MTSFB TR 009:2021

TECHNICAL REPORT

LOW-COST ACOUSTIC SURVEILLANCE INTRUSION DETECTION SYSTEM (ASIDS)

Preface

Malaysian Technical Standards Forum Bhd (MTSFB) has awarded Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, Shah Alam, Selangor the Industry Promotion and Development Grant (IPDG) to implement the Proof of Concept (PoC) of the Low-Cost Acoustic Surveillance Intrusion Detection System (ASIDS). The duration of this PoC is for 12 months starting May 2019. The PoC is carried out in Endau-Rompin, Johor and Sungai Dusun, Selangor.

This Technical Report outlines the objectives, benefits, scope of work, methodology and result analysis.

Prepared by:

Institute for Big Data Analytics and Artificial Intelligence (IBDAAI)
Complex Al- Khawarizmi
Universiti Teknologi MARA
40450 Shah Alam
Selangor Darul Ehsan

In collaboration with:

Wildlife Conservation Society (WCS) Malaysia

Published by:

Malaysian Technical Standards Forum Bhd (MTSFB) 200401016865 (655368-P)
Malaysian Communications and Multimedia Commission (MCMC) Centre of Excellence
Off Persiaran Multimedia, Jalan Impact
63000 Cyberjaya
Selangor Darul Ehsan

Tel : (+603) 8320 0300
Fax : (+603) 8322 0115
Email : admin@mtsfb.org.my
Website : www.mtsfb.org.my

mtsfbcyberjaya

Contents

Page

Abbreviations

For the purpose of this Technical Report, the following abbreviations apply.

ASIDS	Acoustic Surveillance Intrusion Detection System
ASG	ASIDS Gateway
ASN	ASIDS Node
CSS	Chirp Spread Spectrum
FFT	Fast Fourier Transform
LPC	Linear Predictive Coding
MLE	Mel-log Energies
PERHILITAN	Jabatan Perlindungan Hidupan Liar dan Taman Negara
PoC	Proof of Concept
SPI	Serial Peripheral Interface
TTL	Transistor-Transistor Logic
UHF	Ultra High Frequency
WCS Malaysia	Wildlife Conservation Society Malaysia

LOW-COST ACOUSTIC SURVEILLANCE INTRUSION DETECTION SYSTEM (ASIDS)

Executive summary

This project aimed to propose and perform Proof of Concept (PoC) of the Acoustic Surveillance Intrusion Detection System (ASIDS), which is more cost effective. The ASIDS is a sound base audio recognition system that covers an area of approximately 100 m in radius. ASIDS is embedded with a classifier engine using Random Forest classification and feature extraction of Linear Predictive Coding.

ASIDS demonstrates that the probability of intrusion is at 85% accuracy from the data collected by Wildlife Conservation Society (WCS) Malaysia. This data is obtained via audio recognition with less background noise.

The recent machine learning methods such as Convolution Neural Network and a hybrid artificial intelligent method were deployed in this project to ensure the classifier's optimal accuracy. This project required a real testing area and environment to consider all sounds in the forest as a proofing idea prior to proceed with the real implementation. Training and testing datasets of sound activities were obtained from the forest in a real-time environment. This project would be beneficial for the wildlife protection agencies in maintaining security at a lower cost as it is less power-consuming than the camera trapping surveillance technique.

1. Background

Endangered wildlife is protected in remote areas where people are restricted from entering. However, intrusions of poachers and illegal loggers still occur due to a lack of surveillance covering large areas. The challenges in implementing security in remote areas require special equipment and design to endure rainforest conditions. The common usage of the camera trapping system is constrained due to the limitation of the camera angles and reliance to batteries as the main power source. As reported by Wildlife Conservation Society Malaysia (WCS Malaysia), maintenance work such as changing batteries and memory cards were troublesome. In addition, it would be challenging to transmit video data from remote locations with no cellular network access. In some cases, equipment and cameras got stolen or destroyed by trespassers.

In reacting to intrusion promptly, rangers need a system that is effective and at the same time, requiring low maintenance. Thus, lowering power consumption is one of the potential solutions to less frequent maintenance, leading to cost savings.

2. Objective

The objectives of the project are as follows:

a) to develop an autonomous wildlife intrusion detection system (ASIDS) using recent machine learning techniques and the hybrid computing optimisation method for optimal performance;

b) to architect a working prototype for the system in a real forest environment; and

c) to evaluate the performance of ASIDS with different classifiers in a real forest environment.

3. Scope of work

The site selected for this project was Endau-Rompin National Park. This site is selected because it covers a huge area of approximately 489 km^2. Thus, it is expected that the surveillance in this area is very costly and difficult.

Validation exercise was conducted at Sungai Dusun Wildlife Reserve, Selangor.

The scope of work includes:

a) research in collaboration with WCS Malaysia and Endau-Rompin National Park;

b) coordination and placement of Proof of Concept (PoC) in Sungai Dusun Wildlife Reserve through WCS Malaysia; and

c) collaboration with the Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN).

Audio signal is applicable as the main source of input to allow machines to identify a threat or intrusion. The audio signal is useful in scenarios where low power and storage requirements are critical. Audio files require less storage capacity. Audio recordings are 360 degrees over 500 m radius based on the quality of the microphone and the loudness of the source of the sound.

4. Target groups and benefits

4.1 Target group

The target groups for ASIDS are as follows:

a) WCS Malaysia;

b) PERHILITAN; and

c) wildlife sanctuaries.

4.2 ASIDS benefits

Benefits of ASIDS are as follows:

a) low-cost surveillance system allows better cost-to-area coverage;

b) provides real-time intrusion detection alert to notify authorities to act on time; and

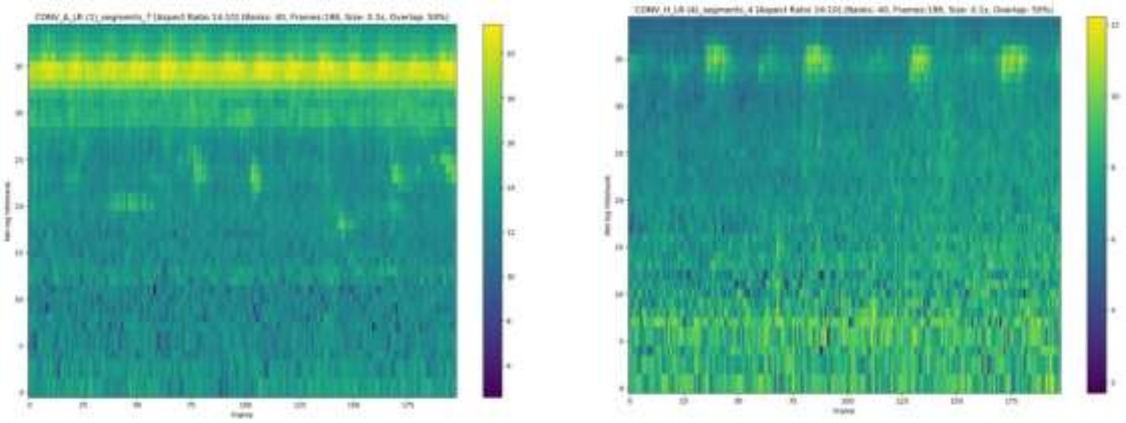
c) novel sound-based wide-area detection system that covers a large radius.

