International Journal of Engineering & Technology, 7(4.16) (2018) 87-91

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Deep Transfer Learning based Acoustic Detection of Rice Weevils, *Sitophilus Oryzae* (L.) in Stored Grains

Stephenn Rabano[1], Melvin K. Cabatuan[1]*, Elmer Jose P. Dadios[1], Edwin J. Calilung[1], Ryan Rhay P. Vicerra[1], Edwin Sybingco[1], and Efren R. Regpala[2]

[1]Gokongwei College of Engineering, De La Salle University – Manila, Taft Avenue, Manila, Philippines
[2]Research and Development, Philippine Center for Postharvest Development and Mechanization, Science City of Muñoz, Nueva Ecija, Philippines
**Corresponding author E-mail:melvin.cabatuan@dlsu.edu.ph*

Abstract

The presence of rice weevils is causing degradation of rice quantity and quality during storage. Classifying rice grade is critical since rice weevils are not easily detected. This study used deep transfer learning on spectrogram images of sounds to recognize the presence or absence of rice weevils in a sound clip. There are 1000 audio files with rice weevil presence and 1000 audio files with the absence of rice weevils in the dataset, each having a duration of 5 seconds. Random environments and random number and age of insects were considered to have models that are less dependent on the environment setting. The dataset was preprocessed to generate the spectrogram image of each audio clip. Features of those images were extracted to train some pre-trained Keras models on the dataset. In the dataset, 1400 images were used for training and 600 were used for testing. Each among the models Xception, ResNet50, InceptionResNetV2, and MobileNet obtained a rank-1 accuracy of 99.17% while VGG16, VGG19, and InceptionV3 all got a rank-1 accuracy of 99.33%. The average precision, average recall, and average F1 score in each trained model are all 99%. These account for the effectiveness of using deep transfer learning on spectrogram images of audio recordings in the detection of rice weevils in stored grains. This is also the first study that used deep transfer learning on spectrogram images in the acoustic detection of rice weevils.

Keywords: *acoustic detection, deep transfer learning, librosa, Keras, Sitophilus oryzae (L.).*

1. Introduction

Grain production in any country varies yearly. It is necessary that grain is strategically stored from years of over-production for use in years of under-production. Another reason for storage is that the time of production is not the time of consumption. Stored grain can have quantity and quality losses. These losses are attributed to infestation by insects, mites, rodents, birds, and microorganisms [1]. These infestations lower crop values. Insects do not only consume grain. With their metabolic byproducts and body parts, they contaminate the grain as well. Some produce heat and moisture leading to growth of microflora and the development of hotspots in grain. When the grains are heavily infested, they are not suitable for seed purposes making them unfit for human consumption.

Grain infestation starts right before the crops are harvested in grain-growing regions [2]. It is a practice that grains are stored year after year in the same bins, not always properly cleaned, attributing to quick infestation of stored fresh grain. Recognition of the presence of insects such as the rice weevil with the scientific name *Sitophilus oryzae* (L.) will aide in the regular bin inspection. If an infestation is determined, measures such as fumigation may be applied right away.

Different pest infestation detection techniques are already discovered. Manual samples, traps, and probes have been used to detect insects on farms [1]. At present, manual inspection, sieving, cracking-floatation and Berlese funnels are used to determine the presence of insects in grain handling facilities. The said methods are considered inefficient and time-consuming. The usefulness of some methods have been demonstrated in research laboratories [3], [4], [1]. These methods include acoustic detection, carbon dioxide measurement, uric acid measurement, near-infrared spectroscopy, and soft X-ray method. They are potential at the industry level to detect insects in grain samples. Image analysis programs have also been developed to automatically scan X-ray images to detect insect infestations. Near-infrared (NIR) spectroscopy has been investigated to detect hidden insects in wheat kernels. But X-ray and NIR spectroscopy methods are costly. Current NIR instrumentation is deemed complex in terms of operating procedures and calibrations. Another method uses a mobile application that locates agricultural pests to limit the utilization of pesticides through pesticide monitoring [20].

Acoustical methods use insect-feeding sounds for automatic monitoring of both internal and external grain feeding insects [1]. Amplification and filtering of movement and feeding sounds of insects aid in acoustical detection of insects hidden inside kernels of grain. A disadvantage though of acoustical methods is the inability to detect dead insects in grain and infestation by early larval stages of insects.

