



Ghost bat call detection using AI: strategies to improve model performance

Benjamin C. Lawrence (1), Dr. Saeed Shehnepoor (1)

(1) Wood, Perth, Australia

ABSTRACT

The declining population of the Australian native ghost bat (*Macroderma gigas*) has resulted in extensive monitoring programs, including passive acoustic monitoring. Detection of the ghost bat's social and echolocation calls can be used to inform management and conservation actions. Analysis of acoustic datasets to identify ghost bat calls is currently performed by a small number of experts who review audio recordings. However, this analysis is time consuming due to the ghost bat's diverse and complex calls and the massive data sets that are gathered. Machine learning (ML) models can efficiently process these data sets to provide significant time and cost benefits over current analysis methods without sacrificing accuracy. This paper explores the process of optimising an ML model using a large dataset of audio recordings from 10 ghost bat roosts spread across the Pilbara region of Western Australia. Average improvements of 16% in precision and 13% in recall were achieved across all sites by expanding the spatial and temporal range of the training datasets and tuning the detection thresholds for each site. The ML model had an average of 95% precision and 80% recall across all sites.

1 INTRODUCTION

It is well documented that humans are responsible for a massive decline in biodiversity across our planet, and that actions must be taken to prevent further loss of species and habitats (Hogue et al., 2022). Understanding the behaviour and interaction of a given species with its environment is key to identifying effective actions (Tobias et al., 2019). For bats, this typically includes the locations of roosting and foraging sites, population dynamics, reproductive ecology, and estimating the species tolerance of anthropogenic emissions such as light, noise, vibration and dust. However, for cryptic species such as the ghost bats in the Pilbara region of Western Australia, studies on these aspects are limited and more research is needed to understand their behaviour (Bradley et al., 2024).

Ghost bats have historically been studied by entering roosts to visually observe or capture bats (Douglas 1967). There are several challenges with this approach. Firstly, many ghost bats roost in old mine shafts and caves in remote locations, so the cost of access and the safety risks limit the frequency and duration of site work. Secondly, the roosting behaviour of ghost bats is not well understood (Hanrahan et al., 2022) so it is not certain that they will be present at a particular site. Thirdly, ghost bats are easily disturbed, and entering roosts has been found to cause mothers to drop pups or trigger colony-scale roost abandonment (Toop 1985; Armstrong 2011).

The challenges with site visits mean that remote surveillance techniques such as passive acoustic monitoring (PAM) are attractive options for the study of ghost bat behaviour (Ruykys et al., 2023). Fortunately, ghost bats produce a variety of social and echolocation calls (Hanrahan 2020), all of which can be readily detected using off-the-shelf ultrasonic acoustic recorders. Deploying acoustic recorders at cave entrances has become an established and accepted method to monitor for ghost bats, as they are known to vocalise at the entrance and inside roosts (Hanrahan 2020). These monitors can be set to record audio at pre-defined trigger thresholds to capture bat calls, enabling recording for weeks to several months or years depending on the amount of activity. Modern PAM systems now include online capabilities which enable data to be uploaded in real time, further reducing monitoring costs while also improving data consistency (Lang et al., 2024, Li et al., 2022).

However, PAM systems have their drawbacks. The main challenge is the analysis due to the complexity and size of the datasets (Ducrettet et al., 2025). Identification of bat calls requires understanding the acoustic call characteristics of the species of interest as well as those of other bat species, insects, birds and other ambient sounds

present in the audio recordings (Brabant et al., 2018). However, the number of recognised ghost bat call identification experts is limited, and the scale of the datasets is vast (many terabytes of recordings), so it is impractical for experts to manually analyse the data.

Automated identification approaches such as machine learning (ML) address the analysis constraints and have become increasingly more common in recent years. Early methods using ML algorithms for human speech recognition demonstrated that ML could outperform traditional methods in detecting and classifying bat echolocation calls (Skowronski et al., 2005). Recent advancements in ML have further reduced the time and effort required for analysis of large datasets (Chalmers et al., 2021). The use of deep learning techniques such as convolutional neural networks (CNNs) have also improved the accuracy of species detection from acoustic data and have been demonstrated to outperform traditional methods (Sharma et al., 2022). Models such as ResNet and YOLOv5 have been employed to detect and classify a range of different vocalising species with high precision and recall rates for (Husain et al., 2023). There is an ever-increasing range of ML methods that can be used to identify bat calls in PAM datasets with varying degrees of effort. It is tempting to use off-the-shelf programs, but this must be done with caution as there are several studies which demonstrate discrepancies between different methods/programs (Lemen et al., 2015). ML model results must be thoroughly validated to provide confidence in the results before replacing industry standard manual analysis methods (Rydel et al., 2017).

