ZoionNET: Low-Power Animal Intrusion Detection in Methagiri Forest Using Acoustic-based Deep Learning Classifier for Cellular Remote Sensing System

PROJECT TITLE: ZoionNET: Low-Power Animal Intrusion Detection in Methagiri Forest Using Acoustic-based Deep Learning Classifier for Cellular Remote Sensing System

1) EXECUTIVE SUMMARY:

1.1) PROBLEM STATEMENT:

Forest Methagiri, Hosur, Tamil Nadu has huge forest divisions and wildlife connected with the human habitat in various aspects. This area is susceptible since it acts as a median line for the forest and village, causing huge locomotion of animals into the village causing disputes in the daily lifestyle of the human habitants.

Wildlife Intrusion into Villages - “Leopard strays into the village, Hosur” was one of the headlines in the news, on Jan 12, 2021, 20:13 IST, the Leopard entered the village killing 23 stray dogs, cattle goats, and cows, attacking 5 people and 37 elephants had damaged the crops. The intrusion was attributed to a lack of detection and monitoring systems for wildlife, leading to human-wildlife conflicts.

Elephant is the flagship species of Melagiris. Migration and crop raiding, mostly triggered by habitat degradation are showing varying patterns of late, resulting in a substantial increase in human-elephant conflict in the area. Every year hundreds of farmers lose their crops and sometimes even their lives to wild elephants around Hosur. As a consequence, these elephants often die unnatural deaths due to shooting, accidents, and electrocution. Keeping wild elephants away from areas of human activity is essential to its conservation.

Previously, back in 2011, the project’s Community Partner, Kenneth Anderson Nature Society NGO, tried various IoT-based sensing techniques to mitigate the intrusion of animals. However, due to the high computational and power supply demands of the IoT system, their trial failed. This occurred due to the video classification-based deep learning model running on the IoT system.

1.2) OBJECTIVES AND SOLUTION:

1.2.1) Acoustic and seismic sensors deployment in the forest to capture animal vocalizations and movement data.

1.2.2) Deep learning models trained on this data to accurately classify different animal species.

1.2.3) Low-power embedded systems to run the classifiers and transmit alerts via cellular networks.

1.2.4) Cloud-based dashboard for real-time monitoring and historical analysis.

