Published in final edited form as:

Proc IEEE Int Conf Acoust Speech Signal Process. 2018 April ; 2018: 970–974. doi:10.1109/ICASSP.2018.8461992.

CLASSIFIER CASCADE TO AID IN DETECTION OF EPILEPTIFORM TRANSIENTS IN INTERICTAL EEG

Elham Bagheri[a], **Jing Jin**[a,b], **Justin Dauwels**[a], **Sydney Cash**[b], **M.Brandon Westover**[b]

[a]Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore 639798

[b]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; and Harvard Medical School, Cambridge, MA, USA

Abstract

The presence of Epileptiform Transients (ET) in the electroencephalogram (EEG) is a key finding in the medical workup of a patient with suspected epilepsy. Automated ET detection can increase the uniformity and speed of ET detection. Current ET detection methods suffer from insufficient precision and high false positive rates. Since ETs occur infrequently in the EEG of most patients, the majority of recordings comprise background EEG waveforms. In this work we establish a method to exclude as much background data as possible from EEG recordings by applying a classifier cascade. The remaining data can then be classified using other ET detection methods. We compare a single Support Vector Machine (SVM) to a cascade of SVMs for detecting ETs. Our results show that the precision and false positive rate improve significantly by incorporating a classifier cascade before ET detection. Our method can help improve the precision and false positive rate of an ET detection system. At a fixed sensitivity, we were able to improve precision by 6.78%; and at a fixed false positive rate, the sensitivity improved by 2.83%.

Index Terms

Classifier cascade; classifier ensemble; interictal spike detection; epilepsy; support vector machine

1. INTRODUCTION

The scalp EEG of a patient having epilepsy is typically characterized by presence of occasional epileptiform transients (ETs), including spikes with 20–70 ms and sharp waves with 70–200 ms duration. The presence of ETs predicts recurrence of seizures following a first seizure [1, 2]. In addition, it supports the diagnosis of epilepsy [3]. However, it is challenging to detect ETs because of the large variety of ET morphologies. ETs can also be similar to waves that are part of normal background activity and to artifacts, such as extracerebral potentials from muscle, eyes, heart, electrodes, etc. [4].

The gold standard for ET detection in clinical practice remains visual inspection, which is tedious and time-consuming. The disagreement among EEG experts in annotating ETs is substantial [5–7]. Misinterpretation of the EEG can lead to misdiagnosis of epilepsy and delay the treatment of the correct underlying cause of the seizure-like events [5, 8]. With all

these underlining problems, we are motivated to develop automated ET detection methods, which increases uniformity in EEG annotation of patients with epilepsy [4].

Many methods have been developed for automated ET detection. These methods include mimetic analysis, template matching, parametric methods, power spectral analysis, artificial neural networks [4], and wavelet transforms [9]. Some methods have combined different techniques of classification as well as artifact removal methods [10–13]. For instance, they have applied template matching combined with clustering [10], template matching combined with support vector machines (SVMs) and independent component analysis (ICA) [11], nonlinear energy operator combined with mimetic analysis and Adaboost classifiers [12], and sequence merging combined by SVMs [13], wavelet transform with machine learning techniques [14, 15]. In a recent study [16], the performance of an ET detection software on a dataset consisting of 40 epileptic patients, was compared with ET detection by three skilled technologists. They reported that the algorithm is noninferior to human experts. However, the common problem with these methods is the lack of a sizable dataset consisting of different ET profiles to validate the performance of the ET detection systems. Reliability for clinical purposes has not been established. There are currently no universally-accepted ET detection systems. Existing ET detection systems either suffer from a low sensitivity [5, 17]; or a low specificity [4,5,18,19].

In this study, a large dataset consisting of 156 interictal scalp EEG recordings is analyzed. We propose a new method to aid in classifying the interictal EEG. Our method is based on a classifier cascade, i.e., an ensemble of weak classifiers. We establish our method to exclude as many background EEG waveforms as possible from the EEG. Each weak classifier is designed to have a high sensitivity to detect ETs while allowing for modest or weak specificity. These classifiers are applied in a cascade; the cascade can then be combined with other ET detection algorithms to improve the performance.

