

EEG-Based Person Verification Using Multi-Sphere SVDD and UBM

Phuoc Nguyen¹, Dat Tran¹, Trung Le², Xu Huang¹, and Wanli Ma¹

¹ University of Canberra, ACT 2601 Australia

² HCM City University of Education, Vietnam
`dat.tran@canberra.edu.au`

Abstract. The use of brain-wave patterns extracted from electroencephalography (EEG) brain signals for person verification has been investigated recently. The challenge is that the EEG signals are noisy due to low conductivity of the human skull and the EEG data have unknown distribution. We propose a multi-sphere support vector data description (MSSVDD) method to reduce noise and to provide a mixture of hyperspheres that can describe the EEG data distribution. We also propose a MSSVDD universal background model (UBM) to model impostors in person verification. Experimental results show that our proposed methods achieved lower verification error rates than other verification methods.

Keywords: Support vector data description, support vector machine, person verification, EEG, universal background model.

1 Introduction

The use of brain-wave patterns for person verification has been investigated at IDIAP in Switzerland [1]. It has been shown that the brain-wave pattern of every individual is unique and that the electroencephalography (EEG) brain signal can be used for person identification [2]. Traditional biometrics such as fingerprint, voice, and retina can be damaged or missing for some people, however EEG biometric exists in every person. EEG has advantageous such as unobtrusive, requiring living person recording, spontaneous signal, individual uniqueness due to different brain configurations [3].

Person verification is different from person identification. Person identification is to match the user biometric data against all the records in a database, while person verification is to accept or to reject a user claimed identity providing his biometric data. EEG-based biometry is an emerging research topic that can open new research directions and applications. Most EEG-based biometry work was focusing mainly on person identification, for example [4] and very little work has been done on person verification [1]. In [5], Manhattan distances on autoregressive (AR) coefficients with PCA were used to compute thresholds for determining test patterns were clients or impostors, the person verification task from 5 subjects were done in 2 stages. In [6], Independent Component Analysis (ICA) was used to

determine dominating brain regions to extract AR features, then a Naive Bayes probabilistic model is employed for person authentication of 7 subjects with Half Total Error Rate (HTER) of 2.2%. In [1], Gaussian mixture models has been applied for person verification task on EEG signal from 9 subjects. Half total error rate of 6.6 % was achieved for imagination left task. Those results were satisfactory but the numbers of subjects were very small.

Most features in BCI research have high dimensionality due to the number of channels. In addition, the training sets are usually small since the training process is time consuming and demanding [7]. As a result, models based on density estimation do not have enough data for training. Usually, maximum a posteriori adaptation is used to overcome this problem as seen in [1]. On the other hand, models based on kernel aim at determining the boundaries of the data instead of probability density, hence they do not require a lot of data for training. In support vector data description (SVDD), a spherically shaped boundary around a normal data set is used to separate this set from abnormal data. The volume of this data description is minimized to reduce the chance of accepting abnormal data. In [8], SVDD was used to model 70 individuals using energy features and attained 0.9913 area under curve (AUC). However SVDD does not guarantee that the single spherically shaped boundary can best describe the EEG data due to complex data distributions and much noise and outliers. A better description is the use of multiple spheres, however there is currently no investigation available for EEG-based person verification.

In speaker verification, the task can be stated as a hypothesis testing between the two hypotheses: the input utterance is from the hypothesis speaker, client, (H_0) or not from the hypothesis speaker, impostors, (H_1). The difficulties are usually in modelling the hypothesis H_1 since it should represent all possible alternatives speakers. There are two approaches to model the alternatives speakers [9]. The first approach uses a set of other speaker models to cover the space of the alter native hypothesis, this is called likelihood ratio sets, cohorts or background speakers. The selection, size and combination of the background speakers have been the subject of much research [9]. The second approach pools speech from several speakers and train a single model, this is called the world model or universal background model (UBM). This approach has become the predominate and has been focused on selection and composition of the speakers [9]. The advantage of this approach is that a single background model can be trained once and shared for all hypothesis testing of individual speakers.

