DETECTING LOGGING ACTIVITY IN FORESTS USING SOUND PROCESSING

Mwaniki Nyaga

School of Computing and Informatics, University of Nairobi,
Nairobi, Kenya
svengali@students.uonbi.ac.ke

Andrew M. Kahonge

School of Computing and Informatics, University of Nairobi,
Nairobi, Kenya
andrew.mwaura@uonbi.ac.ke

Evans K. Miriti

School of Computing and Informatics, University of Nairobi,
Nairobi, Kenya
eamiriti@uonbi.ac.ke

Abstract

Illegal logging and subsequent trade in illegally logged timber is a major problem for many timber-producing countries in Africa. Forestry services in many of the countries lack the machinery and staff sufficient to ensure that each part of the forest is constantly under surveillance. Manning these forests through patrols and guards is logistically impractical to cover due to the vast coverage. Automated efforts in forest monitoring have been suggested that include change detection with the use of aerial photographs. We attempt to tackle this problem through remote sensing by employing sound processing and integrating it with Global Positioning System (GPS) and Global Packet Radio Service (GPRS). Hardware and software tools were designed and developed to record, process and send SMS alerts. We then evaluated the solution under several scenarios. Training of the system was done with Google AudioSet. An accuracy of 88 percent in detection of logging was achieved.

***Keywords*: Forest monitoring, illegal logging, sound processing, Geographical Positioning System (GPS), Global Packet Radio Service (GPRS), Raspberry Pi, Acoustic Audio Detection, Google AudioSet.**

1. Introduction

Monitoring of forests takes place throughout the world for a variety of reasons and at different spatial and temporal scales. Both developing and developed countries have undertaken periodic assessments of forests to ascertain standing timber for many decades [Paivinen R. *et al*., (1989)]. However, concerns about environmental change and its effects on forest ecosystems in the 1980s prompted further monitoring activity. More recently, the need to quantify and examine changes in forest biodiversity and carbon capital has caused new thinking about monitoring, together with growing interest in the capture of periodic information on quality and quantity of forest coverage.

According to [World Forest Organization, (2020)], 15% of greenhouse gas emissions are due to deforestation. This exacerbates the severity of climate change and loss of biodiversity through loss of habitat. Looking at the Kenya scenario, the Kenya Forestry Service (KFS) is in charge of monitoring gazetted forests and ensuring that there is no illegal exploitation of forest resources. It is evident that the agencies mandated with management of forests are understaffed and this impacts negatively on the patrols, monitoring and surveillance of forest areas. They do not have the budget, the machinery, or adequate staff to ensure that this happens. Human activities like logging, charcoal burning and encroachment of forest land continue to be carried out in areas where there is a lack of forest guards and rangers.

Saw millers obtain licenses to permit them to practice logging within plantations. They also have to pay logging fees to the forest department. These collections are usually not sufficient to fund full surveillance in forest reserves, based on the current economic situation. License fees for extracting timber have remained as they were in the 1970s. At the same time, fines for illegal logging and other activities are incredibly low making the timber trade more attractive as it yields very high returns. The saw-millers therefore have no incentive to stick to the guidelines on logging.

The Kenya Forestry Service on the other hand does not have adequate mechanisms to enforce the rules and regulations under the current retrenchment process. Structural Adjustment Programs spearheaded by the World Bank have seen the Government trim the public sector and reduce the budget allocated to its various departments. The Kenya Forestry Service has not been spared from this. Most of the forest guards have been retrenched making it very difficult for the forests to be managed well. This has resulted in weak institutional capacity and poor enforcement of forest laws, which has been identified as a major driver of forest cover change in Kenya [Kariuki S. M., (2016)]. An assessment of the institutional capacity of the local Mt. Kenya forest office to discharge its duties noted that there were over 40 sawmills operating in the Mount Kenya area in 1991 and that these saw mills were equipped with high-tech logging and sawing equipment [Bussmann R., (1993)]. In contrast, none of the forest stations had an adequate number of Lorries, 4 by 4 trucks or tractors to support forest operations. Also noted was that *Chogoria* Forest Station had no vehicle at all, Castle Station only one tractor, while *Chehe* and *Ragati* Forest Stations shared a single old land rover, but could not afford petrol.