2 METHODS

2.1 Defining the objective

The purpose of PAM for fauna detection is to provide insight into animal behaviour (Oestreich et al., 2024). The applications can be broad (e.g. estimating overall biodiversity) or focused on meeting specific rules, regulations and permit requirements. The ML model for this paper was developed to meet regulatory requirements for long-term monitoring ghost bat behaviour near mine sites in Western Australia, including detecting presence, locating maternity roosts, estimating population numbers and identifying disturbance. The aim of the monitoring is to better understand the behaviour of ghost bats and their response to nearby mining, which will inform the planning and management of mining operations.

Ghost bat presence can be determined by reporting positive detections. Potential maternity roosts can be identified by counting the number of calls around sunrise and sunset (Hanrahan et al., 2019). Population sizes can be estimated by factoring in a call rate of the species (e.g. 20 – 40 calls per night may indicate 5 – 10 ghost bats for a given site (Bullen 2020)). A thorough interrogation of the call detection dataset, in conjunction with other datasets, could reveal how ghost bats respond to environmental stressors such as noise and habitat disturbance (Armstrong, 2010). In addition, this ML model is able to differentiate between different ghost bat calls and therefore could be an indicator of behavioural events such as breeding or parturition (Hanrahan 2020), but this is not currently required for this application. The objectives can be achieved by accurately detecting ghost bat calls and reporting the time of detection.

The focus of this ML model was to maximise precision because high numbers of false positives may lead to incorrect conclusions for ghost bat activity. For example, it would be problematic if ghost bats were incorrectly detected where none were present or continuing to report ghost bat presence when a cave or adit had been abandoned. A high precision was targeted for all sites to minimise false positives as far as practicable. A lower recall was considered to be acceptable as the ghost bats are highly vocal and typically trigger several audio recordings, giving the ML model multiple chances for detection.

2.2 ML model overview

This ML model has a standard image processing architecture which uses spectrograms as the acoustic representation of bat calls. A set of classes has been defined to separate ghost bat calls from other bat calls, insects and ambient sounds. The model calculates the probability that a given audio file contains each of the classes, with the sum of probabilities totalling 1 (or 100%). If the probability of a target class is over a defined threshold, then that audio file is classified as a positive detection. For example, if the threshold was 0.7 and an audio file had a 0.75 probability of having a ghost bat call, it would be classed as a ghost bat detection.

The image classification ML model stages are shown on Figure 1. This paper focuses stages 5 and 6. Steps 1 – 4 are not covered as no novel data collection, preprocessing, signal enhancement, feature representation or model architecture have been used. There are numerous publications using machine learning for fauna detection which explain these components in detail.

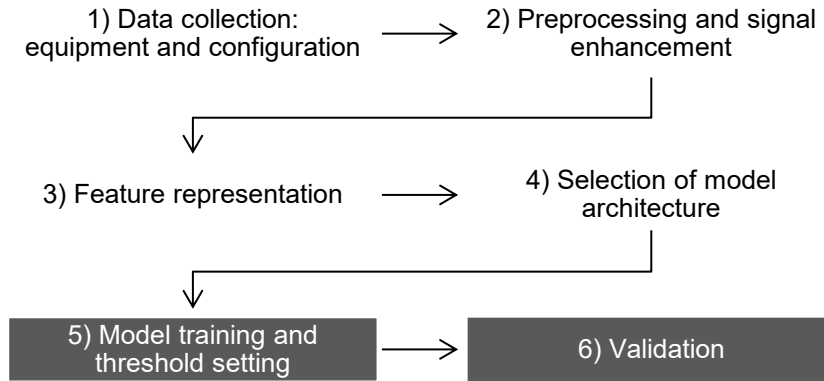


Figure 1: Ghost bat detection ML model stages

2.3 Calculating ML model performance

ML model performance can be quantified in many ways depending on the application (Raja et al., 2024). For the detection of ghost bats, the simple metrics of precision and recall (sensitivity) and their corresponding confidence intervals were considered to be sufficient. The precision and recall formulae are presented in Equation 1 and 2 respectively. The precision metric indicates the likelihood that a detection is a true ghost bat call, while the recall indicates the percentage of ghost bat calls detected versus being missed by the ML model.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

The precision and recall of the ML model are calculated by performing inference (running the model) on a validation set. The validation set must capture the variation in field data for the precision and recall metrics to be indicative of the ML model's performance on real-world data. Calculating confidence intervals is one way of estimating the range in precision and recall based on the size of the validation set. The Wilson Confidence Interval in Equation 3 has been used as it handles edge cases (e.g. no false positives), performs well for small validation sets, is simple to implement and useful for binary classification metrics such as precision and recall (Wilson 1927, O'Neill 2021).

$$Lower\ bound = \frac{\hat{p} + \frac{z^2}{2n} - z \cdot \sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{z^2}{4n}}{n}}}{1 + \frac{z^2}{n}} \quad Upper\ bound = \frac{\hat{p} + \frac{z^2}{2n} + z \cdot \sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{z^2}{4n}}{n}}}{1 + \frac{z^2}{n}} \quad (3)$$

2.4 Expert review

Manual analysis of the spectrograms for training and validation was carried out by the authors using Wildlife Acoustics' Kaleidoscope software. This approach is standard practice for image classification models for detecting fauna. A recognised ghost bat expert reviewed a selection of audio recordings from the training and validation sets to confirm that the calls were correctly labelled.

