Acoustic Event Classification Using Ensemble of One-Class Classifiers for Monitoring Application

Achyut Mani Tripathi
Department of Computer Science & Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India
Email: t.achyut@iitg.ernet.in

Diganta Baruah
Department of Information Technology
Sikkim Manipal Institute of Technology
Majitar, -737136, Sikkim, India
Email: diganta.b@smit.smu.edu.in

Rashmi Dutta Baruah
Department of Computer Science & Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India
Email: r.duttabaruah@iitg.ernet.in

***Abstract*—In this paper we investigate the application of ensemble of one-class classifiers to the problem of acoustic event classification. We present some initial results that are based on acoustic signal emitted by different litter causing material when contacted by human. When a person interacts with an object made with a specific material, a characteristic sound is produced as a result of the interactions. We consider such interactions or activities as atomic events. We propose the application of ensemble of one-class fuzzy rule-based classifier to the problem of identification of activities that can cause possible litter in the public places. The experimental results show that the classifier gives satisfactory results and at the same time has low false alarm rate. The results are comparable to widely used one-class SVM. Moreover, the method is adaptive and suitable for incremental learning.**

I. Introduction

Litter is a growing threat globally, its control, prevention, and monitoring are major challenges faced by most of the countries worldwide. In an attempt to address these issues, we are working on an automated system that would enable us to detect the activities that can possibly cause litter in public places, such as bus stops and park. Upon detection of such activities the system would generate a voice message (through a speaker) so that people who perform littering activity (habitually, deliberately, or accidentally) can be reminded to appropriately bin their trash. The system would also help authorities to take preventive measures and strategic decisions, for example, place sufficient litter receptacles in locations where frequent littering activities are detected or deploy more man-power to keep the area clean.

In this paper we are presenting some initial results that are based on acoustic signals emitted by different litter causing material when contacted by human. Here we are considering two common sources of litter, polymer packets mostly used for packaging of snacks like potato chips, and paper cups. When a person interacts with such objects made with specific material a characteristic sound is produced as a result of interaction. For example, when someone opens a packet of chips it will produce a specific sound. The aim is to recognize such sound producing events or acoustic events from a continuous audio stream. We considered acoustic sensing over image or video sensing for monitoring task, as it has certain distinctive characteristics. First, acoustic sensing in omnidirectional as it can capture information from all directions, and is relatively less sensitive to the position and orientation. Second, it allows for non-intrusive sensing without invading the subject's privacy. Third, processing of acoustic data is relatively faster than image or video. Finally, the cost of the system based on acoustic sensors will be less.

The area of acoustic event detection and classification has recently gained attention due to its relevance to many real-world applications such as surveillance and monitoring [1], ambient assisted living [2], [3], [4], [5], [6], audio indexing and retrieval [7], [8], [9], human robot interaction [10]. While the task of acoustic event classification (AEC) involves determining the type of events that have already been extracted from an audio stream, acoustic event detection (AED) deals with both identifying the type of events and position of those events in time. One of the vital steps in acoustic event classification and detection is audio signal feature extraction .The problem of feature extraction has been addressed by many existing works. Some of the features that have been successfully applied to AEC task are: perceptual features (short-time energy, zero-crossing rate, sub-band energy, spectral-centroid, spectral roll-off, pitch) [11], and conventional automatic speech recognition features (Mel-frequency Ceepstral Coefficients-MFCC [12], Linear Predictive Cepstral Coefficients-LPCC) [12]. The most commonly used approaches for classification are: Bayesian Classifier [13], Gaussian Mixture Model (GMM)[14], [15], Hidden Markov Models (HMM) [16], Support Vector Machines (SVM) [17], Artificial Neural Networks, Decision trees, Random forests, Fuzzy rule-based classifiers [4].