Currently, KFS is the lead agency in the assessment and monitoring of the state of forests and forest resources in Kenya. The most prevalent illegal logging activities in Kenya's gazetted forests are primarily assessed and detected through aerial surveys conducted above mountain forests [Environment M. O., (2018)].
Day-to-day forest patrols are conducted on the ground by forest rangers. Due to limited resources and capacity, a forest ranger usually sticks to the same well-worn footpath for their patrol, and only explores further based on intuition or scarce information. In doing so, they could miss logging activity occurring just beyond their regular route.

2. Methodology

The goal was to come up with a prototype that can achieve near instant detection and reporting of logging activity by listening for the whir of a chainsaw, a motor bike, or the sound of a logging truck is detected and then automatically alert local authorities to take action on the environmental crime being committed.

An exploratory research design was adopted to allow flexibility and easy adaptability to change. That way, results obtained could be vital in further work in the same area. The software development employed was evolutionary prototyping. A hardware device based on Raspberry PI was designed, implemented and tested alongside a web based backend. The entire system, both hardware and software, comprised three main modules: The listening module running in Raspberry PI, and two others, the audio process and alert module, running on the web based backend. These are illustrated in Figure 1. Audio files were cleaned and preprocessed before use. To train and test the sound detection module, three types of audio were used. These would be sent to the audio classifier for classification.

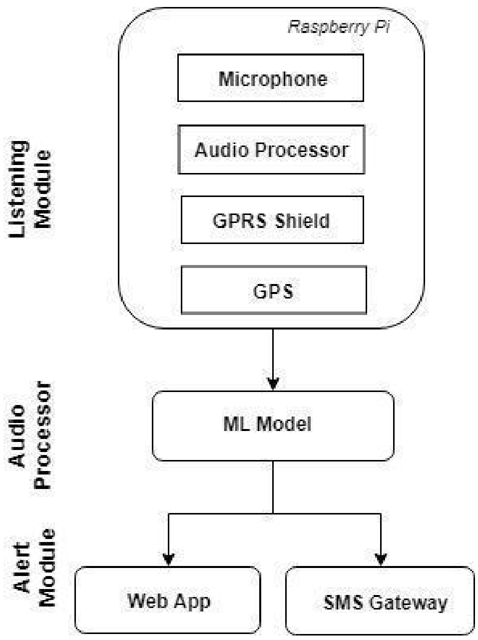


Fig 1. The model

1.1. Listening Module

It consists of a microphone, an audio processor, a GPS module and a GPRS shield all connected to a Raspberry Pi Model 3B. The microphone provides a means to capture environmental sound. The audio processor starts and maintains a continuous livestream, and performs audio segmentation. The GPRS module provides an internet connection for the Raspberry Pi useful for audio upload while the GPS module provides location information for the device. The audio processor is started when the Raspberry Pi boots up. It serves to perform the following functions:

(1) Commence and maintain a live audio stream from the microphone.
(2) Segment the stream into 10 second long PCM wav files.
(3) Upload the files to a cloud server running on Microsoft Azure.

The GPRS shield establishes a Point-to-Point connection with the provided SIM card mobile service provider upon boot up of the Raspberry Pi. The Point-to-Point Protocol provides a standard way to transmit datagrams over a serial link, as well as a way for machines at either end of the link to negotiate various optional characteristics of the link. UNIX distributes a Point-to-Point daemon most commonly used to manage a modem for dial-up or certain kinds of broadband connections [Shimonski R., *et al.,* (2005)]. This is what was leveraged to configure the connection over GPRS to the cloud server. This connection is subsequently utilized in uploading the audio segments created from the livestream to the cloud server for analysis as well as GPS coordinates from the GPS module.

The GPS Module receives global positioning signals from satellites. These signals are processed to identify the geographic location of the GPS receiver according to the well-known GPS technology. The GPS device is powered by the USB interface of the Raspberry Pi Model 3B. It reports its location as GPS coordinates with accompanying timestamps periodically to the web app in order to keep track of the device's whereabouts.