5. Methodology

5.1 Introduction to illegal forest intrusion detection via audio signal approach

Currently the action that prevents forest from unlawful logging relies on patrolling and reporting by witnesses. Using audio signal approach, classifying the difference of intrusion or non-intrusion in real world scenarios with noisy environment is challenging.

ASIDS utilised the Mel-log Energies (MLE) feature that include Fast Fourier Transform (FFT) of the sound signal extracted information. MLE is selected as its computational complexity is relatively low compared to other similar available method and the detection is quick. Figure 3 shows MLE features of variety of audio signal sources.



Natural forest background **Wood cutting sound**

Figure 3. The audio spectrum of natural forest background and wood cutting sound

The chain saws logging sound observed by 3D sound chart analysis research concluded that the main features from logging voice is that the FFT energy is almost exclusively often focused on 2.4 kHz ~ 2.9 kHz, which accounts for about 75% of the total energy. In terms of shape distribution, the rest of the energy evenly distributed in more than 1.1 kHz bands.

Figure 4 shows the extracted feature of FFT audio representation of logging activity that is focused on 2.4 kHz frequency. Difference in focused energy can be ideal for classification purposes as different sources focus on different frequencies.

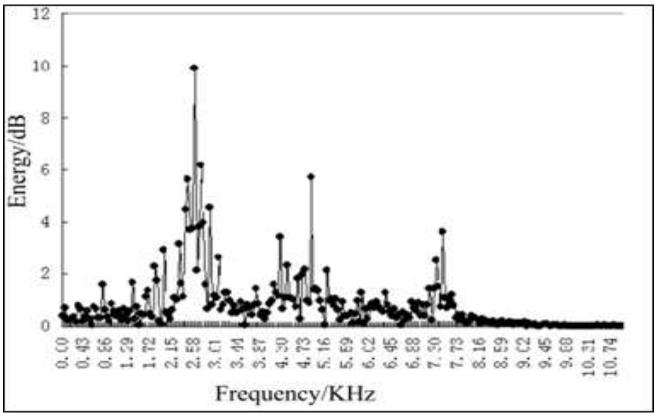


Figure 4. Logging voice signal after FFT

5.2 ASIDS concepts

The ASIDS is an audio recognition system that covers an area of approximately 100 m in radius utilising nodes. Each node is equipped with an on-board computer that conducts processing on site and sends alert messages through a Long-Range communication module using Ultra-High Frequency (UHF) technology.

The system consists of an ecosystem of two major components, which are the ASIDS Node (ASN) and ASIDS Gateway (ASG). ASN acts as the radar for a certain location to detect and send alerts on intrusion. ASG is the concentration point of data where all nodes send status alerts of each node.

Figure 1 is an illustration of the ASIDS ecosystem concept. Each ASN monitors its respective surrounding area and sends data to the ASG for data collection. ASG collects the data from multiple ASNs.

Figure 1. ASIDS ecosystem concept

The ASG is the data center that collects and manages alerts from each individual node. The ASG then displays the node status for the user to observe.

Figure 2 shows the ASIDS data flow, protocols and languages that have been implemented.

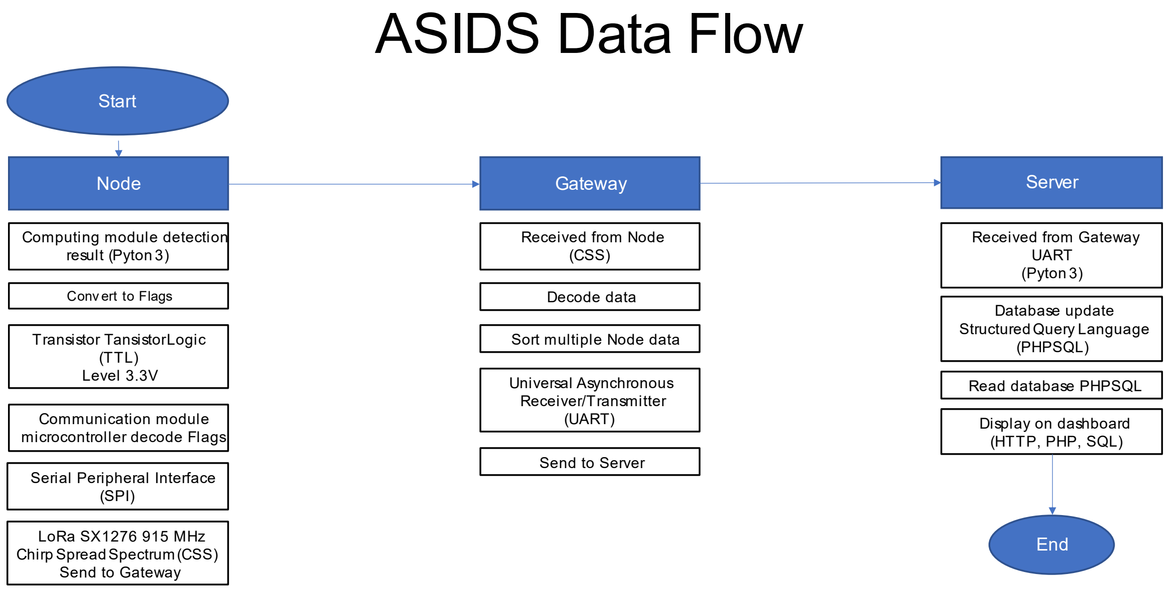


Figure 2. ASIDS data flow

First, the node will compute the detection result and convert them to flags using Transistor-Transistor Logic (TTL) at the level of 3.3 V. Next, the flags will be received by the dedicated microcontroller of the node communications module. The microcontroller will decode the flags via synchronous serial communication of Serial Peripheral Interface (SPI) to communicate with the LoRa SX1276. The LoRa SX1276 will modulate the message using the Chirp Spread Spectrum (CSS).