Even though designed for the task of automatic speech recognition, MFCC features have been shown to work for non-speech environment sound recognition [2]. This motivated us to prefer MFCC for feature extraction. For the classification task, we considered one-class classifiers which are suitable when all data belong to one class(often referred to as target class). In our problem it is not possible to collect and label data for all the human activities apart from activities that may cause litter i.e the available data has only one class that represent activities that can cause litter. We consider four

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.236

activities opening a polymer packet (chips packet or any such kind of wrappers) (Packet Opening-PO), crushing a packet (Packet Crushing-PC), throwing or dropping a paper cup (Cup Dropping-CD), and crushing a paper cup (Cup Crushing-CC). Hereafter,we will be refering to these activities as littering activities (LAs). In near future we will be increasing the number of such activities associated with all kinds of common litter such as cans and PET bottles, and packaging. Therefore, our classification model needs to learn and adapt to new data representing new littering activities without forgetting the previously acquired knowledge. This form of learning where the data arrives in batches and the classifiers needs to learn new data without losing the previous knowledge and at the same time without requiring access to previously seen data is often referred to as incremental learning. Ensemble classifiers can address this issue naturally [18].The idea is to train separate one class classifiers for each such littering activity and generate additional classifiers as new batch of data corresponding to new littering activities becomes available. Finally, their output can be combined to identify the class.

Among the various discriminative classifiers, the rule-based models are able to represent outlier objects in an inherent way. If an instance is not covered by any of the existing rules then it can be considered as an outlier [19]. In literature, it has been shown that fuzzy rule-based model outperforms conventional (non-fuzzy) rule-based model in identifying outliers [19]. As the goal of one-class classifier is to distinguish between objects of target class and all other possible objects (considered outlier objects), fuzzy rule-based models can serve the purpose. So, fuzz rule-based (FRB) approach is selected for the given problem of acoustic event classification. To the best of our knowledge ensemble of one-class classifiers based on fuzzy-rules have not been applied so far in the domain to AEC. Another approach that has been successfully applied to AEC and AED is one-class SVM [5]. In this work, we also compare the performance of one-class FRB classifier and one-class SVM classifier.
The remainder of the paper is organized as follows: section II present brief overview of fuzzy rule-based classifiers, section III represent methodology used for classification, in section IV experiment and results are discussed, and finally the conclusions and future work is presented in section V.

II. BACKGROUND

In this section a brief overview of the key approaches used in our work is presented.

A. Fuzzy Rule-based Classifiers

As defined in [24], a classifier that uses fuzzy sets or fuzzy logic during the training phase, or operation is a fuzzy classifier. A fuzzy rule-based (FRB classifier) uses a set of fuzzy rules and an inference mechanism that generates either a soft or crisp class label for a given input. FRB classifiers can be distinguished into the following three categories depending on the consequent part of fuzzy rules involved in classification [24].

1) TYPE 1: Class label as consequent
Let R be the number of fuzzy rules, n be the number of features, and m be the number of classes, then the rules for this type of classifier can be given as,
$Rule^i$: IF x_1 is A_1^i AND...AND $x_n$isA_n^i THEN y_i is class k, where x_j, $j = 1, 2, ...n$ is the input variable, A_1^i is a fuzzy set , y^i is the output associated with $Rule^i$, and class k, $1 \leq k \leq m$ is the class label.
In this model each rule votes for a class. To determine the output of the classifier, the votes of all rules are aggregated. One of the aggregation method is maximum aggregation method. For the given input **x**, the soft class label for **x** consists of membership values $g_k(\mathbf{x}) \in [0, 1]$, $k = 1, ..., m$.
If $i \rightarrow k$ denotes the rule i that votes for class k. Then $g_k(\mathbf{x})$ is given as,

$$g_k(\mathbf{x}) = \max_{i \rightarrow k} \tau^i(\mathbf{x}) \tag{1}$$

For a crisp label, **(x)** is assigned to the class with the largest $g_k(\mathbf{x})$.

