Intelligent Species Monitoring using Sony Spresense Micro-controller: Integrating Audio and Visual Analysis for Conservation and Biodiversity Management

Shree Harsha Satish, Wenping Jiang, Kai Yan Lo, and Yasamin Makoui

Department of Computer Science, University of Liverpool

August 21, 2023

Abstract

Invasive non-native alien species can pose a significant threat to biodiversity and ecosystems worldwide, causing severe harm to native flora and fauna and have a potential risk of farming. By leveraging the Sony Spresense controller's capabilities and including audio, camera inputs along with the internal GPS, we aim to develop an intelligent system to analyse audio and visual data to identify and track different species in a given environment. We aim to cover invasive plant species with image inputs and monitoring animal species (dog whelks, snails and mosquitoes) through both image and sound inputs.

1 Background and Concept

The predominant trends in machine learning (ML) application for biodiversity management fall into two categories. First, the use of high-end technology, such as drones, to chart the geographical spread of invasive species over large territories, a technique favored in scientific studies. Second, the integration of the Internet of Things (IoT) with tinyML for commercial purposes, broadening the reach of ML techniques in alignment with our objectives.

Our review of AI research within the biological sphere indicated a strong preference for either computer vision and sound detection methods, attesting to their efficacy. However, these generally necessitate high-power devices and don't necessarily combine the two approaches. In contrast, we aim to combine the two approaches using energy-efficient micro-controllers.

Our proposition involves the integration of two models within a single device, capable of adaptive utilization in any environmental context. For the plant mode, we utilize computer vision models and near real-time classification to identify invasive and diseased plants as the device moves. Conversely, the animal mode leverages an omni-directional mic for sound perception to employ near real-time sound classification for moving animal detection. This dual-mode works for two reasons: 1) Plants, being stationary and mostly silent are in contrast with mobile, sound-producing animals, so using computer vision techniques for plant detection is more sensible, while for animals, due to their non-stationary nature, computer vision without more help might not be useful; 2) Moving devices employing computer vision align well, whereas fixed devices yield better detection and classification results without the disturbance from self-generated noise.

2 Data Gathering

Considering the project's timeline and the resources at our disposal, we selected Japanese knot-weed, rose leaves, tomato leaves, cherry laurel leaves, and pontic rhododendron as detection classes for the plant model. For the animal model, we chose to include the sounds of dog whelks, hermit crabs.

Data for this project was accumulated through three primary sources: Online repositories, collaborations with researchers from the Biology Department at the University of Liverpool, and direct in-field photography. As research into the classification of diseased plants is well established, we sourced images of affected rose leaves, potato leaves, and Japanese Knot-weed from nearby gardens and parks in Liverpool. We coordinated with researchers from the Biology Department to acquire diverse animal sound and video data.

Collecting data from these multiple, reputable sources ensures the quality of our research datasets. Simultaneously, we employed image augmentation techniques to fulfill the volume and to match test scenario conditions (for example with image skewing).

3 Model Implementation

Drawing from the resources provided on the Sony Development webpages, we endeavored to utilize Arduino at first. However, the complexity of coding required was beyond our reach within the given time frame. Our original plan was to utilize both short-time and long-duration windows to analyse the audio signals and then build a classifier to classify these windows. We encountered many problems from integrating both audio and visual models into a single Arduino code to trying to convert models into a tflite format. Ultimately, we ended up using the edgeImpulse platform because of its simpler use case.

1. Plant image classification

   For plant image classification, we used a pre-trained MobileNet classifier and collected over 200 images

for each of the eight classes before transfer learning, but because of the space restriction on edgeImpulse we had to use only 1626 images in total. These images comprised a balanced mix of eight classes, representing healthy and diseased leaves from tomatoes, roses, and cherry laurels, in addition to two invasive species: Japanese knot-weed and Pontic Rhododendron.

2. Animal audio classification

   For this purpose, we employed a 1D CNN model, trained on short time (25ms windows) mel-frequency energies of 17 minutes of audio data. This audio data encapsulates periods of silence/background noise, along with sounds from dog whelks and crabs. Originally, we had intended to train our models on approximately 45 minutes of dog whelk and crab noises, coupled with around 120 minutes of mosquito wing-beat sounds. However, due to limitations pertaining to audio quality, label quality and size, we opted to restrict our dataset to the 17 minutes utilized.