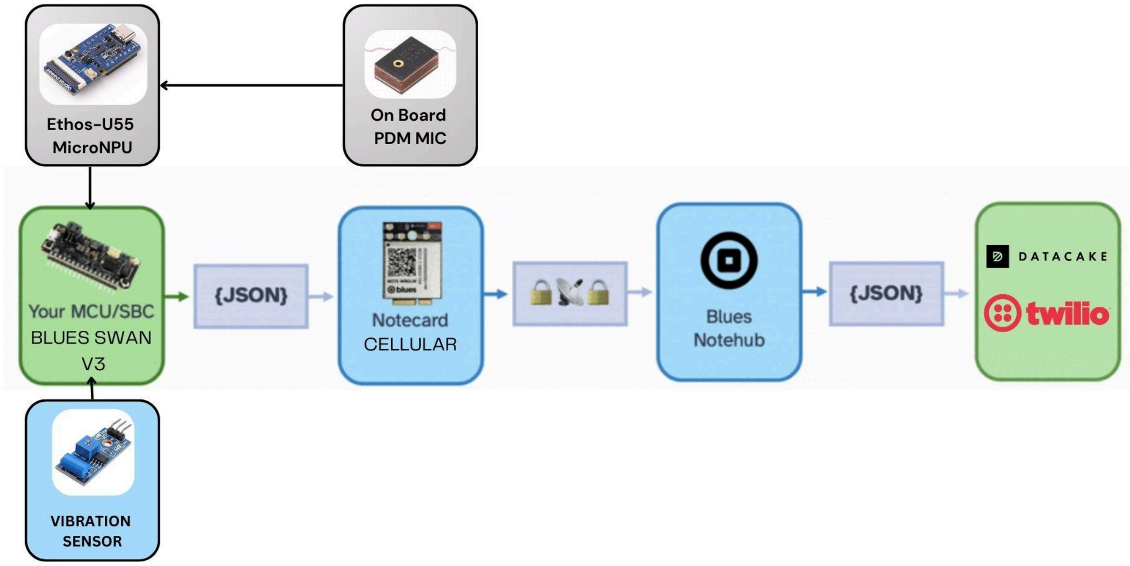
2) GENERAL DESCRIPTION

2.1) PROBLEM BEING ADDRESSED

This project aims to address this challenge by deploying a cellular system with acoustic and seismic sensors in the median area of Methagiri Forest and the village. This area has great coverage of cellular network which makes it suitable for deployment. These sensors will continuously monitor for animal vocalizations and movement patterns, with the data fed into deep-learning models running on low-power embedded systems, a dedicated Neural Processing Unit (NPU) in the cellular system. The classifiers will be trained to identify different animal species with high accuracy, allowing the system to detect intrusions and send real-time alerts via cellular networks. Primarily the project classifies certain animals like elephants, cheetahs, and monkeys since these are the ones that cause 90% of the concerns in the area. The approach aims to enhance wildlife monitoring and mitigate human-wildlife conflicts through real-time detection and alerts.

A cloud-based dashboard will provide forest rangers and local authorities with a unified view of sensor data and alerts. This will enable them to respond quickly to incidents, deploy patrols, and implement mitigation strategies. The system will also collect long-term data on animal behavior and movement patterns, which can inform habitat conservation efforts.

3) TECHNICAL SOLUTION



3.1) WORKING OF THE ZOION-NET CELLULAR SYSTEM

The Blues Swan V3 is an STM32-based microcontroller that acts as the core module of the IoT system and helps in controlling the operations in the embedded system. The MCU is connected to a mechanical vibration sensor and a MicroNPU (Neural Processing Unit). The MicroNPU which is the Seeed Studio Groove Vision V2 module, containing the Ethos U55 NPU chip runs the audio classification Deep Learning Model.

The onboard PDM Microphone module provides the time-domain audio waveforms, representing the amplitude of sound over time, as an input to the MicroNPU. The MicroNPU then converts them into a spectrogram, which is a frequency-domain representation of audio signals.

The conversion process works by sampling the audio at 8 kHz. The audio waveform is divided into discrete samples taken 8,000 times per second. The Real-time continuous audio signals will be fed to the microNPU, where the technique of sliding window is used to make those discrete samples and pass them to the classifier model. The classifier AI model leverages the dual neural network system consisting of active and dormant states for detecting and classifying the animals, providing a low power- high accuracy intelligent system, since most of the time the NPU will be in a sleep state due to the active state of detection model which activates the entire system upon anomaly.

The Short-Time Fourier Transform (STFT) is applied to each window converting the time-domain data (audio signals) into frequency components, generating the spectrograms. This resulting spectrogram, which will be a 2D representation with time on one axis and frequency on the other, is then fed to the Ethos U55 MicroNPU. The sliding window technique captures both short-term fluctuations and long-term trends in the audio signal, ensuring a more accurate classification of animal sounds.

After the class of animal detected using its respective animal sound, the model expels an interference containing the confidence score of detection and class of animal detected to the Blues swan V3 MCU via I2C interface through TX RX pins. Then the firmware code running in the MCU will analyze these interferences and determine whether to initiate a payload to the cloud system based on the prescribed thresholds.

Upon successful JSON payloads to the Blues Notehub proxy server, these data will be re-routed to the end cloud applications such as Datacake and Twilio for further operations.

3.2) CNN-BASED DEEP LEARNING AUDIO CLASSIFIER FOR EMBEDDED SYSTEM

State-of-the-art Convolutional Neural Networks (CNNs) were fine-tuned on the mel-spectrogram data extracted from the sounds of various animals. These fine-tuned CNNs were used for both the detection and further classification of the detected animals.

To enable real-time inference on low-power embedded systems, the models were optimized using quantization. This significantly reduced their computational and memory footprint without sacrificing accuracy.