GMM UBM has been the state-of-the-art high performance probabilistic model for representing speaker model and alternative speaker models because of its capability to approximate arbitrary densities [10]. Recently, Support Vector Machine (SVM) has been used for speaker verification. Because SVM draws an optimal hyper-plane to separate two classes, its application to speaker verification is to separate the client from impostors. The score of a vector is its the distance to the hyperplane and the threshold can be changed by adding some amount to the decision value causing the hyperplane moves nearer or further the class H_0 . In [11] SVM was combined with GMM for speaker verification by stacking the

means of the adapted mixture components into a GMM super-vectors. Then a linear kernel is derived based on an approximation to KL divergence between two GMM models. In [12] and [8] the authors use SVDD for speaker verification task the distance to the sphere centers are used as scores, the former author convert score to probability and the later author use percentage of rejection data as threshold.

In this paper, we propose a multi-sphere SVDD (MSSVDD) person verification in which an optimisation problem and an iterative algorithm are proposed to determine model parameters for MSSVDD to provide a better data description to EEG data for a person. An SVM UBM model is proposed to represent the person and background models as a set of sphere distributions, each sphere can represent some characteristics of a person with well-trained decision boundaries. Hence the collection of spheres can hopefully represent the varieties of feature space.

Experimental results on 4 large data sets show that the proposed multi-sphere SVDD can perform as the GMM UBM and the SVM UBM out perform GMM UBM.

2 Multi-Sphere SVDD

2.1 Problem Formulation

Consider a set of m hyperspheres $S_j(c_j, R_j)$ with center c_j and radius R_j , $j = 1, \dots, m$. This hypersphere set is a good data description of the normal data set $X = \{x_1, x_2, \dots, x_n\}$ if each of the hyperspheres describes a distribution in this data set and the sum of all radii $\sum_{j=1}^m R_j^2$ should be minimised.

Let matrix $U = [u_{ij}]_{n \times m}$, $i = 1, \dots, n$, $j = 1, \dots, m$ where u_{ij} is the hard membership representing the belonging of data point x_i to hypersphere S_j , $u_{ij} = 0$ if x_i is not in S_j and $u_{ij} = 1$ if x_i is in S_j . The optimisation problem of multi-sphere SVDD can be formulated as follows

$$\min_{R, c, \xi} \left(\sum_{j=1}^m R_j^2 + C \sum_{i=1}^n \xi_i \right) \quad (1)$$

subject to

$$\begin{aligned} \sum_{j=1}^m u_{ij} \|\phi(x_i) - c_j\|^2 &\leq \sum_{j=1}^m u_{ij} R_j^2 + \xi_i \quad i = 1, \dots, n \\ \xi_i &\geq 0, \quad i = 1, \dots, n \end{aligned} \quad (2)$$

where $R = [R_j]_{j=1, \dots, m}$ is vector of radii, C is a constant, $\xi = [\xi_i]_{i=1, \dots, n}$ is vector of slack variables, $\phi(\cdot)$ is the nonlinear function related to the symmetric, positive definite kernel function $K(x_1, x_2) = \phi(x_1)^T \phi(x_2)$, and $c = [c_j]_{j=1, \dots, m}$ is vector of centres.

Minimising the function in (1) over variables R , c and ξ subject to (2) will determine radii and centres of hyperspheres and slack variables if the matrix U

is given. On the other hand, the matrix U will be determined if radii and centres of hyperspheres are given. Therefore an iterative algorithm will be applied to find the complete solution. The algorithm consists of two alternative steps: 1) Calculate radii and centres of hyperspheres and slack variables, and 2) Calculate membership U .

For classifying a data point x , the following decision function is used

$$f(x) = \text{sign} \left(\max_j \left\{ R_j^2 - \|\phi(x) - c_j\|^2 \right\} \right) \quad (3)$$

The unknown data point x is normal if $f(x) = +1$ or abnormal if $f(x) = -1$.

2.2 Calculating Radii, Centres and Slack Variables

The Lagrange function for the optimisation problem in (1) subject to (2) is as follows

$$L(R, c, \xi, \alpha, \beta) = \sum_{j=1}^m R_j^2 + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i \left(\|\phi(x_i) - c_{s(i)}\|^2 - R_{s(i)}^2 - \xi_i \right) - \sum_{i=1}^n \beta_i \xi_i \quad (4)$$

where $s(i)$ is index of the hypersphere to which data point x_i belongs and satisfies $u_{is(i)} = 1$ and $u_{ij} = 0 \forall j \neq s(i)$.

