Using the eVorta platform to develop a machine learning automated classification/detection model for stoats in the New Zealand environmental context: A preliminary trial.

Em Oyston, May 2022
Pure Salt Ltd NZ

Background to trail camera use in the NZ conservation context

The use of trail cameras for the detection and monitoring of native and invasive fauna has become a widespread and common method in New Zealand conservation circles. The method has largely been adopted due to the relative sensitivity of detection of target species; the resolution and diversity of targets it is able to detect; the ability for the tool to monitor for long periods without maintenance; the ability to use it in low density qualitative or larger scale quantitative methodologies; and the accessibility of the equipment to acquire.

As more users around the country are using trail cameras, and projects are using them on a larger scale – a challenge has developed around the cost/time required to process the increasing amount of information (footage) captured by the cameras, as well as the expertise to process raw data captured and analyse it in a consistent and meaningful way. There is also significant potential to use cameras as an eradication surveillance tool, utilising cameras at a density that saturate every home range of the target species, however, the amount of footage produced requiring processing makes this prohibitive with currently used methods and technology.

Current options available for processing data

Options available for processing data have largely been based on 'manual classification' where a user works through individual images or video and identifies ('tags') the image with what species is present in the footage. The results of this are usually compiled into metadata in the form of a database for the purpose of analyses via summary statistics, statistical modelling, or spatial temporal modelling. Manual classification can occur on a number of different interfaces through different software platforms that improve the efficiencies of manual classification through features such as the use of 'hot keys' for tagging, effective layouts for improving workflows, and automatic compilation of metadata based on manual classification output of footage.

Some of the better platforms for manual classification available (based on my personal experience) include DigiKam, the Zero Invasive Predators (ZIP) classifier, and an MS Access database designed by DOC's Joris Tinnemans. Although these platforms improve the efficiency of manual classification, maximum processing rates achieved are still only around 3000 images per person hour, with the quality of detection (especially for cryptic species) often degrading over time due to human fatigue.

In recent years, machine learning algorithms have been developing for processing images, and the detection of wildlife has been one application for this technology. Machine learning is effectively the ability for computer algorithms to improve automatically through the experience and processing of data to improve results/objectives over time. This technology is in its infancy for wildlife applications but holds significant potential in costs savings for data processing.

Until recently, using machine learning algorithms to start developing models that could automatically detect and classify footage required a relatively high proficiency in computer programming languages. One of the more well-known opensource machine learning APIs (Application Programming Interface) is Megadetector (using Microsofts AI for Earth framework). The Megadetector AI model has the ability to detects animals, people, and vehicles in footage but does not identify individual species – just detects whether an animal is present. A MegaDetector GUI (Graphical User Interface) version has had an open-source release (last release 2/21, Petar Gyurov) which has been trialled by some conservation applications in NZ with varying levels of success. Feedback is that it can be a useful tool for 'weeding out' empty images (false triggers), but requires significant computing resources, and even with higher specification machines with dedicated high end Graphical Processing Unit (GPU) it can still take between 5 – 12 seconds to process an image, with significant precision issues still at hand (false positives and negatives in detection, and basic classification errors between vehicle, animal, and human).

The eVorta platform

In August 2021 a small informal working group from DOC received a video presentation by an Australian company, eVorta, on a cloud-based machine learning automated classification and detection platform they had developed. Originally, the platform had been developed for use within Australia but it was growing in use internationally.

eVorta requires users to upload their footage through a portal in the browser-based cloud interface. Footage is then automatically classified using an iteration of architecture (a model) that has been developed through a combination of users manually tagging and training architecture that then uses machine learning to improve on. The results are a cloud database of images that have detections and target species classified with bounded boxes accompanied with confidence scores relating to their respective classification (see **Figure 1**). The eVorta platform is a commercial model, and the costs of the platform is on a per photo basis at $0.01 AUD (ie to processes 100 photos would be $1 AUD). As well as detection and classification, the platform also provides the user with analytics such as spatial/temporal mapping, summary statistics, and the ability to export results and metadata in a variety of formats.

The user group were impressed with the eVorta presentation of real life application/data and saw the potential for a New Zealand model (trained for NZ native fauna and introduced mammals species) to be developed in ordered to provide a solution to significantly reducing the costs of processing large trail camera datasets.

There was an opportunity identified to trial the development of a stoat detection model using eVorta, on a programme of work using a network of trial cameras (n = 84) to detect what was assumed to be low densities of stoats on Five Fingers Peninsula, Resolution Island. The programme was being run by Pure Salt NZ Ltd in collaboration with DOC, and funded by Jobs For Nature.