LoRa SX1276 is also used to receive data from the node at the gateway. It will then decode the data and sort multiple incoming data from multiple nodes. The gateway will send the data to the server using the Universal Asynchronous Receiver/Transmitter (UART) protocol. The server will receive the UART data using Python 3 and update it to the MySQL database. The database will be continuously updated as new information is received. The server will also pull the data and display them on a dashboard for the user.

5.3 Phase 1: Data collection, recording operation and data preparation

Collection of real-world recording of environmental, wildlife and intrusions sound (signal datasets) are conducted in forest areas (National Parks). This phase will involve field trips.

Figure 5 shows the recording site and its specifications.

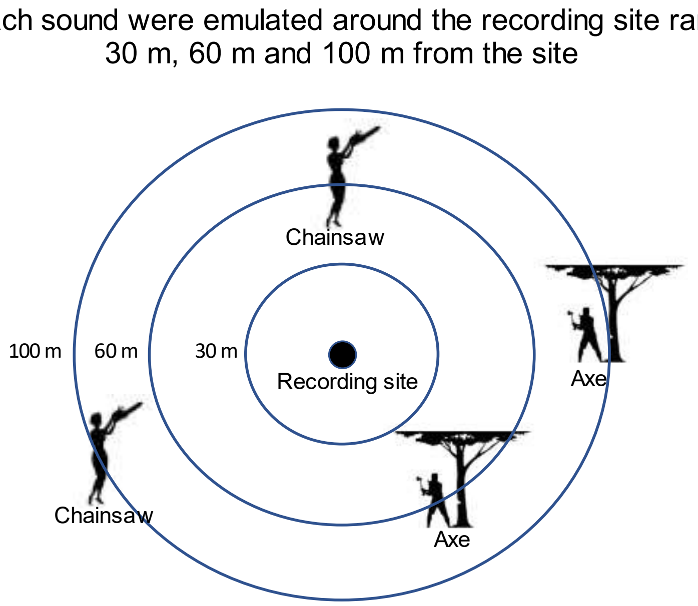


Figure 5. Recording site and its specifications

Figure 6 shows the data collection sites.

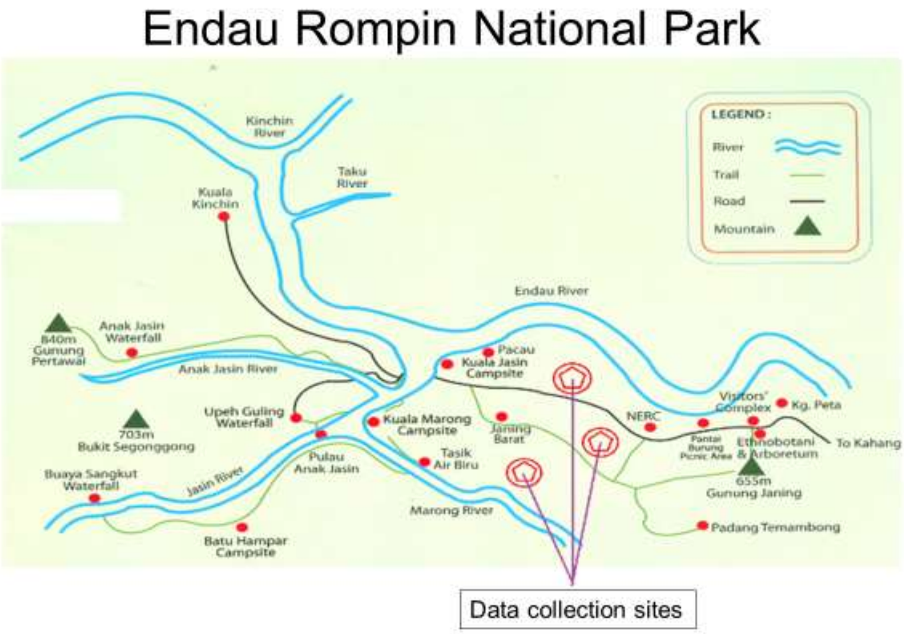


Figure 6. Data collection sites

5.4 Phase 2: Algorithm exploration and evaluation for improvement

Phase 2 involves the exploration of the potential enhancement of the current methods (Random Forest) and recent methods such as Convolution Neural Network and hybrid methods with computing optimisation methods. A high-level computing software language is used. The new algorithm and engine will be evaluated and embedded in the current ASIDS.

5.5 Phase 3: Development of low-cost ASIDS

This phase demonstrates the development of ASIDS in capturing audio input in real time and classifying it as either intrusion or non-intrusion. When the data captured is classified as intrusion, the system gives a warning to rangers via a communication module to prevent further damage to the wildlife reserve.

Figure 7 displays the ASIDS system architecture and explains the main components of the system generally in block diagram form. The system should be able to classify the audio as an intrusion or non-intrusion to allow accurate alarms of intrusions to notify rangers. This is an example of system flow based on Random Forest and Linear Predictive Coding (LPC) as the engine.

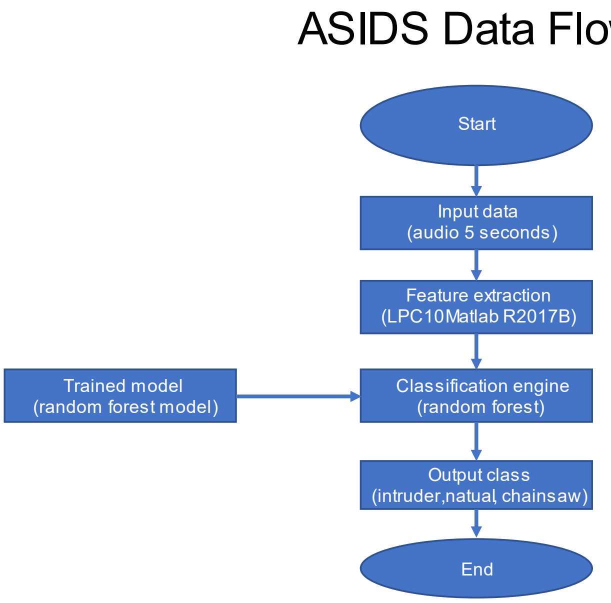


Figure 7. ASIDS system architecture

Figure 8 displays the end product system flow diagram consisting of a loop of real time recording of audio, classifying them and warning the rangers of any intrusions that is detected. In other words, the system is an endless loop of audio monitoring supervised by machine learning and computing technique to allow detection of intruders via audio signal recordings.

