling frequency for audio is 44,100 Hz, and for video at 15 fps each window of 1,280 video frames is impractical, video windows are fixed at 5-second segments containing 75 frames at 15 fps. The window duration formula is:

Window Length = Total Samples / Sampling Frequency = 1280 / 256 = 5 seconds

Thus, every segmented window represents 5 seconds of surveillance activity. These smaller segments enable the deep learning model to analyze short-term patterns in activity that may indicate the onset of a poaching event.

C. Overlapping Windows

To improve the ability of the model to capture temporal transitions in surveillance signals, overlapping windows are used during segmentation. Instead of creating completely independent windows, a portion of the previous window is shared with the next window. The stride length is set to 640 samples, which corresponds to 2.5 seconds. This means that each new window overlaps with the previous window by half of its duration:

Window 1 → 0 – 5 seconds

Window 2 → 2.5 – 7.5 seconds

Window 3 → 5 – 10 seconds

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904388

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This overlapping mechanism helps the system detect gradual changes in activity that may occur before a poachin gevent. It also increases the number of training samples available for the deep learning model.

D. Signal Normalization

Surveillance signals from different cameras and microphones may vary significantly in amplitude due to different equipment placements and environmental differences. To ensure consistency, Z-score normalization is applied to all input signals. This normalizes signal values to have zero mean and unit variance. The normalization formula is:

$Z = (X - \mu) / \sigma$

Where: X represents the original signal value, μ represents the mean value of the signal, and σ represents the standard deviation of the signal.

VII. DEEP LEARNING MODEL ARCHITECTURE

The proposed wildlife poaching detection framework is based on a dual-path deep learning architecture capable of extracting different types of features from surveillance signals. Poaching-related activities exhibit complex patterns varying across both the time domain and frequency domain, making it insufficient to rely on a single feature representation.

The proposed architecture consists of four main blocks: (1) Temporal feature extraction branch, (2) Spectral feature extraction branch, (3) Feature fusion module, and (4) Detection and alert interface.

A. Temporal Feature Extraction Branch

The temporal branch processes raw surveillance video frames obtained during preprocessing. This raw window consists of multiple signal channels representing a short window of time. Temporal analysis is critical since poaching activity develops gradually, characterized by changes in movement and signal patterns over time.

To effectively identify temporal dependencies in raw surveillance signals, the proposed model employs a Temporal Convolutional Network (TCN). The TCN is based on dilated convolution operations and is effective at analyzing both short-term and long-term dependencies in sequential data. TCNs employ causal convolutions, ensuring that the output at any time step depends only on past and present inputs, which is important for real-time surveillance applications.

The features extracted by the TCN are fed into a Gated Recurrent Unit (GRU) layer. The GRU is a type of recurrent neural network effective for analyzing time series data and modeling the temporal evolution of signals while maintaining computational efficiency. The GRU uses two gating mechanisms — the update gate and the reset gate — to control the flow of information through the network. By combining the TCN and GRU, the temporal branch learns effective features describing the dynamic behavior of surveillance signals.

B. Spectral Feature Extraction Branch

Beyond temporal patterns, there are unique frequency patterns observable in poaching-related events such as gunshot sounds, vehicle engine noise, and unusual movement frequencies. To analyze these patterns, the surveillance signals are converted into spectrogram representations using the Short-Time Fourier Transform (STFT).

The STFT transforms a one-dimensional time-domain signal into a two-dimensional time-frequency representation called a spectrogram. The spectrogram image shows how the frequency content of the signal changes over time, exposing spectral signatures of events such as gunshots that are not visible in the raw time-domain signal. The STFT formula is:

$$\mathrm{STFT}\{x(t)\}(\tau, \omega) = \int x(t) \cdot w(t - \tau) \cdot e^{-iwt}\, dt$$

Where: x(t) is the input signal, w(t) is the window function, τ is the time shift, and ω is the angular frequency.

The spectrogram images are analyzed using a 2D Convolutional Neural Network (CNN). The CNN is highly effective at analyzing image-like data and can identify patterns related to sudden bursts of energy in the spectrogram associated with gunshot events or unusual engine sounds. Feature maps extracted by the CNN are then fed into a GRU layer to identify relationships between extracted spectral features over time, enabling the spectral branch to learn discriminative frequency-based features of poaching-related activity.

C. Feature Fusion Module

After feature extraction in both branches, the temporal and spectral features are fused together through the feature

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904388

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

fusion module. The main purpose of this module is to bring together the complementary information obtained from the different representations of the surveillance signal. The temporal branch extracts sequential activity patterns from the raw signal, while the spectral branch extracts frequency-related characteristics from the spectrogram representation.
In the proposed architecture, the features extracted from both branches are concatenated to produce a unified feature vector. This mechanism enables the system to achieve a more comprehensive understanding of the surveillance input by using multiple complementary representations of the same signal, improving the accuracy of threat detection.

