Environmental monitoring by sound source detection using machine learning

Hery Tina RAMANAN'HAJA
Ecole Doctorale Thématique Energies
Renouvelables et Environnement
University of Antsiranana,
Antsiranana, Madagascar

Maheritiana Jonathan Jérémie RANDRIARISON
Ecole Supérieure Polytechnique d'Antsiranana
University of Antsiranana,
Antsiranana, Madagascar

Rakotobe Tefy
RAOELIVOLOLONA Ecole Doctorale
Thématique Energies Renouvelables et
Environnement University of
Antsiranana,
Antsiranana, Madagascar

Odette FOKAPU
Université de technologie de Compiègne,
UMR CNRS 7338 Biomécanique et Bioingénierie.
University of Picardie Jules Verne,
IUT Aisne , Cuffies-Soissons, France

Youssef KEBBATI
Laboratoire de Physique et Chimie
de l'Environnement et de l'Espace
University of Orleans,
Orleans, France

Jean Marie RAZAFIMAHENINA
Ecole Doctorale Thématique Energies
Renouvelables et Environnement
University of Antsiranana,
Antsiranana, Madagascar

Abstract—**In this study, we aim to detect ecological violations tied to deforestation, especially in locations like Montagne d'Ambre National Park. Our method involves recognizing sounds produced during tree cutting with an axe. To achieve this, we've implemented a proactive monitoring system based on the detection of axe blows. In this initial phase, our focus is on sound processing. We collected a variety of sounds from the monitored area, including lemur calls, bird songs, cicadas, water flow, and waterfalls. Additionally, we included sounds associated with human activities, such as stone breaking, hammering, and sawing. In total, we gathered 108 minutes of sound data, which we divided into 5-second segments, resulting in 1299 segments. These segments underwent preprocessing steps, which included data normalization, sound peak detection, and applying a 186-millisecond window around the detected peaks. This process allowed us to create a database containing 5007 windows. Next, we extracted temporal, spectral, and cepstral features from this data to use in our algorithms. We trained various algorithms, including Random Forest, k-nearest neighbors, naive bayes, AdaBoost, Support Vector Machine, and logistic regression. Our results indicated that the logistic regression algorithm performed the best, achieving a precision of 99.47 percent, a recall of 98.98 percent, and an F1 score of 99.15 precent. With the successful development of a model capable of detecting tree-cutting sounds, our next step involves expanding the monitoring area and providing power to the monitoring nodes.**

Keywords — *Environmental Monitoring, Sound Source Identification, Machine Learning, Logistic Regression, Signal Processing.*

I. INTRODUCTION

Aware of the problems of climate change and regularly suffering damage from natural disasters, Madagascar is strongly committed to protecting the environment. Actions are being implemented for massive reforestation of the country. Policies are adopted for biodiversity, natural resource management and protected areas. Among the major factors in deforestation in most Malagasy lands is the illegal exploitation of forests for the production of charcoal, the use of firewood in households and excessive use in carpentry. The source from the MNP Montagne d'Ambre Association specified that the cuts are made with an ax.

Currently, as part of environmental monitoring by technological means, the SMART (Spatial Monitoring and Reporting Tools) and GFW (Global Forest Watch) control tools are used by several organizations in Antsiranana Madagascar, such as the MBG (Missouri Botanical Garden) Ankoriakely and the SAGE (Environmental Management Support Service) Antsiranana. They use satellite data or technologies based on participatory detection and patrolling. Thus, the response time for detection reaches a minimum of six hours or a delayed time. In addition, the results are uncertain and the process requires a high level of human resources. In this context, the detection of irregularities always happens after the destruction of the environment. For example, for a case of tree cutting, the previous methods do not make it possible to detect the offense before the tree is cut down. This is how we propose to carry out preventive surveillance by detecting the sound of a tree cutting at the start or during the offense.

The literature takes us towards combining the network of wireless sensors in order to be able to carry out remote monitoring [1], and artificial intelligence which will make it possible to distinguish the possibility of cuts [2].