2. METHODS

In clinical application of ET detection, there is limited tolerance for false alarms (or false positive rate), so that no healthy subjects will be diagnosed as having epilepsy. On the other hand, the true positive rate (sensitivity) does not need to be as high, since even with slightly lower sensitivity we can still detect ETs in sufficiently long EEG recordings. The main objective in ET detection is to determine whether any ETs exist in a patient's EEG, and if so to find their channel locations [20]. Therefore, we aim to develop our algorithm to lower the false positive rate and increase the precision. We consider a large EEG dataset in contrast to several existing studies; and by applying cross-validation, we ensure the results generalize well.

2.1. EEG dataset

We consider recordings of 93 patients with epilepsy, and 63 subjects with non-epileptic EEG. The diagnosis was based on the clinical read which was done independent of this study. Average length of each EEG is 28.5 minutes. Most previous studies have only used the EEG from patients with epilepsy containing ETs. By contrast, we include spike-free EEGs for both training and validation as well. This is beneficial since we need the algorithm

to perform well on all types of patients. The data was acquired at Massachusetts General Hospital (MGH) using the 10–20 standard system of electrode placement. The sampling frequency is 128Hz, and signal is high-pass filtered at 1 Hz. Data is cross-annotated by two trained clinical neurophysiologists using the NeuroBrowser software [21]. Only the waveforms annotated by both experts are considered as ETs. There are a total number of 18,164 ETs in the dataset, with 143 ETs per patient on average, and minimum 1 and maximum 1987 ETs per patient. 5-fold cross-validation was used to split the data into training and test sets. We divided the EEG in segments of 0.5s, and casted the ET detection as a binary classification problem for each segment.

2.2. Classifier cascade

We develop a model ensemble by considering different single classification models. In ensemble learning, rather than creating a single model, a set of models are produced and predictions are made by aggregating the outputs of these models. A prediction model which consists of a set of models is called a model ensemble. The motivation behind using ensemble methods is the idea that a committee of experts working together on a problem are more likely to solve it successfully than a single expert working alone [22].

In our method, we generate several classification models having a high sensitivity. We compare these models in terms of their specificity and select the best one in each step of the algorithm. The training set shrinks at each step. The selected classifiers at every step are applied on the dataset in the form of a cascade. We employ the Support Vector Machine (SVM) with Gaussian radial basis kernel as the basic unit of the classification cascade [22]. Our hypothesis is that by sequentially eliminating background waveforms using SVMs applied as a cascade, while retaining possible ETs, we can increase the precision and reduce the number of false positive detections of an ET detection method.

2.3. Training phase

ETs are considered as 0.5s segments at the specific EEG time and channel annotated by experts. All waveforms overlapping in time with any ET, and 0.5s before and after any ET, on all channels are eliminated from background candidates. We first split data in 5 cross-validation folds. For each training round, we sample one background waveform for each ET, so that the training set is balanced. The average number of ETs is 14,531 per fold. We train classifiers using each background set (from each EEG in training set) and all ETs. For each fold, the number of trained SVMs equals to the number of training EEGs. Thus, around 124–125 SVMs for each fold, and 624 SVMs over all folds are trained.

We then select one of the trained SVMs for the first stage, as follows: All trained SVMs are first applied on the whole training set, including all ETs and the sampled background from all training subjects. Next, we adjust the threshold on the output scores of the classifiers in such a way that sensitivity is greater or equal to 0.999. The classifier having the highest specificity is selected, and is applied on the whole training set.

Hence, all waveforms in the training set which are labeled “non-ET” after applying the selected SVM are eliminated from the dataset. In the next step, again all classifiers are applied on the new training waveforms labeled “ET” in the previous step, and the same

procedure is repeated. The optimum number of steps can be determined based on desired performance metrics.

One purpose for training several classifiers with different training sets is to avoid unbalanced training. Since in interictal EEG, there are only few ETs typically, the training set would be highly unbalanced if using one single classifier. In our approach, each SVM is trained with a balanced input.

2.4. Testing phase

In the testing phase, we apply the classifier cascade to all the EEG waveforms in the test set. In other words, in this phase all ETs and background waveforms are considered. By contrast, in the training phase, one background waveform is randomly selected for each ET, leading to a balanced training set. The test set contains an average of 3633 ETs per fold.

2.5. Combining the cascade with an ET detection method

We test a single SVM as benchmark and compare it to the classifier cascade followed by the same SVM. In this way, we can assess the benefits of the classifier cascade. The two approaches for ET detection are illustrated in Fig. 1. We propose using the classifier cascade to improve the performance compared to one single SVM.