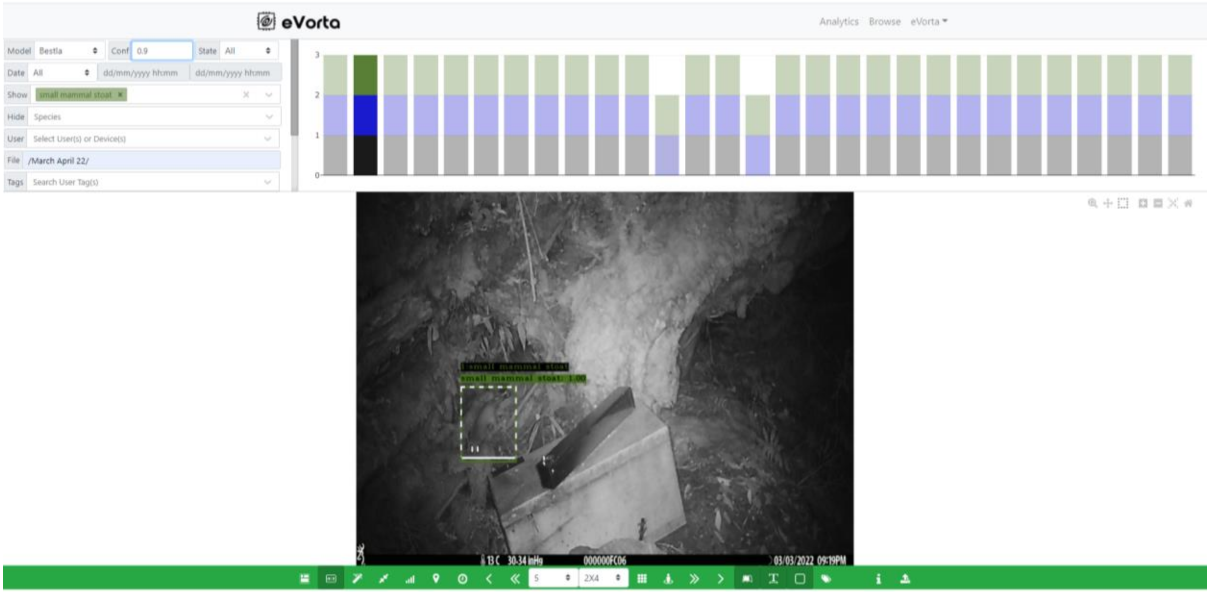


Figure 1: *Example of the eVorta GUI interface, showing the bounded box around a detected and classified stoat.*

Trial of eVorta for the automated detection and classification of stoats on five fingers peninsula

Development of an iterative model that would distinguish stoats began in November 2021. Using pre-classified trail camera footage from a programme of work in the Perth Valley provided by Zero Invasive Predators (ZIP), around 3000 pre-classified stoat images were uploaded to the platform, of which 1270 images were tagged with bounded boxes around them in order to train stoats into the then current iterative model - 'Saturn' .

A new model ('Aegir) was produced which incorporated the new stoat dataset in addition to all trained species data from Australia and other international datasets. The first batch of 'real' field data from Five Fingers Peninsula was tested on the model. This dataset consisted of 24,000 images from 62 cameras produced between November and December 2021. The 'Aegir' model found 1,400 possible stoats, which covered all 4 manually detected stoats, plus one additional that was not found by manual human classifications.

The early results from the 'Aegir' model were promising, however, it was identified that the model would significantly benefit from human training using stoat footage from a more diverse range of camera models and locations. Stoat images from the Perth Valley that were used to train 'Aegir' came from fewer than 20 camera locations, and used the same model of camera.