Setting derivatives of $L(R, c, \xi, \alpha, \beta)$ with respect to primal variables to 0, we obtain m individual optimisation problems as follows

$$\min \left(\sum_{i \in s^{-1}(j)} \alpha_i K(x_i, x_i) - \sum_{i, i' \in s^{-1}(j)} \alpha_i \alpha_{i'} K(x_i, x_{i'}) \right) \quad j = 1, \dots, m \quad (5)$$

subject to

$$\sum_{i \in s^{-1}(j)} \alpha_i = 1 \quad \text{and} \quad 0 \leq \alpha_i \leq C \quad j = 1, \dots, m \quad (6)$$

After solving all of these individual optimization problems, we can calculate the updating radii $R = [R_j]$ and centres $c = [c_j]$, $j = 1, \dots, m$ using the equations in SVDD.

2.3 Calculating Membership U

We use radii and centres of hyperspheres to update the membership matrix U . The following algorithm is proposed:

For $i = 1$ to n do

 If x_i is misclassified then

 Let $j_0 = \arg \min_j \left\{ \|\phi(x_i) - c_j\|^2 - R_j^2 \right\}$

```

        Set  $u_{ij_0} = 1$  and  $u_{ij} = 0$  if  $j \neq j_0$ 
    End if
Else
    Denote  $J = \{j : x_i \in S(c_j, R_j)\}$ 
    Let  $j_0 = \arg \min_{j \in J} \{||\phi(x_i) - c_j||^2\}$ 
    Set  $u_{ij_0} = 1$  and  $u_{ij} = 0$  if  $j \neq j_0$ 
End Else
End For
    
```

2.4 Iterative Learning Process

The proposed iterative learning process for multi-sphere SVDD will run two alternative steps until a convergence is reached as follows

```

Initialise  $U$  by clustering the normal data set in the input space
Repeat the following
    Calculate  $R$ ,  $c$  and  $\xi$  using  $U$ 
    Update  $U$  using  $R$  and  $c$ 
Until a convergence is reached.
    
```

3 MSSVDD UBM

The MSSVDD UBM models the background space by using multi-hypersphere distribution approach. The set of K background people is trained using multi-sphere SVDDs, each person is represented by n spheres resulting nK spheres in total. Then this single background model will be shared for all hypothesis testing of individuals. When testing hypothesis for a person, his/her model will be removed from the background model.

Below are widely used normalisation methods in person verification [13]:

$$L_0(x) = \log P(x|\lambda_0) \quad (7)$$

$$L_1(x) = \frac{P(x|\lambda_0)}{P(x|\lambda)} \quad (8)$$

$$L_1(x) = \log P(x|\lambda_0) - \log P(x|\lambda) \quad (9)$$

$$L_2(x) = \log P(x|\lambda_0) - \max_{\lambda \neq \lambda_0} \log P(x|\lambda) \quad (10)$$

The simplest method of scoring is to use the absolute likelihood score or in its log domain (7). The score (8) and its log domain (9) uses normalisation, the term $\log P(X|\lambda)$ is called the normalisation term and requires calculation of all impostors likelihood functions. An approximation of this method is to use only the closest impostor model for calculating the normalisation term (10)

Let x be a feature vector, $S = S(c, R)$ be a hyper-sphere, we can consider the probability of x belongs to the sphere S as

$$P(x|\lambda_S) = e^{-\|x-c\|}$$

Then the above scores for SVM will become

$$L_0(x) = \log P(x|\lambda_S) = -\|x - c_S\| \quad (11)$$

$$L_1(x) = \frac{P(x|\lambda_S)}{P(x|\lambda)} = \frac{e^{-\|x-c_S\|}}{\sum_{T \neq S} e^{-\|x-c_T\|}} \quad (12)$$

$$L_1(x) = \log P(x|\lambda_S) - \log P(x|\lambda) = -\|x - c_S\| - \log \sum_{T \neq S} e^{-\|x-c_T\|} \quad (13)$$

$$L_2(x) = \log P(x|\lambda_S) - \max_{\lambda \neq \lambda_S} \log P(x|\lambda) = -\|x - c_S\| + \max_{T \neq S} \|x - c_T\| \quad (14)$$

The above scores for SVM have simple interpretations. The score (11) with a radius threshold R_S checks whether x is inside or outside sphere S . The score (12), (13) and (14) check whether x is nearer to the sphere S than other sphere T .