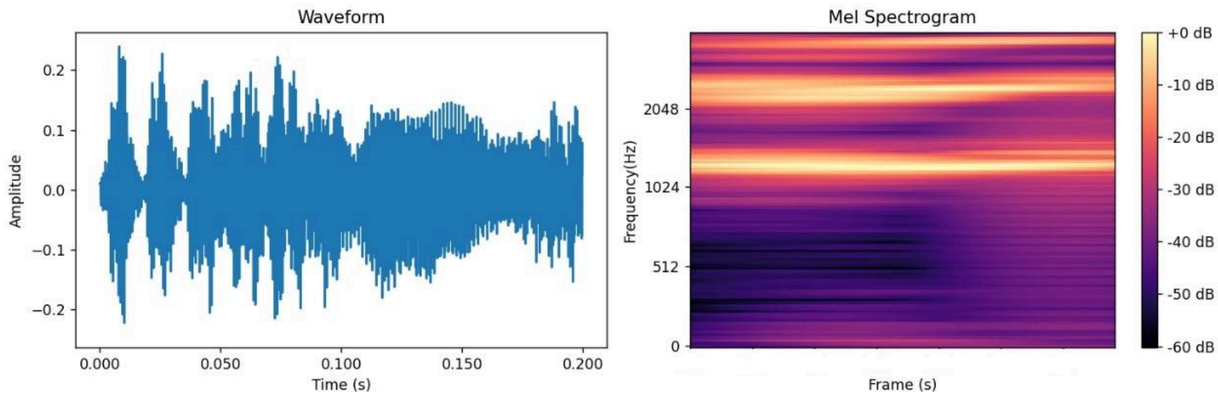
Animal sound data was extracted from various sources to mimic natural sources as much as possible. The major data classes were Monkey Sounds, Elephant Sounds, Cheetah Sounds, and background noises (including Birds).

3.2.1) DATA PREPROCESSING

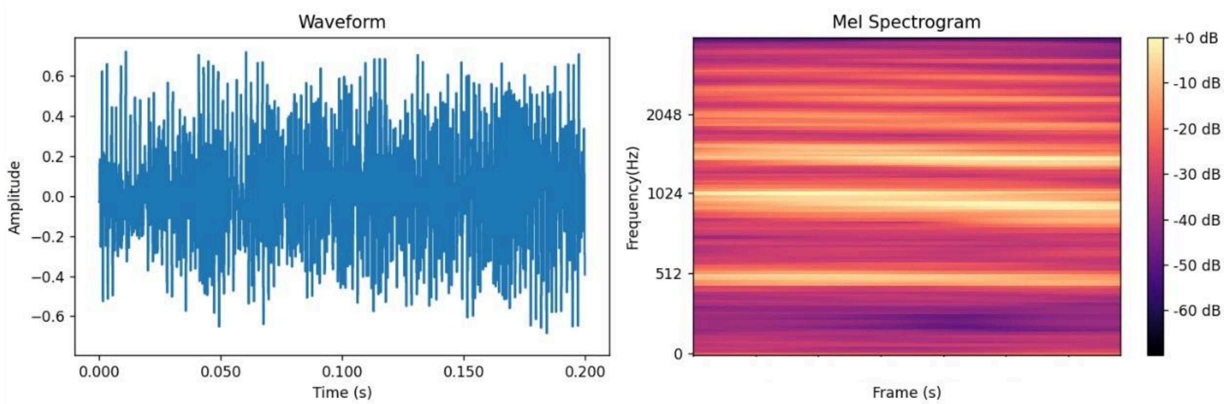
The audio data obtained from various sources was preprocessed into mel-spectrograms to fine-tune the CNN. 0.2-second long segments were sampled every 0.1 seconds from the audio, to make mel-spectrograms with 128 frequency bands and 1600 frames.

During inference, 0.2 seconds of audio signals are obtained from the onboard PDM MIC of Ethos U55 MicroNPU running with a sampling rate of 8000Hz. These audio signals are then converted to mel-spectrograms on Ethos U55 MicroNPU to resemble the images the model was trained on.