We propose a series of processes to detect tree cutting by identifying the sound emitted by axe blows, and to transmit signaling to a central station based on the steps shown in "Fig.1":
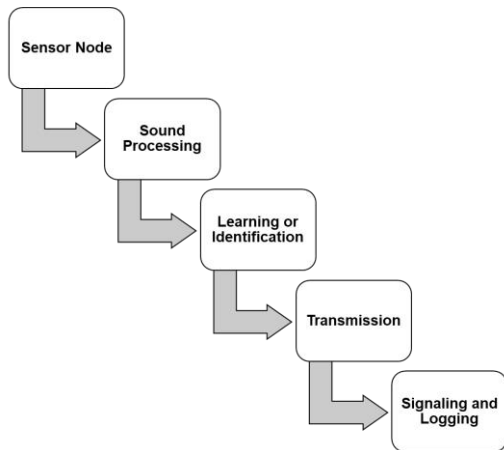

Fig. 1. Steps for reporting detection

In this article, we focus on the second and third blocks linked to data processing which are:

- Sound processing
- Learning or identification

The objective is to create a learning model that differentiates the sound of cutting a tree from any sounds that could be encountered on the Montagne d'Ambre site.

II. METHODS

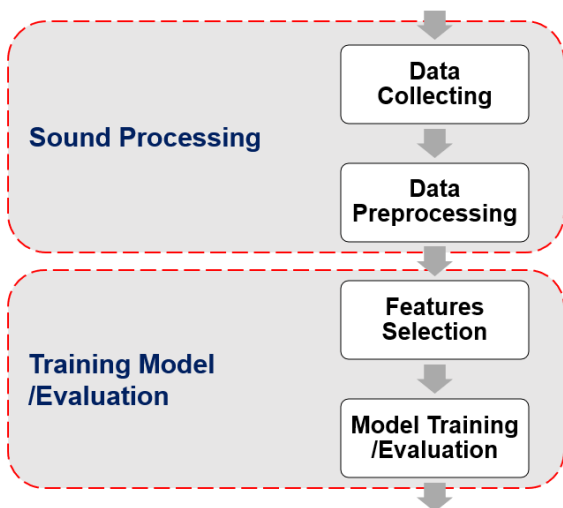For the creation of the classification model, "Fig.2" presents the following steps were adopted:

Fig. 2. Steps for data learning

A. Sound processing

1) Data collecting

To carry out machine learning, sound data likely to exist in the Montagne d'Ambre National Park were collected.

Tree cutting sounds and other sounds such as cicada singing or cymbalization, sounds emitted by the flow of river water and waterfalls, song of different species of birds present on the site are collected to form a database. Then, we added sound of stone breaking, sound of saw cutting and hammer blowing in order to strengthen the generalization capacity of our machine.

A total of 108 minutes of audio was collected and subsequently segmented into 5-second segments. We define by "other" sounds which are not cutting sound as shown in "Table 1":

TABLE 1. NUMBER OF SEGMENT OF SOUND COLLECTED

Sounds	5 seconds sound segment amount
Tree cutting sound	399
Other	900
Total	1299

These segments will undergo preprocessing to create the dataset.

2) Peak detection

In order to minimize computing time, it's important to consider that tree cutting with an axe typically involves a series of short-duration blows. With this in mind, our approach involves the initial detection of peaks in the audio signal.

This process is preceded by the sampling and filtering of sounds using a low-pass filter, with a sampling rate set at 22kHz [3]. Subsequently, we normalize the audio by dividing its amplitude by the maximum amplitude, ensuring consistency in our data.

Following the normalization process, we employ threshold detection, which entails identifying moments when the audio signal surpasses a predefined threshold value. In our case, the reference threshold value is set at 0.25 on the normalized signal.