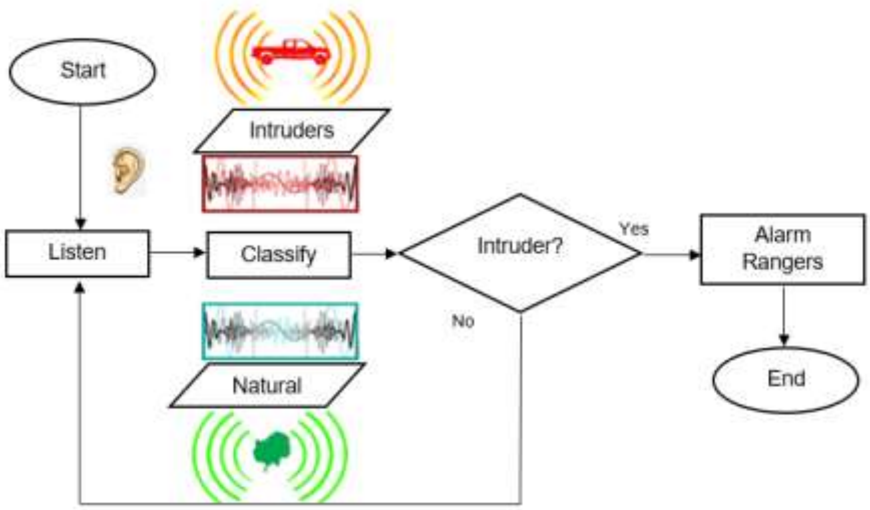


Figure 8. System flow diagram

5.6 Phase 4: Evaluation and analysis and overall proof of concept in a real forest environment

The evaluation and PoC in a real environment were performed in Taman Negara Endau- Rompin, Johor. This phase involved field trips for testing of the novel ASIDS (the improved version of ASIDS) and also comparisons of the current version of ASIDS (built with Random Forest and LPC). Figure 9 shows the testing plan.

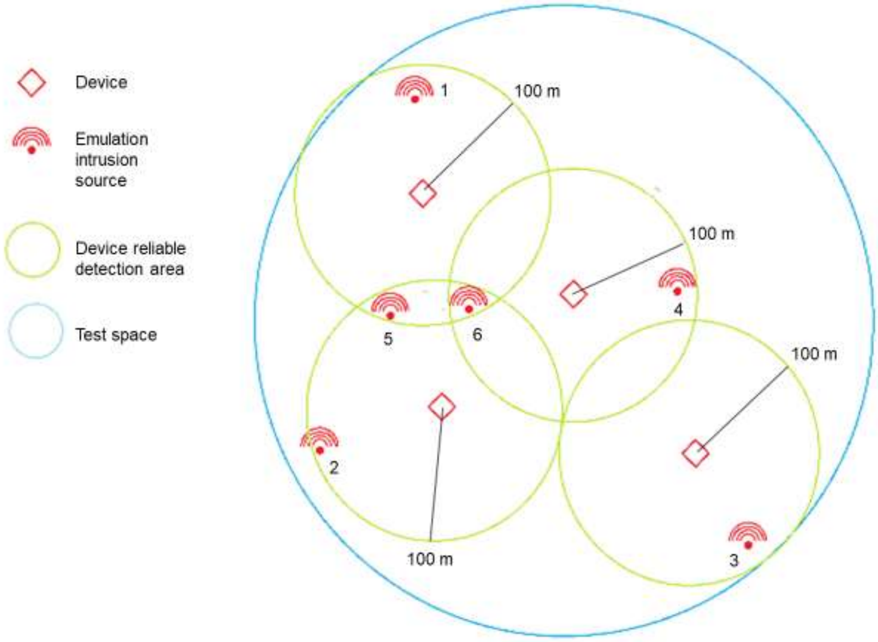


Figure 9. Testing site plan

The result will be compared with existing approaches and benchmarked data. Analysis of results will be based on several criteria such as accuracy rate at every category of intruders.

Figure 10 demonstrates the alert received by the user/rangers of intruders (in this case, vehicle) based on the capturing of the sound via the sensors in the forest. User/rangers can access the ASIDS interface on Android or Web Apps.

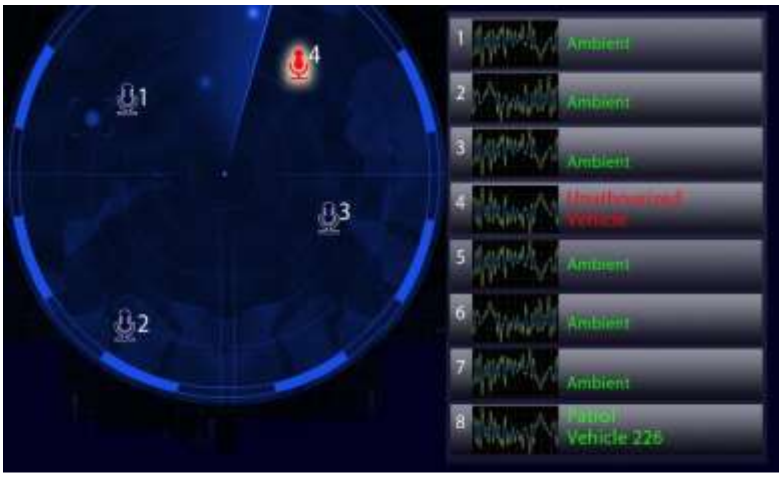


Figure 10. Interface of ASIDS

5.7 Phase 5: Reporting and standard development.

Based on the findings and results analysis of the PoC, this report is prepared to provide an insight of the project itself and furnish recommendations for future standards development in the area of Green ICT.

6. Result analysis

6.1 Validation results

Validation was conducted at Sungai Dusun Wildlife reserve. The results are recorded and analysed in this report. Figures 11-16 show the simulated intrusion of chainsaw activity and cutting a tree using an axe (intruder) against the natural sound of surroundings such as wind, wild animals and insects (no intruder). The results are displayed in ratio percentage of ambience and intruder for three identified distances i.e. 30 m, 60 m and 100 m.