In this paper, we define the probability of x belonging to the sphere S as $P(x|\lambda_S) = e^{-\|x-c\|-R}$ to incorporate the sphere size.

4 Gaussian Mixture Model (GMM)

Since the distribution of feature vectors in X is unknown, it is approximately modelled by a mixture of Gaussian densities, which is a weighted sum of K component densities, given by the equation

$$P(X|\lambda) = \prod_{t=1}^T P(x_t|\lambda) = \prod_{t=1}^T \sum_{i=1}^K w_i N(x_t, \mu_i, \Sigma_i) \quad (15)$$

where λ denotes a prototype consisting of a set of model parameters $\lambda = \{w_i, \mu_i, \Sigma_i\}$, w_i , $i = 1, \dots, K$, are the mixture weights and $N(x_t, \mu_i, \Sigma_i)$, $i = 1, \dots, K$, are the d -variate Gaussian component densities with mean vectors μ_i and covariance matrices Σ_i .

5 GMM-UBM

The background model for GMM-UBM is trained in a similar way to MSSVDD-UBM. The set of K background people is trained using GMMs, each person is represented by a mixture of n Gaussian resulting nK Gaussian in total. Then this single background model will be shared for all hypothesis testing of individuals. When testing hypothesis for a person, his/her model will be removed from the background model. The score (14) will be used for testing.

6 Hypothesis Testing

The verification task can be stated as a hypothesis testing between the two hypotheses: the input is from the hypothesis person, (H_0) or not from the hypothesis person (H_1) [13].

Let λ_0 be the claimed person model and λ be a model representing all other possible people, i.e. impostors. For a given input X and a claimed identity, the choice is between the hypothesis H_0 : X is from the claimed person λ_0 , and the alternative hypothesis H_1 : X is from the impostors λ . A claimed person's score $L(X)$ is computed to reject or accept the person claim satisfying the following rules

$$L(X) \begin{cases} > \theta_{L\text{accept}} \\ \leq \theta_{L\text{reject}} \end{cases} \quad (16)$$

where θ_L are the decision thresholds.

7 Experiments

7.1 Datasets

The Australian EEG Database used in this research consists of EEG recordings of 40 patients. This database consists of EEG records recorded at the John Hunter Hospital [14], near University of Newcastle, over an 11-year period. The recordings were made by 23 electrodes placed on the scalp sampled at 167 Hz for about 20 minutes.

The EEG motor movement/imagery (EEGMMIDB) dataset contain recordings of 109 subjects performed different motor/imagery tasks¹ [15]. Each subject performed 14 experimental runs in one or three minutes. The tasks include opening and closing fists or imagine opening and closing fists. The recordings used 64 electrodes 10-10 system.

The Alcoholism datasets come from a study to examine EEG correlates of genetic predisposition to alcoholism. The datasets contain EEG recordings of control and alcoholic subjects. Each subject was exposed to either a single stimulus (S1) or two stimuli (S1 and S2) which were pictures of objects chosen from the 1980 Snodgrass and Vanderwart picture set. When two stimuli were shown, they were presented in either a matched condition where S1 was identical to S2 or in a non-matched condition where S1 differed from S2. The 64 electrodes placed on the scalp sampled at 256 Hz for 1 second. The Alcoholism large dataset contains training and test data for 10 alcoholic and 10 control subjects. The Alcoholism full dataset contains 120 trials for 122 subjects. The summary of those datasets is listed in Table 1

¹ Available online at <http://www.physionet.org/pn4/eegmmidb/>

Table 1. Dataset descriptions

Dataset	#subjects	#tasks	#trials	length (seconds)
Australian EEG	40	free	1	1200
EEGMMIDB	109	4	14	60 or 180
Alcoholism (large)	20	2	120	1
Alcoholism (full)	122	2	120	1

Table 2. Identification rate vs number of Gaussians in GMM cross validation training