1.2. Audio Processor

This module contains an audio classification model capable of recognizing a signed 16-bit PCM wav file as input, generate embeddings, apply PCA transformation/quantization, use the embeddings as input to a multi-attention classifier and output top 5 class predictions and probabilities. The model currently supports 527 classes which are part of the AudioSet Ontology. The model was trained on AudioSet as described by [Yu C. *et al.,* (2018)].

In training the model, audio required to be featurized. The featurization process involved extraction of input features for the model from audio waveforms and to post-process the model embedding output. VGGish was used as the feature extractor [Simonyan K. and Zisserman A., (2015)]. It converts audio input features into a semantically meaningful, high-level 128-D embedding which can be fed as input to a downstream classification model. The downstream model thus can be shallower than usual because the VGGish embedding is more semantically compact than raw audio features. The architecture of VGGish used had:

(1) The input size as 96x64 for log mel spectrogram audio inputs.
(2) Four groups of convolution/maxpool layers.
(3) A 128-wide fully connected layer.

VGGish was trained with audio features computed based on [Hershey S. *et al.,* (2017)] as follows:

(1) All audio is resampled to 16 kHz mono.
(2) A spectrogram is computed using magnitudes of the Short-Time Fourier Transform with a window size of 25 ms, a window hop of 10 ms, and a periodic Hann window.
(3) Mel spectrograms are computed by mapping the spectrogram to 64 mel bins covering the range 125-7500 Hz.
(4) A stabilized log mel spectrogram is computed by applying log (mel-spectrum + 0.01) where the offset is used to avoid taking a logarithm of zero.
(5) These features are then framed into non-overlapping examples of 0.96 seconds, where each example covers 64 mel bands and 96 frames of 10 ms.

The released AudioSet embeddings were post-processed before release by applying a PCA transformation (which performs both PCA and whitening) as well as quantization to 8 bits per embedding element. The wav files from the Raspberry Pi are uploaded to the cloud server on Microsoft Azure which runs the model. Each file is analyzed and the top five class predictions and their probabilities are output.

1.3. Alert Module

This is a prototype web application consisting of a database, the Internet, web clients and associated job queues. The database is hosted on Microsoft Azure for access by all registered Raspberry Pi devices. The registration is done by each corresponding owner of a device and involves specifying its name as well as assigning a forest ranger to it. The server maintains a job queue that filters out the audio processor's predictions in search of *chainsaw* or *vehicle* classifications. If a prediction is found to belong to either of these classes, the alert module is triggered. The Raspberry Pi's associated details, that is its assigned forest ranger's number as well as its location coordinates are retrieved from the database. An SMS is generated consisting of a description of the type of sound filtered out, its probability of having occurred as well as a link to Google Maps which, if launched, plots a pin of the actual location of the device. This SMS message is sent out to the assigned forest ranger. Authorized clients receive logging occurrence details immediately they are sent out as SMS on their account dashboards showing the details of the sound detected, probability of that sound having occurred, ranger in-charge of the device, the location of the device and the state of the alert (whether it is still *pending* or has been *addressed*). The authorized client has the ability to mark an alert as *addressed* after appropriate action is taken. The location is displayed on a GIS digital map similar to the one shown on figure 2 below.

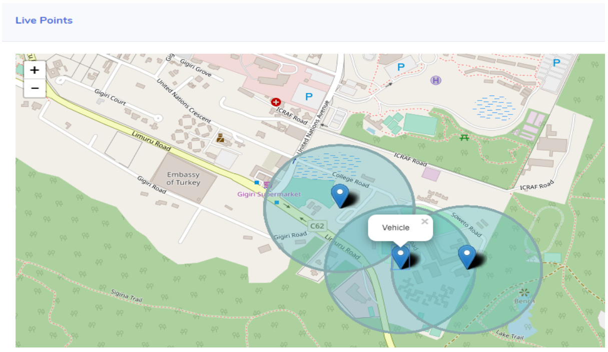


Fig 2. Sample logging site display on GIS map

In a practical scenario trees with thick and high canopies will be installed with listening modules. Therefore the database must be able to manage and distinguish information sent by all listening modules. For this purpose, information must be available to the web server about all listening modules that are installed and to whom they send out notifications to. Whenever a listening module is installed, information associated with its name and to which ranger it is assigned are added to the database. The front end software has a module for this maintenance through a web interface. Access to the database is restricted to authorized users only and therefore information about all owners or managers of devices registered on the system must be maintained.