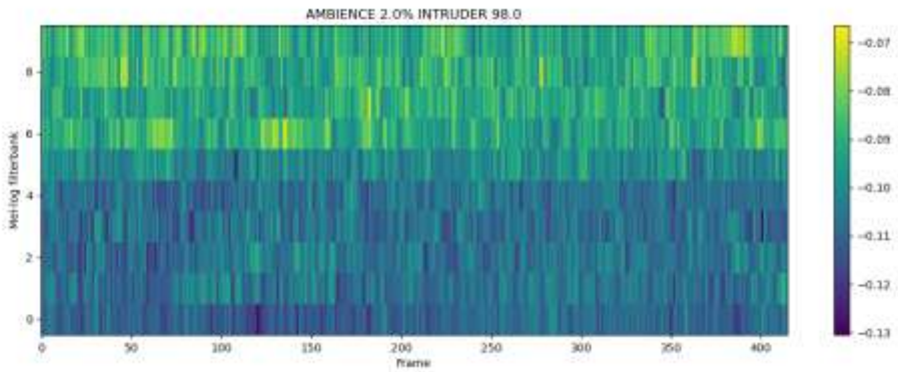


Figure 11. Sound features chainsaw at 30 m

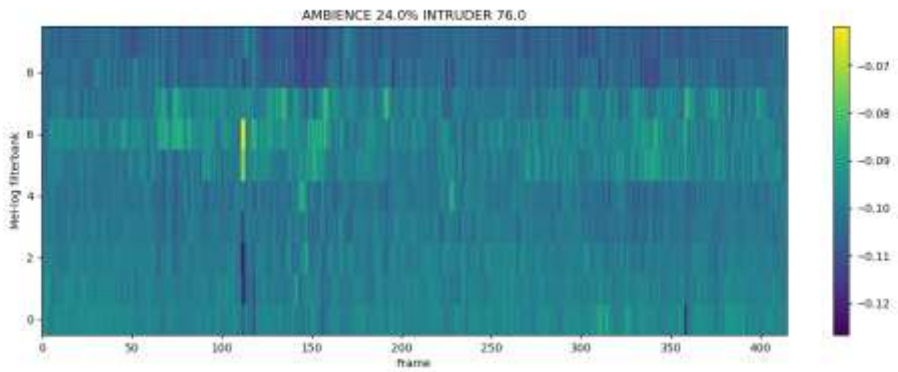


Figure 12. Sound features chainsaw at 60 m

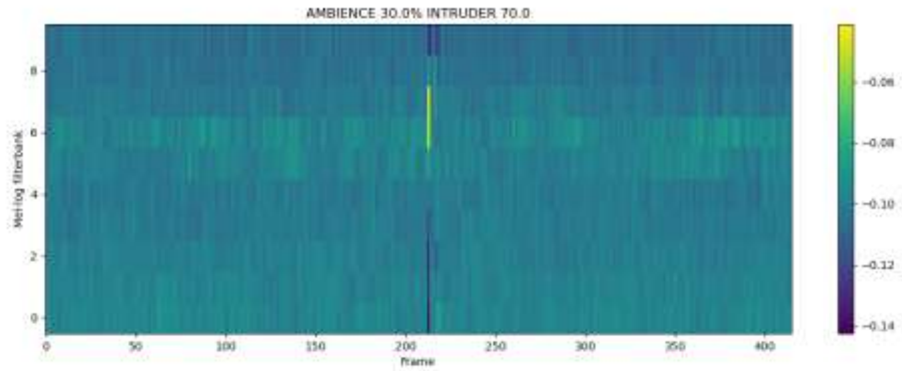


Figure 13. Sound features chainsaw at 100 m

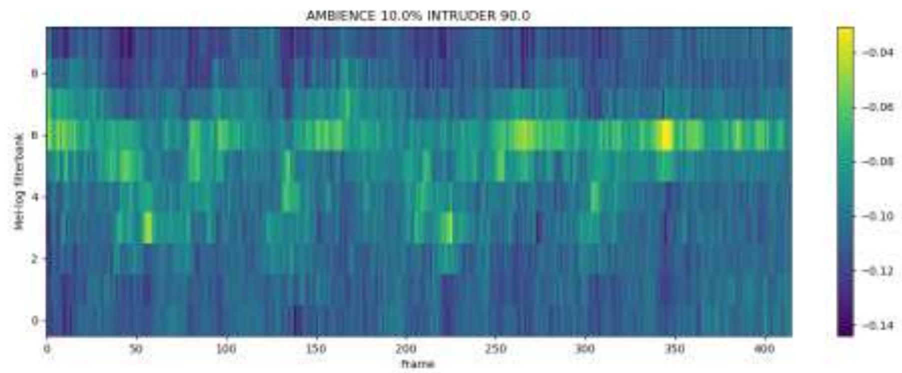


Figure 14. Sound features of cutting tree using an axe at 30 m

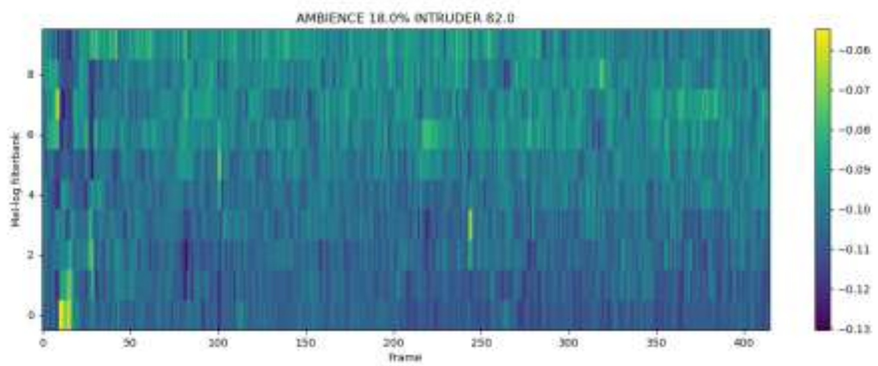


Figure 15. Sound features of cutting tree using an axe at 60 m

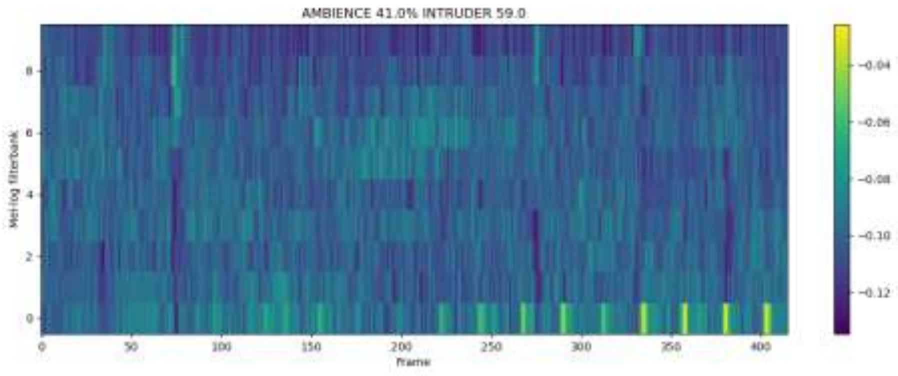


Figure 16. Sound features of cutting tree using an axe at 100 m

The prediction results produced by the algorithm is a probability distribution of class predicted for the specific sound event. In this case, a class prediction which produced more than 50% confidence will be assumed as intruder or non-intruder.