#Gaussians	1	2	4	8	16	32	64	128	256
Australian EEG	73.82%	86.11%	90.05%	91.60%	90.98%	91.86%	92.07%	92.12%	91.60%
EEGMMIDB	66.49%	71.92%	71.96%	71.71%	70.97%	70.63%	71.05%	70.01%	70.58%
Alcoholism (large)	86.72%	87.57%	85.73%	82.63%	79.94%	70.62%	75.85%	78.39%	77.54%
Alcoholism (full)	50.13%	50.55%	50.16%	49.17%	47.72%	45.72%	45.89%	45.41%	45.68%

7.2 Preprocessing and Feature Extraction

EEG signals were first divided into 15-millisecond segments. The Alcoholism large and full datasets are downsampled to 128 Hz. Then the raw EEG signals were spatially filtered using Surface Laplacian (SL) with spline interpolation constant of 4. This filter can represent better the cortical activity of local sources below the electrodes [1]. Then the power spectral density (PSD) in the band 8-30 Hz was estimated. The Welch’s averaged modified periodogram method was used for spectral estimation. Hamming window was set to 1 second and 50% overlap. 12 power components in the frequency band 8-30 Hz were extracted.

Besides PSD features, autoregressive (AR) model parameters were extracted. In AR model, each sample is considered linearly related with a number of its previous samples. The AR model has the advantage of low complexity and has been used for person identification and verification [16] [17] [5]. Burg’s lattice-based method was used with the AR model order 21, as a previous study [17] suggested when there were many subjects and epochs.

The electrodes C3, C4, Cz, P3, P4 and Pz were selected to extract PSD and AR features that result in $6 \times (12+21) = 198$ features. Those electrodes are available in both datasets and were used in previous study [1] based on expert knowledge.

7.3 Experimental Results

All experiments were conducted using 3-fold cross validation training and the best parameters found were used to train models on the whole training set and test on a separate test set. Table 2 shows the identification rate of GMM training. The best parameter number of Gaussian is 128 for Australian EEG dataset and 4 for EEGMMIDB dataset.

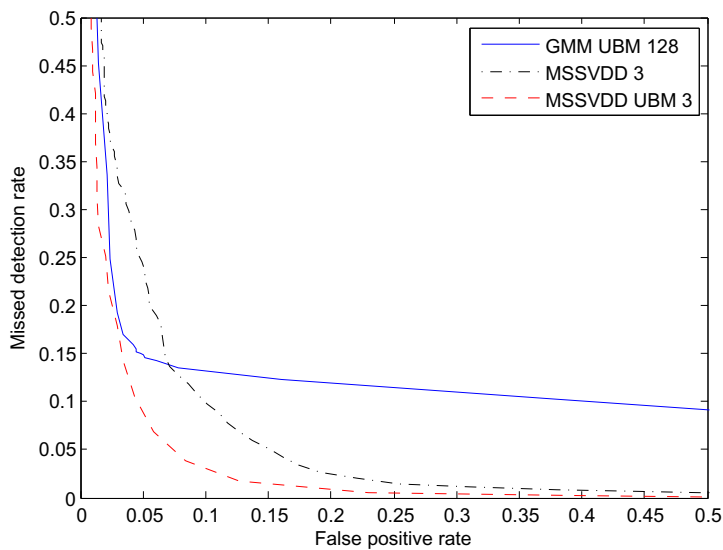


Fig. 1. DET curve of GMM UBM, 3-sphere SVDD, 3-sphere SVDD UBM on the Australian EEG dataset

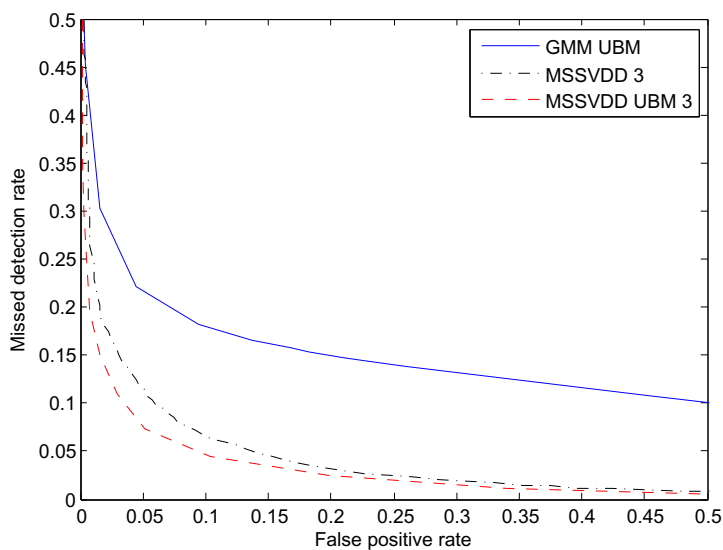


Fig. 2. DET curve of GMM UBM, 3-sphere SVDD, 3-sphere SVDD UBM on the EEGMMIDB dataset

