**A novel group learning of high dimensional data for rare event detection**

**Proposed method (**Estimate the threshold by observing the statistics of training data)

* Positive classification: If the mean value of the classifier outputs of a test sample is *larger* than the threshold.
* Negative classification: If the mean value of the classifier outputs of a test sample is *smaller* than the threshold.
* Threshold estimation: Grid search the optimal quantile [1, 0.95, 0.90, …, 0.10, 0.05, 0] of the mean values of the classifier outputs of all training negative samples (sample: a 800-digit matrix in section 1 and 2, a 4-hr iEEG segment in section 3) which can achieve the highest sensitivity of training data. If multiple quantiles can achieve the same highest sensitivity, the largest one would be selected as the threshold.

**Compared methods**

1. **Majority Voting**

* Positive classification: if the number of positive outputs is greater than the number of negative outputs (considering a test sample).
* Negative classification: if the number of negative outputs is greater than the number of positive outputs (considering a test sample).

1. **Rule-based method**

* Positive classification: if there is at least a positive output (considering a test sample).
* Negative classification: if there is no positive output (considering a test sample).

1. **Non-group-learning (classify the sparse data directly using linear SVM classifier)**

- Positive classification: if the classifier gives a positive output.

- Negative classification: if the classifier gives a negative output.

**Empirical performance results**

* 1. **Classification of digit matrix (balanced training data)**
* Positive class: the artificial digit matrix constituted by 720 even digits, including digits ‘0, 2, 4, 6, 8’, and 80 odd digits, including only digit ‘1’ (as shown in Fig. 1).
* Negative class: the artificial digit matrix constituted by 800 even digits, including digits ‘0, 2, 4, 6, 8’ (as shown in Fig. 2).
* Problem formalization: binary classification of artificial digit matrix (positive class vs. negative class)

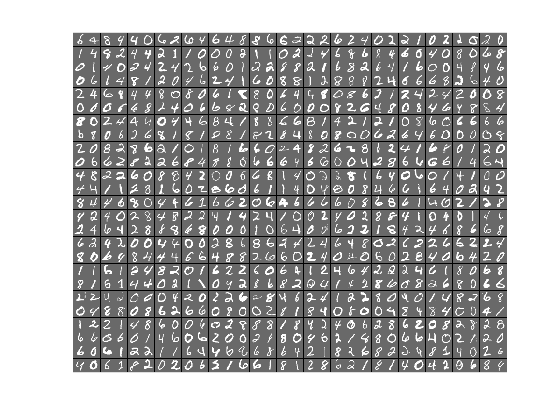


Fig 1. Example of the digit matrix in the positive class

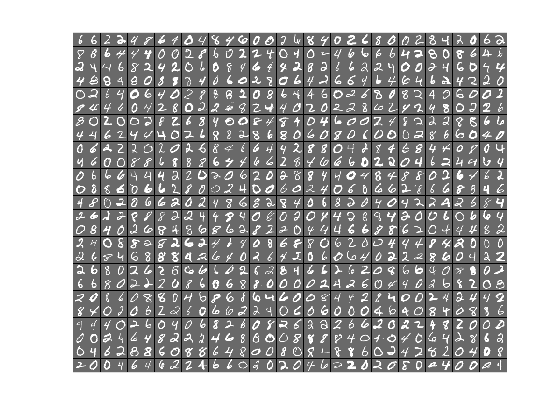


Fig 2. Example of the digit matrix in the positive class

**1.2 Data**

* Training data: 40 positive and 40 negative digit matrices
* Validation data: 40 positive and 40 negative digit matrices
* Test data: 500 positive and 500 negative digit matrices

Note that, training, validation, and test data mutually disjoint. Validation data is used to select the optimal *C* = [2-4, 2-3, 2-2, 2-1, 20, 21, 22, 23, 24,] of linear SVM.

* Downsampled digit data: the handwritten digits in the digit matrix are downsampled from 28×28 pixels to 14×14 pixels in order to have the feature dimension closer to iEEG data (feature dimension of 20s iEEG window = 96).
* Feature dimension of a digit matrix: 156800 (800 digits \* 14×14 pixels).

