KNN and SVM Classification for Chainsaw Sound Identification in the Forest Areas

N'tcho Assoukpou Jean GNAMELE[1]
Yelakan Berenger OUATTARA[2]
Laboratoire des Signaux et Systèmes
UFHB, Abidjan, Cote d'Ivoire

Geneviève BAUDOIN[4]
ESIEE-PARIS
UPEM
Marne-la-vallée, France

Toka Arsene KOBEA[3]
Laboratoire de Physique de l'Atmosphère et Mécanique des
Fluides, UFHB, Abidjan, Cote d'Ivoire

Jean-Marc LAHEURTE[5]
ESYCOM (FRE 2028)
UPEM, Marne-la-vallée, France

Abstract—**We present in this paper a comparative study of two classifiers, namely, SVM (support vector machine) and KNN (K-Nearest Neighbors), which we combine to MFCC (Mel-Frequency Cepstral Coefficients) in order to make possible the detection of chainsaw's sounds in a forest environment. Optimization's calculation of the relevant characteristics of the sounds recorded in the forest and the judicious choice of the key parameters of the classifiers allows us to obtain a true positive rate of 95.63% for the SVM-LOG-KERNEL and 94.02% for the KNN. The SVM-LOG-KERNEL classifier offers a better classification result and a processing time 30 times faster than KNN.**

Keywords—*KNN Algorithm; SVM Algorithm; MFCC; sound recognition; forest monitoring; machine learning*

I. INTRODUCTION

From 16 million hectares at the beginning of the last century, the Ivorian forest decreased to 7 850 864 hectares in 1990 then to 3 401 146 in 2015. Today, the Ivory Coast has around 2 million hectares of natural forest. This situation is largely due to human activities, in particular the overexploitation of the forest with timber and wood energy. To ensure the forest protection, LANDSAT images or military patrols are used. Unfortunately, these solutions are mostly a post observation tool for the evolution of the forest disappearance. A real-time forest monitoring approach is undoubtedly the most appropriate solution to prevent trees cutting. In this work, we propose a monitoring system based on acoustic sensors to classify in-situ forest sounds in real-time.

To date, there is very little work done in sensor networks for real-time monitoring of forest areas. Reference [1] used LANDSAT and RADARSAT images to evaluate the deforestation rate. This work was based on Hidden Markov Models for the multi-source time series analysis. Results showed that the detection of spatial changes in the forest coverage is possible using satellites. However, this method was limited as satellite images have low spatial and temporal resolutions resulting in relatively weak sensitivity. Wang et al. introduced in [2] a Received Signal Strength Indication (RSSI) based method for the detection of deforestation. Sensor nodes were mounted on each guarded tree, and the change in the radio signal strength was used to detect the displacement of the node. It basically means that if a tree falls, the radio signal strength will change between neighboring sensors. While this method is able to detect logging in real-time, it is inefficient for covering large areas since lots of sensors are needed. In [3], it was demonstrated that chainsaw logging can be detected by the analysis of the spectral characteristics of the chainsaw sound signal, the calculations of similarity values and signal-to-noise ratio. However, this approach was not a sound classification method and there was no information about the measurement setup. Authors in [4] presented an approach based on a one-class classifier using a kernel density approach. They show how to find the best combination of parameter σ (Kernel bandwidth) with parameters used in the Mel-Frequency Cepstral Coefficients calculation (MFCCs), and how to choose the optimal decision threshold. With optimized parameters, an accuracy of 98% was achieved.

In [5] Mel-frequency cepstral coefficients are used for audio feature extraction and supervised classification algorithms (support vector machine, k-nearest neighbors, bootstrap aggregation, and random forest) for noise classification. Their data-set is of about 3000 sound samples without animal and chainsaw sound. They explore the parameter of the four algorithms to estimate the optimal parameter values for classification of sound samples in the data-set under study. They achieve a noise classification accuracy in the range 88% – 94%.