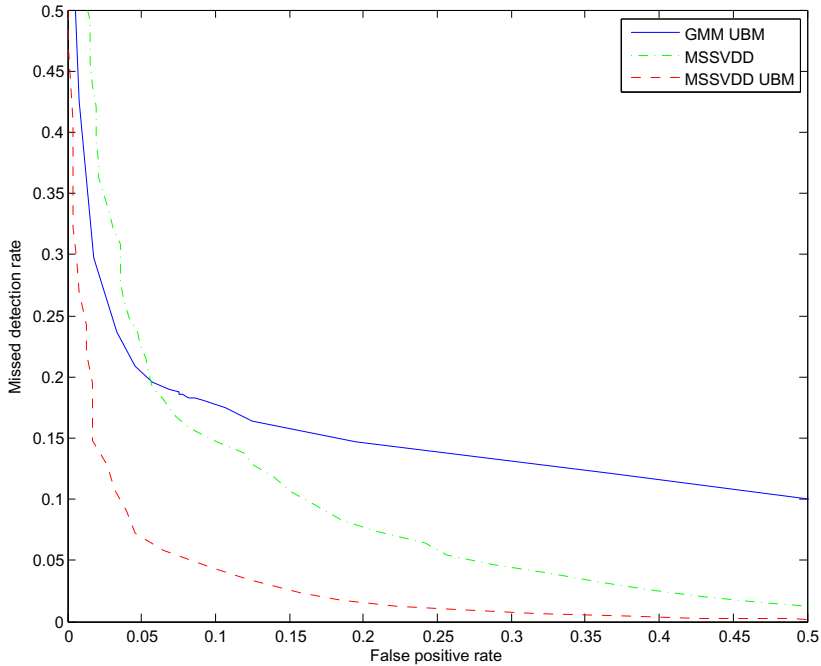


Fig. 3. DET curve of GMM UBM, 3-sphere SVDD, 3-sphere SVDD UBM on the Alcoholism EEG dataset

For MSSVDD training the RBF kernel function $K(x, x') = e^{-\gamma \|x - x'\|^2}$ was used. The parameter for MSSVDD training is γ and ν . The parameter γ was searched in $\{2^k : k = 2l + 1, l = -8, -7, \dots, 2\}$. The parameter ν is considered proportional to the rejection percentage of the positive data when training the smallest hypersphere enclosing the positive data. The best parameters found for the first two datasets are $\gamma = 0.031$ and $\nu = 0.031$ with EER=0.221 and the latter two datasets are $\gamma = 0.125$, $\nu = 0.125$ with EER=0.25 and 0.35 respectively.

Figures 1 and 2 show DET curves of GMM-UBM, MSSVDD, MSSVDD-UBM on Australian EEG dataset and EEGMMIDB datasets, respectively. Over all, the MSSVDD outperforms SVDD, and the MSSVDD-UBM outperforms MSSVDD and GMM-UBM.

8 Conclusion

We have presented a Multi-Sphere Support Vector Data Description approach to solving person verification problem. Experiments on the Australian EEG, EEG motor movement/imagery and Alcoholism datasets show a lower verification error rates comparing with SVDD and Gaussian mixture model-based universal background model (GMM-UBM).