3 FINDINGS

3.1 Understanding the data

The acoustic monitors are typically deployed at mine shaft (known as adits) or cave entrances. The acoustic environment at these locations is highly complex, varies from site to site, and changes seasonally as well as gradually over longer time scales. The following sections describe the characteristics of the sounds that are present in audio recordings from cave entrances in the Pilbara.

3.1.1 Acoustic characteristics of bat calls

Bats have one of the widest vocalisation frequency ranges of any animal, rivalled only by toothed whales (Fitch 2006). Håkansson et al. (2022) explains that bats have specialised vocal structures that make this vocalisation frequency range possible. Echolocation calls are typically high-frequency, rapid calls produced by vocal membranes, which are adapted for catching fast-moving prey. Social communication calls, on the other hand, often involve lower frequencies produced by ventricular folds. This is similar to the technique used in human

vocalizations like the growls in death metal music. This adaptation allows bats to extend their vocal range and produce distinct social calls. The lower frequency characteristics of social calls enable them to travel longer distances and penetrate through obstacles, making them more effective for communication in various environments (Bohn et al., 2006). Jakobsen et al (2013) investigated the intensity and directionality of bat echolocation signals, finding that bats are capable of altering the loudness of their calls using frequency, duration, intensity and directivity. Open space hunting bats were recorded to have echolocation signals above 135 decibel (dB) sound pressure level, around 30 dB above initial estimates. Loud bat calls can pose an issue for acoustic monitors as the signal can distort, which creates unwanted artefacts that complicate ML model training.

Ghost bats have a complex vocal repertoire consisting of echolocation and social calls which cover a frequency range of 1 – 60 kHz (McKenzie and Bullen 2009; 2011). Various call descriptions such as 'Chirp', 'Chitter', 'Twitter' and 'Squeak' have been used by researchers over the years (Douglas 1967, Guppy et al. 1985, McKenzie and Bullen 2009). Hanrahan (2020) expanded these descriptions by using an unbiased characterisation method to group the vocalisations into five ultrasonic-type calls and nine social-type calls. Hanrahan noted that the most common calls were the 'chirp trill' produced while foraging, the 'squabble' during agonistic interactions, and 'ultra-sonic social' calls.

3.1.2 Target ghost bat classes

Ghost bat calls have been divided into three classes for this study:

- 'Ultrasonic ghost bat calls': defined by pulses which are steeply frequency modulated and typically consist of three or four harmonics. The fundamental harmonic is 12 kHz and up to four harmonics reaching up to 80 kHz. These calls are used for both social and echolocation purposes. They are relatively quiet and can therefore only be detected within a close range.
- 'Audible range ghost bat calls': ranging from 2 – 45 kHz with a frequency modulated section that has energy from 5 to 15 kHz. This relatively low frequency modulation or 'warble' is a defining characteristic of the social call and is unique to the ghost bat. It is audible to the human ear and often described as being like the sound of rubbing two coins together. These calls are relatively loud and can be measured over a wide distance compared to the ultrasonic ghost bat call.
- 'Distorted ghost bat calls': these are primarily social calls that are produced by ghost bats close to the microphone, resulting in distortion of the acoustic signal. These calls retain the fundamental characteristics of the audible range social calls but include significant acoustic artefacts at higher frequencies. Distorted ghost bat calls have a distinct acoustic character which can be identified by the ML model and separated from distorted calls produced by other bat species.

3.1.3 Faint ghost bat calls

Faint ghost bat calls are either ultrasonic or audible range calls produced by ghost bats that are not close enough to the microphone to produce a clear recording. The ML model cannot reliably classify faint calls as positive detections because the images of the calls are lacking the characteristics of the target classes. A human expert has the same challenge and must use additional methods such as listening to the audio and using contextual information to make a judgment with some residual uncertainty. By definition, a faint call has inherent uncertainty, so the probability should fall within a range either side of the threshold. Faint calls will be present in the field data but should be evaluated on a case-by-case basis and not included in the precision and recall calculations.

3.1.4 Other bats

The presence of multiple bat species with overlapping call frequencies requires the models to be highly sensitive and specific to accurately identify ghost bats. Care should be taken to correctly identify ghost bat echolocation and social calls during field surveys as, while they are unique, there is potential to confuse them with some similar Taphozous call types as well as some cave-insect sounds (Bullen 2022).

3.1.5 Ambient noise

The Pilbara region has a range of ambient sounds such as wind, rain, and varying levels of background noise. This diverse range of sound can have similar characteristics to ghost bat calls and result in false positives (e.g. identifying an insect as a ghost bat), or it can overlap ghost bat calls and result in false negatives (incorrectly identifying a ghost bat as a non-ghost bat).