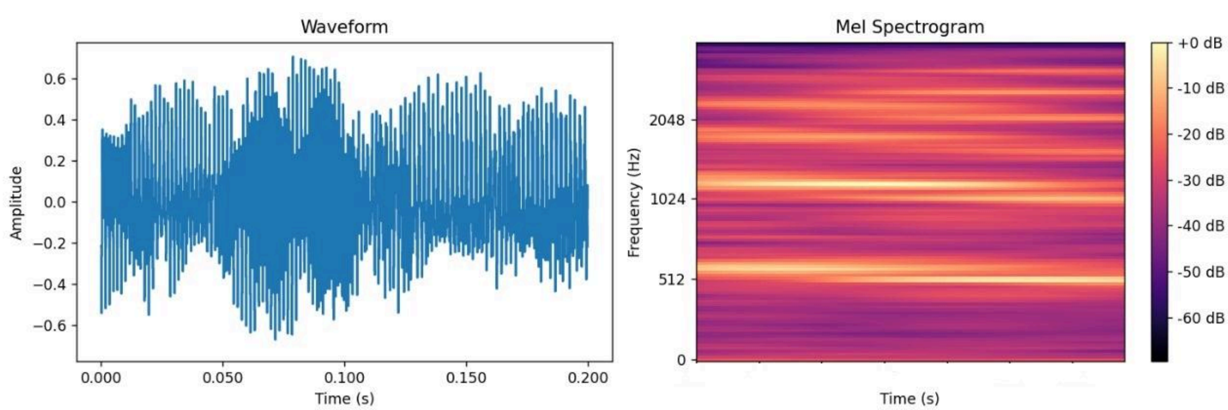
3.2.1.1) Monkey Audio Preprocessing:



3.2.1.2) Elephant Audio Preprocessing:



3.2.1.3) Cheetah Audio Preprocessing:



3.2.2) FINE-TUNING OF THE CLASSIFIER MODEL

A pre-trained YOLOv8 model was fine-tuned on the preprocessed dataset of mel-spectrograms.

YOLOv8's detection and classification models use a traditional convolution network for a backbone for feature extraction, and a neck with Feature Pyramid or path aggression for multi-scale feature aggregation. It further uses a global average pooling layer along with some fully connected layers to make the final predictions. The last layer is a softmax layer for the classification model and an anchor-free head in the case of the detection model.

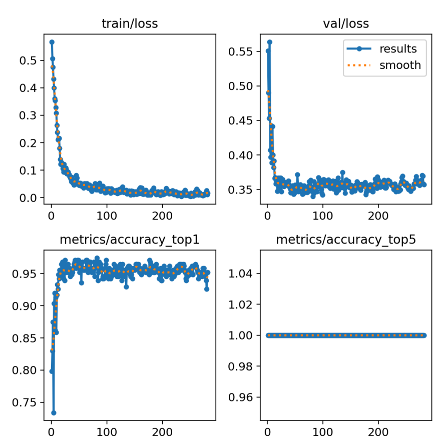
Two YOLOv8 models were fine-tuned, a detection model and a classification model. The models were fine-tuned for 1000 epochs, with 200 epochs of patience and int8 quantization.

For fine-tuning the detection model, the dataset was partitioned as follows: Negative Class (Birds, Background Noises), Positive Class (Cheetah, Elephant, Monkey). Additional Bird data was added so that the model could learn to distinguish between Bird and Cheetah sounds because of their high correlation. For fine-tuning the classification model, the dataset was partitioned into 4 classes: Class 0 (Background Noise + Bird Sounds), Class 1 (Cheetah), Class 2 (Elephant), and Class 3 (Monkey).

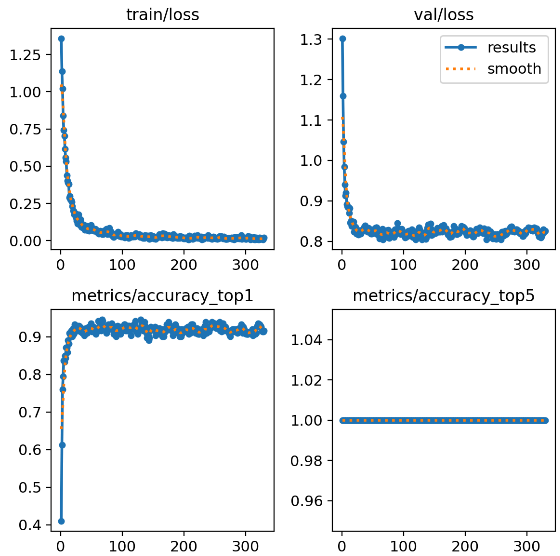
3.2.2.1) Fine-Tuning Results

Both models had really good training and testing accuracies. The detection model had an accuracy of 0.97436 and the classification model had an accuracy of 0.94551.