For an adult rice weevil, there is a major peak of energy in the frequency spectrum, moving in a frequency range from 1.8 kHz to 3.0 kHz, with a small resumption peak in the range 3.3 kHz to 3.8 kHz [5]. For the rice weevil in the larval stage, the peak of energy is in a narrow frequency range from 1.3 kHz up to 2.0 kHz.

A combination of sound parameterization and neural network was used in a study for identification of insect sounds [6]. Each acous-

tic signal was preprocessed and segmented. MFCC values were used in the training of the probabilistic neural network (PNN). Another study used sound parameterization with MFCCs but the classification is through the hidden Markov model (HMM) [7]. There is also a study that used a combination of linear predictive cepstral coefficients (LPCCs), line spectral frequencies (LSFs), and MFCCs to extract features, with the combination of MFCC and LSF giving the best results. In a certain study, MFCC and linear frequency cepstral coefficient (LFCC) methods were used in a support vector machine (SVM) for successful insect classification of 88 species [8]. Successful acoustic methods were also employed in [9] and [10].

A classification system was developed detecting sounds in recordings and classifies them as one of four types: background noise, whistles, pulses, and combined whistles and pulses [11]. A database of underwater recordings made off the Spanish coast during 2011 was used to test the classifier. A sound detection rate of 87.5% was achieved for a 23.6% classification error rate through the use of cepstral-coefficient-based parameterization. Two parameters computed using the multiple signal classification algorithm and an unpredictability measure were included in the classifier to improve the said results. The parameters helped to classify the segments containing whistles, increasing the detection rate to 90.3% and reducing the classification error rate to 18.1%.

This study utilized transfer learning on spectrogram images of audio files to detect the presence of rice weevil sound. The following Keras pre-trained deep neural models were used: Xception, VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, and MobileNet.

2. Deep Transfer Learning: Keras Pre-Trained Models

Transfer learning uses pre-trained models and makes small changes in the architecture [12]. The network weights from the pre-trained model are extracted and transferred to another network instead of training this network from scratch, that is, learned features are transferred. In deep learning, since large amount of data is needed to achieve good results, transfer learning is popular to avoid expensive training of new deep neural networks [13].

Image classification using deep convolutional networks is the most popular transfer learning application [13]. Such models include the Keras pre-trained models. Transfer learning allows the use of deep learning with a small amount of data and lower computational capabilities.

The Keras pre-trained models are used for prediction, feature extraction, and fine-tuning [14]. These are models made available with pre-trained weights. The weights are trained on ImageNet for image classification. These models include Xception, VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, MobileNet, DenseNet, NASNet, MobileNetV2. The last three models stated are not used in this study.

The said architectures are compatible with TensorFlow, Theano, and CNTK backends [12]. When these models are instantiated, they will be built according to the image data format set in the Keras configuration file. The following table shows each model's performance on the ImageNet validation dataset, specifically the top-1 accuracy.

Table 1: Some Pre-trained Keras Models' Performance on the ImageNet Validation Dataset

Model	Size	Top-1 Accuracy
Xception	88 MB	0.79
VGG16	528 MB	0.715
VGG19	549 MB	0.727
ResNet50	99 MB	0.759
InceptionV3	92 MB	0.788
InceptionResNetV2	215 MB	0.804
MobileNet	17 MB	0.665

Xception has a 299x299 default input size, VGG16 has 224x224, VGG19 has 224x224, ResNet50 has 224x224, InceptionV3 has 299x299, InceptionResNetV2 has 299x299, and MobileNet has 224x224 [15].

3. Materials and method

3.1. Audio Dataset

The audio dataset is composed of two clusters, the positive dataset (with rice weevil sound) and the negative dataset (without rice weevil sound), each set containing 1,000 audio clips. The audio clips are of .wav extension, each having a 5-second duration.

Ten positive audio clips and nine negative audio clips were trimmed from [16]. In those ten positive audio clips, the larval stage of the insects in the recordings is from 16 days to 18 days old. Those clips were recorded noise-free using the following acoustic sensors: Bruel and Kjaer accelerometer, piezoelectric disk sensor, PVDF film sensor, 30 kHz ultrasonic sensor, and 40 kHz ultrasonic sensor. The nine negative audio clips, recorded using accelerometers, are environmental sounds without the presence of the rice weevil.