2) TYPE 2: Function as consequent
The rules for this classifier model can be of the following type,
$Rule^i$: IF x_1 is A_1^i AND...AND $x_n$is A_n^i THEN

$$y_1^i = \sum_{j=0}^{n} a_{j1}^i x_j \text{ AND...AND } y_m^i = \sum_{j=0}^{n} a_{jm}^i x_j \tag{2}$$

where x_j,$j = 1, 2, ..n$ is the j^{th} input variable, A_1^i is an antecedent fuzzy set of rule i, $y_k^i, k = 1, 2, ..m$ is the output of rule i for each class, and a_{jk}^i is the j^{th} consequence parameter of the output k of the rule i.
In this model every rule votes for all the classes. Among several methods, the output can be determined using the weighted sum aggregation method,

$$g_k(\mathbf{x}) = \frac{\sum_{i=1}^{R} y_k^i \tau^i(\mathbf{x})}{\sum_{i=1}^{R} \tau^i(\mathbf{x})} \tag{3}$$

As in TYPE 1 classifier, for a crisp label, X is assigned to the class with maximum value of $g_k(\mathbf{x})$

3) Type 3: Linguistic label as consequent
The rules for this classifiers model are of the following form,

$$\begin{aligned} Rule^i : IF x_1 \text{ is } A_1^i \text{ AND...AND } x_n \text{ is } An^i \text{ THEN} \\ \textit{class k is } B_1^i \text{ AND } \textit{class l is } B_2^i \end{aligned} \tag{4}$$

where $x_j, j = 1, 2...n$is the input variable, A_1^i, and B_1^i are linguistic terms for $Rule^i$ (e.g Small, Large, Medium etc.) defined by fuzzy sets, and class k and class l, $1 \leq k \leq m$, $1 \leq l \leq m$, and $k \neq l$ are the class labels.
The classifier in this case operates as a Mamdani-type fuzzy system, and the output is a soft label.

There are several methods to learn rules for fuzzy classifiers, among which are fuzzy neural networks [21], genetic algorithms [22] and clustering-based approaches [23], [24].

In this work we have used Type 1 fuzzy classifier with Gaussian membership functions, where the firing strength of a rule is given as:

$$\tau_i(x) = \prod_{j=1}^{n} \mu_{ij}, i = [1, R], j = [1, n] \text{ and } \tag{5}$$

$\mu_{ij} = e^{\frac{\|x-x_i^*\|}{2(r_{ij})^2}}$ is the membership function and r_{ij} is the spread of membership function.

The rules of the FRB classifier are extracted using subtractive clustering [23]. In subtractive clustering , every data point is considered to be a potential cluster centre. This potential is quantified in terms of a value (potential value). The potential value depends on the distance of the data point to all other data points, the larger the number of neighbouring data points the higher is the potential. The neighbourhood of a data point is defined by a constant (radius); data points outside the neighbourhood do not have significant influence on the potential value. Initially the potential value is determined for every data point and the point with the highest potential value is selected as the first cluster center. After identifying the first cluster center, the potential of all data points is reduced by an amount that is dependent on their distance to the cluster center. So, the points closer to the cluster center have less chance to be selected as next cluster center. Now, the next cluster center is the point with the next maximum potential. The termination of the process is controlled by two threshold values. If the ratio of potential of the current data point **(x)** and the potential of the first cluster center is greater than an upper threshold value then **(x)** is accepted as cluster center and the process continues. If this ratio is less than a lower threshold value then **x** is rejected and the process terminates (in [23] the author used 0.5 as upper threshold value and 0.15 as lower threshold value). If the ratio lies between the two threshold values then it is checked if the data point provides trade-off between having a sufficient potential and not in proximity to the existing cluster centers. Each of the estimated cluster centers forms the basis of the fuzzy rules. Considering equation 5, x_i^* are the cluster centers estimated by sub cluster and r_i is the radius of the cluster. In subtractive clustering the radius is user-defined input parameter for which the suggested value is in the range 0.2-0.5[23].