For classifying two forest type along with one mixed class of forest of Japanese Cedar Japanese Cypress, based on image classification, authors in [6] **e**valuate four classifiers: Neural Network, Support Vector Machine, Random Forest and NN-GA (Neural Network -Genetic algorithm) . The comparative analysis has revealed that the NN-GA has performed significantly well than all the other classifiers in this study with an accuracy of 95.54 %. The authors in [7] propose a sound recognition method dedicated to environmental sounds, designed with its main focus on embedded applications which is loosely based on the human hearing system, while a robust set of binary features permits a simple k-NN classifier to be

used. Sensitivity and specificity were evaluated, and chainsaw sound values, without additional noise, was, respectively, 0.978 and 0.121. With the addition of +6 dB of pink noise, sensitivity and specificity were 0.979 and 0.117, respectively. However, we don't have any information about the recording distances of sounds and the number of samples per category of sounds evaluated, and the database of 29 different environmental sounds, contains only anthropic sounds.

The work in [8] compared six of the most commonly used machine learning algorithms, including the KNN and SVM algorithms. Twelve (12) sound classes were considered without including chainsaw sounds. The number of samples taken per class being 10, remains insignificant and for the classification based on the twelve classes, neither the value of nearest neighbors nor the kernel function of the SVM is specified. The classification rate is less than 54 % for the six algorithms.

In [9], authors proposed an analysis of the autocorrelation function for the recorded acoustic signals with the help of decision trees, support vector machine (SVM), and K-Nearest Neighbor (KNN). Three sensors were combined to confirm their approach with a best hit rate of 97.7 %.

In this paper, we propose a technique to accurate the detection of the chainsaw sounds in the forest environment by evaluating two method already used in the application of automatic speech recognition, namely the MFCCs analysis. MFCCs are combined with two classifier algorithms: KNN and SVM whose parameters have been optimized. The concerned parameters are: the frame duration of the recorded sounds, the number of cepstral coefficients, the filters band for the MFCCs calculations and for KNN and SVM classifier the parameters are respectively the number K of nearest neighbors and the value of degree of SVM Kernel function. This work could be a real contribution for real time monitoring for protected forest by connecting the chainsaw sound identification system to a wireless network.

The paper is organized as follows: In Section II, MFCC, KNN and SVM algorithms are presented. In Section III, the proposed method is presented, and in Section IV we discuss about the results.

II. PRESENTATION OF MFCCs AND CLASSIFIERS MODEL

A. Mel Frequency Cepstral Coefficient (MFCCs)

MFCCs [10] [11] are coefficients used in automatic speech recognition (ASR) to describe the short-term power spectrum of a speech signal. With MFCCs analysis, the dominant features of different sounds are extracted and are defined as coefficients. These coefficients result from the discrete cosine transform (DCT) of the log power spectrum calculated on a non-linear perceptual Mel-frequency scale of the acoustic signal. The preliminary processing steps for extracting the Cepstral Coefficients from various sounds are presented in Fig. 1.
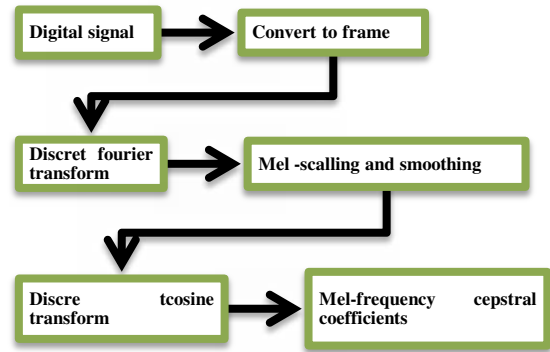

Fig. 1.   Diagram Block Showing the MFCC Processing Steps for Extracting Cepstral Coefficients.

The sound signal is converted to frames of 20 ms to 40 ms with overlapping durations between successive frames. Then pre-emphasis and weighting windows are applied on these frames [12]. The discrete Fourier transform (DFT) is calculated for each frame and the log of the amplitude spectrum is taken. The Mel's frequency spectrum is obtained by applying a bank of M band-pass filters using Mel-scaling center frequencies on the previous signal [13], [14], [15]. The Mel's frequency spectrum is given by:

$$P_p = \sum_{v}^{N/2-1}\left(d_{p,v} \times F_v\right), \ p = 0,1,\dots,M \tag{1}$$

where N is the number of sampling points within a speech frame, $d_{p,v}$ is the amplitude of the band-pass filter with index p at frequency v and F_v the Fourier transform of the signal [16].