1.4. The detection process

The general procedure designed is as depicted in figure 3.

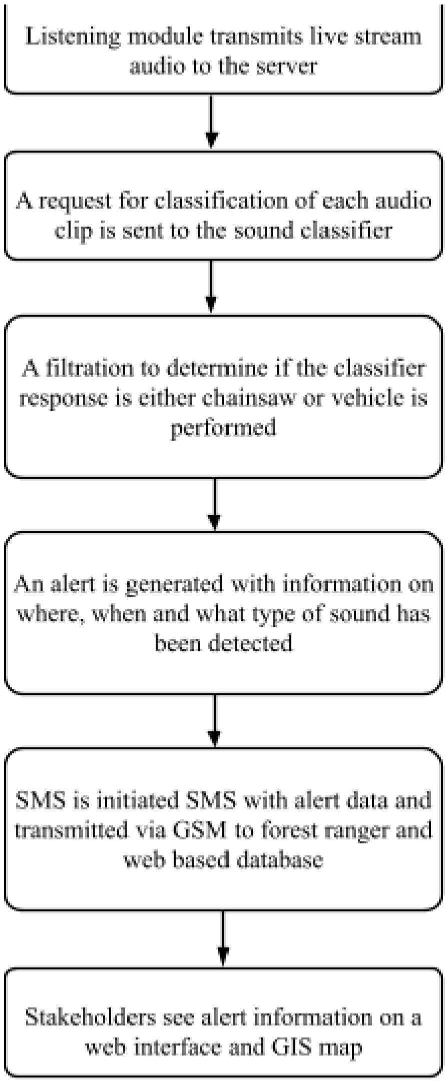


Fig 3. Flowchart of the general procedure

The listening module transmits audio from the livestream to the server. A request is then sent to the sound classifier in order to classify each audio clip. A filtration of the sound classifications is done to determine whether there exists any chainsaws or vehicles detected within the livestream. As soon as either sound is detected an alert is generated which retrieves the most recently reported location of the listening module, the type of sound detected, its probability of having occurred and the time of its detection.

An SMS is then initiated with the aforementioned alert data, inclusive of a link to the location capable of being plotted on Google Maps. It is transmitted via GSM to a pre-configured forest ranger contact as well as a web application database on the server. This message is received at a mobile device containing the SIM card number provided on listening module registration. Aspects of it are also displayed on the dashboard of the web application user who registered the device.

3. Results

After full implementation and testing of the system, evaluation of the prototype was done with the aim to determine if the developed system is delivering the expected results. The following areas were evaluated to provide answers to the research questions set during feasibility study of the project, which are in line with the project objectives and requirements.

1.5. Determining possibility of audio event detection

To ascertain this, three types of audio files were sent to the audio classifier for classification. The files used were chosen among clean pre-processed audio files, live audio from the listening module and filtered audio from the listening module. The predictions were recorded as seen in Table 1 with varying distances from the microphone being the key parameter.

ID	Test Cases	Label	Probability	Distance
3.1	Clean truck	Vehicle	0.71	-
3.2	Clean chainsaw	Chainsaw	0.53	-
3.3	Clean chop	Chop	0.38	-
3.4	Live log007	Chainsaw	0.56	15cm
3.5	Live log009	Vehicle	0.16	15cm
3.6	Live log011	Chainsaw	0.51	15cm
3.7	Filtered log011	Chainsaw	0.30	15cm
3.8	Filtered log 009	-	-	-
3.9	log020	-	-	-
4.0	log022	-	-	-
4.1	log023	Vehicle	0.19	45cm
4.2	log025	Vehicle	0.34	38cm
4.3	log026	Chainsaw	0.24	30cm
4.4	log350	Vehicle	0.34	15cm
4.5	log356	-	-	-
4.6	log495	Chainsaw	0.56	7cm
4.7	log496	Chainsaw	0.52	10cm
4.8	log499	Chainsaw	0.22	40cm
4.9	log500	Vehicle	0.31	50cm