III. METHODOLOGY

The following steps describe the methodology for acoustic event classification. The four littering activities discussed in section I are considered as atomic events.

Step 1. In the first step the audio input signal is read, digitized, and normalized.

Step 2. Short-time energy calculation: First the audio signal is divided into F frames so that each frame contains a samples. The number of samples in each frame is given by

$$s = w_{(size)} * f_{(sampling)} \text{ and } F = l_{(signal)}/w_{(size)} \tag{6}$$

where, $l_{(signal)}$ is the total length of signal, $f_{(sampling)}$ is the sampling frequency of audio input signal, $w_{(size)}$ is the window size.
After dividing the signal into F frames, energy is calculated for each frame. short time energy (per frame) is calculated for the a normalized signal.Short time energy calculation of each frame is given by

$$E_q = \sum_{m=q-w_s+1}^{q} [x(m)w(q-m)]^2 \tag{7}$$

Where $w(q-m)$is the window, q is the sample that the analysis window is centered on, and w_s is the window length.

Step 3. To isolate the atomic events energy of each frame is used. If the energy of a frame is greater than pre-defined threshold energy, then the frame is considered as event causing frame. Such frames are considered for further processing and other frames are discarded. The threshold energy is determined experimentally.

Step 4. After the identification of all the event causing frames, frame-wise feature extraction is performed by using MFCC. Steps to calculate the MFCC feature of an input signal are given below [12].
a) Calculate the Fourier transform of audio signal.

$$\begin{aligned} X_{FF}(p) &= \sum_{t=0}^{N-1} x(t)exp^{(\frac{-2t\pi p}{N})} \\ \text{where, } x(t) &= t^{th} \text{ sample of signal} \\ X(p) &= p^{th} \text{ FFT coefficient} \\ N &= \text{total number of sample in signal} \\ \text{where,} 0 \le p &\le N-1 \\ 0 \le n &\le N-1 \end{aligned} \tag{8}$$

b) Map the powers of spectrum obtained above onto the mel scale, using triangular overlapping windows. Mel scale is represented as

$$\begin{aligned} mel_{(scale)} &= 1127 \log_e \left(1 + \frac{f_{(sampling)}}{100}\right) \\ f_{(sampling)} &= \text{frequency in hertz} \end{aligned} \tag{9}$$

c) Take the logs of powers at each of the mel-frequency.
d) Take the Discrete Cosine Transform (DCT) of each mel log power obtained in above step.

$$X_{DC}(k) = \sum_{t=0}^{N-1} x_k cos\left[\frac{\pi}{N}\left(t+\frac{1}{2}\right)k\right] \tag{10}$$

e) The MFCCs are the amplitude of resulting spectrum.
Here we are extracting 13 MFCC features only for event causing frames instead of extracting the MFCC for whole audio signal. Among the all DCT coefficients obtained after step b (here it is 26) only first 13 are used because Higher DCT coefficients represent fast changes in filterbank energies and degrades ASR (automatic speech recognition)

performance.

Step 5. The extracted 13 MFCC features of each frame are used for training purpose. If P is the total number of event causing frames in training audio signal then the size of training data matrix will be P $\times$ 13. Now, one-class FRB and one-class SVM are trained using this data.
Step 6. For preparing the test data, steps 1-4 are repeated for all test audio signals and a data matrix of size Q $\times$ 13 is obtained, where Q is the number of events causing frames in test signals.