"Fig. 3(1)" on illustrates an example of the captured sounds, while "Fig. 3(2)" at the bottom displays an overlay of normalized sound (in blue) and detected peaks during each axis stroke (in red).
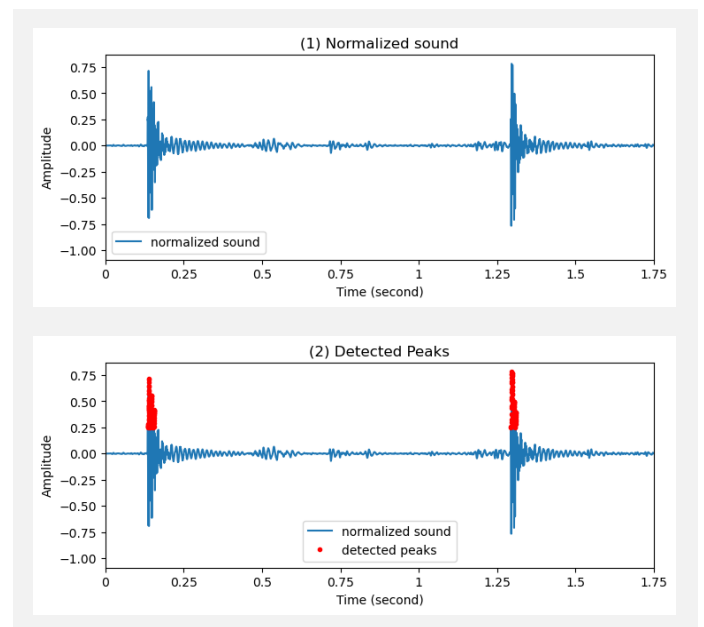

Fig. 3. Highlighting of detected peaks

Still in the principle of minimizing the computation, only the positive alternation is taken into account for the detection of peaks.

Peaks detected are used in triggering windowing.

3) Windowing

a) Windowing description

After peak detection, we proceed to window samples to reduced duration. These windowed data, from 5s segment constitute our dataset.

Windowing involves selecting samples after the first peak. The window size is set to 4096 samples, a value obtained from the visualization of the temporal characteristic of the tree cutting sound.

"Fig. 4" displays two overlapping curves: the blue section represents normalized sound, and the red section represents windowed samples following peak detection.
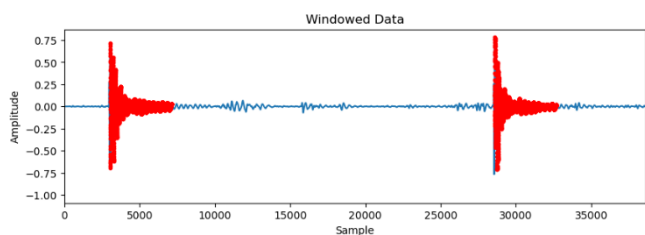

Fig. 4. Windowed data highlighting

After windowing, we obtain from the 5s segments a total of 5007 samples presented in "Table 2".

TABLE 1. CATALOG OF COLLECTED DATA

Sounds	5 seconds sound segment amount	Windowed Data Amount
Tree cutting sound	399	1468
Other	900	3539
Total	1299	5007

b) Triggering windowing

When training the model, windowing is done automatically just after peak detection. On the other hand, in order to make the use of our system practical, the following algorithm is used before windowing and identification:

- Detection of a first peak with a first load
- Confirmation by detecting a second similar peak with a second loading
- If the duration detected between two successive similar peaks is at a value between 2 to 5 seconds, we proceed to windowing and identification of a third loading

Thus, the total time before ensuring identification or not is a maximum of 15 seconds.

B. Training / identification

1) Establishment of the dataset

To design the learning model, the next steps are data separation and labelling, feature selection, and algorithm choice. To establish the model, we share this dataset into a training set to train the machine and into a test set for evaluating its performance.

80% of the data is used as a training set and 20% as a testing set. We define as Positive, an entry corresponding to a tree cut, and Negative the others are. We show on "Table 2" this repartition.

TABLE 2. DATASET REPARTITION

Dataset	Training Set (80%)	Test Set (20%)	Total
Positive	1174	294	1468
Negative	2881	658	3539
Total	4055	952	5007

2) Features selections

We selected 26 features from the literature to describe the sound characteristics: Short-Time Fourier Transform with chroma, Root Mean Square, Spectral Centroid, Spectral Bandwidth, Spectral Roll-off and 20 Mel Frequency Cepstral Coefficients in sound processing [4][5][6].