Table 1. Summary of the audio event detection evaluation

3.1 - 3.3 represent clips that have been sourced from AudioSet. Truck and chainsaw score relatively highly based on their probability scores. Chops, however, yield a very low probability score yet the clips are clean, lack noise and are of the proper format.
3.4 - 3.6 represent live audio captured 15cm from the Listening Module's microphone. The probability scores are seen to drop drastically. A misclassification on log009 as Vehicle yet it is a Chainsaw can be attributed to the distance from the microphone the sound is being captured from.
3.7 - 3.8 represent live clips filtered with a high pass/low pass filter. They reveal that filtering yields a lower predictive accuracy, or no classification altogether. This can be attributed to the disparate range of frequencies that chainsaws and revving engines resonate on. For chainsaws their resonance frequency lies between 3800-4200 Hz, while revving engines peak around 300Hz.
4.0 - 4.2 show the misclassification of chainsaws as vehicles when the source moves further away from the microphone. This is further supported by 4.9
4.3 - 4.8 further illustrate the direct proportionality of proximity to probability score. The closer the source is, the more accurately it is classified, and the converse is true.
During the tests carried out in evaluation of audio event detection it was determined reasons why the model may fail to classify the audio event correctly are:-

(1) The range of the microphone used is too low to pick up clear recordings beyond 35cm.
(2) When the physical interface has been interfered with causing disconnection.
(3) When the server goes momentarily offline.
(4) If the listening module is not powered or experiences rapid power fluctuations.

1.6. Measuring success rate of GPS device transmitting spatial data to web server

In carrying out this evaluation, jobs were scheduled to run at intervals of five minutes on the listening module, each time initiating transmission of spatial data to the web server's GPS API. The success rate of this test was recorded in a table in SQLite database. Using SQLite analysis tools, the data was analyzed and the following table presents a summary of the statistics taken.

Number of Jobs Run	Spatial Data Transmission Failure	Spatial Data Transmission Success
50	9	41

Table 2. Evaluation results of Integrating impact data and GPS spatial data

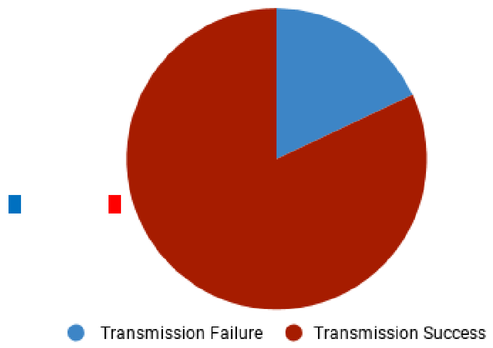


Fig 5. Pie Chart showing the success rate of Transmitting GPS spatial data

Possible reasons for experiencing incidents of failure of the GPS module was evaluated as well. It was found that the following are possible reasons:-

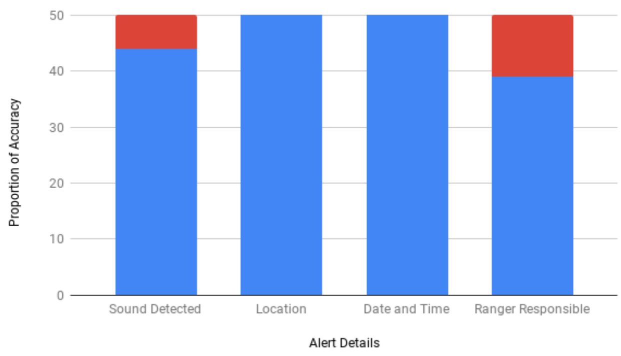
(1) When the GPS device has just started after power on. It takes a mean time of 2 minute and 30 seconds for the GPS device to start functioning.
(2) When the GPS antenna is not attached in place.
(3) When the GPS device is hindered from the sky by obstructing objects like buildings etc.

1.7. Determining success rate of web server receiving logging data

The time taken and the success rate of logging incident update were recorded in a table in SQLite server database. Successful update is in respect to the type of sound detected, its probability of occurrence, the listening module location in geographical coordinates of longitudes and latitudes, date and time of alert generation, and the ranger responsible. Using data that was generated by initiating sound events associated with logging, 50 instances were analyzed. The following table presents a summary of the statistics taken in this experiment.