For comparison of results, we are using widely used One-class SVM classifier. The one-class classification discovers the hyperplane that separate the desired function of training pattern from the origin of featured space.For classification, four one class FRB classifiers are trained for each LA's. These four classifiers constitutes the ensemble of FRB classifiers, the decision of all the classifiers in the ensemble is combined using the following methods:
Mean vote rule[25]: Suppose the number of one-class FRB classifiers is c and let us assume that classifier k has r_k rules where $k = 1..c$

$$y\left(\mathbf{x}\right) = \frac{1}{c}\sum_{k=1}^{c} I_k(\mathbf{x}) \tag{11}$$

where,

$$I_k(\mathbf{x}) = \begin{cases} 1, & if \max_l \left(\tau_k^l(\mathbf{x})\right) \geq \theta_k \\ 0, & if \max_l \left(\tau_k^l(\mathbf{x})\right) < \theta_k \end{cases} \tag{12}$$

and $l = 1, ..r_k$, τ_k^l is the firing strength of l, and θ_k is the threshold for classifier k.
The instance **x** is the predicted to be of target class (LA) if $y(\mathbf{x})$ is greater then a threshold value (θ)
Maximum rule[18]: It first determines the maximum firing strength within the rules of classifiers λ_k, $k = 1..c$. It then takes the maximum among the individual classifiers outputs and if this value is greater than a threshold (θ) then the class of **x** predicted as LA.

$$y(\mathbf{x}) = \begin{cases} 1, & if max_k \left(\lambda_k(\mathbf{x})\right) \geq \theta \\ 0, & if max_k \left(\lambda_k(\mathbf{x})\right) < \theta \end{cases} \tag{13}$$
$$\text{where, } \lambda_k(\mathbf{x}) = max(\tau_k^l(\mathbf{x}))$$

IV. EXPERIMENTS AND RESULTS

As mentioned in section I, the four littering activities are performed by various people, and sound produced by each of the activities is captured in regular intervals using a microphone in a semi open indoor environment. Activities are performed within a distance of around 2 meters from the recording device and the recording is performed at 16KHz. The training data consists of 55 short audio recordings of 5 seconds long (10 PC + 10 PO + 15 CD + 20 CC). Fig.1, Fig.2, Fig.3, Fig.4 shows sample audio signals for the four different LAs. The set of test data consists of 25, seconds long recordings with 5 recordings for each activity and 5 recordings for random activities other than LAs. The random activities are referred here as Non-Littering activities (NLAs).The first two models (M1 and M2) are based on 4 one-class FRB classifiers that are trained on data with label PC, PO, CD, CC respectively. The third one-class FRB model (M3) is developed by combining all the data from different classes resulting into a single one-class classifier. The decision is taken using the maximum rule. The fourth model (M4) is similar to M1 where 4 different one-class classifiers are trained using one-class SVM and mean vote rule is used. Finally, the fifth model (M5) is built in the same way as M3 using one-class SVM and considering entire training data as belonging to one class. Table-1 summarizes the five models.

TABLE I: MODEL DESCRIPTION

Models	Description
M1	Ensemble of four one-class FRB classifiers (mean vote rule)
M2	Ensemble of four one-class FRB classifiers (maximum rule)
M3	Single one-class FRB classifier
M4	Ensemble of four one-class SVM classifiers
M5	Single one-class SVM classifiers

TABLE II: DISTRIBUTION OF EVENT FRAMES

Type of Audio Recording	Audio Recordings		Events Frames	
	Training	Testing	Training	Testing
Packet Crushing	10	5	238	245
Packet Opening	10	5	188	200
Cup Drop	15	5	248	91
Cup Crushing	20	5	284	137
NLAs	-	1	-	173

As discussed on section III, the processing is performed on a frame-by-frame basis. When the energy of a frame exceeds the defined threshold, it indicates occurrence of an event (possibly littering activity). Such frames are identified and considered for both training and testing. The distribution of high energy frames (event frames) are given in Table II. We performed experiments by varying the threshold parameter and computed the true positive rate (TP) and false positive rate (FP). True positive rate is the proportion of positive cases (LAs) that were correctly identified and false positive rate is the proportion of negative cases (NLAs) that were incorrectly classified as positive (LAs). For our task, we require high TP and low FP. In Table III, the TP and FP values of all the classifier models along with the various threshold values is presented. It shows