Mel's cepstral coefficients are finally obtained by the DCT of the Mel's frequency spectrum:

$$MFCC_k = \sum_{j=1}^{M}\left(L_p \times \cos\left(\frac{k(2j-1)\pi}{2M}\right)\right) \tag{2}$$

With $L_p = \log_{10}\left(P_p\right)$, $= 0,1,\dots,m$, where m is the number of cepstral coefficients chosen for subsequent treatment $(m < M)$.

For Automatic Speech Recognition, $m \in [\ 2;13]$ and the rest is discarded.

B. KNN Algorithm (K-Nearest Neighbors) for Sounds Classification

KNN is a supervised nonparametric instance-based learning algorithm [17]. The classification of a new individual is based on its similarity with the K nearest neighbors. These nearest neighbors are themselves members of predefined classes with given label. The similarity between the new individual X and the constituents $x_{i,j}$ (i being the class and j the individual in this class) of the classes is obtained by calculating the distance between them. This requires that X and $x_{i,j}$ be either scalars or multidimensional vectors [18]. A learning phase on these labeled classes makes it possible to determine the parameter K. In binary classification, it is useful to choose odd K to avoid egalitarian votes. Although several heuristic techniques exist to define the value of K, we keep its value which minimizes classification errors.

C. SVM Algorithm

SVM is an automatic learning algorithm that belongs to the class of linear classifiers. In its approach, a SVM will seek to find an affine hyperplane that separates into two categories processed data. At the end of the training phase, the SVM - consider the sign of Equation (3) to determine the nature of the new individual $x \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ to be classified.

$$h(x) = w_1 x_1 + \cdots + w_n x_n + b = \sum_{i=1}^{n}(w_i \cdot x_i) + b \quad (3)$$

Where the vector $w \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$ (the weight vector) and the scalar b (the bias) are the two quantities defining the affine hyperplane and they are determined during the training phase. In practice, there is no separating hyperplane in the study area, to remedy this, the SVM uses "kernel functions". These functions allow the passage of the study space to another space of larger dimension where the determination of the hyperplane will be possible. The most popular kernel functions are shown in [19].

III. DESCRIPTION OF THE PROPOSED MODEL

The proposed method to detect the chainsaw sound among the other sounds consists in the combination of MFCCs to KNN and to SVM leading to MFCCs- KNN and MFCCs-SVM models. Fig. 2 shows the different steps to generate the True Positif Rate with these two classifiers.

Records of chainsaw sounds were carried out in three different forests areas. A first measurement campaign was made in Armainvilliers national forest (Gretz-Armainvilliers, France), the second one in Yapo-Abbé protected forest (Agboville, Côte d'Ivoire) and the third one in the man-made National Floristic Center (Abidjan, Cote d'Ivoire). The sound recordings were made from a distance of 10 meters to 100 meters in WAV format. A sampling rate recording is 44.1 kHz. The training sounds consist in recording sounds : three different STIHL chainsaw machines with simple sounds or cutting trees sounds ; and website sound bank [20], [21] composed by other chainsaw sounds and forest sound (bird; insect, some animal). So that we obtained 3265 sound records of 5 seconds each as presented in Table I.

In the traditional MFCC's calculation for ASR, only the first twelve cepstral coefficients (when using a 8 kHz sampling frequency) are taken into account while rejecting the first coefficient. This is because the first coefficient is regarded as somewhat unreliable and the higher DCT coefficients represent fast changes in the filter-bank energies. In our approach, we take into account this first coefficient considered as the generalized frequency band energy (FBE) [22].

We pay particular attention to the percentage of sounds correctly scored according to the two classes in the presence. This percentage corresponds to the TPR (True Positive Rate) of each classifier. This size is important because it allows both the ability of the classifier to label the sounds and gives us information on the error rate incurred.
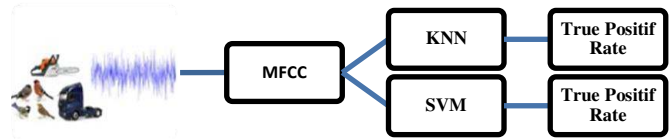

Fig. 2. MFCCs-KNN and MFCCs-SVM Algorithm Blocs for Sound Detection.