The other positive audio clips, recorded using the X-NUCLEO-CCA02M1 expansion board through the STM32 NUCLEO-F401RE development board, were taken from the Research and Development Department of the Philippine Center for Postharvest Development and Mechanization (PHilMech). The X-NUCLEO-CCA02M1 is an evaluation board with two digital MP34DT01-M MEMS microphones [19] soldered onto the board. The following figure shows the actual image of the X-NUCLEO-CCA02M1 board. The microphones are the two gold components at the lower part of the board.

Fig. 1: X-NUCLEO-CCA02M1 expansion board

The board supports audio streaming to a personal computer through a USB connector. MP34DT01-M is an ultra-compact, low-power, omnidirectional, digital MEMS microphone built with a capacitive sensing element and an IC interface. When this board is flashed and connected to the computer, it will be recognized as a standard multichannel USB microphone.

A variable number of rice weevil were present during the recording with different noise present across clips. The next figure shows the data gathering setup in PHilMech.

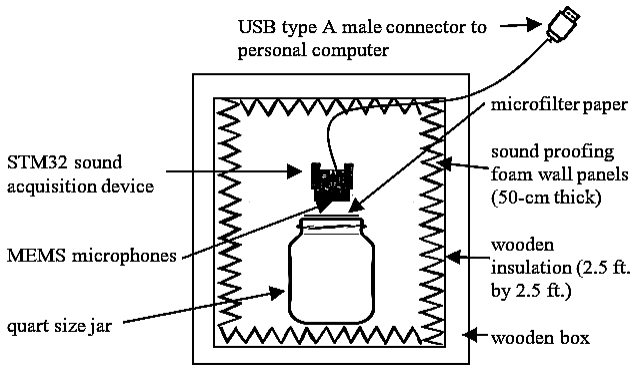

Fig. 2: Audio acquisition setup

In the previous figure, a layer of vinyl sheet, a layer of ordinary foam, and another layer of vinyl sheet were constructed between the wooden insulation (plywood) and the external frame (plywood). The total size of the box is 3 ft. by 3 ft. A layer of sound proofing foam covers the wooden insulation inside the box for audio recording. The jar contained 100 grams of rice grains with rice weevils. The jar was covered on top with a microfilter paper to prevent the insects from escaping the jar. The two built-in microphones of the STM32 sound recording device was positioned approximately 1 cm above the microfilter paper. The recording was done in a personal computer where the STM32 device is connected via a USB cord. The recording application used was Audacity 2.2.2.

The other negative audio clips are some environmental sounds randomly taken from the ESC-50 dataset that has 2,000 labeled environmental recordings of 50 classes with 40 clips in each class [17]. There are 10 classes per category, the categories being animal sounds, natural soundscapes and water sounds, human (non-speech) sounds, interior/domestic sounds, and exterior/urban noises.

The following table shows the summary of the dataset parameters:

Table 2: Some Pre-trained Keras Models' Performance on the ImageNet Validation Dataset

Parameters	Value
number of positive dataset	1000
number of negative dataset	1000
sampling rate	44.1 kHz
duration of an audio clip	5 s
bit rate	705.6 kbps
number of bits	16
number of channels	1
dataset extension	.wav

All dataset audio clips were filtered using a high pass filter set at 1.3 kHz since the frequency range for larval and adult stage weevil is 1.3 kHz to 3.8 kHz [5]. After filtering, the clips were normalized using the Peak Loudness (RMS) method.

The spectrogram images of the resulting clips were generated using the librosa python package [18]. The sampling rate was set at 44.1 kHz, the same sampling rate of the clips. The window length is set at 500 samples and hop length is set as ¼ of the window length at 125 samples. This is to make sure that at least one period of the minimum rice weevil sound frequency will be covered. The output images are of .jpg extension, each with a size of 503 pixels by 376 pixels. The following figures show sample spectrogram images generated:

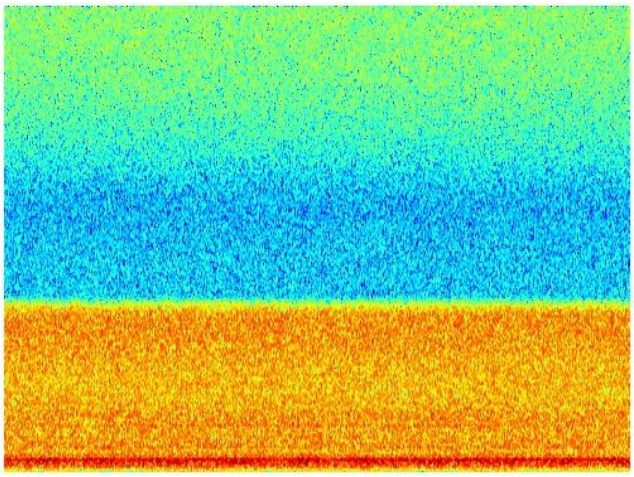

Fig. 3: Spectrogram image of a positive audio clip

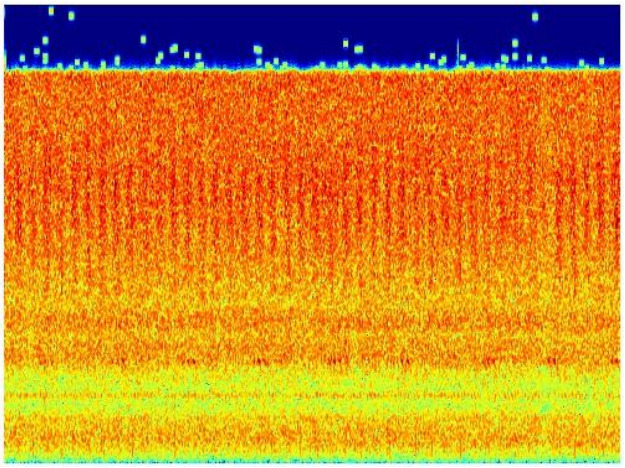

Fig. 4: Spectrogram image of a negative audio clip

3.2. Feature Extraction and Training

For transfer learning, the following pre-trained Keras models were used as feature extractors: Xception, VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, and MobileNet. Extraction of features in the spectrogram image dataset was done by taking only the activations available before the last fully connected layer of each network, more specifically before the final softmax classifier. Those activations were taken as the feature vector of the classifier model.

After the feature extraction using each model, the training of the Logistic Regression classifier was done using the extracted features and labels: 70% of the data was allotted for training and 30% for testing. These correspond to 1400 training data and 600 testing data.

The following figure shows the processes involved in each model.

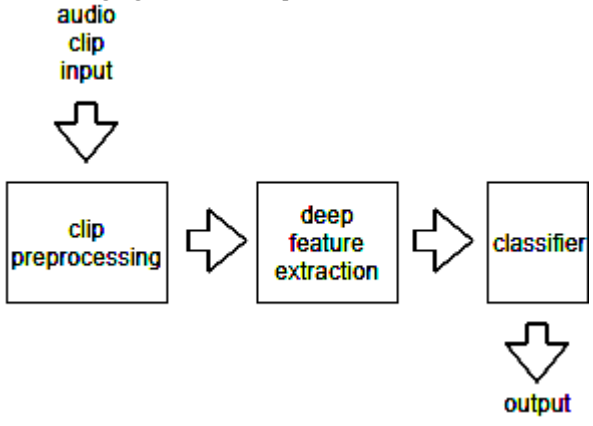

Fig. 5: Model processes

Under pre-processing, audio clips were filtered and normalized. The spectrogram images of the resulting audio files were generated. Deep features of these images were then extracted for classifier training and testing. The classifier output will finally provide the prediction.

4. Experimental Results and Discussion

After feature extraction using the Keras pre-trained models and training the resulting model of new features, Xception, ResNet50, InceptionResNetV2, and MobileNet obtained a rank-1 accuracy of 99.17% each while VGG16, VGG19, and InceptionV3 all got a rank-1 accuracy of 99.33%.

Figure 6 shows the confusion matrix of Xception, ResNet50, InceptionResNetV2, and MobileNet while Figure 7 shows the one of VGG16, VGG19, and InceptionV3. Label 0 is for the negative data while label 1 is for the positive data.