Details	Correct	Wrong
Sound Detected	44	6
Location	50	0
Date and Time	50	0
Ranger Responsible	39	11

Table 3. Results on the accuracy of various aspects of logging event detection as updated on the server database



Correct Wrong

Fig 6. Accuracy of various aspects of accidents updated on the database

The GPS Device transmitted to the database its unit ID that corresponds with the one that already exists for each Listening Module on the database, enabling the system to attach all other details of the Listening Module supposedly located near logging activity. Most important information on display at each logging incident is the type of sound detected and its location. The evaluation achieved 88% and 100% success rate in this test respectively.

To determine the accuracy of the location coordinates submitted to the database on each impact, a different GPS device – a cell phone handset with GPS capabilities – was used. Coordinates of the logging site as updated on the database were compared to details on the cell phone handset. The following table is a sample of the spatial data comparison on the two devices.

GPS Device Coordinates		Cell phone GPS Coordinates		Variations	
Latitude	Longitude	Latitude	Longitude	Latitude	Long
1°14'10.2"S	36°48'44.0"E	-1.236157	36.812212	Nil	Nil
1°14'14.6"S	36°48'55.2"E	-1.237380	36.815329	-0.19	-0.13
1°14'23.4"S	36°49'00.3"E	-1.239836	36.816745	Nil	Nil
1°14'25.8"S	36°48'59.7"E	-1.240512	36.816584	Nil	Nil
1°14'32.3"S	36°48'51.8"E	-1.242314	36.814374	Nil	Nil
1°14'23.0"S	36°48'40.4"E	-1.239724	36.811211	0.59	0.17

Table 4. Sample of the spatial data of the two GPS devices

It was ascertained from review of GPS literature that one reason why a GPS device could give wrong location data is due to errors introduced on the location data by the owners of the GPS System. However, the variations are minimal and would not affect successful implementation of the system.

1.8. Measuring turnaround time for logging event detection

Based on the data recorded in the table 5 below, time variations of less than three minutes were considered as successful cases of logging event detection. This is because it takes about two and a half minutes to upload audio clips from the listening module to the server via GPRS.

Live Audio Clip Recorded on		Live Audio Clip Analysed on		Variations	
Date	Time	Date	Time	Date	Time
20/05/2019	13:15:00	20/05/2019	13:18:05	Nil	00:03:05
20/05/2019	13:15:10	20/05/2019	13:18:06	Nil	00:02:56
20/05/2019	13:15:20	20/05/2019	13:18:07	Nil	00:02:47
27/05/2019	13:15:30	27/05/2019	13:18:08	Nil	00:02:38
27/05/2019	13:15:40	27/05/2019	13:18:09	Nil	00:02:29
27/05/2019	13:15:50	27/05/2019	13:20:50	Nil	00:05:00
27/05/2019	13:16:00	27/05/2019	13:20:51	Nil	00:04:51

Table 5. Time comparison between recording and analysis

Cases where time variation exceeded three and a half minutes were considered failures. Factors that caused the Listening Module to relay audio with delay are:-

(1) The model cannot analyse partial files and therefore significant time is given to write complete files.
(2) To reduce the number of jobs run for uploads one minute is given for batch processing.
(3) Congestion on the GPRS Network.

The GSM/GPRS Network used in this research is Safaricom, which has a wide coverage in Kenya most significantly in remote areas. Therefore, determining the GSM provider to use is critical in achieving desired results.

4. Conclusion

This project was indeed a good opportunity to unveil what an innovation using microphones, GPS, GSM/GPRS networks, machine learning, databases, and cloud services could achieve. Literature cited alludes that there exists a gap in prompt illegal logging detection and reporting mechanisms especially in this country. Indeed this research comes in handy as a technology that will instantly detect and report logging activity in unauthorized areas will undoubtedly save trees, forests and the natural ecosystem at large. Results from the evaluations carried out verify that the Listening Module could put together spatial data received from GPS satellites together with audio data from the surrounding. The results indicate an excellent success rate of 88% logging event detection and as such asserting this configuration. This demonstrates the viability of detecting logging activity and determining its location and time – vital aspects in performing intervention activities to save trees.