The MLE features are used as descriptors to indicate patterns from distinct sounds. The algorithm can recognise these patterns and employ it for sound class prediction. The yellow areas are the higher MLE valued features while the darker areas are lower in value.

Table 1 shows the summary of results between distances.

Table 1. Summary of results between distances.

Sound	Distance (m)	Prediction probability distribution (%)	
		Ambience	Intruder
Chainsaw	30	2	98
Chainsaw	60	24	76
Chainsaw	100	30	70
Axe	30	10	90
Axe	60	18	82
Axe	100	41	59

It was found that accuracy of a prediction degrades as the distance widens. This is shown by the higher detection results for both simulated sounds nearer to the node, and lower result quality for sounds farther from the node.

It can also be concluded that the results vary between type of sounds. The detection of the chainsaw was easier compared to detecting the sound of tree cutting using an axe. However, the confidence level is still above 50%, which is a positive indication of an intruder being detected at the farthest distance i.e. 100 m.

In addition, considerations shall be made to reduce the rate of false detections. Hence, an increased threshold of confidence is suggested to avoid false detections.

In order to measure the performance of the algorithms in detecting intruders or non-intruders, a confusion matrix is used.

Figure 17 shows a labelled confusion matrix.

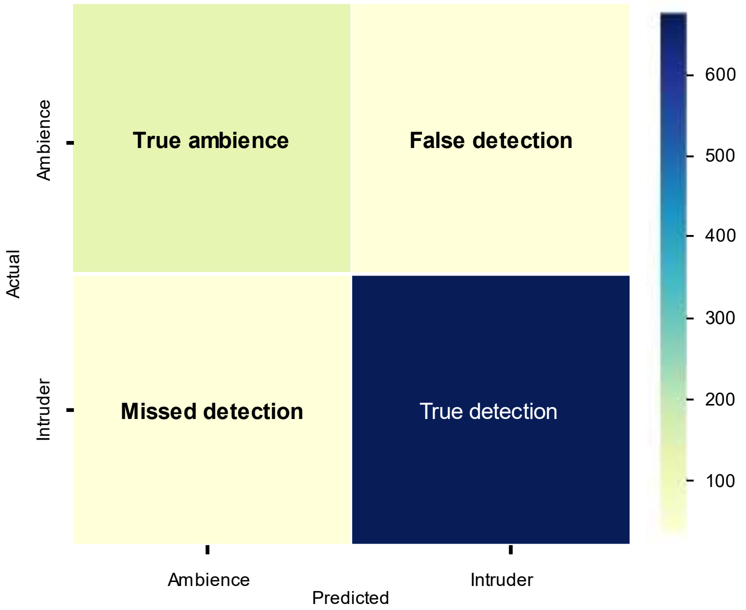


Figure 17. Confusion matrix simplified

The categories of prediction are labelled as follows:

a) True ambience - a prediction of non-intruder when there is no intruder;

b) Missed detection - a prediction of non-intruder when there is intruder;

c) False detection - a prediction of intruder when there is no intruder; and

d) True detection - a prediction of intruder when there is intruder.

The results of the prediction probability distribution are applied to the confusion matrix as shown in Figure 18.

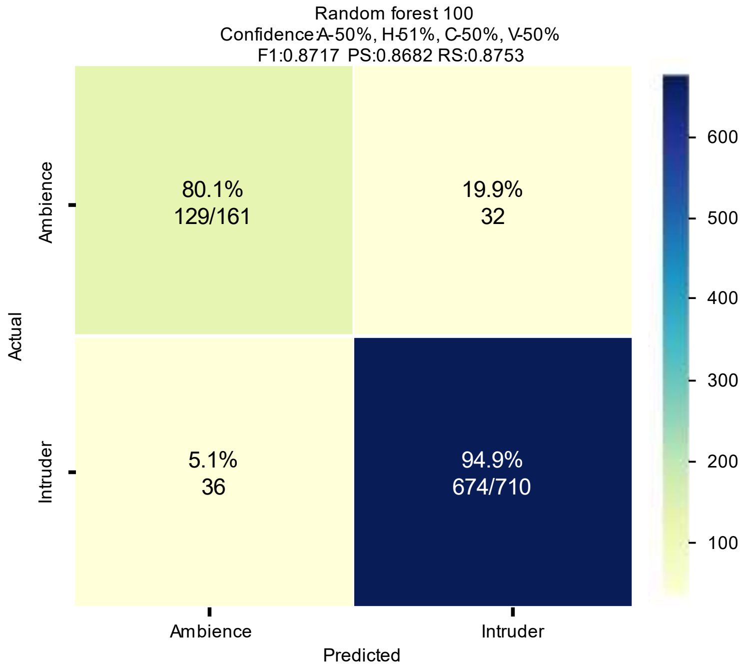


Figure 18. Prediction results (unrefined)

The figure demonstrates the unrefined results with 94.9% correctly predicted as intruders. However, one of the drawbacks is that 19.9% of the ambience is also considered as an intruder. This is considered as a high percentage of false detections. Thus, a thresholding subroutine has been implemented to reduce these false detections.

Figure 19 shows the refined results with the application of increased threshold of 64%, which is producing an approximately of 85% prediction accuracy. This reduces the false detection rate significantly from 19.9% to 6.2%.