Training Results of Detection Model:



Training Results of Classification Model:



3.2.3) INFERENCE PIPELINE

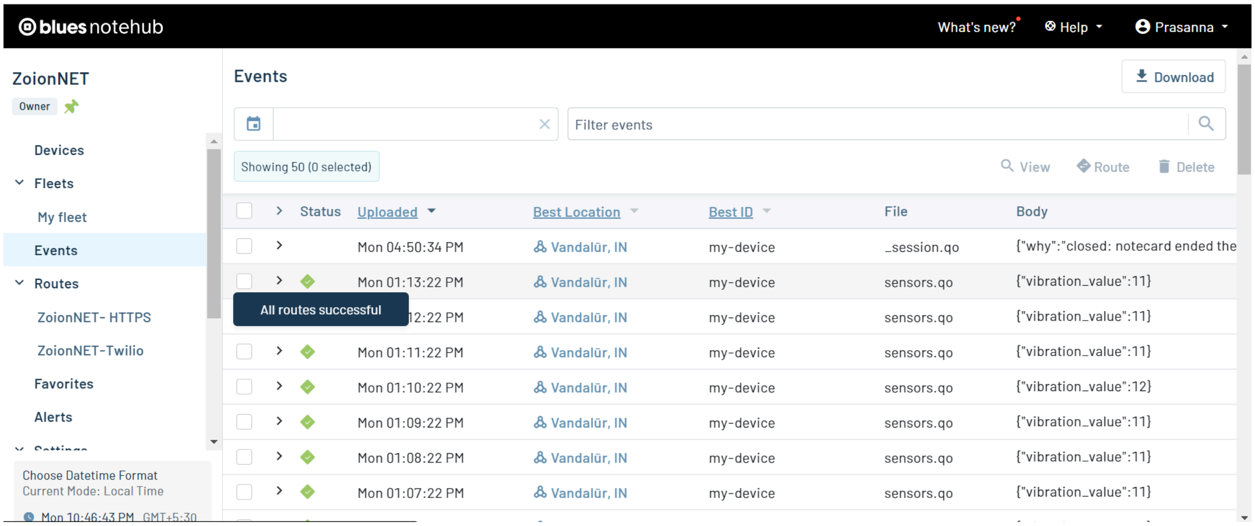
The detection model operates during the "dormant" state of the sensor, all around the clock. The classification model operates during the "active" state when there is a high chance of an animal intrusion as suggested by the detection model.

An exponential moving average is maintained over the previous detection model's output, if it passes a certain threshold, the sensor switches to active mode. Here too another exponential moving average is maintained over the classification model's 'none' class, once its average exceeds a threshold, the model switches back to a dormant state.

As the detection model is lightweight, it is bound to produce false values, the exponential moving average is used to decrease the impact of occasional false outputs on the state switch. At the same time the detection model uses much less power when run constantly, than its classification counterpart.

Hence the two-model approach aims to bring down energy costs while still maintaining the sensor accuracy.

3.3) IN-BOUND COMMUNICATION SYSTEM WITH CELLULAR NOTECARD



The inbound communication of the project ZoionNET refers to the forward communication from the cellular-based embedded system to the cloud service. In the project instance, the Blues cellular notecard acts as the communication module and does a device-to-cloud data pump to the Blues Notehub proxy server with specific interval updates. The Notecard will synchronize data with Notehub where the data payload includes requests like open and close sync connections, brown-out and hard reset, end and new session instances, and outgoing data payloads like seismic data, interference of the classifier AI model - the class of the animal detected, with some optional environmental data facilitated by the in-chip sensors. These data payloads are in JSON (JavaScript Object Notation) structured lightweight data format, enabling Blues Notecard to send and receive information over cellular networks at the lowest data rate possible ensuring the successful transmission of data packets with minimal overhead.