3.1.6 Acoustic artefacts

Like all microphones, the Wildlife Acoustics UM-1 has a limited dynamic range and sounds above the upper limit will cause the signal to distort. Microphones installed at cave entrances are particularly susceptible to signal distortion when bats such as the ghost bat and Taphozous produce loud calls as they enter and exit the cave. Distorted bat calls pose a challenge to the ML model as the acoustic signature of the distortion is similar for all bat species.

3.2 Model classes and training data

The training set for an ML model is fundamental to its accuracy. The general principles for developing an effective training set are to ensure that the data captures the variety in each class, is correctly labelled, does not include irrelevant noise, and contains sufficient data in each class (Koshute et al. 2021). A clean training dataset will ensure the ML model learns the patterns for each class and can identify these features in unseen data. The ML model must be robust to noise in the real world, so it is important that noise is also represented in the training data.

The primary focus for a ghost bat detection ML model is maximising the target class probabilities where ghost bat calls are present. This means that the target classes should have clearly defined characteristics and the training data should consist only of clear calls without the presence of sounds from other classes. Creating separate classes for calls with different acoustic characteristics (e.g. the audible range, ultrasonic range and distorted ghost bat calls) can reduce confusion with other similar sounds by providing clear classification boundaries. However, creating too many classes should be avoided as it can result in low performance on unseen data.

The data for the classes for other bats, insects and ambient sounds is highly variable, so the training sets should include a range of clear and unclear training data (i.e. including noise, faint calls, sounds from multiple non-target classes). This will improve the ML model's ability to generalize and will therefore assign higher probabilities to these classes when encountering ambiguity in real-world data. The resulting confusion between non-target classes is not of consequence to the model performance, provided that there is minimal impact on the target classes. The exception is other bat calls which have similar characteristics to the target classes. Creating separate classes with clean training data for these calls introduces a classification boundary which reduces confusion with the target classes.

3.3 Validation sets

Effective validation datasets are fundamental to the development of a high-performance ML model for detecting fauna in acoustic datasets (Sharma et al., 2022). The validation sets are used to determine the precision and recall of the ML model through the development stage to identify areas for improvements. For applications such as ghost bat detection, it is important to determine the ML performance on a site-by-site basis using location specific validation datasets over different seasons. This will identify underfitting and overfitting of the ML model, which can typically be addressed through retraining.

One year's worth of data for each of the ten sites was provided for this project. The dataset totalled over 2 million audio recordings, with most sites containing approximately one ghost bat per thousand files. The complexity of this dataset required iterative development of the validation set to ensure it contained a range of challenging files for evaluation of the ML model performance. Initially, inference was performed using a proof-of-concept model on a selection of data from each site and positive ghost bat classifications were manually analysed. The proof-of-concept model results were predominantly false positives due to other bats, insects and ambient sounds with similar acoustic characteristics to ghost bat calls and were not present in the training data. The false positives were added to the training and validation sets, and inference was performed using the updated model. This process was repeated until the model precision was satisfactory. The ML model recall was validated by manually analysing all audio files which had probabilities slightly below the ghost bat call classification threshold. The false negatives were incorporated into the training and validation sets and the process was repeated until the model recall was satisfactory.

A summary of the final validation set is shown in Table 1. This validation set was created from a dataset consisting of a randomly selected day each month at each site (a total of 120 days of data), capturing the spatial and temporal variations in the real-world dataset. The level of activity fauna activity differed at each site, resulting in significant differences in the valuation set classes and overall number of files.

3.4 Threshold tuning

Threshold tuning is a step at the end of the ML model optimisation process which can improve site-specific performance. Threshold tuning involves adjusting the target class threshold to achieve a desired precision or recall. The benefit of this technique is that the ML model performance can be optimised for specific sites without having to carry out retraining. For example, a ML model deployed at site with low numbers of ghost bats and very high numbers of Taphozous is more likely to report false positives than sites without Taphozous. Increasing the threshold of the ghost bat ultrasonic call class (which is similar to the Taphozous ultrasonic call) will reduce the number of false positives and improve precision.

Care should be taken when threshold tuning using validation datasets with small numbers of the target class (e.g. Site 8). The ML model probabilities vary based on small changes in the acoustic data. Adjusting the threshold to change a single ghost bat call from a false negative into a true positive to improve recall unlikely to have the same effect on other ghost bat calls and may have unintended consequences on precision.