TABLE. I. SOUNDS AND NUMBER OF RECORDS

Sound	Number of records
Chainsaws	301
Other sounds	2964

The accuracy of these approaches is relative to the optimum parameters of these to be found.

IV. MFCCS-KNN AND MFCCS- SVM PARAMETERS OPTIMIZATION AND RESULTS

The optimum parameters yielding a TPR are determined. The parameters to be optimized are:

MFCCs calculation:

- The frequency band for filter arrangement varies from 3170 Hz to 11170 Hz with a frequency sweep of 1000 Hz

- The frame duration (FD) varying from 25 ms to 1000 ms with a sample time of 25 ms

- The cepstral coefficients number (CCN) varying from 9 to 13.Subsequently we examine the need to abandon the cepstral coefficients higher than 13 as is customary in ASR.

The number of filters in the MFCC calculation has been fixed to M= 35 by the fact that it is the best compromise between the calculation time of the algorithm and its accuracy as it is suggested by authors in [20].

KNN (K-Nearest Neighbors)

- The number K of nearest neighbors from 3 to 15 by step of 2.

SVM-LOG-KERNEL

There are a variety of kernel functions as shown [19]. The most used are Linear, Polynomial, Radial Basis, Gaussian Radial, and Sigmoid kernel functions. We are interested, however, by the LOG-KERNEL function [23]. This choice is motivated by the use of the logarithm function in the calculation of MFCCs and also by the fact that the impact of the LOG-KERNEL function in the classification of sounds remains to be demonstrated. Indeed, The LOG-KERNEL seems to be particularly interesting for images. This function is in the following form:

$$k(x, y) = -\log(\|x - y\|^d + 1) \quad (4)$$

In this function, we vary the key parameter d, from 1 to 7 to find its optimum value.

A. Impact of MFCCs Filters Band Arrangement

The curves in the Fig. 3 are comparative TPRs obtained by KNN and SVM algorithms relative to the frequency indexes given in Table II. For each frequency index, we retained only the best TPR after analysis.

TABLE. II.   A: FREQUENCY INDEXATION

Frequency Band	3170 Hz-4170 Hz	4170 Hz - 5170 Hz	5170 Hz - 6170 Hz	6170 Hz - 7170 Hz
Index	1	2	3	4

B: FREQUENCY INDEXATION

Frequency Band	7170 Hz - 8170 Hz	8170 Hz - 9170 Hz	9170 Hz - 10170 Hz	10170 Hz-11170 Hz
Index	5	6	7	8

The best TPR is obtained for both classifier at very high-frequency: 94.02 % of TPR at the frequency band [8170 Hz-9170 Hz] for the KNN classifier and 95.63 % at the frequency band [7170 Hz -8170 Hz] for SVM-LOG-KERNEL classifier. That means that the frequency band of Mel's cepstral coefficients impact subsequently the accuracy of TPR. The appropriate frequency of MFCCs must be the audio high frequency band. Therefore, even if the MFCCs are calculated in this frequency, MFCCs-SVM offered a better TPR than MFCCs-KNN.

B. Impact of Frame Duration (FD)

We study the impact of frame duration on the TPR by keeping for each classifier its optimized frequency band obtained in the section IV-A.

In Fig. 4, the classification rate depends on the chosen frame duration. For the KNN, the increase of the frame duration does not improve the classification rate. The best rate obtained is for the smallest frame duration namely 25 ms. On the other hand, the increase of the frame duration improves the classification rate of the SVM-LOG-KERNEL. The maximum value for the SVM is reached for a frame duration of 450 ms.

C. Impact of Cepstral Coefficients Number

In this section, we evaluate the impact of the number of cepstral coefficients on the classification rate. We consider the frame duration for which the classification rate is the most promising following the study of Section IV-B. In Fig. 5, the TPR is better for larger number of cepstral coefficients for both classifiers.