3. RESULTS AND DISCUSSION

- **Classifier cascade**

  We determined the classifiers and the thresholds to be applied on their output by keeping the sensitivity at 0.999 at every stage. After learning the classifiers, we applied them on the test set for every fold. The testing sensitivity and specificity of the algorithm for all 5 folds as well as the average value over all folds are listed in Table 1 and Table 2, respectively. After 10 steps, we achieve specificity of 0.919 at sensitivity of 0.883. The overall sensitivity decreases as we add more steps, while the overall specificity is increased, as expected. The choice of number of stages to be used depends on the application.

  The ratio of retained ETs (sensitivity) and excluded background waveforms (specificity) for individual subjects is illustrated in Fig. 2. For the majority of subjects the sensitivity and specificity values are high. However, for a few subjects the algorithm has low performance. The subjects with low sensitivity typically have very few number of ETs (l,4,14), while one subject with sensitivity of 0.26 has 226 ETs.

- **ET detection by SVM**

  We take 2 approaches for ET detection: (1) We train and test an ET detection method using a single SVM; (2) We train and test an SVM on the data retained from the classifier cascade, considering a 4-step cascade which includes 4 SVMs.

  We compute the sensitivity, precision, and false positive rate (FPR) per minute of EEG recording. The results for the overall algorithm are listed in Table 3. We compute and compare the performance on parts of ROC where the overall specificity of ET detection

equals to 0.99, 0.999, and 0.9999. We consider the high values of specificity, since we require the overall specificity of the system to be very high in order to control the false positives.

As shown in Table 3, sensitivity declines by applying the classifier cascade to eliminate background EEG before applying the ET detection method. This is expected, since at every stage of the cascade we lose a small portion of ETs. As mentioned earlier, we are more interested in high precision than high sensitivity. Hence, we can compromise on sensitivity to some extent. On the other hand, we observe that precision is significantly improved, and false positive rate per minute is reduced by adding steps to the cascade.

The precision versus sensitivity of the SVM detector, with and without the cascade, is shown in Fig. 3. This figure shows the precision versus sensitivity (recall) for the specific values of specificity=0.99, 0.999, 0.9999 similar to the values included in Table 3. The sensitivity versus false positive rate per minute on a logarithmic scale (with base 10), corresponding to each of the three mentioned specificity values, is shown in Fig. 4.

From Fig. 3 we observe that the precision-recall curve for ET detection with the classifier cascade is shifted upwards compared to the case in which only a SVM detector is applied to the whole dataset. This shift in the precision-recall curve indicates that by including the cascade of classifiers as the first step, we can achieve higher precision for the same sensitivity. Fig. 4 shows that the false positive rate is significantly reduced by including the cascade before applying SVM ET detector, and by increasing the number of classifiers included in the cascade.

According to Fig. 4, for a fixed false positive rate of 1.2 per minute (−0.081 in logarithmic scale), sensitivity is increased from 39.70% to 42.53% which indicates a total increase of 2.83% in sensitivity.

Fig. 3 shows that for a fixed sensitivity of 42.53%, precision is increased from 59.38% to 66.16%, indicating a total increase of 6.78% in precision.

In summary, the results show that our method can successfully increase the precision and reduce the number of false positive detections of an ET detection system. The classifier cascade can be incorporated with any ET detection system as an initial step.

4. CONCLUSIONS

In this study, we propose a method to reject background EEG activity from interictal EEG of patients with epilepsy by applying an ensemble of classifiers as a cascade. The purpose is to improve the performance of ET detection systems in terms of precision and false positive rate. We use a large dataset consisting of 93 epileptic subjects as well as 63 subjects with ET-free EEG.

In future work, we hope to apply this method on an even larger dataset to provide more reliable performance. In addition, we will employ other classifiers as basic units of the system, in order to increase the background rejection rate and further improve precision and

false positive rate. We will consider various sizes for cascade as well, to pick the optimum for our application. We will also investigate the algorithm with more ET detection methods.

Furthermore, we will perform more analysis on individual subjects to determine why the algorithm does not perform very well for certain patients. Understanding the characteristics of the patients on which our algorithm fails to perform well could give clues to tune the ET detection algorithms to handle these difficult cases.

Finally, we plan to analyze ETs that are lost at the early stages of the cascade. In this way, we can learn more about morphologies of ETs that can easily be misinterpreted by ET detection algorithms. We will also analyze the background waveforms which are retained after the late stages of the algorithm. These waveforms comprise difficult-to-reject false positive detections; therefore, studying their morphologies and trying to eliminate them could yield further reduction in the false positive rate of ET detection methods in the future.