The Blues NoteHub which is the centralized data hub and also a proxy server, acknowledges the session instances, grants the connection for communication, accepts the data payload, stores it for further data processing and securely re-routing to other cloud service applications like Datacake and Twilio messaging services.

The device will work in a continuous or periodic mode where it syncs data based on the operational model in our case every 10 minutes or when a specific condition (Animal intrusion) is met.

3.3.1) FURTHER DATA ROUTING

A Cloud Platform called Datacake is used primarily for data visualization, device status monitoring, and advanced data analysis. And Twilio messaging service is used to provide seamless SMS alerts to the forest rangers and the end users (NGO).

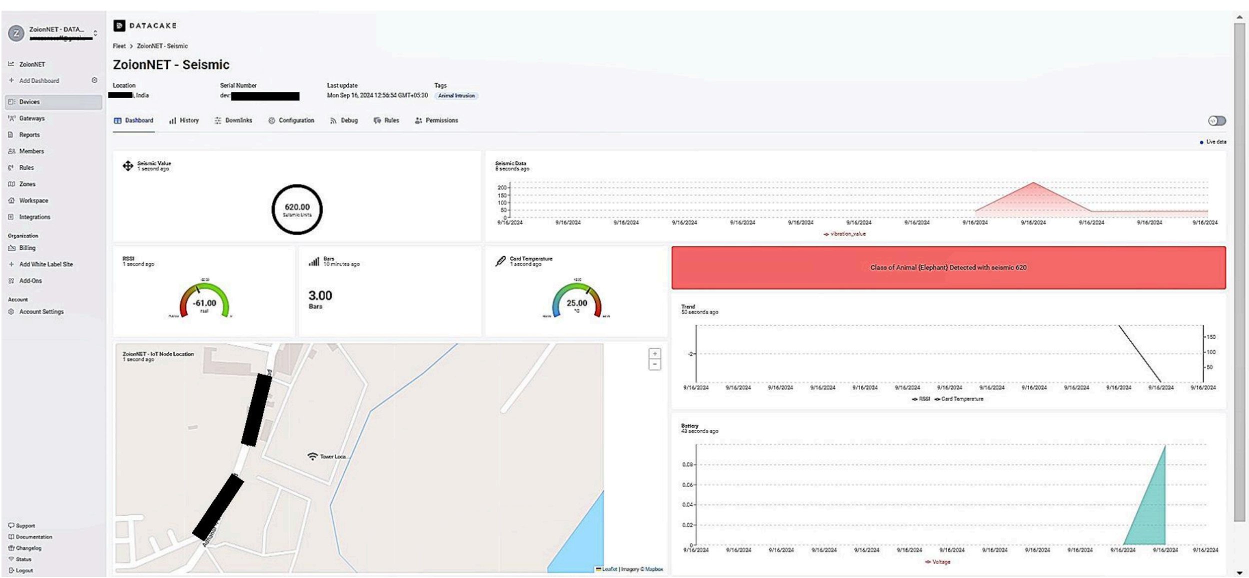
The Notecard uses JSON-based requests to interact with external APIs. It allows outbound communication via secure HTTP(S), making it easy to integrate with cloud platforms like Datacake and Twilio.

3.3.2) Alert System and Dashboard

Forest rangers and local authorities will be able to access the dashboard via web and mobile interfaces to monitor the situation in real time. This will help identify high-risk areas and periods, allowing for more targeted mitigation strategies like fencing, early warning systems, or habitat management.

3.3.2.1) DATACAKE WEB DASHBOARD

The Cellular Notecard, after collecting data from the Microcontroller Unit via the connected sensors, sends it in the form of JSON payloads to Datacake's API endpoint as "sensors.qo". The outbound communication from the Notecard is triggered by defining an HTTP webhook in its settings, which points to Datacake's API resulting in successful data routing.