TABLE III: EXPERIMENTAL RESULTS

Models	Threshold values									
	0.1		0.2		0.3		0.4		0.5	
	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
M1	0.95	0.04	0.95	0.04	0.81	0	0.81	0	0.81	0
M2	0.98	0.15	0.95	0.01	0.86	0	0.60	0	0.42	0
M3	1	0.87	0.95	0.49	0.88	0.13	0.75	0.01	0.53	0
M4	0.93	0.10	0.95	0.07	0.92	0.09	0.89	0.06	0.88	0.04
M5	0.98	0.19	0.98	0.08	0.93	0.05	0.91	0.02	0.89	0.01

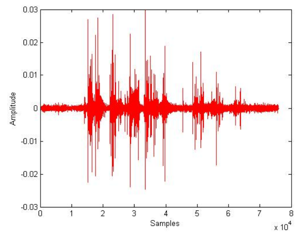

Fig. 1: Cup Crushing

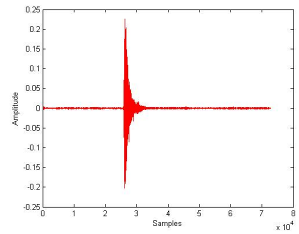

Fig. 2: Cup Drop

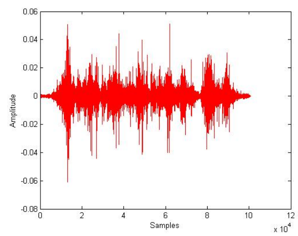

Fig. 3: Packet Crushing

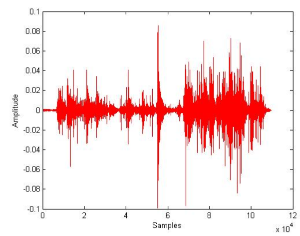

Fig. 4: Packet Opening

that the maximum TP is achieved by M3 when threshold is 0.1, but it also has very high value of FP which is not desirable. The models M1, M2, and M3 achieved minimum value of FP when threshold is 0.4 and 0.5, but at the same time they show inferior TP values. Models M1, M2, and M5 give acceptable values for TP and FP. Thus, it shows that the results of ensemble of one-class FRB classifiers has very low false alarm rate and recognition rate is also comparable to one-class SVM classifiers.

V. Conclusion and Future Work

In this paper we investigated ensemble of one-class classifiers as one of the solutions to the acoustic event classification problem when data for negative class is not available. We developed five classification models based on one-class FRB and one-class SVM to identify litter activities in public spaces. The initial results obtained from the classifiers suggest that ensemble of one-class FRB classifier can correctly identity littering activities and at the same time has a low false alarm rate. Additionally, they are suitable for incremental learning.

In the work presented here, we performed experiments on recordings based on atomic events. We are currently experimenting with audio recordings based on mixed events. The models described here perform per frame classification. As an activity can last more than a frame so in future we would consider classification based on per event that may consist of multiple frames. Finally, this litter activity recognition system will be hardware implemented on low power devices, so we would like to investigate light weight feature extraction methods compared to MFCCs.

References

[1] A. Shirkhodaie and A. Alkilani, "A survey on acoustic signature recognition and classification techniques for persistent surveillance systems," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83 920U–83 920U.

[2] Y. Zhan, S. Miura, J. Nishimura, and T. Kuroda, "Human activity recognition from environmental background sounds for wireless sensor networks," in *Networking, Sensing and Control, 2007 IEEE International Conference on*. IEEE, 2007, pp. 307–312.

[3] P. Zhang, Z. Sun, and A. Purohit, "Coughloc: Location-aware indoor acoustic sensing for non-intrusive cough detection," in *International Workshop on Emerging Mobile Sensing Technologies, Systems, and Applications*, 2011.