Our ultimate goal is to develop an efficient ET detection system for clinical applications. We plan to process the waveforms that are retained after the cascade by using other machine learning algorithms.

5. REFERENCES

[1]. van Donselaar Cees A, Schimsheimer Robert-Jan, Geerts Ada T, and Declerck August C, “Value of the electroencephalogram in adult patients with untreated idiopathic first seizures,” Archives of neurology, vol. 49, no. 3, pp. 231–237, 1992. [PubMed: 1536624]

[2]. Seidel Stefan, Pablik Eleonore, Aull-Watschinger Susanne, Seidl Birgit, and Pataraia Ekaterina, “Incidental epileptiform discharges in patients of a tertiary centre,” Clinical Neurophysiology, vol. 127, no. 1, pp. 102–107, 2016. [PubMed: 25802204]

[3]. Fountain Nathan B and Freeman John M, “Eeg is an essential clinical tool: pro and con,” Epilepsia, vol. 47, no. s1, pp. 23–25, 2006. [PubMed: 17044821]

[4]. Halford Jonathan J., “Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ecg interpretation,” Clinical Neurophysiology, vol. 120, no. 11, pp. 1909–1915, 2009. [PubMed: 19836303]

[5]. Halford Jonathan J, Schalkoff Robert J, Zhou Jing, Benbadis Selim R, Tatum William O, Turner Robert P, Sinha Saurabh R, Fountain Nathan B, Arain Amir, Pritchard Paul B, et al., “Standardized database development for eeg epileptiform transient detection: Eegnet scoring system and machine learning analysis,” Journal of neuroscience methods, vol. 212, no. 2, pp. 308–316, 2013. [PubMed: 23174094]

[6]. Halford JJ, Arain A, Kalamangalam GP, LaRoche SM, Bonilha L, Basha M, Azar NJ, Kutluay E, Martz GU, Wolf BJ, et al., “Characteristics of eeg interpreters associated with higher inter-rater agreement.,” Journal of Clinical Neurophysiology, 2017.

[7]. Bagheri Elham, Dauwels Justin, Dean Brian C, Waters Chad G, Westover M Brandon, and Halford Jonathan J, “Interictal epileptiform discharge characteristics underlying expert interrater agreement,” Clinical Neurophysiology, vol. 128, no. 10, pp. 1994–2005, 2017. [PubMed: 28837905]

[8]. Benbadis Selim R, “Errors in eegs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies,” Epilepsy & Behavior, vol. 11, no. 3, pp. 257–262, 2007. [PubMed: 17719853]

[9]. Güler Inan and Derya Übeyli Elif, “Adaptive neurofuzzy inference system for classification of eeg signals using wavelet coefficients,” Journal of neuroscience methods, vol. 148, no. 2, pp. 113–121, 2005. [PubMed: 16054702]

[10]. Nonclercq Antoine, Foulon Martine, Verheulpen Denis, De Cock Cathy, Buzatu Marga, Mathys Pierre, and Van Bogaert Patrick, “Cluster-based spike detection algorithm adapts to interpatient and intrapatient variation in spike morphology,” Journal of Neuroscience Methods, vol. 210, no. 2, pp. 259–265, 2012. [PubMed: 22850558]

[11]. Lodder Shaun S., Askamp Jessica, and van Putten Michel J.A.M., “Inter-ictal spike detection using a database of smart templates,” Clinical Neurophysiology, vol. 124, no. 12, pp. 2328–2335, 2013. [PubMed: 23791532]

[12]. Liu Yung-Chun, Lin Chou-Ching K., Tsai Jing-Jane, and Sun Yung-Nien, “Model-based spike detection of epileptic eeg data,” Sensors, vol. 13, no. 9, pp. 12536–12547, 2013. [PubMed: 24048343]

[13]. Zhang Jian, Zou Junzhong, Wang Min, Chen Lanlan, Wang Chunmei, and Wang Guisong, “Automatic detection of interictal epileptiform discharges based on time-series sequence merging method,” Neurocomputing, vol. 110, pp. 35–43, 2013.

[14]. Chavakula Vamsidhar, Fernández Iván Sánchez, Peters Jurriaan M, Popli Gautam, Bosl William, Rakhade Sanjay, Rotenberg Alexander, and Loddenkemper Tobias, “Automated quantification of spikes,” Epilepsy & Behavior, vol. 26, no. 2, pp. 143–152, 2013. [PubMed: 23291250]

[15]. Bagheri Elham, Jin Jing, Dauwels Justin, Cash Sydney, and Westover M Brandon, “Fast and efficient rejection of background waveforms in interictal eeg,” in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016, pp. 744–748.