Table 1: Summary of files contained in validation dataset

Site	Ghost bat calls	Taphozous calls	Other bat calls	Insects and ambient sounds	Total audio files
Site 1	51	11	259	43	364
Site 2	72	1	185	2	260
Site 3	21	1	19	39	80
Site 4	112	3	93	577	785
Site 5	52	0	14	23	89
Site 6	41	373	266	112	792
Site 7	354	6	62	72	494
Site 8	9	1	125	145	280
Site 9	40	0	108	33	181
Site 10	449	0	62	487	998
Totals	1,201	396	1,193	1,533	4,323

4 MODEL PERFORMANCE EVALUATION

The recall and precision results have been calculated for three iterations of the ghost bat detection ML model to illustrate the effect of improving the training and validation sets. The first iteration of the ML model was trained on data from a two-month period from Sites 1 - 4. The second iteration of the ML model included training data from all ten sites covering a full year. The third iteration included data training data from two other sites, one of which had very high numbers of other bat species.

Table 2 presents the predicted precision and recall with the associated 95% confidence intervals for each iteration of the ML model. The same detection threshold has been used for all ghost bat call classes at all sites (i.e. no threshold tuning). These results are a good measure of the change in the model performance with each iteration. The inclusion of training data that is representative of the complexity and variation in real-world data has a significant improvement to the ML model precision but little improvement to the recall. The average improvement in precision from the first to the third iteration was 18%, while the average improvement in recall was just 2%. The changes in performance were minimal at some sites (e.g. Site 2 and Site 3) which is due to the training data for the first iteration being representative of those sites.

Table 3 presents the same ML model results with the inclusion of threshold tuning with a target precision of 95%. The ghost bat call class thresholds at each site have been adjusted to maximise the recall while keeping the precision at 95%. For the final iteration of the ML model, the threshold tuning had an average 2% reduction in precision and a 13% improvement in recall across all sites. The final performance was an average 95% precision and 80% recall across all sites.

Table 2: Predicted precision and recall of three stages of model development (no threshold tuning)

Site	First iteration (limited training data)		Second iteration (training data from all sites)		Third iteration (training data from all sites and additional data from other sites)	
	Precision	Recall	Precision	Recall	Precision	Recall
Site 1	76% CI: 62%-86%	82% CI: 68%-90%	91% CI: 75%-97%	55% CI: 37%-70%	97% CI: 83%-99%	55% CI: 37%-70%
Site 2	82% CI: 71%-89%	76% CI: 61%-85%	88% CI: 78%-94%	72% CI: 59%-81%	95% CI: 86%-97%	69% CI: 56%-79%
Site 3	88% CI: 67%-95%	88% CI: 67%-95%	93% CI: 67%-98%	54% CI: 29%-77%	93% CI: 67%-98%	54% CI: 29%-77%
Site 4	54% CI: 44%-62%	78% CI: 70%-84%	97% CI: 91%-99%	63% CI: 53%-71%	100% CI: 96%-100%	64% CI: 54%-72%
Site 5	93% CI: 80%-97%	59% CI: 43%-73%	100% CI: 92%-100%	73% CI: 58%-83%	100% CI: 92%-100%	68% CI: 53%-79%
Site 6	50% CI: 34%-65%	83% CI: 67%-92%	76% CI: 61%-87%	93% CI: 80%-97%	85% CI: 68%-93%	79% CI: 62%-89%
Site 7	96% CI: 92%-97%	54% CI: 47%-60%	94% CI: 90%-95%	79% CI: 74%-83%	96% CI: 93%-97%	75% CI: 70%-79%
Site 8	34% CI: 13%-64%	77% CI: 46%-92%	81% CI: 54%-94%	100% CI: 77%-100%	86% CI: 57%-96%	92% CI: 65%-98%
Site 9	53% CI: 36%-69%	70% CI: 53%-83%	100% CI: 87%-100%	60% CI: 41%-75%	100% CI: 86%-100%	53% CI: 37%-70%
Site 10	92% CI: 88%-93%	70% CI: 65%-73%	99% CI: 97%-99%	68% CI: 63%-71%	99% CI: 97%-99%	67% CI: 62%-71%
Average across all sites	79% CI: 76%-81%	67% CI: 64%-69%	95% CI: 93%-96%	71% CI: 68%-73%	97% CI: 95%-97%	69% CI: 66%-71%

Table 3: Predicted precision and recall of three stages of model development (with threshold tuning)