ASR based on the calculation of the cepstral coefficients of MEL, imposes an abandonment of the coefficients number higher than 13. This is because they represent fast changes in the filter bank coefficients and these fine details do not contribute to Automatic Speech Recognition. In our study, we extend their number up to 30 and we evaluate its impact on the TPR value. For the SVM, the increase in the number of MFCCs causes a decrease of about 2% in the classification rate (Fig. 6b) .The values in Fig. 6b are certainly lower than those obtained in Fig. 5b, but they offer TPR at least 93.45%, which is an appreciable rate.

For the KNN (Fig. 6a), where the increase in the number of MFCCs beyond 13 does not offer better results than those in Fig. 5a. The abandonment of the MFCCs above 13 as suggested in framework of the ASR is verified in our study.

D. Impact of the Classifier Parameters

We evaluate the impact of the key parameter of each classifier. For the best frequency band, we take for each frame duration, the best classification rate according to the value of the key parameter of the classifier (Fig. 7).

The number of nearest neighbors that optimizes the classification is 3 (K = 3) while the optimum degree for the SVM-LOG-KERNEL is 1 (d = 1). The increase of the metric dimension in the context of the use of the *LOG-KERNEL* does not favor the exactness of the labeling of sounds. Just as increasing the number of nearest neighbors does not affect positively the ability of the KNN. This increase in the value of the key parameters has the disadvantage of making the calculations more cumbersome.
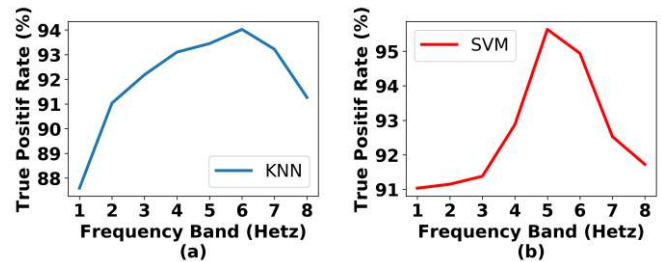

Fig. 3.   TPR According to the Frequency Band of Calculated MFCC.
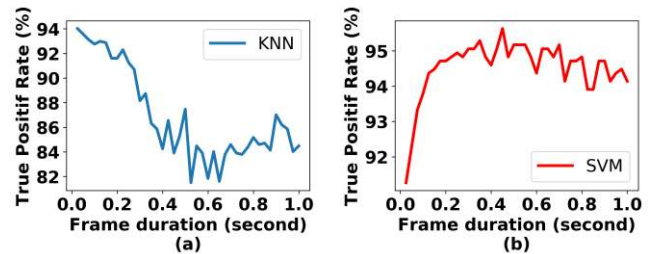

Fig. 4.   TPR According to Frames Duration.
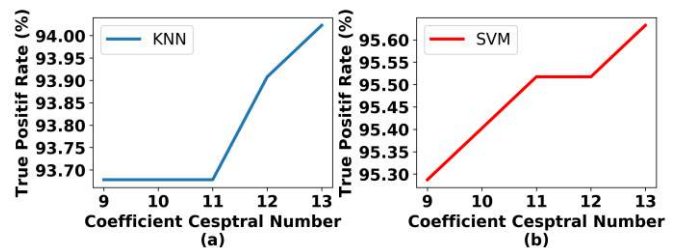

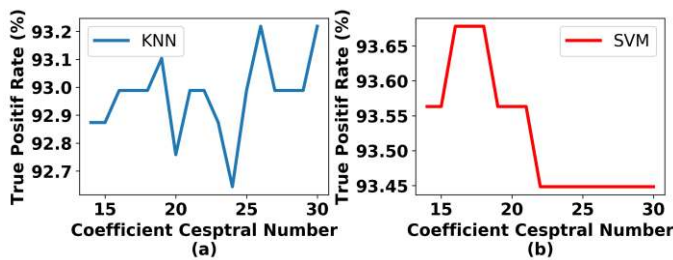Fig. 5.   Impact of the Number of Mel-Frequency Cepstral Coefficients on TPR.