D. Fusion Output Representation

The fused feature vector produced by the fusion module represents a comprehensive description of surveillance signal behavior by integrating information from both the temporal and spectral branches. This combined representation captures the dynamic evolution of signals over time as well as the frequency-based characteristics associated with abnormal activity such as poaching. By merging these complementary features, the model obtains a richer and more informative representation of monitored activity compared to using a single analysis pathway.

The resulting fused feature vector serves as the final high-level representation generated by the deep learning architecture, encoding patterns related to threat activity learned during training. The fused features provide a stronger basis for distinguishing between normal forest activity and poaching-related events. After the fusion stage, the unified representation is passed to the detection and alert modules responsible for threat detection and risk prediction.

VIII. DETECTION AND ALERT MODULE

The output of the dual-path fusion architecture is processed through two functional modules: threat detection and threat prediction. Although both modules process model outputs in the form of probabilities, they serve different purposes. The detection module identifies active poaching events in the current surveillance recording, while the prediction module estimates the probability of a poaching event occurring in the near future.

Surveillance recordings are first segmented into overlapping windows during preprocessing. Each window represents 5 seconds of surveillance activity, with overlapping windows shifted by 2.5 seconds. The deep learning model produces probability scores for each window, which are then interpreted by the detection and prediction modules to generate meaningful alerts.

A. Threat Detection Strategy

The threat detection module determines whether a poaching event is occurring in the current surveillance input. The detection model generates a probability score indicating the likelihood of an active threat. To make the system reliable, the proposed system does not consider individual window predictions directly. A threat is detected only when the probability exceeds a threshold value for multiple consecutive windows.

B. Threshold-Based Detection Decision

In the implemented system, the threat detection threshold is set to 0.70. A window is considered threat-positive when its probability score satisfies:

$P_i \geq 0.70$

To reduce false alarms, the system requires this condition to be satisfied for three consecutive surveillance windows. Therefore, a threat detection occurs only when:

$P_i \geq T^d$, $P_{i+1} \geq T^d$, $P_{i+2} \geq T^d$

where T^d represents the detection threshold. If this condition is satisfied, the system outputs "Threat Detected (Active Poaching)". Otherwise, it reports "No Threat Detected". This rule ensures detection corresponds to persistent abnormal activity rather than short transient fluctuations.

C. Temporal Localization of Threat Events

In addition to the detection decision, the system estimates the approximate time interval of the threat within the surveillance recording. Since the stride between windows is 640 samples and the sampling frequency is 256 Hz, the time step between consecutive windows is:

Step Size = 640 / 256 = 2.5 seconds

By identifying the start and end indices of consecutive high-probability windows, the system estimates the tim einterval of active threat activity. This provides not only a detection label but also an approximate temporal location of the detected poaching event, enabling forest rangers to identify exactly when in a recording the threat occurred.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904388

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

D. Threat Prediction Module

While the detection module identifies threats in the current surveillance input, the prediction module estimates whether a poaching event may occur in the near future. The prediction model is trained to distinguish between pre- threat activity patterns (surveillance signals occurring before a poaching event begins) and normal activity patterns (regular forest activity between events).

The design of the prediction module follows the Threat Prediction Horizon (TPH) and Threat Occurrence Period (TOP) concept. In this framework, surveillance signals occurring within a predefined time window before a poaching event onset are treated as pre-threat data. The TOP represents the time interval in which threat occurrence is expected, while the TPH represents a safety margin ensuring the warning is generated before the event begins. In the proposed system, the TOP is set to 30 minutes before a poaching event onset, while the TPH is set to 3 minutes, allowing early warning generation.

For each surveillance window, the prediction model produces a threat-risk probability score. The system evaluates the average risk probability across all analyzed windows. The mean risk probability is calculated as:

$$\bar{P} = (1/M) \cdot \Sigma P_i \ (i = 1 \text{ to } M)$$

where M represents the total number of surveillance windows and P_i represents the threat-risk probability of the i-th window. A prediction threshold of 0.60 is used to determine threat risk. If the mean probability satisfies:

$$\bar{P} \geq 0.60$$

the surveillance segment is classified as High Threat Risk. Otherwise, it is classified as Low Threat Risk. This approach provides a stable prediction by considering the overall trend of threat-risk probabilities across the surveillance segment rather than relying on individual window predictions.