JSON DATA UNITS

"file": "sensors.qo",

"body": {

"animal": "elephant",

"description": "Vibration threshold exceeded, data of an elephant sent.",

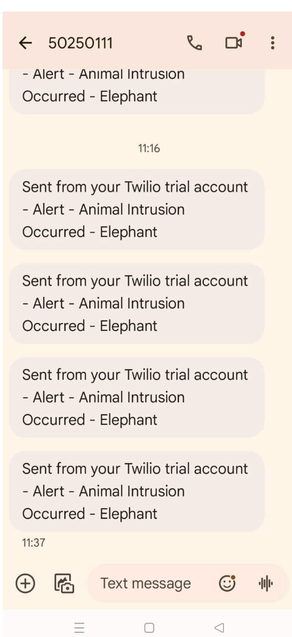
"vibration_value": 968

}

Once data reaches Datacake, it can be visualized in real-time through graphs, charts, and customizable dashboards. This integration provides long-term data storage and analytics.

3.3.2.2) TWILIO MESSAGING SERVICE SYSTEM

Twilio provides a platform that enables messaging service and alerting via SMS and pre-recorded voice calls. The Cellular Notecard pushes data to Twilio as “twilio.qo” by sending HTTP POST requests in the Blues Notehub so that the server takes care of routing an initialization of an Alert SMS from Twilio. This is typically used to trigger alerts or notifications based on predefined thresholds (i.e., vibration exceeding a set limit resulting in an abnormal sensor reading, or inference from the AI model detecting any class of animal).



JSON DATA UNITS,

"file": "twilio.qo",

"body": {

"customFrom": "+125661XXXXX",

"customMessage": "Alert - Animal Intrusion Occurred - Elephant",

"customTo": "+9199627XXXXX"

}

4) PROJECT DETAILS

4.1) TECHNOLOGIES USED (HARDWARE)

4.1.1) ONBOARD PULSE DENSITY MODULATION MICROPHONE

The onboard PDM microphone in the Seeed Studio Groove Vision V2 module encodes the audio signal as a high-frequency bitstream. Instead of directly outputting an analog signal, it uses a one-bit digital representation of the sound wave.

4.1.2) VIBRATION SENSOR

A vibration sensor is used to detect the physical vibrations exhibited by heavy animals like elephants, bison, and rhinos in the forest environment.

4.1.3) ETHOS U55 MICRONPU

The Arm Ethos-U55 microNPU (Neural Processing Unit) is the one that handles the neural network layers that are used in audio classifying CNN Model. The Ethos-U55 MicroNPU provides up to 90% energy reduction for ML workloads for audio classification which makes it ideal for battery-powered devices.

4.1.4) BLUES SWAN V3 MICROCONTROLLER

The Blues SWAN V3 Microcontroller is an STM32-based development board that is used to connect with Blues Wireless Notecard through the notecarrier board seamlessly.

4.1.5) NOTECARRIER A

Notecarrier A is a carrier board for Blues Cellular Notecard also an expansion board designed to work with Blues Wireless Notecard. It is a low-power cellular IoT module.

4.1.6) BLUES IO CELLULAR NOTECARD WITH IN-BUILT GPS

The Cellular Notecard is a System on Module (SoM) that integrates a cellular module, GPS, and an onboard accelerometer with a temperature sensor. It comes with a built-in SIM card, offering 10 years of cellular service and 500MB of data.

4.2) TECHNOLOGIES USED (ALGORITHMS)

4.2.1) CELLULAR TRIANGULATION, (WORKS WHEN GPS FAILS DUE TO UNCERTAIN WEATHER CONDITIONS)

Cellular Triangulation is a method used to estimate the geographic location of a device using signals from cellular towers. Unlike GPS, which requires satellite signals, cellular triangulation only requires cellular connectivity.

In scenarios where GPS signals may be weak or unavailable, the Notecard employs cellular triangulation to estimate its location. This technique uses information from nearby cell towers to determine the device's position.

4.2.2) SHORT TIME FOURIER TRANSFORMER (STFT) FOR SPECTROGRAM GENERATION

4.1.7) BILL OF MATERIAL AND COMPARATIVE ANALYSIS

HARDWARE USED:

A single Raspberry Pi 5 with 4GB RAM, priced at $178.82, costs more than our entire project, which comes in at just $133.57. This makes our solution not only more cost-effective but also more power-efficient, without compromising on performance.