Site	First iteration (limited training data)		Second iteration (training data from all sites)		Third iteration (training data from all sites and additional data from other sites)	
	Precision	Recall	Precision	Recall	Precision	Recall
Site 1	95% CI: 83%-98%	73% CI: 57%-84%	95% CI: 82%-98%	67% CI: 51%-80%	95% CI: 84%-98%	76% CI: 61%-86%
Site 2	95% CI: 86%-98%	69% CI: 56%-79%	95% CI: 87%-98%	74% CI: 62%-83%	95% CI: 86%-97%	81% CI: 70%-88%
Site 3	95% CI: 76%-99%	83% CI: 62%-93%	94% CI: 70%-98%	63% CI: 37%-81%	94% CI: 72%-98%	67% CI: 42%-84%
Site 4	95% CI: 88%-97%	60% CI: 50%-69%	95% CI: 90%-97%	96% CI: 91%-97%	95% CI: 90%-97%	96% CI: 91%-97%
Site 5	93% CI: 80%-97%	59% CI: 43%-73%	95% CI: 85%-98%	91% CI: 80%-95%	96% CI: 87%-98%	97% CI: 89%-99%
Site 6	95% CI: 81%-98%	83% CI: 67%-92%	96% CI: 77%-99%	55% CI: 35%-72%	96% CI: 81%-99%	64% CI: 45%-79%
Site 7	95% CI: 92%-97%	68% CI: 62%-72%	96% CI: 93%-97%	77% CI: 72%-81%	95% CI: 92%-96%	80% CI: 76%-84%
Site 8	100% CI: 64%-100%	54% CI: 22%-82%	100% CI: 70%-100%	69% CI: 37%-89%	100% CI: 70%-100%	69% CI: 37%-89%
Site 9	94% CI: 72%-98%	34% CI: 16%-58%	94% CI: 80%-98%	68% CI: 50%-81%	94% CI: 79%-98%	66% CI: 48%-80%
Site 10	95% CI: 92%-96%	68% CI: 63%-72%	95% CI: 92%-96%	76% CI: 72%-79%	95% CI: 92%-96%	78% CI: 74%-81%
Average across all sites	95% CI: 93%-96%	67% CI: 63%-69%	95% CI: 94%-96%	78% CI: 75%-79%	95% CI: 93%-96%	80% CI: 78%-82%

5 LEARNINGS

5.1 ML model trade-offs

The precision and recall metrics can be influenced by the training data, number of model classes and the threshold. The interaction of these components can become complex when high precision and recall is required. Table 4 provides a general summary of the trade-offs between each aspect as observed during the development of this ML model.

5.2 Efficient development of ML models for complex applications

Developing an ML model can take significant time depending on the complexity of the data and the performance requirements. There is additional time to factor in with the deployment of ML models in the field, ongoing validation and management of data. Efficient development is therefore an important consideration.

The early stages of the development of the ghost bat detection ML model were relatively inefficient due to a limited understanding of the data. Considerable effort was spent developing training and validation sets that were not representative of the variation and complexity of real-world data. A more efficient approach was then used, which involved using a separate simplified model to perform inference on a random selection of data that was spatially and temporally diverse. This approach enabled the data to be filtered for samples with similar

characteristics to ghost bat calls that could be manually analysed for validation and retraining. This approach is effective for complex applications such as detecting ghost bat calls.

Table 4: Factors observed to affect precision and recall for a ghost bat detection ML model

Model component	Advantages	Disadvantages
Smaller training sets for target species	Lower complexity of dataset can improve probabilities provided there are strong similarities	Data does not capture the full range of calls of acoustically complex calls, lowering precision and recall
Larger training sets for target species	Improves recall by capturing the variation in acoustically complex calls	Large target species training data or unclear training data result in underfitting, reducing the precision of the model.
Small training sets for extraneous sounds	Reducing the amount of extraneous sound training data can reduce underfitting of the model and improve recall, particularly if the extraneous sounds are from other sites.	Too little extraneous training data lowers the precision because unseen data is likely to generate false positives
Larger training sets for extraneous sounds	Improves precision by increasing the probabilities of correctly identifying extraneous sounds	Too much extraneous training data can result in overfitting, lowering recall
Few model classes	Model is able to generalise well and avoids overfitting. There is less confusion between classes, potentially resulting in higher recall	Too few classes for acoustically complex target species can introduce too much variation within the class. Simplifying all calls into a single class also limits the information collected
Many model classes	Improve precision by having clearly defined classes which increase the probabilities. This is particularly important where there are other bats that have similar calls to the target species. Another advantage is that more information is collected	Too many classes can lead to overfitting, resulting in lower recall
Low threshold	Improves recall	Lowers precision
High threshold	Improves precision	Lowers recall

6 CONCLUSIONS

Using PAM to detect ghost bat calls is an effective method for providing insight into the behaviour of this species to inform management and conservation. Automatic identification methods such as machine learning are necessary to efficiently analyze the large datasets, but care must be taken to ensure the models are accurate as the datasets can be highly complex. Effective validation using data that is representative of the complexity and variation in real world data is fundamental to providing confidence in the model’s performance. Average improvements of 16% in precision and 13% in recall were achieved across all sites by expanding the spatial and temporal range of the training datasets and tuning the detection thresholds for each site. The ML model had an average of 95% precision and 80% recall across all sites.

ACKNOWLEDGEMENTS

The authors would Peter Glorie, Samuel Turner and the team at Wood for the opportunity to develop the ghost bat detection ML model and present the results. The authors would also like to thank Rio Tinto for providing the dataset. This dataset has been fundamental to the success of the model.