In order to reduce the complexity of the algorithm, the k-best estimator method was used to choose the influential factors. The following histogram shown in "Fig. 5", illustrates the significance of each feature in our model.
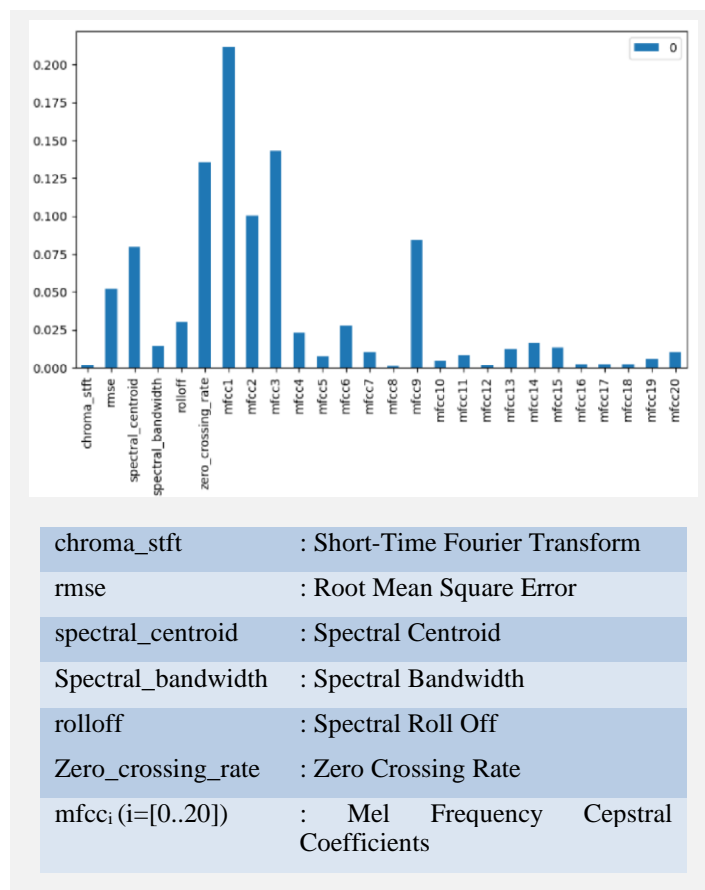

chroma_stft	: Short-Time Fourier Transform
rmse	: Root Mean Square Error
spectral_centroid	: Spectral Centroid
Spectral_bandwidth	: Spectral Bandwidth
rolloff	: Spectral Roll Off
Zero_crossing_rate	: Zero Crossing Rate
mfcc$_i$ (i=[0..20])	: Mel Frequency Cepstral Coefficients

Fig. 5. Overview of features importance

Based on "Fig.5", we've selected the first 10 influential factors. Increasing the number of non-influential factors can

impact our model, making it more complex and raising the risk of overfitting.

3) Algorithm selections

Six common machine learning algorithms were used for training, including Random Forest [9], K-Nearest Neighbors (KNN) [8], Support Vector Machine (SVM)[10], Naive Bayes[7], AdaBoost, and Logistic Regression [12].

All programs were processed using the Python programming language.

4) Evaluation and metrics

a) Confusion matrix

In order to evaluate our learning models, we will use the elements of the confusion matrix. The confusion matrix uses the following values to perform the evaluation [11]:

- Positive P: Indicating the number of real positive cases in the data. In our case, it is the number of windowed samples corresponding to an ax sound.

- Negative N: Indicating the number of real negative cases in the data. In our case, it is the number of windowed samples designating a sound other than an ax blow.

- True Positive TP: Indicating the number of positive cases that are correctly classified by the classifier. In our case, it is the number of cutting data input and detected by the machine.

- False Positive FP: Indicating the number of positive cases that are incorrectly classified by the classifier. In our case, this is the number of other sounds which are detected as cutting or which are not.

- True Negative TN: Indicating the number of negative cases that are correctly classified by the classifier. In our case, this is the number of other sounds in input and detected as such by the machine.

- False Negative FN: Indicating the number of negative cases that are incorrectly classified by the classifier. In our case, this is the number of cases of other sound in input but detected as tree cutting sound by the machine.