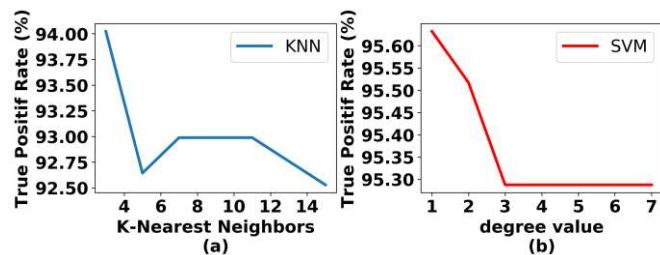Fig. 6.   Impact of the Number of Mel-Frequency Cepstral Coefficients on TPR.

Fig. 7.   TPR According to Key Parameter of Classifiers.

E. Comparison of the MFCC-KNN and MFCC-SVM Performances

In Table III, we summarize the TPR for the complete classification of all recorded sounds (general) then individually the TPR for the chainsaw sounds (chainsaw) classification and the TPR for the other sounds (forest) classification are given. The process time for each classification is also display in Table III. An Ubuntu virtual machine installed on a PC running at 2.4 GHz with an 8 GB RAM is used for the tests. The virtual machine has a RAM of 3 GB and 1 processor.

The SVM-LOG-KERNEL classifier offers the best classification rates for a processing time more than 30 times lower than KNN. This process time does not include the MFCCs calculation time because it is similar for both classifiers. As shown, SVM -LOG-KERNEL has a definite advantage in its use for sound labeling.

TABLE. III.   RESULT FOR EACH CLASSIFIER

Classifiers	True Positif Rate (%)			Process Time (second)
	general	chainsaw	forest	
KNN	94.02	40.51	99.37	67
SVM-LOG-KERNEL	95.63	53.16	99.87	2

V. CONCLUSION AND DISCUSSION

The choice of a classifier for the real-time recognition of chainsaw sound in the forest environment has required the comparison of SVM and KNN classifiers with MFCCs calculations. The best TPR we obtained to classify the whole forest sounds are 95.63% and 94.02%, respectively for SVM classifier and KNN classifier. These values are the result of the choice of judicious parameters: number of near Neighbors K=3 for KNN algorithm and the order of the Log-Kernel function in SVM algorithm d=1. The presented results

concerned a most extensive database with sounds spectrums near the chainsaw sound i.e. aircraft, motorbike reducing SVM and KNN classification particularly the chainsaw sound. Therefore, we obtained 53.16% and 40.50% respectively for SVM and KNN classifier. For the forest environment, SVM and KNN classifiers give similar results with the more complex algorithm NN-GA but they are simple to implement and fast processing.

The difference between processing time of MFCC-KNN and MFCC-SVM can be understood from the analysis of the classification approach for each of them. The KNN algorithm, focusing on calculating the distance between the new individual and the reference samples, to detect similarities, sees its execution time increases linearly with the size and amount of reference patterns. In contrast, the SVM, whose approach is based only on the provision in a metric space of the new individual in order to identify it, performs less calculation. SVM is most suitable when considering a large number of samples constituting the reference base, which is our case study.

It is important to determine a method able to increase the chainsaw detection rate. This other method will be a complement to MFCC-SVM algorithm. Thus the decision to label a new individual will take into account the results coupled with the two methods.

REFERENCES

[1] A.-B. Salberg and Ø. D. Trier. (2012).Temporal analysis of multisensor data for forest change detection using hidden Markov models. Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 6749–6752. doi: 10.1109 / IGARSS.2012.6352556.

[2] S. F. Wang, K. Y. Wang, X. J. Wang, Z. Q. Liu. (2012). A Novel Illegal Logging Monitoring System Based on WSN. Advanced Materials Research, 1417-1421. https://doi.org/10.4028/www.scientific.net/AMR. 518-523.1417.

[3] Tang Y., Han P., Wang Z., Hu L., Gao Y., Li H. (2012). Based on intelligent voice recognition of forest illegal felling of detecting methods.   2nd International Conference on Cloud Computing and Intelligent Systems, 1153 – 1156. doi: 10.1109/CCIS.2012.6664564.

[4] Colonna, J. G., Gatto, B., Dos Santos, E. M., & Nakamura, E. F. (2016). A framework for chainsaw detection using one-class kernel and wireless acoustic sensor networks into the amazon rainforest. Mobile Data Management (MDM), 17th IEEE International Conference, 2, 34–36. doi: 10.1109/MDM.2016.86.