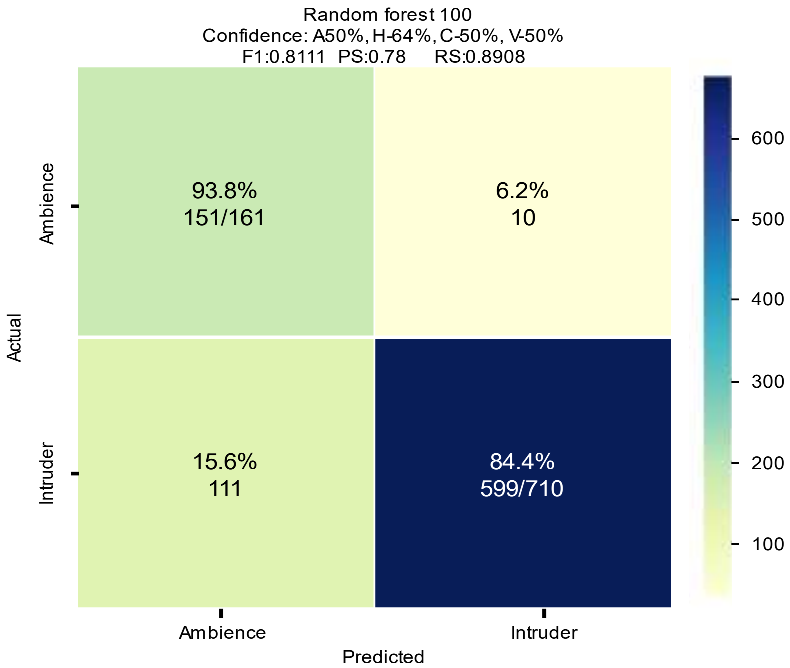


Figure 19. Prediction results (refined)

In order to project the results to include class prediction, the simplified confusion matrix with class prediction is being used. Figure 20 shows the simplified confusion matrix with class prediction.

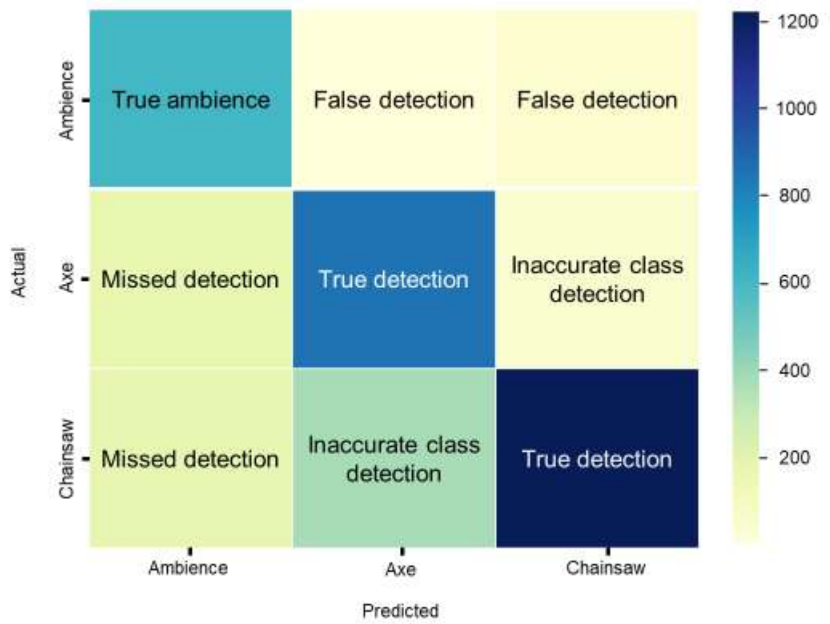


Figure 20. Simplified confusion matrix with class prediction

Figure 21 shows the detailed prediction performance of individual classes.

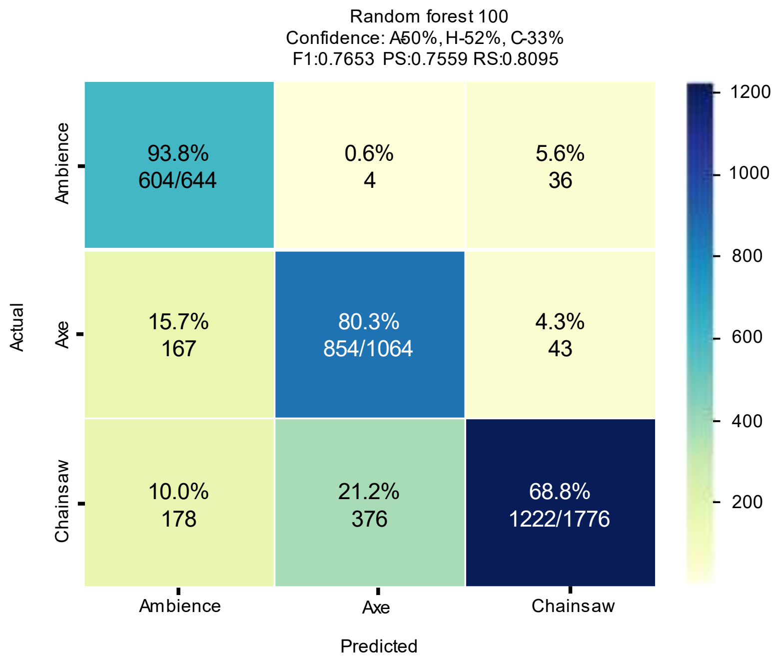


Figure 21. Detailed prediction results of different classes

Based on the performance of algorithm specified, there are mix-ups between prediction of axe and chainsaw of approximately 20%. A thorough scanning of the 11 seconds recording was carried out by splitting the recording into 20 separate sections as a countermeasure.

The final result is based on class intensity of each section. Therefore, the predictions of these two different classes are classified as intruders.

6.2 Node power consumption

The power consumption of the ASN is crucial for the concept to work. The power consumption is measured using an ammeter. It is found that the power consumption varies between the processes of the node.

Figure 22 demonstrates the ammeter used to measure the power consumption of the ASN. It was indicated that the voltage is at 5.10 V at the rate of 0.11 A or 110 mAh. This information can be used to derive the expected battery life of the ASN.

Table 2 displays the data collected using the ammeter. The power consumption reading varies by the type of processes.

The average time in state reflects the amount of relative time used between processes on a percentage basis.

Table 2. Approximate power consumption and computing activities

Type of process	Power consumption reading (mAh)	Watt/hour (W)	Average time in state (%)
Idle	20	0.1	Nil
Recording	80	0.4	80
Detection algorithm	140	0.7	20
Approximate Average	100	0.5	100

It is found that 80% of the time, the ASN will be recording while only 20% is spent on running the detection algorithm. The nature of distribution of time and power consumption of each process is approximately 100 mAh, which is considered as the average consumption of an ASN.