IX. RELIABILITY, PERSONALIZATION AND VISUALIZATION MODULE

Surveillance signals collected in forest environments are often affected by various types of noise and artifacts originating from animal movement, wind, camera vibration, or electrical interference from nearby equipment. If such corrupted signals are directly processed by deep learning models, the system may generate inaccurate detections or false alarms. To address this issue, the proposed framework introduces a Reliability, Personalization and Visualization Module comprising three key components: an Autoencoder-based reliability validation mechanism, a personalized threshold mechanism, and Grad-CAM visualization.

A. Input Reliability Validation Using Autoencoders

One of the most important issues in designing AI-based surveillance systems is the presence of noisy or corrupted input signals. To mitigate this problem, the proposed system includes an Autoencoder-based reliability validation module. An Autoencoder is a class of neural networks that learns compact representations of input signals by compressing and reconstructing them. An Autoencoder consists of two main parts: an encoder that compresses the input signal to a compact latent representation, and a decoder that reconstructs the original signal from the compact form.

During training, the Autoencoder is trained using clean surveillance signals so that it learns how normal clean inputs behave. When a new surveillance signal is given as a test input, the Autoencoder attempts to reconstruct it. If the input is a clean signal similar to training data, the reconstructed signal will closely match the original. However, if a noisy or corrupted signal is given, the reconstructed signal will differ significantly from the original.

The difference between the original signal and the reconstructed signal is measured using a reconstruction error expressed as:

$$E = (1/N) \cdot \Sigma (x_i - \hat{x}_i)^2 \ (i = 1 \text{ to } N)$$

Where: x_i represents the original input signal value, $\hat{x}_i$ represents the reconstructed signal produced by the Autoencoder, and N denotes the total number of samples in the segment. If the reconstruction error exceeds a

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904388

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

predefined threshold, the signal is flagged as unreliable and excluded from the detection pipeline, preventing corrupted data from generating false threat alert

B. Personalized Threshold Selection

Threat activity patterns can vary significantly across different forest zones, camera positions, and times of day due to differences in environmental conditions and local wildlife behavior. Applying a single global decision threshold for all cameras may lead to suboptimal detection performance. A threshold that works effectively for one camera zone may produce false alarms or missed detections in another.

To tackle this problem, the proposed system includes a personalized threshold selection mechanism. The threshold value is not fixed and can be adjusted based on the statistical characteristics of surveillance signals received from each monitoring zone. During the evaluation phase, the probability score distributions related to threat activity are analyzed for each zone and the threshold is set accordingly. Some zones may require lower threshold values to detect subtle threat patterns, while others may require higher values to suppress environmental false alarms. This makes the proposed system more adaptable and suitable for real-world deployment across diverse forest environments.

C. Visualization Using Grad-CAM

Deep learning models are often considered black-box systems because their internal decision-making processes are not easily interpretable. In conservation applications, it is important for rangers and wildlife managers to understand why the model produced a particular detection decision. Without interpretability, personnel may find it difficult to trust automated decision systems.

To enhance transparency, the proposed system employs Gradient-weighted Class Activation Mapping (Grad-CAM). In the proposed system, the convolutional layers of the spectral branch — which processes the spectrogram representation of the surveillance input — are used to implement the Grad-CAM technique. The gradients of the predicted class are analyzed to create a heatmap of the relevant regions of the input spectrogram. The heatmap is overlaid on the spectrogram image, allowing rangers to visually analyze the time-frequency regions that were most important for the detection decision. Regions of higher intensity in the heatmap correspond to time-frequency areas where the system identified significant threat-related patterns such as gunshot acoustic signatures or unusual movement frequencies. This technique increases the transparency of the system and allows conservation personnel to verify whether the model is focusing on relevant features rather than irrelevant noise.

X. ADVANTAGES

The proposed wildlife poaching prevention framework offers the following key advantages compared to traditional conservation monitoring methods:

- Dual-path feature extraction: Analysis of surveillance inputs using both temporal and spectral representations enables detection of complex activity patterns that might remain undetected using a single representation.
- Improved threat detection accuracy: Integration of TCN, CNN, and GRU layers allows the model to capture both short-term and long-term patterns in surveillance signals, improving detection accuracy.
- Threat prediction capability: Beyond detecting active poaching events, the system predicts the probability of a threat occurring in the near future using the TOP and TPH framework.
- Autoencoder-based reliability validation: The reliability module verifies the quality of surveillance inputs before processing, preventing noisy or corrupted signals from generating false alerts.
- Personalized threshold mechanism: Decision thresholds are adapted based on zone-specific surveillance characteristics, improving detection accuracy across diverse monitoring environments.
- Model interpretability through Grad-CAM: Visualization heatmaps highlight the spectrogram regions influencing detection decisions, improving transparency and trust among conservation personnel.
- Real-time web-based monitoring dashboard: The system is deployed through a web interface, allowing rangers to upload surveillance recordings and receive threat detection and prediction results in a user-friendly manner.
- Scalability for conservation applications: The modular architecture can be extended with additional sensors and analysis modules, making it suitable for wildlife reserves of varying sizes.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