REFERENCES

- Armstrong, Kyle N (2010). Assessing the short-term effect of minerals exploration drilling on colonies of bats of conservation significance: a case study near Marble Bar, Western Australia. *Journal of the Royal Society of Western Australia* 93: 165–174.
- Armstrong, Kyle N (2011). The current status of bats in Western Australia. Pp. 257–269. In: *The Biology and Conservation of Australasian Bats*, (eds B Law, P Eby, D Lunney and L Lumsden). Royal Zoological Society of NSW, Mosman, NSW, Australia
- Bohn KM, Moss CF, Wilkinson GS. Correlated evolution between hearing sensitivity and social calls in bats. *Biol Lett.* 2006 Dec 22;2(4):561-4. doi: 10.1098/rsbl.2006.0501. PMID: 17148288; PMCID: PMC1833989.
- Brabant, R., Laurent, Y., Dolap, U., Degraer, S., & Poerink, B. J. (2018). Comparing the results of four widely used automated bat identification software programs to identify nine bat species in coastal Western Europe. *Belgian Journal of Zoology*, 148(2). <https://doi.org/10.26496/BJZ.2018.21>
- Bradley, Holly & Bullen, R. & Cross, Sophie & Knuckey, Chris & O'Connell, Morgan & Ottewell, Kym & Reiffer, Scott & Wild, Suzi & van Leeuwen, Stephen. (2024). A Critical Review of Survey Techniques and Perceived Threats for the Threatened Pilbara Ghost Bat (*Macroderma Gigas*). *Journal of the Royal Society of Western Australia*. 106. 10.70880/001c.128232.
- Bullen, R. D. & McKenzie, Norman. (2011). Recent developments in studies of the community structure, foraging ecology and conservation of Western Australian bats. *Australian Zoologist*. 35. 31-43. 10.7882/FS.2011.007.
- Bullen, R. D. (2020). Rio Tinto, Paraburdoo Western Range, Pilbara WA, Acoustic Survey of Ghost Bat Activity, July 2018 to February 2020. Report prepared for Rio Tinto.
- Bullen, R. D. (2022). A review of ghost bat ecology, threats and survey requirements. Prepared for the Department of Agriculture, Water and Environment
- Chalmers, D., MacKenzie, N. G., & Carter, S. (2020). Artificial Intelligence and Entrepreneurship: Implications for Venture Creation in the Fourth Industrial Revolution. *Entrepreneurship Theory and Practice*, 45(5), 1028-1053. <https://doi.org/10.1177/1042258720934581> (Original work published 2021)
- Douglas, A M. (1967). The Natural History of the Ghost Bat, *Macroderma Gigas* (Microchiroptera Megadermatidae), in Western Australia. *The Western Australian Naturalist*, 10(6), 125--138. <https://www.biodiversitylibrary.org/part/311676>
- Ducretet, M., Linossier, J., Sueur, J., Mathevon, N., Sèbe, F., & Hauptert, S. (2025). *Bridging Passive Acoustic Monitoring and Essential Biodiversity Variables with detectability*. <https://doi.org/10.22541/au.174308491.17489353/v1>
- Ethogram of Ghost Bat (*Macroderma gigas*) Behaviours and Associated Social Vocalisations. *Acta Chiropterologica*, 24(1). <https://doi.org/10.3161/15081109acc2022.24.1.016>
- Fitch, T. (2006). *Production of Vocalizations in Mammals* (pp. 115–121). <https://doi.org/10.1016/B0-08-044854-2/00821-X>
- Guppy, A., Coles, R.B. & Pettigrew, J.D. 1985. Echolocation and acoustic communication in the Australian ghost bat, *Macroderma gigas* (Microchiroptera: Megadermatidae). *Aust. Mammal*. 8: 299–308.
- Håkansson, J., Mikkelsen, C., Jakobsen, L. H., & Elemans, C. P. H. (2022). Bats expand their vocal range by recruiting different laryngeal structures for echolocation and social communication. *PLOS Biology*, 20(11), e3001881. <https://doi.org/10.1371/journal.pbio.3001881>
- Hanrahan, N., Turbill, C., Armstrong, K. N., Armstrong, K. N., Dalziell, A. H., & Welbergen, J. A. (2019). Ghost bats exhibit informative daily and seasonal temporal patterns in the production of social vocalisations. *Australian Journal of Zoology*, 67(6), 305–315. <https://doi.org/10.1071/ZO20055>
- Hanrahan NM (2020) The acoustic ecology of the Ghost Bat (*Macroderma gigas*): form, function and applied uses of vocalisations. PhD thesis. Hawkesbury Institute for the Environment, Western Sydney University, Sydney, Australia
- Hanrahan, Nicola & Dalziell, Anastasia & Turbill, Christopher & Armstrong, Kyle & Welbergen, Justin. (2022). Ethogram of Ghost Bat (*Macroderma gigas*) Behaviours and Associated Social Vocalisations. *Acta Chiropterologica*. 24. 195-208. 10.3161/15081109ACC2022.24.1.016.
- Hogue, A. S., & Breon, K. (2022). The greatest threats to species. *Conservation Science and Practice*, 4(5), e12670. <https://doi.org/10.1111/csp2.12670>
- Husain, Badrul & Osawa, Takahiro & Frianto, Dodi & Nandika, Muhammad & Agustine, Dewi & Putri, Sagung. (2024). Deteksi Lubang Jalan Secara Otomatis dari Rekaman Drone Menggunakan Model Machine Learning Berbasis YOLOv5 Instance Segmentation di Kota Pekanbaru, Provinsi Riau, Indonesia. *Jurnal Zona*. 8. 137-146. 10.52364/zona.v8i2.126.
- Koshute, P., Zook, J., & McCulloh, I. (2021). Recommending Training Set Sizes for Classification. *arXiv: Learning*. <https://arxiv.org/pdf/2102.09382.pdf>