We will use these parameters to assess the performance of the algorithms.

b) Metrics and model perfomance

As metrics, we will use accuracy, recall, precision, and F1 score [11]. The definition and formula of each metric are :

- Accuracy metric: Measures the ratio of correctly predicted instances to the total number of instances in the dataset. In other words, accuracy tells you how many of the predictions made by your model were correct. It quantifies how well the machine can correctly identify or classify different patterns.

$$Accuracy = \frac{TP+TN}{P+N} \qquad (1)$$

- Recall metric: quantifies sensitivity and assesses the model's ability to accurately identify instances of the positive class.

$$Recall = \frac{TP}{P} \qquad (2)$$

- Precision metric: Correlates with the model's "specificity" or its capacity to accurately identify negative instances. It is a measure of the model's ability to avoid false positives.

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

- F1-Score: The F1-score is the harmonic mean of precision and recall. It provides a balance between these two metrics and is useful when you want to consider both false positives and false negatives.

$$F1Score = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (4)$$

III. RESULTS AND INTERPRETATION

1) Confusion matrix result

"Table 3" presented here illustrates the results obtained from our experimentation for P= 294 and N=658, remembering that P+N=952. This is the amount of evaluation data or Test Set:

TABLE 3. CONFUSION MATRIX RECAPITULATION

Classifier	TP	FN	TN	FP
RANDOM FOREST	293	1	649	9
KNN	288	6	639	19
SVM	289	5	605	53
ADABOOST	293	1	650	8
NAIVE BAYES	290	4	623	35
LOGISTIC REGRESSION	291	3	656	2

2) Metrics comparison

Considered the confusion matrix result in "Table 3" and equation of metrics "(1)", "(2)", "(3)" and "(4)", we can have the results of on "Table 4":

TABLE 4. METRICS RESULT

Algorithm	Accuracy	Recall	Precision	F1-Score
RANDOM FOREST	98,95	99,66	97,02	98,32
KNN	97,37	97,96	93,81	95,84
SVM	93,91	98,30	84,50	90,88
ADABOOST	99,05	99,66	97,34	98,49
NAIVE BAYES	95,90	98,64	89,23	93,70
LOGISTIC REGRESSION	99,47	98,98	99,32	99,15

Notably, all models demonstrate a good performance. Particularly, when examining the "recall" metric, it becomes evident that most models yield high scores, with the exceptions being SVM and Naïve Bayes. Furthermore, it is worth highlighting that all tested models show notably high "precision".

In the context of our application, where maximizing the detection rate of deforestation is of paramount importance, on the other hand, minimizing false detections is essential to reduce operational costs. The choice of model is based on these specific goals of our application and the observed results of all models.

Thus, we prioritize selecting a model with greater sensitivity to align with our objectives. Moreover, given that our system will operate in remote natural environments without access to the electrical grid, energy efficiency is a critical consideration.

"Fig. 6" displays a comparison of all the tested methods and highlights the balance of metrics in logistic regression.
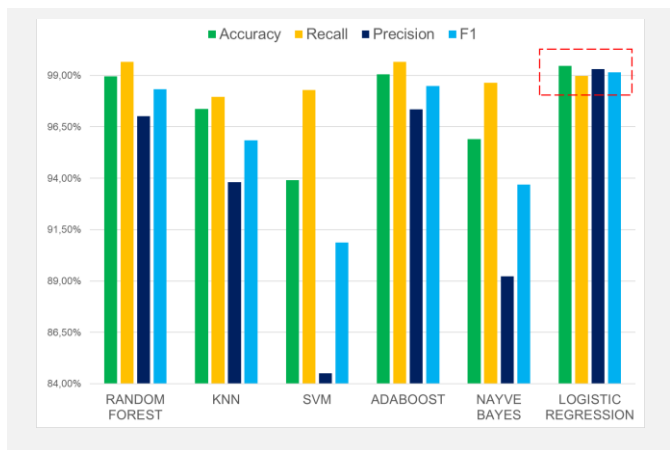

Fig. 6. Comparison of all tested methods

After analysis, it becomes apparent that Logistic Regression aligns well with our constraints. This choice is justified by its excellent precision performance, which is crucial for our goals, as well as its acceptable recall. The advantage is that we both accurately detect the presence of a cut and avoid false detections when there is no cut. Additionally, Logistic Regression offers straightforward implementation, particularly for the inference step, and boasts lower complexity compared to alternative models.