Further experiments with the prototype confirmed that live audio and spatial data were successfully transmitted from the Listening Module to a remote web based server using GPRS wireless network. Although other systems have demonstrated this capability before, it has been proven in this project that audio and location data can be transmitted to a web server on the cloud where further processing can be achieved. The evaluations conducted in this area were overwhelmingly encouraging with sound detected correctly at the rate of 88%, the logging location coordinates at 100%, impact date and time at the rate of 100% and alerts sent to the correct forest ranger responsible at the rate of 78%. Therefore, these results demonstrate the viability of this solution.

The ultimate objective is to assist monitoring managers as well as forest rangers to know the presence and location of logging activity as it happens anywhere within the protected sections of a forest reserve. Using an acoustic event detection model learned from Google AudioSet data, a Listening Module, an SMS sending API and an authorized user dashboard, the detection and consequent alerting of relevant authorities on logging activity has been realized. The low probabilities predicted for live audio data were as a consequence of noise emanating from the procured microphone. The potential errors primarily relate to the audio quality realized from the microphone used. However, these results are encouraging and consistent with the accuracy goal as very little misclassification has been observed to occur from the use of live audio data. This in upshot means that logging activity can instantly be reported to authorized personnel and agencies as and when it occurs with all the necessary details to guide the authorities in rescue operations.

The various evaluations carried out in the project facilitated accumulation of up to 50 records of simulated logging events. The sample data on acoustic logging detection is similar to what researchers in this area may query and carry out data mining exercises with, discovering new knowledge on logging activity; knowledge which in turn would be useful in mitigating further deforestation.

1.9. Further work

There is room for further research to be carried out to enable some of the following. Logging activity reporting by authorized users through mobile phones; Use of low cost (audio) recorders (LCRs) represents a new opportunity to investigate the sonic complexity of both natural and urban ecosystems; Reduction of the hiss, low-hum and ambient noise of the microphone using XY cardioid condenser microphones can be used as soundscape recorders; Data warehousing and data mining models on logging data to discover new knowledge and logging patterns that would help in reducing deforestation; Determinations on how audio filtering can be used to improve clarity of recorded audio and the range the device can listen to; Use of solar in order to power the Listening Module beneath the canopy; Determinations on how well the system would function in case of listening modules involving many logging events; Future testing should include other soundscapes that were unable to be picked up due to microphone quality such as hand saws, wood axe chopping, and car horns; Cryptography can be used to encrypt the transmitted live audio from the Listening Module, and decrypt it at the audio processor, thus reducing brute-force attacks.

5. Acknowledgment

The authors wish to thank the School of Computing and Informatics at The University of Nairobi for availing their laboratories for assembly of the system hardware prototype.

6. References

[1] Bussmann, R. (1993). *The forests of Mount Kenya (Kenya) - A phytosociological approach with special reference to ecologial problems*. Universität Tübingen.

[2] MEF (2018). *Republic of Kenya Ministry of Environment And Forestry Taskforce Report on Forest Resources Management and Logging Activities in Kenya*. Repository of Kenya Government Documents

[3] Hershey, S. et al. (2017) *CNN Architectures for Large-Scale Audio Classification*. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 131-135, doi: 10.1109/ICASSP.2017.7952132.

[4] Kariuki, S. M. (2016). *The Role of the Kenya Forest Service in the Management of Land Degradation and Environmental Conflict in The Mau Forest Complex*. The University of Nairobi.

[5] Paivinen, R., Van Brusselen, J. and Schuck, A. (1989) "The growing stock of European forests using remote sensing and forest inventory data", Forestry. Narnia, 82(5), pp. 479–490, doi: 10.1093/forestry/cpp017.

[6] Shimonski, R., Cross, M. and Laura, H. (2005). "'*Network Operating Systems', Network Study Guide & Practice Exams*". Syngress, pp. 493–553, doi: 10.1016/B978-193183642-5/50012-6.

[7] Simonyan, K. and Zisserman, A. (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition". international conference on learning representations

[8] World Forest Organization (2020). "*The Impact Of Burning Charcoal On Kenya's Ecosystems And The Transition To A Sustainable Future*", The World Forest Organization.

[9] Yu, C. et al. (2018). *"Multi-level Attention Model for Weakly Supervised Audio Classification"*. Electrical Engineering and Systems Science