**1.3 Classification results**

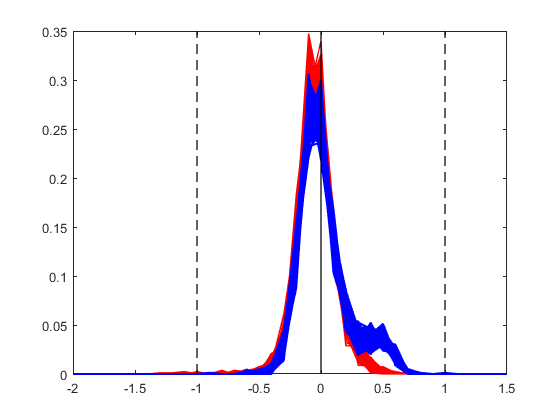
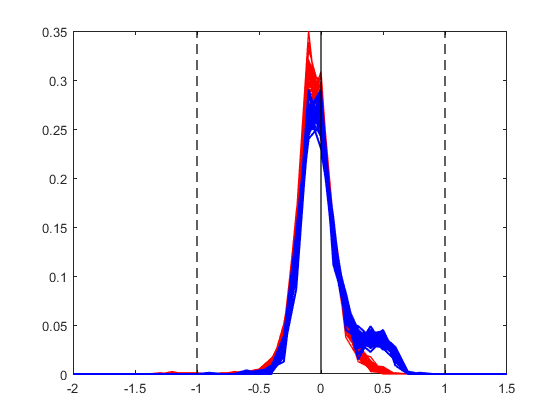


Fig 3. Histograms of training (left) and test (right) series, blue: positive class, red: negative class

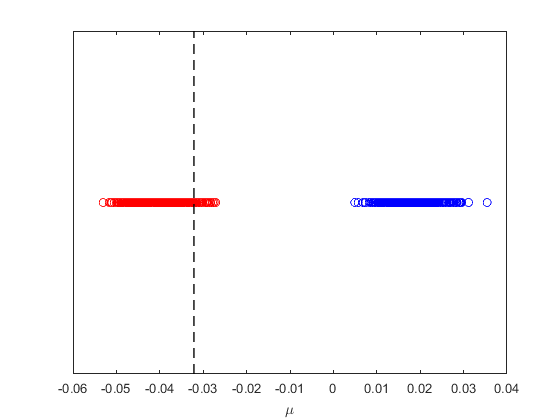
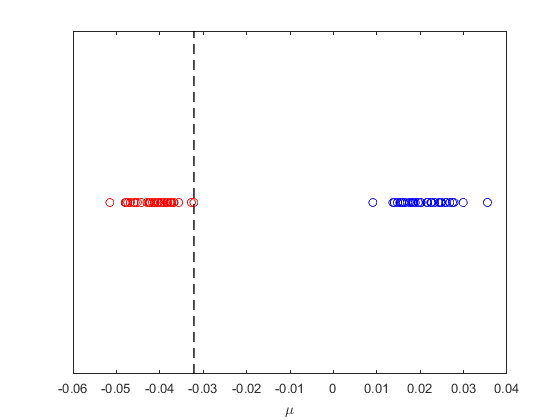


Fig 4. Mean values of training (left) and test (right) series, blue: positive class, red: negative class, both training and test series apply the same threshold (indicated as black dash line) estimated using training negative data.

Table 1. Performance indices (SS and SP) of proposed method and other methods

|  |  |  |
| --- | --- | --- |
| Method | SS | SP |
| Proposed method | 1.00 | 0.95 |
| Majority voting | 0 | 1.00 |
| Rule-based method | 1.00 | 0 |
| SVM | 0.80 | 1.00 |

* 1. **Classification of digit matrix (unbalanced training data)**
* The digit matrix of positive and negative classes and the problem formalization in this section are same as section 1.1.
  1. **Data**
* Training data: 8 positive and 40 negative digit matrices
* Validation data: 40 positive and 40 negative digit matrices
* Test data: 500 positive and 500 negative digit matrices

Note that, training, validation, and test data mutually disjoint. Validation data is used to select the optimal *C* = [2-4, 2-3, 2-2, 2-1, 20, 21, 22, 23, 24,] of linear SVM.