[4] M. Popescu, S. Coupland, and S. Date, "A fuzzy logic system for acoustic fall detection." in *AAAI Fall Symposium: AI in Eldercare: New Solutions to Old Problems*, 2008, pp. 78–83.

[5] M. Popescu and A. Mahnot, "Acoustic fall detection using one-class classifiers," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3505–3508.

[6] A. Temko, R. Malkin, C. Zieger, D. Macho, C. Nadeu, and M. Omologo, "Acoustic event detection and classification in smart-room environments: Evaluation of chil project systems," *Cough*, vol. 65, no. 48, p. 5, 2006.

[7] G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machines," *Neural Networks, IEEE Transactions on*, vol. 14, no. 1, pp. 209–215, 2003.

[8] L. Lu, H.-J. Zhang, and H. Jiang, "Content analysis for audio classification and segmentation," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 7, pp. 504–516, 2002.

[9] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *Speech and Audio Processing, IEEE transactions on*, vol. 10, no. 5, pp. 293–302, 2002.

[10] J. Stork, L. Spinello, J. Silva, K. O. Arras *et al.*, "Audio-based human activity recognition using non-markovian ensemble voting," in *RO-MAN, 2012 IEEE*. IEEE, 2012, pp. 509–514.

[11] B. Liang, H. Yaali, L. Songyang, C. Jianyun, and W. Lingda, "Feature analysis and extraction for audio automatic classification," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, vol. 1. IEEE, 2005, pp. 767–772.

[12] M. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition," *Speech Communication*, vol. 54, no. 4, pp. 543–565, 2012.

[13] X. Zhuang, X. Zhou, T. S. Huang, and M. Hasegawa-Johnson, "Feature analysis and selection for acoustic event detection," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 17–20.

[14] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson, "Acoustic fall detection using gaussian mixture models and gmm supervectors," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 69–72.

[15] X. Zhuang, X. Zhou, M. A. Hasegawa-Johnson, and T. S. Huang, "Real-world acoustic event detection," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1543–1551, 2010.

[16] X. Zhou, X. Zhuang, M. Liu, H. Tang, M. Hasegawa-Johnson, and T. Huang, "Hmm-based acoustic event detection with adaboost feature selection," in *Multimodal technologies for perception of humans*. Springer, 2008, pp. 345–353.

[17] A. Temko and C. Nadeu, "Classification of acoustic events using svm-based clustering schemes," *Pattern Recognition*, vol. 39, no. 4, pp. 682–694, 2006.

[18] R. Polikar, "Ensemble based systems in decision making," *Circuits and Systems Magazine, IEEE*, vol. 6, no. 3, pp. 21–45, 2006.

[19] J. C. Huhn and E. Hullermeier, "Fr3: a fuzzy rule learner for inducing reliable classifiers," *Fuzzy Systems, IEEE Transactions on*, vol. 17, no. 1, pp. 138–149, 2009.

[20] O. U. Obot and F. Uzoka, "Experimental study of fuzzy-rule based management of tropical diseases: Case of malaria diagnosis," *Studies in health technology and informatics*, vol. 137, p. 328, 2008.

[21] D. Nauck, F. Klawonn, and R. Kruse, *Foundations of neuro-fuzzy systems*. John Wiley & Sons, Inc., 1997.

[22] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *Fuzzy Systems, IEEE Transactions on*, vol. 3, no. 3, pp. 260–270, 1995.

[23] S. L. Chiu, "Fuzzy model identification based on cluster estimation." *Journal of intelligent and Fuzzy systems*, vol. 2, no. 3, pp. 267–278, 1994.

[24] L. Kuncheva *et al.*, "How good are fuzzy if-then classifiers?" *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 30, no. 4, pp. 501–509, 2000.

[25] D. M. Tax and R. P. Duin, "Combining one-class classifiers," in *Multiple Classifier Systems*. Springer, 2001, pp. 299–308.