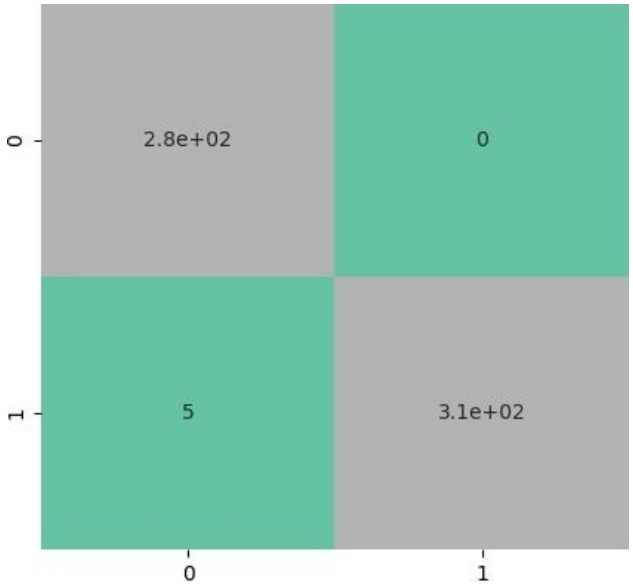

Fig. 6: Confusion matrix of Xception, ResNet50, InceptionResNetV2, and MobileNet

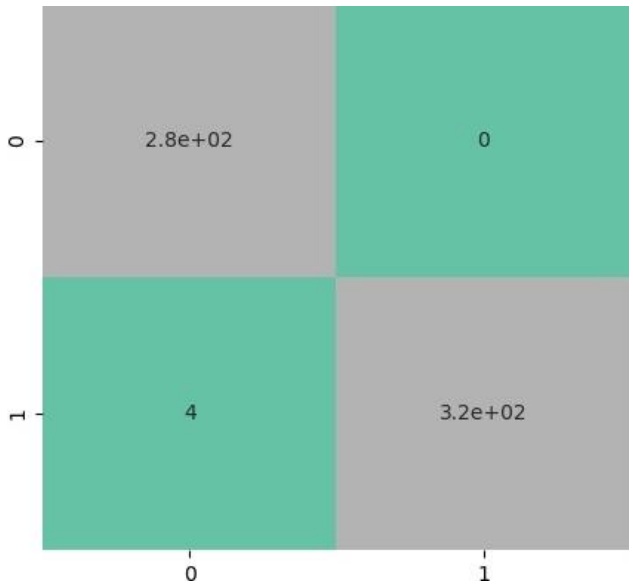

Fig. 7: Confusion matrix of VGG16, VGG19, and InceptionV3

There are 5 negative test data predicted as positive in each model in Figure 6. In Figure 7, there are 4 negative test data predicted as positive. All other test data were predicted correctly.

The following table shows the precision, recall (sensitivity), and f1 score of each model. Note that the average precision, average recall, and average F1 score of each model are all 0.99 or 99%.

Table 3: Performance Parameters of the Trained Models

Model	Classification	Precision	Recall	F1 Score
Xception	negative data	0.98	1.00	0.99
	positive data	1.00	0.98	0.99
VGG16	negative data	0.99	1.00	0.99
	positive data	1.00	0.99	0.99
VGG19	negative data	0.99	1.00	0.99
	positive data	1.00	0.99	0.99
ResNet50	negative data	0.98	1.00	0.99
	positive data	1.00	0.98	0.99
InceptionV3	negative data	0.99	1.00	0.99
	positive data	1.00	0.99	0.99
InceptionResNetV2	negative data	0.99	1.00	0.99
	positive data	1.00	0.99	0.99
MobileNet	negative data	0.98	1.00	0.99
	positive data	1.00	0.98	0.99

5. Conclusion

Results show that given the data used, VGG16, VGG19, and InceptionV3 outperform Xception, ResNet50, InceptionResNetV2, and MobileNet based on their rank-1 accuracy by a difference of 0.16%. That mere difference is not significant that any of the said models are said to be effective in rice weevil sound detection. The detection rates of the models based on rank-1 accuracy are higher than those of other studies. This study is also the first one to use deep learning on spectrogram images of acoustic files in the detection of rice weevil presence.

It is recommended that other Keras pre-trained models such as DenseNet, NASNet, and MobileNetV2 be utilized as well. The dataset may still be expanded to include clips with rice weevil sound in other environments.

Acknowledgement

The authors would like to deeply thank De La Salle University, Manila, specifically the Gokongwei College of Engineering Graduate School, the Philippine Center for Postharvest Development and Mechanization, Science City of Muñoz, Nueva Ecija, Philippines, and the Commission on Higher Education, Philippines for the untiring support to researches of engineering graduate students.