[5] Y. Alsouda,S. Pllana,A. Kurti (2019). IoT-based Urban Noise Identification Using Machine Learning: Performance of SVM, KNN, Bagging, and Random Forest. In Proceedings of the International Conference on Omni-Layer Intelligent Systems. ACM, 62–67.

[6] Chatterjee S., Ghosh S., Dawn S., Hore S., Dey N.(2016) Forest type classification: a hybrid NN-GA model based approach. In: Satapathy, S.C., Mandal, J.K., Udgata, S.K., Bhateja, V. (eds.) Information Systems Design and Intelligent Applications. AISC, vol. 435, pp. 227–236. Springer, New Delhi. doi: 10.1007/978-81-322-2757-1_23.

[7] Kugler M.; Tossavainen T.; Nakatsu M.; Kuroyanagi S.; Iwata (2016). A. Real-time hardware implementation of a sound recognition system with in-field learning. IEICE Trans. Inf. Syst, 99, 1885–1894.

[8] Bountourakis, V.; Vrysis, L.; Papanikolaou, G. Machine Learning Algorithms for Environmental Sound Recognition: Towards Soundscape Semantics. In Proceedings of the Audio Mostly 2015 on Interaction with Sound, Thessaloniki, Greece, 7–9 October 2015.

[9] L. Czúni, and P. Z. Varga. (2014). Lightweight Acoustic Detection of Logging in Wireless Sensor Networks. The International Conference on Digital Information, Networking, and Wireless Communications (DINWC2014). The Society of Digital Information and Wireless Communication.

[10] J. G. Colonna, A. D. Ribas, E. M. d. Santos, and E. F. N. (2012). Feature subset selection for automatically classifying anuran calls using sensor networks. International Joint Conference on Neural Networks (IJCNN). 1–8doi: 10.1109 / IJCNN.2012.6252794.

[11] D. O'Shaughnesssy. (2008). Invited paper: Automatic speech recognition: History, methods and challenges. Pattern Recognition, 10, 2965–2979. https://doi.org/10.1016/j.patcog.2008.05.008.

[12] Fitch, W. T. (2000a). The evolution of speech: A comparative review. Trends in Cognitive Sciences, 4(7) ,258–267. https://doi.org/10.1016 /S1364-6613(00)01494-7.

[13] Md. Sahidullah ,Goutam Saha. (2012). Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Communication 54,4(2012),543–565. https://doi.org/10. 1016/j.specom.2011.11.004.

[14] H. Fastl and E. Zwicker. (2006).Psychoacoustics - Facts and Models. (3rd ed.). Springer.

[15] B. C. Moore. (2013). An Introduction to the Psychology of Hearing. (6th ed.). BRILL.

[16] K. Kido. (2014). Digital Fourier Analysis: Advanced Techniques. Springer.

[17] K. Lan, D, Wang, S. Fong, L. Liu, K. Wong and N. Dey, "A survey of data mining and deep learning in bioinformatics", Journal of Medical Systems, vol. 42, no. 8, pp. 1-20, 2018.

[18] R.O. Duda, P.E. Hart, D.G. Stork. (2011). Pattern Classification. (2nd ed.). WileyBlackwell.

[19] Souza, César R. "Kernel Functions for Machine Learning Applications." Accessed on 17 October 2019. http://crsouza.blogspot.com/2010/03/ kernel-functionsfor-machine-learning.html.

[20] http://bbcsfx.acropolis.org.uk/ Accessed on 17 october , 2019.

[21] UNIVERSAL-SOUNDBANK Bruitages & sons gratuits (2019) http://www.universal-soundbank.com/forets2.htm / Accessed on 6 May 2019.

[22] Zheng Fang, Zhang Guoliang, Song Zhanjiang.(2001). Comparison of different implementations of MFCC. Journal of Computer Science and Technology,6, 582-589, https://doi.org/10.1007/BF02943243.

[23] S. Boughorbel, J. Tarel, and N. Boujemaa.: Conditionally Positive Definite Kernels for SVM Based Image Recognition.In: IEEE International Conference on Multimedia and Expo (ICME2005), pp.113-116. (2005).