The expected runtime for the ASN will depend on the battery size. The expected runtime of the node can by determined by the formula shown as below:

The formula is used to calculate the expected runtime of the ASN. The formula is shown as below:

$$\frac{Power\ consumption\ (watt/hour)}{Battery\ capacity\ (watt/hour)} = runtime$$

Where,

$Power\ consumption\ (Watts) = Amps\ x\ Volts$

Voltage = 5 V
Amps = mAh/1000

The voltage is calculated based on the standard requirement of a single board computer.
In ideal condition, a 20,000 mAh battery can reach up to 200 hours of runtime.

7. Conclusion

In conclusion, the project has successfully developed an autonomous wildlife intrusion detection system (ASIDS) using recent machine learning techniques and the hybrid computing optimisation method for optimal performance such as such as MLE, FFT, Random Forest and LPC. A PoC has been built to test the system in the real forest environment.

Based on the evaluation:

a) The detection of intruders is improved with the novel ASIDS with a detection rate of 85% in a radius of 100 m surveillance area.

b) Near real-time detection speed is possible, even on limited computing resources.

c) The low power consumption of the ASIDS nodes is considered practical for actual use.

Therefore, this PoC has resulted in a cost-effective surveillance option for wildlife protection.

It is highly recommended that this research is to be continued to fully utilise this novel technology on more applications. This can also be further enhanced for commercialisation.

Bibliography

[1] PERHILITAN. Retrieved April 04, 2018, http://www.wildlife.gov.my/index.php/2016-04-11-03-50-17/2016-04-11-03-57-37/latar-belakang

[2] Bulletin of Electrical Engineering and Informatics:2020, Afendi, A. S. M., Yusoff, M., & Omar, M - *Mel-log energies analysis of authentic audible intrusion activities in a Malaysian forest*

[3] IAES International Journal of Artificial Intelligence:2019, Affendi, A. S. M., & Yusoff, M - *Review of anomalous sound event detection approaches*

[4] International Conference on Soft Computing in Data Science:2018, Yusoff, M., & Afendi, A. S. M. - *Acoustic Surveillance Intrusion Detection with Linear Predictive Coding and Random Forest*

[5] 2017 IEEE International conference on acoustics, speech and signal processing (ICASSP), J. Li et al.:2017 - *A comparison of deep learning methods for environmental sound detection*

[6] 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP):2020, R. Serizel, et al. - *Sound Event Detection in Synthetic Domestic Environments*

[7] AIMS Public Health:2021, K. K. Lella and A. Pja - *Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice*

[8] arXiv preprint arXiv:1708.03211, H. Phan et al. - *DNN and CNN with weighted and multi-task loss functions for audio event detection*

[9] arXiv preprint arXiv:2007.03931 2020N, Turpault and R. Serizel - *Training sound event detection on a heterogeneous dataset*

[10] arXiv preprint arXiv:2101.02919:2021, Q. Wang et al. - *A Four-Stage Data Augmentation Approach to ResNet-Conformer Based Acoustic Modeling for Sound Event Localization and Detection*

[11] Automation in Construction:2020, Z. Liu and S. Li - *A sound monitoring system for prevention of underground pipeline damage caused by construction*

[12] BMC genetics:2018, B. F. Darst et al. - *Using recursive feature elimination in random forest to account for correlated variables in high dimensional data*

[13] Davis, E. (2018, February 28). *New Study Shows Over a Third of Protected Areas*

[14] DCASE 2017-Workshop on Detection and Classification of Acoustic Scenes and Events:2017, Mesaros et al., *A - DCASE 2017 challenge setup: Tasks, datasets and baseline system and classification using extended R-FCN Approach*

[15] Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017:2017, H. Lim et al - *Rare sound event detection using 1D convolutional recurrent neural networks*

[16] E. Cakır and T. Virtanen:2019, *Convolutional recurrent neural networks for rare sound event detection, Deep Neural Networks for Sound Event Detection*

[17] Electronics:2019, M. Z. Alom et al. - *A state-of-the-art survey on deep learning theory and architectures*
https://www.thestar.com.my/news/nation/2019/10/22/10-poachers-detained-almost-rm1mil-wortah-of-wildlife-parts-seized

[18] IEEE Access:2020, T. K. Chan and C. S. Chin - *A Comprehensive Review of Polyphonic Sound Event Detection*

[19] IEEE international conference on acoustics, speech and signal processing:2017, S. Hershey wt al. - *CNN architectures for large-scale audio classification*

[20] IEEE Transactions on Neural Networks and Learning Systems:2018D, W. Otter et al. - *A survey of the usages of deep learning for natural language processing*

[21] *IEEE Winter Conference on Applications of Computer Vision*:2020, B. Barz and J.Denzler - *Deep learning on small datasets without pre-training using cosine loss*

[22] In International Conference on Advances in Computing and Data Sciences:2019 S. Joshi et al. - *Issues in Training a Convolutional Neural Network Model for Image Classification*

[23] Journal of communication disorders, K. Chung - *Perceived sound quality of different signal processing algorithms by cochlear implant listeners in real-world acoustic environments*

[24] Journal of the American association for laboratory animal science(2010), R. P. Reynolds et al. - *Noise in a Laboratory Animal Facility from the Human and Mouse Perspectives*

[25] Neural Networks:2020C, Tian et al. - *Image denoising using deep CNN with batch renormalization*

[26] Proceedings of the Global AI Congress:2019, S. Banerjee and R. Pamula - *Random Forest Boosted CNN: An Empirical Technique for Plant Classification*

[27] Surveyed are Severely at Risk of Losing Tigers. Retrieved April 04, 2018, https://www.worldwildlife.org/press-releases/new-study-shows-over-a-third-of-protected-areas-surveyed-are-severely-at-risk-of-losing-tigers

[28] Workshop on DCASE2017 Challenge, Tech. Rep Dang, et al., Vu, T. H:2017, *Deep learning for DCASE2017 challenge*

[29] Zolkepli, F.:2019 - Ten poachers detained, almost RM1mil worth of wildlife parts seized

Publication of:

Malaysian Technical Standards Forum Bhd (MTSFB)
Malaysian Communications and Multimedia Commission (MCMC) Centre of Excellence
Off Persiaran Multimedia, Jalan Impact, Cyber 6
63000 Cyberjaya, Selangor Darul Ehsan
Tel: (+603) 8320 0300
Fax: (+603) 8322 0115
Website: www.mtsfb.org.my

In collaboration with

Malaysian Communications and Multimedia Commission (MCMC)
MCMC Tower 1
Jalan Impact, Cyber 6, 63000 Cyberjaya
Selangor Darul Ehsan
Tel: (+603) 8688 8000
Fax: (+603) 8688 1000
Website: www.mcmc.gov.my