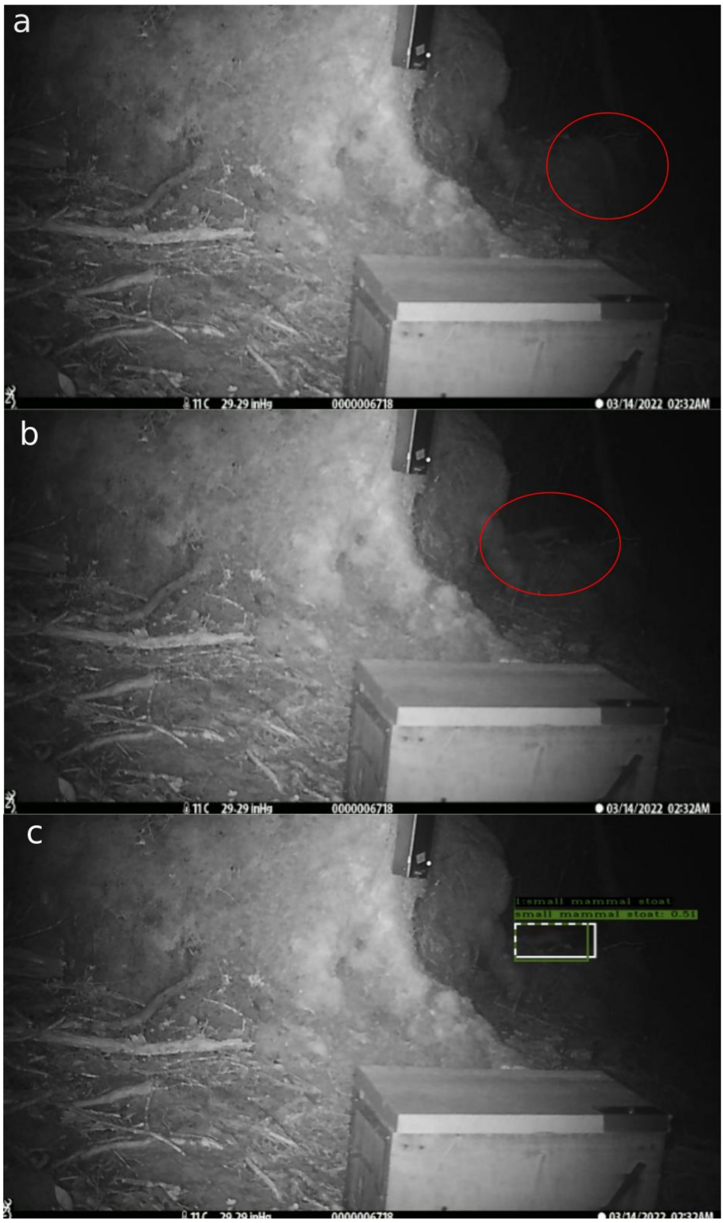
In March 2022, another 1100 stoat images were uploaded to and manually tagged in order to train and improve the model. Effort was made to obtain images from a more diverse range of camera models and locations – obtained and provided by the Te Manahuna Aoraki project.

In April 2022, an updated model was created ('Bestla') based on the recent additional training. The March dataset from Five Fingers Peninsula was uploaded (88,246 images) in order to compare how 'Bestla' would perform against manual classification for the detection of stoats.

Results

From the March 2022 dataset from Five Fingers Peninsula (n=88,246), 'Bestla' processed 1031 images as having stoat detections at a 0.8 confidence interval. 999 of the 1031 were false positives, with 32 images being true positives (capturing 10 independent 'events' where a stoat visited a camera station).

At 0.8 confidence, 'Bestla' detected one event than manual classification did not, and manual classification picked up an event that 'Bestla' did not at 0.8 confidence. The extra event that manual classification picked up could not be verified independently from a single image: ie it required the ability to look across a sequence of photos to determine the likelihood the detection was a stoat (see **Figures 2a-c)**.



Figures 2a-c: *Example of sequential images where validation of independent images as having stoat detections would be difficult. The movements of the target between Figures 2a and 2b provides inference that the movement is similar to that of a stoat - allowing a manual classification with a higher level of confidence. As* *eVorta analyses each image as independent images, this was picked up at a lower confidence level of 0.51 (Figure 2c)*

The eVorta platform makes reviewing false positives relatively efficient, providing a ‘grid view’ interface zoomed in on a bounded box of the detection in question (see **Figure 3**). To review the 999 false positives produced by ‘Bestla’ at a 0.8 confidence interval, it took me around 10 minutes – which was significantly less time than reviewing and classifying the same number of images on any other interface I have used.