3. Web API Interface for Data Posting and Querying

   Our API web interface facilitates both online and offline data inputs from the monitoring device. All monitored occurrences can be displayed either in a list format or as a map view. Additionally, it offers various filtering options based on different query criteria, such as confidence level for each classification type or specific time periods, catering to the users' specific interests.

   Query Interface Link. Example image in Figure 3

4 Limitations and Modifications

Our original strategy for this project included programming a device for autonomous operation. This involved auto-capturing images and audio files and storing the results on SD cards. We planned to upload data to the web interface using an API and to conduct a field test in a real-life environment with invasive plant species, diseased plants, and invasive animal species.

As stated before, the previous constraints resulted in our use of the edgeImpulse platform and the demos all require a laptop connection instead of complete autonomous operation. Instead of conducting a field test, we decided to perform classification tasks in a laboratory environment. For this stage of the project, we chose not to export the results to the web server. Despite these constraints and the modifications made to our initial strategy, the models performed as expected with some difficulties in two scenarios: when it encounters unseen noises in animal mode/poor lighting conditions in plant mode.

5 Device Installation and Testing at an Indoor Lab Environment

The device installation and testing phase was conducted in an indoor laboratory environment. Our setup included a Spresense Micro-controller equipped with a microphone and a laptop running Edge Impulse software.

For the image classification component, we utilized leaves from a Cherry Laurel, Japanese Knot-weed, and Pontic Rhododendron as subjects for detection. These plant leaves were positioned in front of the Spresense camera to evaluate the classification accuracy. The confusion matrix for the overall plant mode on a validation set (balanced set of 608 images) is in Figure 1.

Turning our focus to audio classification, we played audio of invasive animal species, which had not been used in the training phase, and recorded the Intensity levels (dB). We captured audio input in real-time to classify the sound. The background noise was 37dB. These are the settings for the demo in the video. However, the classification results on collected recordings and a validation set (balanced set of 6 minutes for each of the 3 classes) are in Figure 2.

6 Future Work

Looking ahead, we have several enhancements planned for the project. One of the obvious ones is to program complete autonomous operation of the micro-controller.

In addition to this, we're planning to upgrade the animal model by integrating audio detection with camera activation. This means that upon detection of a specific species' sounds, the device will automatically trigger the camera to capture images of the surrounding environment. This integration would not only verify the presence of the detected species but also provide comprehensive information about its habitat and potential interactions with other species. This additional data could be invaluable for further research and analysis, contributing to our understanding of local ecosystems and the impacts of invasive species.

Further, these future improvements will bring us closer to our ultimate goal of creating a device that offers seamless, real-time monitoring of biodiversity. By harnessing the power of machine learning, we aim to support the management of biodiversity, aiding in the identification and control of invasive species and the conservation of local ecosystems.

Confusion matrix (validation set)

	CLASS1	CLASS2	CLASS3	CLASS4	CLASS5	CLASS6	CLASS7	CLASS8
CLASS1	**70.3%**	5.4%	0%	0%	0%	16.2%	8.1%	0%
CLASS2	0%	**90.5%**	0%	9.5%	0%	0%	0%	0%
CLASS3	0%	0%	**88%**	0%	0%	0%	12%	0%
CLASS4	0%	33.3%	6.7%	**53.3%**	0%	0%	6.7%	0%
CLASS5	0%	0%	8.3%	0%	**79.2%**	0%	4.2%	8.3%
CLASS6	7.7%	3.8%	0%	0%	0%	**88.5%**	0%	0%
CLASS7	0%	3.4%	3.4%	0%	0%	0%	**93.1%**	0%
CLASS8	0%	3.0%	3.0%	0%	0%	0%	0%	**93.9%**
F1 SCORE	0.80	0.76	0.85	0.64	0.88	0.84	0.84	0.94

Figure 1: Confusion matrix of the validation set (608 images) in plant mode

Confusion matrix (validation set)