[16]. Scheuer Mark L, Bagic Anto, and Wilson Scott B, “Spike detection: Inter-reader agreement and a statistical turing test on a large data set,” Clinical Neurophysiology, vol. 128, no. 1, pp. 243–250, 2017. [PubMed: 27913148]

[17]. Ver Hoef Lawrence, Elgavish Rotem, and Knowlton Robert C, “Effect of detection parameters on automated electroencephalography spike detection sensitivity and false-positive rate,” Journal of Clinical Neurophysiology, vol. 27, no. 1, pp. 12–16, 2010. [PubMed: 20087204]

[18]. De Lucia Marzia, Fritschy Juan, Dayan Peter, and Holder David S, “A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis,” Medical & biological engineering & computing, vol. 46, no. 3, pp. 263–272, 2008. [PubMed: 18071771]

[19]. Indiradevi KP, Elias Elizabeth, Sathidevi PS, Nayak S Dinesh, and Radhakrishnan K, “A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram,” Computers in biology and medicine, vol. 38, no. 7, pp. 805–816, 2008. [PubMed: 18550047]

[20]. Webber WRS and Lesser Ronald P., “Automated spike detection in eeg,” Clinical Neurophysiology, vol. 128, no. 1, pp. 241–242, 2017. [PubMed: 27940048]

[21]. Jin Jing, Dauwels Justin, Cash Sydney, and Westover M Brandon, “Spikegui: Software for rapid interictal discharge annotation via template matching and online machine learning,” in Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE, 2014, pp. 4435–4438.

[22]. Kelleher John D, Namee Brian Mac, and D’Arcy Aoife, Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies, MIT Press, 2015.

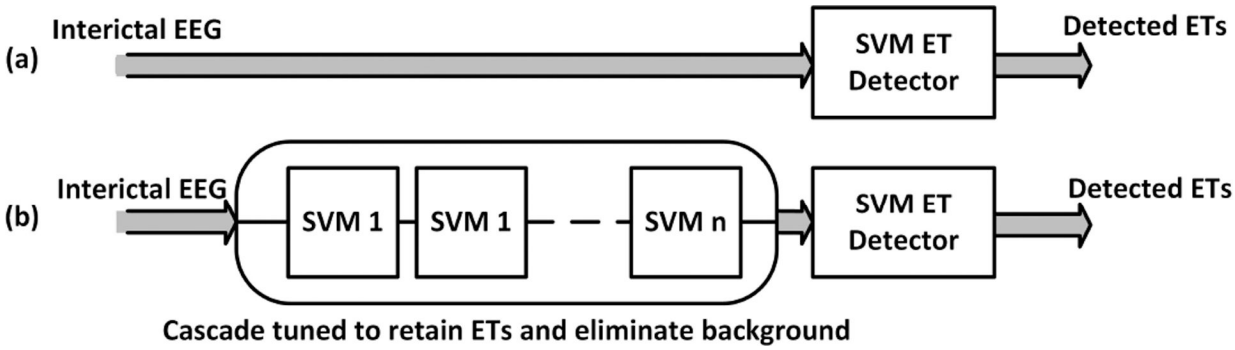

Fig. 1:

Two approaches for ET detection: (a) One single SVM detector (b) An SVM cascade followed by one SVM detector.