References

[1] Neethirajan S, Karunakaran C, Jayas DS & White NDG (2007), Detection techniques for stored-product insects in grain. *Food Control* 18(2), 157-162.

[2] Burks CS, Yasin M, El-Shafie HA & Wakil W (2015), Pests of stored dates. *Sustainable pest management in date palm: current status and emerging challenges*, 237-286.

[3] Eliopoulos PA, Potamitis I & Kontodimas DC (2016), Estimation of population density of stored grain pests via bioacoustic detection. *Crop Protection* 85, 71-78.

[4] Eliopoulos PA, Potamitis I, Kontodimas DC & Givropoulou EG (2015), Detection of adult beetles inside the stored wheat mass based on their acoustic emissions. *Journal of economic entomology* 108(6), 2808-2814.

[5] Fleurat-Lessard F, Tomasini B, Kostine L & Fuzeau B (2006), Acoustic detection and automatic identification of insect stages activity in grain bulks by noise spectra processing through classification algorithms. *Conference Working on Stored Product Protection*.

[6] Le-Qing Z (2011). Insect sound recognition based on mfcc and pnn. *Multimedia and Signal Processing (CMSP), 2011 International Conference* 2, 42-46.

[7] Yazgaç BG, Kırcı M & Kıvan M. (2016), Detection of sunn pests using sound signal processing methods. *Agro-Geoinformatics (Agro-Geoinformatics), 2016 Fifth International Conference*, 1-6.
[8] Noda JJ, Travieso CM, Sánchez-Rodríguez D, Dutta MK & Singh A (2016), Using bioacoustic signals and support vector machine for automatic classification of insects. *Signal Processing and Integrated Networks (SPIN), 2016 3rd International Conference*, 656-659.
[9] Flynn T, Salloum H, Hull-Sanders H, Sedunov A, Sedunov N, Sinelnikov Y, Sutin A & Masters D (2016), Acoustic methods of invasive species detection in agriculture shipments. *Technologies for Homeland Security (HST), 2016 IEEE Symposium*, 1-5.
[10] Pan W, Kong X, Xu J & Pan W (2016), Measurement and analysis system of vibration for the detection of insect acoustic signals. *Electromagnetic Compatibility (APEMC), 2016 Asia-Pacific International Symposium* 1, 1090-1092.
[11] Peso Parada P & Cardenal-López A (2014), Using Gaussian mixture models to detect and classify dolphin whistles and pulses. *The Journal of the Acoustical Society of America* 135(6), 3371-3380.
[12] Gupta D, Jain S, Shaikh F & Singh G (2017), Transfer learning & the art of using pre-trained models in deep learning. *Analytics Vidhya*, available online: https://www.analyticsvidhya.com/blog/2017/06/transfer-learning-the-art-of-fine-tuning-a-pre-trained-model/, last visit: 07.07.2018.
[13] Chollet F, *Deep learning with python*, Manning Publications Co. (2017).
[14] Gulli A & Pal S, *Deep Learning with Keras*, Packt Publishing Ltd., (2017).
[15] Kieffer B, Babaie M, Kalra S & Tizhoosh HR (2017), Convolutional neural networks for histopathology image classification: training vs. using pre-trained networks. *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 1-6.
[16] Mankin RW & Fisher JR (2002), Acoustic detection of black vine weevil, Otiorhynchus sulcatus (Fabricius)(Coleoptera: Curculionidae) larval infestations in nursery containers. *Journal of Environmental Horticulture*, 20(3), 166-170.
[17] Piczak KJ (2015), ESC: Dataset for environmental sound classification. *Proceedings of the 23rd ACM international conference on Multimedia*, 1015-1018.
[18] McFee B, Raffel C, Liang D, Ellis DP, McVicar M, Battenberg E & Nieto O (2015), librosa: Audio and music signal analysis in python. *Proceedings of the 14th python in science conference*, 18-25.
[19] Oliva SL, Palmieri A, Invidia L, Patrono L & Rametta P (2018), Rapid Prototyping of a Star Topology Network based on Bluetooth Low Energy Technology. *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, 1-6.
[20] Cortez D, Molina C, Abante M, Santos M, and Sarmiento M (2018), Nullpest: a mobile application of agricultural pest locator using sonar sensor set-up. *International Journal of Engineering & Technology* 7(3.29), 107-109.