IV. CONCLUSION AND FUTURE SCOPE

In conclusion, we have a device and model capable of detecting with an accuracy of 99.47 percent for a possible tree cutting by adopting a recognition model based on logistic regression. Having considered the balance of sensitivity and specificity, we can effectively distinguish the existence or not of a tree cut having an F1-score of 99.15 percent. The preprocessing and transmission time are of the order of milliseconds, and the time of the identification process is around 15 seconds. This is considered effective compared to the method currently used by Madagascar National Parks.

Although the data processing part is ensured, it is now to be considered as a work package, the modeling and implementation of the network topology, the optimization of the capture device in terms of range and security, the optimization on the coverage and deployment capacity of the sensor nodes, then study the power supply and energy optimization.

REFERENCES

[1] Sheikh Ferdoush, Xinrong Li, "Wireless Sensor Network System Design Using Raspberry Pi and Arduino for Environmental Monitoring Applications", Procedia Computer Science, Volume 34, 2014, Pages 103-110, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2014.07.059.

[2] Al Qundus, J., Dabbour, K., Gupta, S. et al. "Wireless sensor network for AI-based flood disaster detection". Ann Oper Res 319, 697–719 (2022). https://doi.org/10.1007/s10479-020-03754-x

[3] Duan, S., Towsey, M., Zhang, J., Truskinger, A., Wimmer, J., & Roe, P. (2011, December). "Acoustic component detection for automatic species recognition in environmental monitoring". In 2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (pp. 514-519). IEEE.

[4] Ahmad, Sheikh Fahad & Singh, Deepak. (2019). Automatic Detection of Tree Cutting in Forests using Acoustic Properties. Journal of King Saud University - Computer and Information Sciences. 34. 10.1016/j.jksuci.2019.01.016.

[5] Soto-Murillo, M.A.; Galván-Tejada, J.I.; Galván-Tejada, C.E.; Celaya-Padilla, J.M.; Luna-García, H.; Magallanes-Quintanar, R.; Gutiérrez-García, T.A.; Gamboa-Rosales, H. "Automatic Evaluation of Heart Condition According to the Sounds Emitted and Implementing Six Classification Methods." Healthcare 2021, 9, 317. https://doi.org/10.3390/healthcare9030317

[6] Tusar Kanti Dash, Soumya Mishra, Ganapati Panda, Suresh Chandra Satapathy, Detection of COVID-19 from speech signal using bio-inspired based cepstral features, Pattern Recognition, Volume 117, 2021,107999, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2021.107999.

[7] Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H.-P. (2014). Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications. IEEE Communications Surveys & Tutorials, 16(4), 1996–2018. doi:10.1109/comst.2014.2320099

[8] Jia-Ching Wang, Jhing-Fa Wang, Kuok Wai He and Cheng-Shu Hsu, "Environmental Sound Classification using Hybrid SVM/KNN Classifier and MPEG-7 Audio Low-Level Descriptor," The 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 2006, pp. 1731-1735, doi: 10.1109/IJCNN.2006.246644.

[9] T. Kojima, T. Ijiri, J. White, H. Kataoka and A. Hirabayashi, "CogKnife: Food recognition from their cutting sounds," 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, 2016, pp. 1-6, doi: 10.1109/ICMEW.2016.7574741.

[10] Ahmad Taher Azar, Hanaa Ismail Elshazly, Aboul Ella Hassanien, Abeer Mohamed Elkorany, "A random forest classifier for lymph diseases", Computer Methods and Programs in Biomedicine, Volume 113, Issue 2, 2014, Pages 465-473, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2013.11.004.

[11] Kurdi, H.; Al-Aldawsari, A.; Al-Turaiki, I.; Aldawood, A.S. "Early Detection of Red Palm Weevil, Rhynchophorus ferrugineus (Olivier), Infestation Using Data Mining." Plants 2021, 10, 95. https://doi.org/10.3390/plants10010095.

[12] Wang QQ, Yu SC, Qi X, et al. [Overview of logistic regression model analysis and application]. Zhonghua yu Fang yi xue za zhi [Chinese Journal of Preventive Medicine]. 2019 Sep;53(9):955-960. DOI: 10.3760/cma.j.issn.0253-9624.2019.09.018. PMID: 31474082.