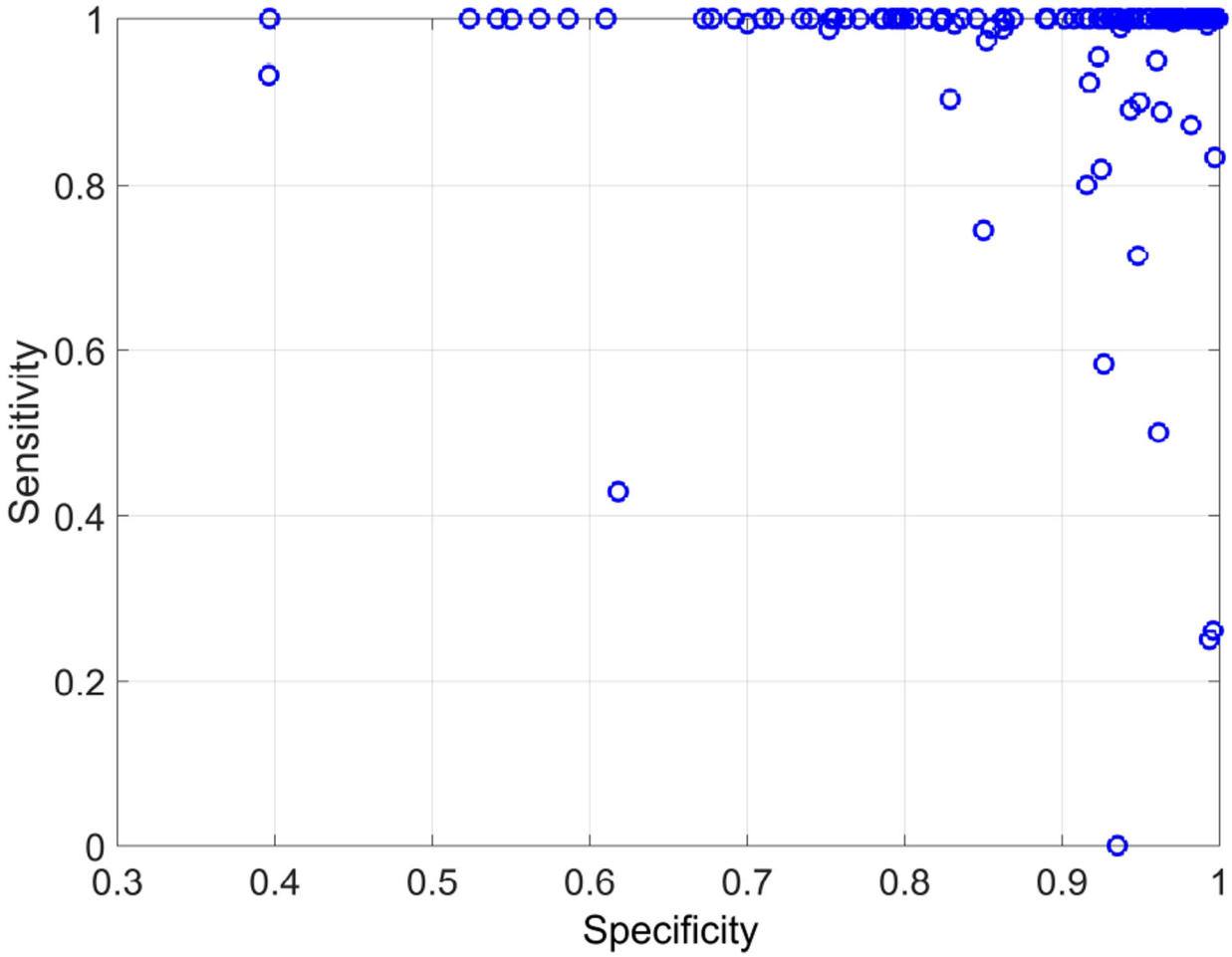

Fig. 2:
Sensitivity versus specificity for each subject after a 10-stage classifier cascade.