	CRAB	DOG WHELK	SILENCE
CRAB	**98.0%**	0%	2.0%
DOG WHELK	0%	**83.1%**	16.9%
SILENCE	0.2%	22.9%	**76.9%**
F1 SCORE	0.99	0.76	0.81

Figure 2: Confusion matrix of the validation set (18 minutes) in Animal mode

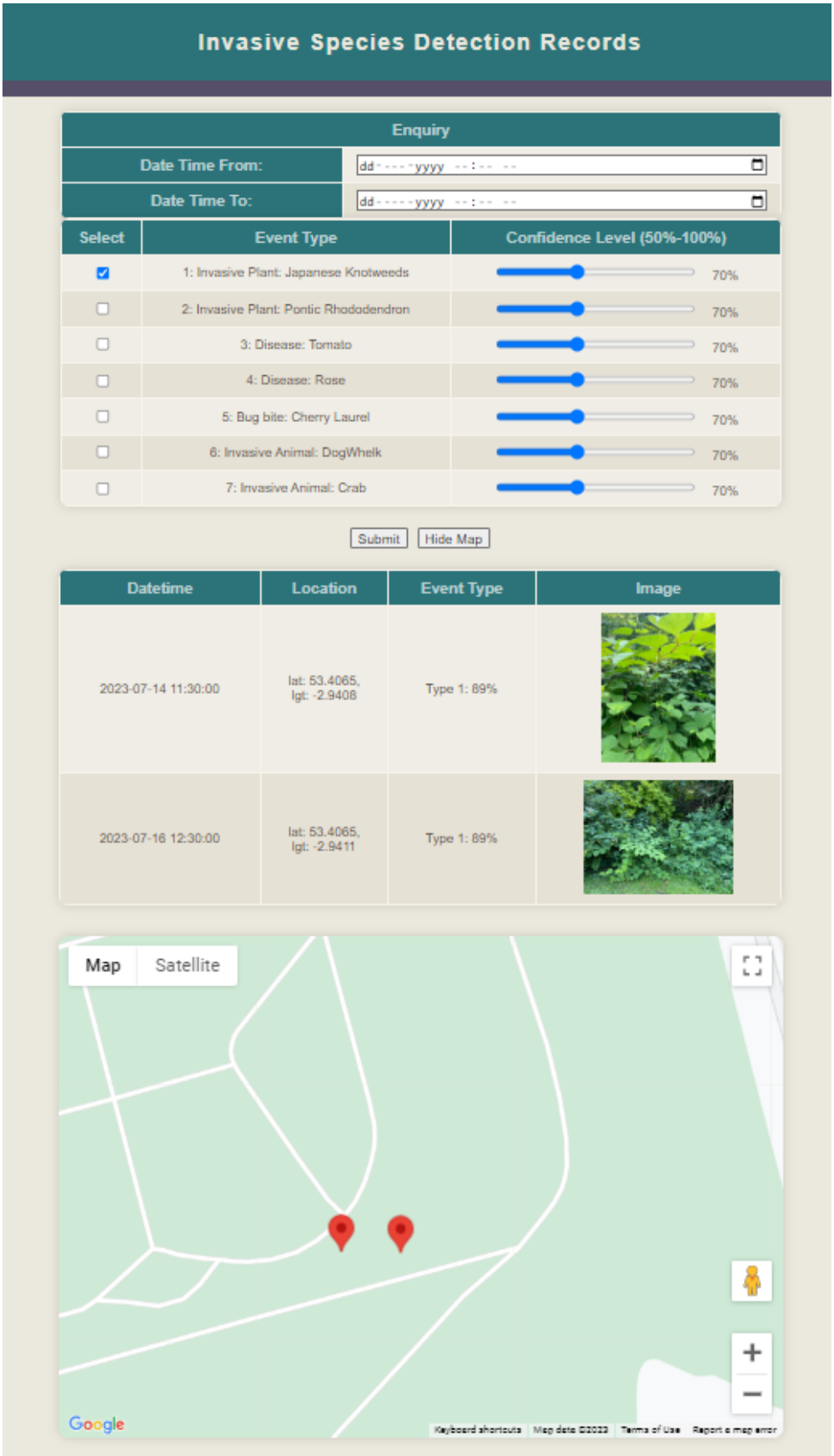

Figure 3: Web interface