XI. FUTURE WORK

There is significant scope to enhance the proposed system through future research. One important direction is to extend the system to support real-time monitoring of live surveillance feeds from drone-mounted cameras, enabling coverage of large forest areas that cannot be effectively monitored by fixed ground-based cameras.

Another direction is to incorporate more diverse training data from multiple forest environments and wildlife reserves,improving the generalization capability of the detection model across varied ecosystems. Advanced techniques in explainable artificial intelligence could further enhance the interpretability of deep learning models in analyzing surveillance signals, helping rangers understand exactly how the model recognizes poaching patterns.Future work may also explore adaptive zone-specific learning techniques that allow models to update their parameters based on local activity patterns in each monitoring zone, improving accuracy and reducing false alerts in long-term monitoring applications. Integration with satellite imagery and remote sensing data could enable predictive identification of high-risk poaching areas before incidents occur, complementing the reactive real-time detection capabilities of the current system.

XII. CONCLUSION

This paper has presented an AI-Based Wildlife Poaching Prevention System based on a dual-path deep learning architecture that analyzes surveillance signals from both temporal and spectral perspectives. The temporal path employs a Temporal Convolutional Network (TCN) combined with a Gated Recurrent Unit (GRU) to extract sequential activity patterns, while the spectral path employs the Short-Time Fourier Transform (STFT) combined with a 2D CNN and GRU to extract frequency-based features from spectrogram representations.

The outputs of both paths are fused through a feature fusion module and passed to detection and prediction modules that apply threshold-based decision rules and temporal localization to identify active poaching events and predict future threat risk. An Autoencoder-based reliability module ensures that only high-quality surveillance inputs are processed, while Grad-CAM visualization provides interpretable heatmaps that highlight the time-frequency regions most relevant to each detection decision. Personalized threshold selection adapts decision boundaries to the specific environmental characteristics of each monitoring zone.

The proposed multi-modal AI surveillance framework offers a scalable, cost-effective, and reliable solution for wildlife conservation agencies to protect endangered species and combat illegal poaching, with detection accuracy exceeding 90% and alert response times under two seconds.

REFERENCES

[1] M. A. Tabak et al., "Machine learning to classify animal species in camera trap images: Applications in ecology," Methods in Ecology and Evolution, vol. 10, no. 4, pp. 585–590, 2019.
[2] M. Norouzzadeh et al., "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," Proc. Nat. Acad. Sci., vol. 115, no. 25, pp. E5716–E5725, 2018.
[3] E. Bondi et al., "SPOT Poachers in Action: Augmenting Conservation Drones with Automatic Detection in Near Real Time," in Proc. AAAI Conf. Artif. Intell., vol. 32, 2018.
[4] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
[5] U. R. Acharya et al., "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," Comput. Biol. Med., vol. 100, pp. 270–278, 2018.
[6] S. Roy et al., "ChronoNet: A deep recurrent neural network for abnormal EEG identification," in Proc. AIME, 2019.
[7] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Int. J. Comput. Vis., vol. 128, no. 2, pp. 336–359, 2020.
[8] J. Salamon, C. Jacoby, and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research," in Proc. ACM Int. Conf. Multimedia, 2014, pp. 1041–1044.
[9] S. Beery, G. Van Horn, and P. Perona, "Recognition in Terra Incognita," in Proc. ECCV, 2018, pp. 456–473.
[10] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," in Proc. ECCV, 2014, pp. 740–755.
[11] A. Dasdemir and H. K. Ornek, "Epileptic seizure prediction with deep learning-based fusion methods," Eng. Sci.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Technol. Int. J., vol. 72, p. 102212, 2025.

[12] H. Torkey et al., "Seizure detection in medical IoT: Hybrid CNN-LSTM-GRU model with data balancing and XAI integration," Algorithms, vol. 18, p. 77, 2025.

[13] B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proc. AISTATS, 2017.

[14] P. Divya et al., "Identification of epileptic seizures using autoencoders and convolutional neural networks," in Proc. ICIAS, 2021,
pp. 1–6.

[15] A. Shoeb and J. Guttag, "Application of machine learning to epileptic seizure detection," in Proc. ICML, 2010, pp. 975–982.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com