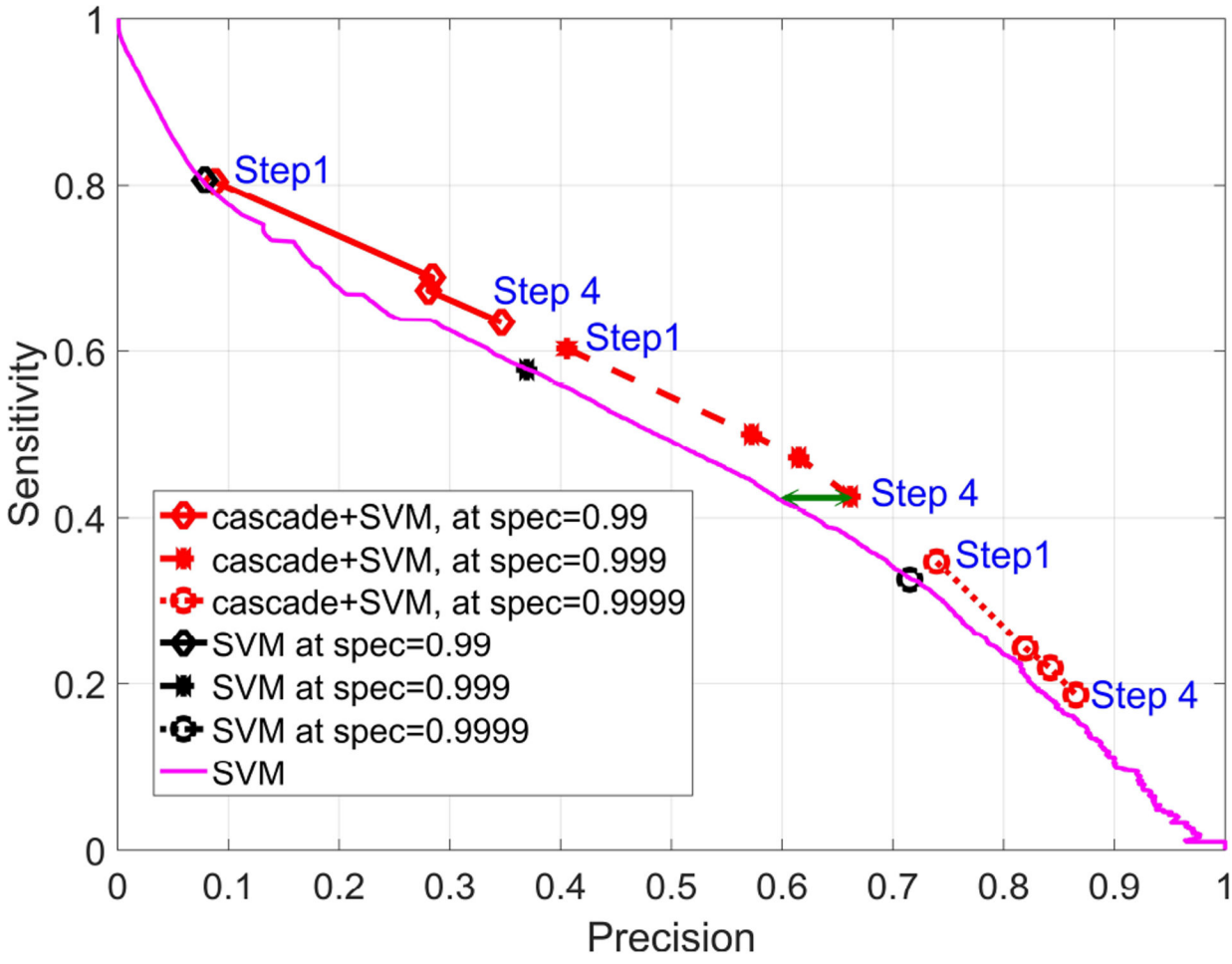

Fig. 3:
Precision versus sensitivity at certain specificity values, with and without the classifier cascade.