* 1. **Classification results**

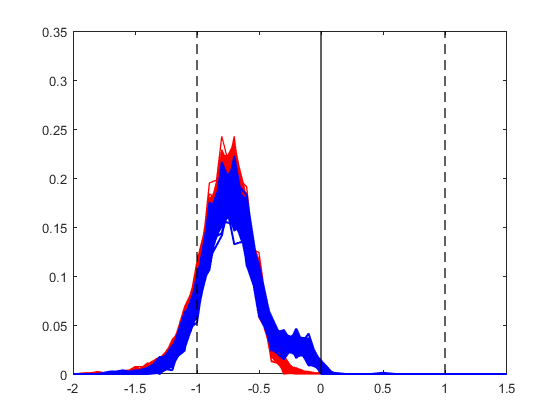
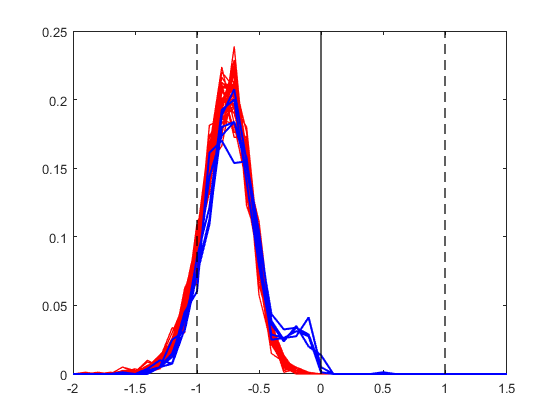


Fig 3. Histograms of training (left) and test (right) segments, blue: positive class, red: negative class

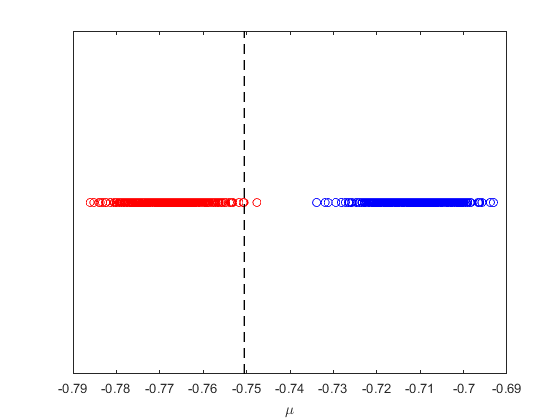
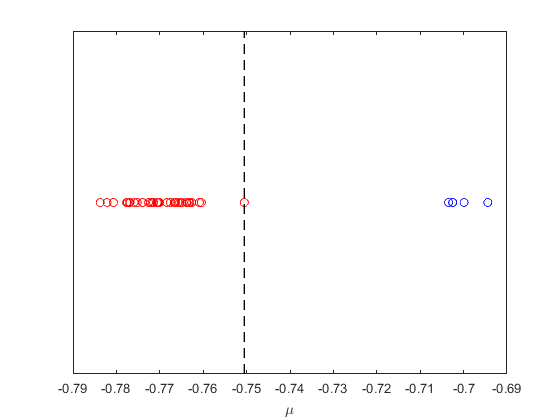


Fig 4. Mean values of training (left) and test (right) segments, blue: positive class, red: negative class, both training and test series apply the same threshold (indicated as black dash line) estimated using training negative data.

Table 2. Performance indices (SS and SP) of proposed method and other methods

|  |  |  |
| --- | --- | --- |
| Method | SS | SP |
| Proposed method | 1.00 | ~ 1.00 (0.998) |
| Majority voting | 0 | 1.00 |
| Rule-based method | 0.48 | 1.00 |
| SVM | 0 | 1.00 |

* 1. **Classification of 4-hr consecutive iEEG segments**
* Preictal segments: 4-hr consecutive iEEG segments from the period of 0.5 to 4.5 hr before lead seizures.
* Interictal segments: any other available 4-hr consecutive iEEG segments 4.5 hours away from a seizure.
* Problem formalization: binary classification of 4-hr consecutive iEEG segments (preictal vs. interictal)

Note: lead seizures must be preceded by at least 3-days seizure free period.

* 1. **Data**
* We include all available preictal segments (during ~ 1 year recording period) of these four canines in this experiment.
* The interictal segments are select from a random region of the recording period. We keep the imbalanced ratio 8:1 (interictal vs. preictal) used in [Shiao et al., 2017].

Table 3. Available data of four canines

|  |  |  |
| --- | --- | --- |
| Dog | # 4-hr preictal segments | # 4-hr interictal segments |
| L2 | 6 | 48 |
| L7 | 7 | 56 |
| M3 | 18 | 144 |
| P2 | 5 | 40 |

* We use the same experiment design [Shiao et al., 2017], using Dog P2 as an example (see Table 4).
* The model selection of each experiment is perform using only the training set. n fold cross-validation (n equal to the number of preictal in the training set) is applied in each experiment to select the optimal model of linear SVM. In each fold, one different pair of preictal and interictal segments are used as the validation data (balanced validation set) and the rest segments are used to train the linear SVM model.