- Lang, L., Corceiro, A., Antunes, R., Lima de Aguiar, M., Alves, P., Agostinho, M., Pereira, N., Lopes, C. M., & Gaspar, P. D. (2024). *Autonomous Bat Echolocation Monitoring System with Machine Learning for Real-Time Pest and Biodiversity Management*. 1–5. <https://doi.org/10.1109/dasa63652.2024.10836283>
- Lemen, C. A., Freeman, P. W., White, J. A., & Andersen, B. R. (2015). The problem of low agreement among automated identification programs for acoustical surveys of bats. *Western North American Naturalist*, 75(2), 218–225. <https://doi.org/10.3398/064.075.0210>
- Li, W., Qiu, J., Lei, P., Chen, X., Fan, F., Deng, X., Dai, Y., Deng, Y.-B., Wang, K., & Mei, Z. (2022). A real-time passive acoustic monitoring system to detect Yangtze finless porpoise clicks in Ganjiang River, China. *Frontiers in Marine Science*, 9. <https://doi.org/10.3389/fmars.2022.883774>
- McKenzie, Norman & Bullen, R.. (2009). The echolocation calls, habitat relationships, foraging niches and communities of Pilbara microbats. *Records of the Western Australian Museum Supplement*. 78. 121-153. 10.18195/issn.0313-122x.78(1).2009.123-155.
- Oestreich, W. K., Oliver, R. Y., Chapman, M., Go, M. C., & McKenna, M. F. (2024). Listening to animal behavior to understand changing ecosystems. *Trends in Ecology and Evolution*. <https://doi.org/10.1016/j.tree.2024.06.007>
- O'Neill, B. (2021). Mathematical properties and finite-population correction for the Wilson score interval. *arXiv: Statistics Theory*. <https://arxiv.org/abs/2109.12464v1>
- Raja, V. J., Solaimalai, D. M. G., Rani, D. L., Deepa P. and Vidhya, R. G. "Machine Learning Revolutionizing Performance Evaluation: Recent Developments and Breakthroughs," *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Coimbatore, India, 2024, pp. 780-785, doi: 10.1109/ICSCSS60660.2024.10625103.
- Ruykys, L., Hanrahan, N., & Stokeld, D. (2024). Novel acoustic lure methodology facilitates detection of the cryptic ghost bat at a landscape scale. *Wildlife Research*, 51(1), 1-15. Article WR22189. <https://doi.org/10.1071/WR22189>
- Rydell, Jens & Nyman, Stefan & Eklöf, Johan & Jones, Gareth & Russo, Danilo. (2017). Testing the performances of automated identification of bat echolocation calls: A request for prudence. *Ecological Indicators*. 78. 416-420. 10.1016/j.ecolind.2017.03.023.
- Sharma A, Virmani T, Pathak V, Sharma A, Pathak K, Kumar G, Pathak D. Artificial Intelligence-Based Data-Driven Strategy to Accelerate Research, Development, and Clinical Trials of COVID Vaccine. *Biomed Res Int*. 2022 Jul 6;2022:7205241. doi: 10.1155/2022/7205241. PMID: 35845955; PMCID: PMC9279074.
- Sharma, S., Sato, K., & Gautam, B. P. (2022). Bioacoustics Monitoring of Wildlife using Artificial Intelligence: A Methodological Literature Review. *International Conference on Networking and Network Applications*, 1–9. <https://doi.org/10.1109/NaNA56854.2022.00063>
- Skowronski, Mark D., Harris, John G.; Automatic detection of microchiroptera echolocation calls from field recordings using machine learning algorithms. *J. Acoust. Soc. Am.* 1 April 2005; 117 (4_Supplement): 2552. <https://doi.org/10.1121/1.4788490>
- Tobias, J., & Pigot, A. L. (2019). Integrating behaviour and ecology into global biodiversity conservation strategies. *bioRxiv*, 566406. <https://doi.org/10.1101/566406>
- Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, 22(158), 209–212. <https://doi.org/10.1080/01621459.1927.10502953>