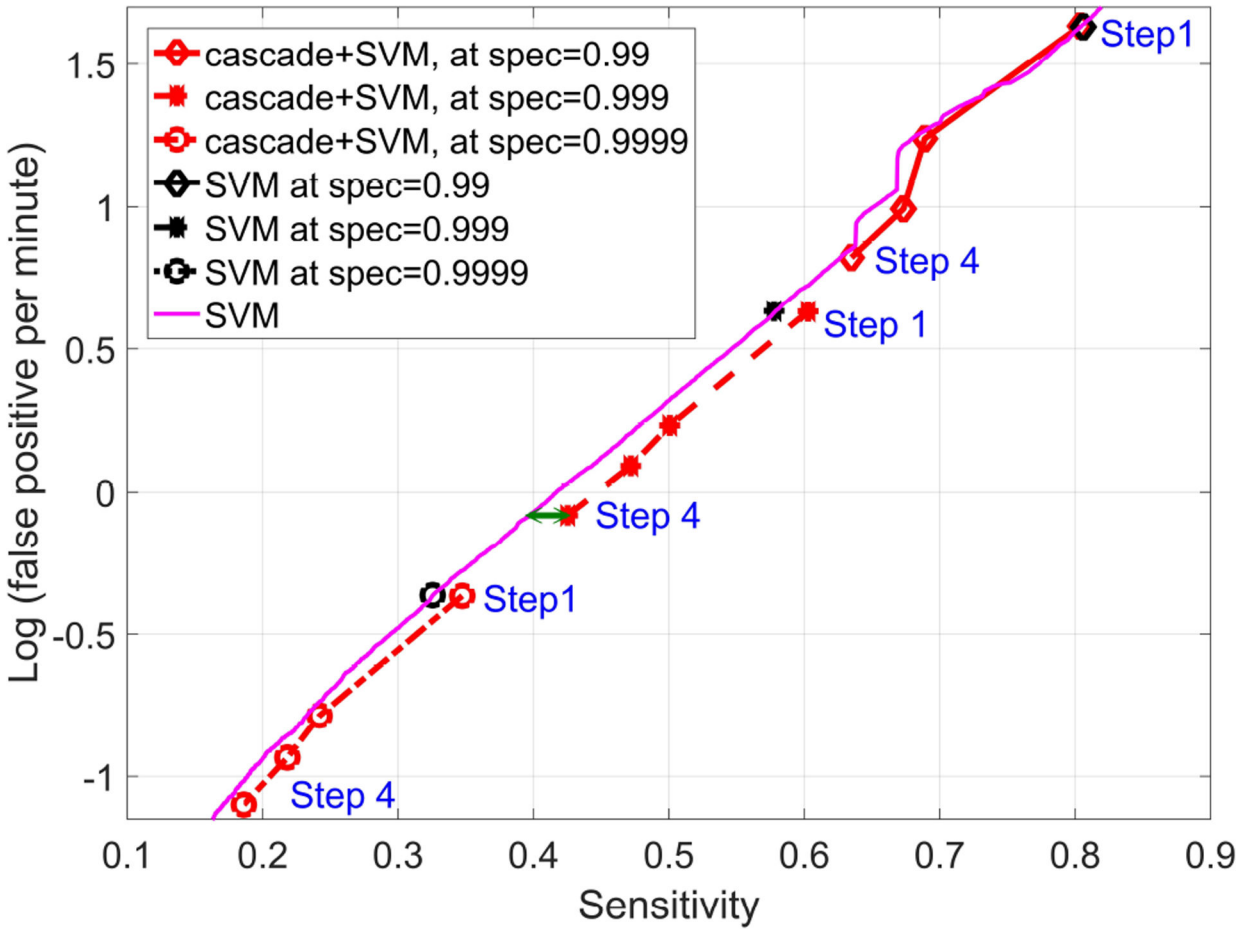

Fig. 4:

False positive rate versus sensitivity at certain specificity values, with and without classifier cascade.

Table 1:

Sensitivity of the classifier cascade after each step.

Stage	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
1	0.999	0.982	1	1	1	0.996
2	0.996	0.971	0.998	0.998	0.908	0.974
3	0.973	0.962	0.991	0.986	0.908	0.964
4	0.970	0.962	0.991	0.985	0.852	0.952
5	0.962	0.849	0.991	0.970	0.848	0.924
6	0.959	0.847	0.991	0.970	0.824	0.918
7	0.930	0.847	0.984	0.962	0.795	0.904
8	0.907	0.843	0.984	0.953	0.794	0.896
9	0.904	0.842	0.952	0.953	0.789	0.888
10	0.904	0.841	0.952	0.9487	0.771	0.883

Table 2:

Specificity of the classifier cascade after each step.

Step	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
1	0.533	0.784	0.323	0.577	0.586	0.561
2	0.738	0.848	0.408	0.730	0.754	0.696
3	0.836	0.888	0.556	0.828	0.792	0.780
4	0.851	0.908	0.668	0.861	0.843	0.826
5	0.875	0.915	0.720	0.886	0.866	0.852
6	0.887	0.927	0.762	0.904	0.879	0.872
7	0.903	0.933	0.792	0.918	0.896	0.888
8	0.910	0.940	0.825	0.927	0.915	0.904
9	0.914	0.945	0.828	0.934	0.923	0.909
10	0.925	0.950	0.851	0.941	0.930	0.919

Table 3:

Overall performance of ET detection by applying SVM with and without the initial classifier cascade.

Method	Specificity=0.99			Specificity=0.999			Specificity=0.9999		
	Sensitivity	Precision	FPR	Sensitivity	Precision	FPR	Sensitivity	Precision	FPR
SVM	0.806	0.078	42.76	0.577	0.369	4.289	0.325	0.715	0.432
1 stage+SVM	0.803	0.088	43.041	0.602	0.405	4.277	0.347	0.739	0.430
2 stages+SVM	0.689	0.284	17.254	0.501	0.572	1.705	0.242	0.819	0.163
3 stage+SVM	0.673	0.281	9.823	0.472	0.615	1.232	0.218	0.842	0.116
4 stage+SVM	0.634	0.346	6.635	0.425	0.661	0.828	0.186	0.865	0.079

Proc IEEE Int Conf Acoust Speech Signal Process. Author manuscript; available in PMC 2019 October 03.