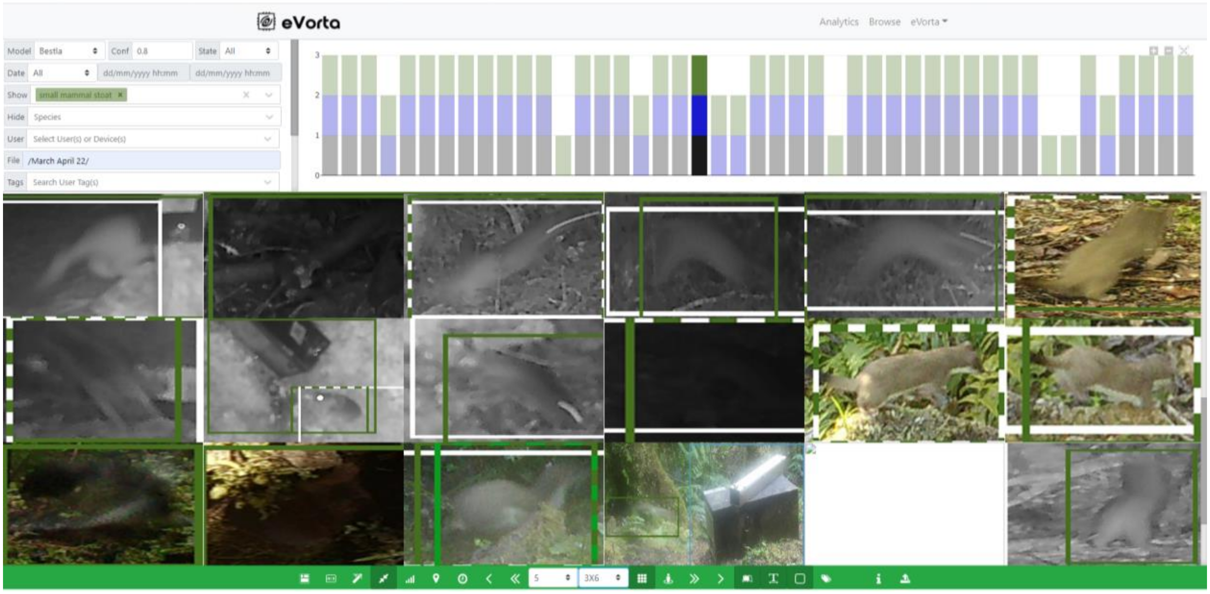


Figure 3: *eVorta platform interface showing ‘grid view’ verification of bounded box detections which was used to validate false positives*

At 0.5 confidence, ‘Bestla’ processed 2424 images as having stoat detections, with 2388 positives, and capturing 11 stoat events. At 0.5 confidence, ‘Bestla’ picked up one more stoat event than manual classification did.

Table 1: Summary output of ‘Bestla’ stoat detections on five finger peninsula dataset (n=88,246)

	Bestla @ 0.9 conf	**Bestla @ 0.8 conf**	**Bestla @ 0.5 conf**	**Manual classification**
Number of independent events where stoats were detected	9	10	11	10
Number of images where stoats were detected (true positives)	28	32	36	44
Number of false positives	Unknown	999	2388	-

Discussion

For the limited training that has gone into developing a 2nd iteration of a model that can automatically detect and classify stoats, the eVorta platform has performed well. Given that the model improves with every dataset uploaded through validation and machine learning, the platform has significant potential for cost saving in relation to the classification and processing of trail camera data. The ability to cost-effectively and efficiently process data also enables significant upscaling of camera networks – further enabling monitoring and detection abilities in the New Zealand conservation context.

Keeping in mind that any further use of the current model is likely to bring improvements in recall and reducing false positives: at 0.8 confidence, using 'Bestla' meant that a user would only need to review 1.17% of the original data to detect what was detected with manual classification of 88,246 images. Using 'Bestla' at 0.5 confidence meant that a user would only need to review 2.75% of the dataset and would detect more than manual classification did of 88,246 images.

In addition to detecting and classifying stoats, with no further training, the current model 'Bestla' seems to have done a good job at non-target species – detecting and classifying birds, deer, and rodents (based of training and machine learning from Australian and other international datasets). These observations are purely qualitative, as the focus of the trial has been on the precision and recall of stoat detections – but it is worth noting that with additional training, eVorta could potentially also be used as detection tool for a wide-range of species in New Zealand at no additional processing cost.

Both the first model 'Aegir' as well as the 2nd iteration 'Bestla' outperformed manual classification in terms of stoat detections, with both of these models detecting additional stoat events in the November/December 2021 dataset (n=~24,000) and March 2022 dataset (n=86,246) that manual classification had missed. With further training, the next iterations of models are likely to improve precision and recall and reduce false positives further.

Using the March 2022 Five Fingers peninsula data set as an example, it took approximately 30 person hours to classify 88,246 images, while the commercial model of eVorta would cost $882. If someone was manually classifying large datasets on a regular basis, it would not be unreasonable to assume an hourly rate of $25-30 per hour which would work out at a similar cost to using the eVorta platform, with a lower efficacy and without any analytics of the data.

Conclusion

To date, the eVorta platform appears to be a promising tool for automated classification and detection of trail camera footage. I believe with further use and training it could provide cost savings, improve our ability to detect cryptic species, and broaden the abilities of trial cameras as a detection tool. The caveats to this are that investment be needed to train and improve target species; and that the current platform requires high speed internet to pragmatically upload and manipulate large datasets.

The more users that use the platform, the stronger the models will become. At present, the costs are likely to be cost neutral in many scenarios, but provide the added bonus of providing analytics of data, cloud storage of data, and what looks to be better detection efficacy than manual classification will provide on large datasets.