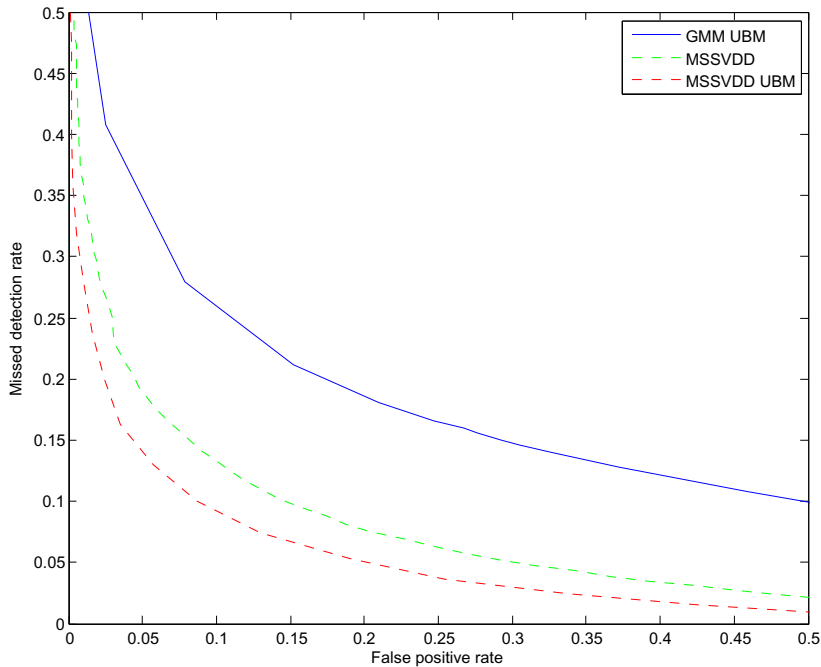


Fig. 4. DET curve of GMM UBM, 3-sphere SVDD, 3-sphere SVDD UBM on the full Alcoholism dataset

References

1. Marcel, S., Millán, J.R.: Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(4), 743–752 (2007)
2. Xu, Q., Zhou, H., Wang, Y., Huang, J.: Fuzzy support vector machine for classification of EEG signals using wavelet-based features. *Medical Engineering & Physics* 31(7), 858–865 (2009)
3. Riera, A., Soria-Frisch, A., Caparrini, M., Grau, C., Ruffini, G.: Unobtrusive biometric system based on electroencephalogram analysis. *EURASIP Journal on Advances in Signal Processing* 2008, 18 (2008)
4. Poulos, M., Rangoussi, M., Alexandris, N., Evangelou, A.: Person identification from the EEG using nonlinear signal classification. *Methods of Information in Medicine* 41(1), 64–75 (2002)
5. Palaniappan, R.: Two-stage biometric authentication method using thought activity brain waves. *International Journal of Neural Systems* 18(1), 59 (2008)
6. He, C., Wang, J.: An independent component analysis (ICA) based approach for EEG person authentication. In: *3rd International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2009*, pp. 1–4 (2009)
7. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain computer interfaces. *Journal of Neural Engineering* 4, R1 (2007)

8. Zúquete, A., Quintela, B., Cunha, J.P.S.: Biometric authentication using brain responses to visual stimuli. In: *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pp. 103–112 (2010)
9. Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-García, J., Petrovska-Delacrétaz, D., Reynolds, D.A.: A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing* 2004, 430–451 (2004)
10. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker verification using adapted gaussian mixture models. *Digital Signal Processing* 10(1), 19–41 (2000)
11. Campbell, W.M., Sturim, D.E., Reynolds, D.A.: Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Processing Letters* 13(5), 308–311 (2006)
12. Brew, A., Grimaldi, M., Cunningham, P.: An evaluation of one-class classification techniques for speaker verification. *Artificial Intelligence Review* 27(4), 295–307 (2007)
13. Tran, D.T.: *Fuzzy Approaches to Speech and Speaker Recognition*. PhD thesis, University of Canberra (2000)
14. Hunter, M., Smith, R.L., Hyslop, W., Rosso, O.A., Gerlach, R., Rostas, J.A.P., Williams, D.B., Henskens, F.: The Australian EEG database. *Clin. EEG Neurosci.* 36(2), 76–81 (2005)
15. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering* 51(6), 1034–1043 (2004)
16. Poulos, M., Rangoussi, M., Chrissikopoulos, V., Evangelou, A.: Person identification based on parametric processing of the EEG. In: *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems, ICECS 1999*, vol. 1, pp. 283–286 (1999)
17. Paranjape, R.B., Mahovsky, J., Benedicenti, L., Koles, Z.: The electroencephalogram as a biometric. In: *Canadian Conference on Electrical and Computer Engineering*, vol. 2, pp. 1363–1366 (2001)