Table 4. Experimental design for dog L2 using the unbalanced setting (the decimal label encode 4 hrs segment)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experiment | Training set | | Test set | |
| Preictal | Interictal | Preictal | Interictal |
| 1 | 2-5 | 2-40 | 1 | 1 |
| 2 | 1, 3-5 | 1, 3-40 | 2 | 2 |
| 3 | 1, 2, 4, 5 | 1, 2, 4-40 | 3 | 3 |
| 4 | 1-3, 5 | 1-3, 5-40 | 4 | 4 |
| 5 | 1-4 | 1-4, 6-40 | 5 | 5 |

**3.3 Classification results**

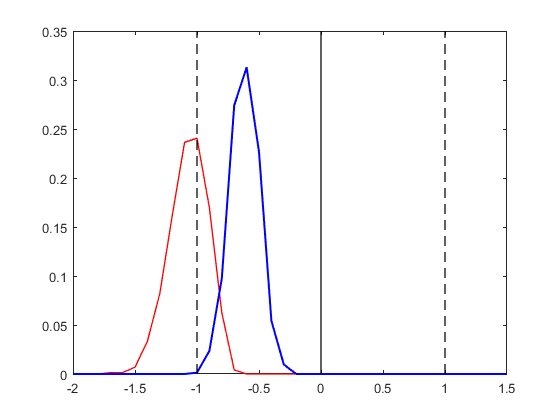
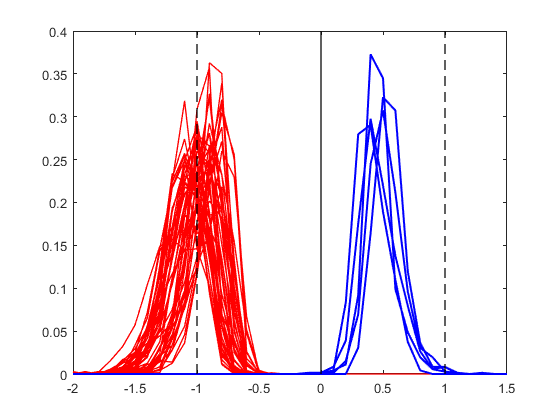


Fig 7. Histograms of training (left) and test (right) segments, blue: positive class, red: negative class

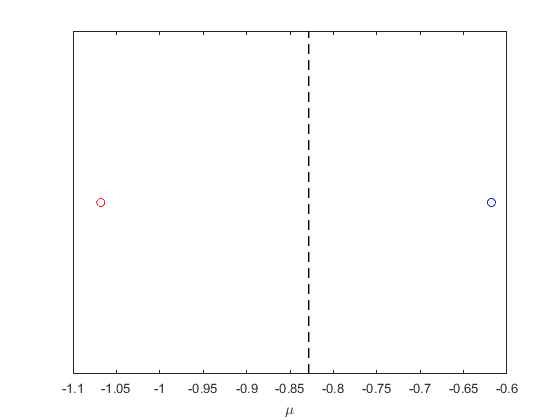
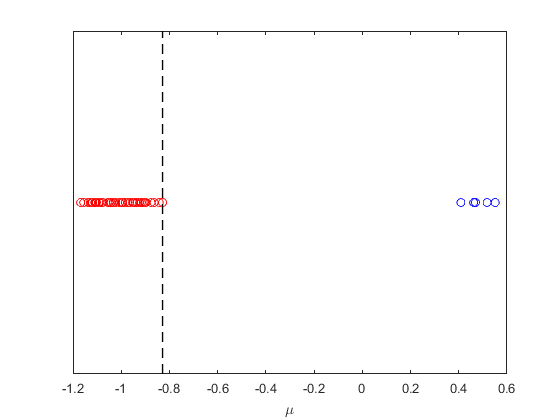


Fig 8. Mean values of training (left) and test (right) segments, blue: positive class, red: negative class

Table 5. Performance indices (SS and SP) of proposed method and other methods

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dog | Proposed method | | Majority Voting | | Rule-based method | |
| SS | SP | SS | SP | SS | SP |
| L2 | 1.00 | 1.00 | 0.83 | 1.00 | 0.83 | 1.00 |
| L7 | 0.86 | 1.00 | 0.71 | 1.00 | 0.86 | 0.29 |
| M3 | 0.89 | 0.89 | 0.67 | 1.00 | 0.89 | 0.56 |
| P2 | 0.80 | 1.00 | 0.40 | 1.00 | 1.00 | 0.40 |