S.NO	COMPONENT	USAGE	BILL OF MATERIAL
1	Blues NoteCarrier-A v1.7	Carrier Board + Antenna	$25.00
2	Blues NoteCard-Cellular	Cellular + GPS Module	$49.00
3	Blues Swan V3	Microcontroller - STM32	$24.95
4	Seeed Studio Groove Vision AI V2 Module with on-board MIC (AKA ETHOS U55 MICRONPU)	Neural Processing Unit (5 TOPS)	$15.99
5	Digital Vibration Sensor	Seismic Data	$4.00
6	SAMSUNG EVO Plus 32GB MicroSD Card	Dataset storage	$3.57
7	3.7v 12500mah 18650 Li-Ion Battery	Power Supply	$13.00
	TOTAL COST		$133.57

SOFTWARE USED:

S.NO	SOFTWARE	USAGE	SERVICE COST
1	Matlab	Simulation - Deep Learning Toolkit	Educational License - Free
2	Twilio	Messaging System - Alert	**N/A**
3	Datacake	Visualization and Graphing	**N/A**
4	Ultralytics	Yolov8 convolutional networks for classification and detection	**N/A**

5) SOCIAL IMPACT ON HUMANITY OR LOCAL COMMUNITY:

8000 affected people: The Melagiri forest Area along with the nearby villages (only closed-edged areas) has 8,000 people of various age groups, who are instantly and directly affected by the problems caused by the forest. So the Project will mitigate those factors.

This project has the potential to significantly improve the coexistence between wildlife and human communities in Methagiri Forest and similar regions. By providing an early warning system for animal intrusions, it can help prevent conflicts and protect both human lives and livelihoods.

It will also contribute to the conservation of endangered species like elephants and leopards by enabling better monitoring and management of their habitats. The long-term data collected by the system can inform policy decisions and guide conservation efforts.

Moreover, the project will serve as a model for how communication technologies like IoT and deep learning can be leveraged to address pressing environmental challenges. It demonstrates the potential for scalable, low-cost solutions that can be replicated in other biodiversity hotspots around the world.

5) IMPLEMENTATION STATUS, TESTING AND TRIAL

5.1) Ongoing work, testing, and trials conducted

5.1.1) Conducted field trials of the sensor hardware and connectivity in Methagiri Forest.

5.1.2) Collected and annotated real-world data for training the deep learning models.

5.1.3) Implemented the alert system and dashboard with integration into the sensor network.

5.1.4) Deployed a pilot system in a high-risk area of the forest for testing and validation

5.2) FUTURE PLANS

5.2.1) Usage of LoRaWAN system for covering deep inside forest. This helps to scale the solution to the places of cellular network absence.

We aim to have a fully functional prototype ready for deployment by the end of 2025.

Acknowledgment

The Team would like to thank the Kenneth Anderson Nature Society (KANS) and Cauvery North Wildlife Sanctuary for their valuable support.

CONTACT INFORMATION OF TEAM:

S.NO	NAME	AFFILIATION	OFFICIAL MAIL
1	Prasanna Vijay	Sri Sairam Engineering College	prasannavijay@ieee.org
2	Adarsh Gupta	IIT Guwahati	adarsh.gupta@iitg.ac.in
3	Vansh Panwar	IIT Guwahati	p.vansh@iitg.ac.in
4	Vatsalya Singh	IIT Guwahati	s.vatsalya@iitg.ac.in
5	Shubham Singhal	IIT Guwahati	s.singhal@iitg.ac.in
6	Anubhav Saha	IIT Guwahati	anubhav.saha@iitg.ac.in

INFORMATION OF THE ADVISOR:
NAME: Dr. Teena Sharma, Assistant Professor
AFFILIATION: Mehta Family School of Data Science and Artificial Intelligence, Indian Institute of Technology, Guwahati, India
EMAIL: teena@iitg.ac.in

INFORMATION OF THE COMMUNITY PARTNER:
NGO: Kenneth Anderson Nature Society, Cauvery North Wildlife Sanctuary (TN)
Email: kans@kans.org.in