IEICE TRANS. INF. & SYST., VOL.E102–D, NO.4 APRIL 2019

PAPER

Feature Selection of Deep Learning Models for EEG-Based RSVP Target Detection

Jingxia CHEN[†a)], **Zijing MAO**[††b)], **Ru ZHENG**[†c)], **Yufei HUANG**[††d)], ***Nonmembers***, ***and*** **Lifeng HE**[†,†††e)], ***Member***

SUMMARY Most recent work used raw electroencephalograph (EEG) data to train deep learning (DL) models, with the assumption that DL models can learn discriminative features by itself. It is not yet clear what kind of RSVP specific features can be selected and combined with EEG raw data to improve the RSVP classification performance of DL models. In this paper, we tried to extract RSVP specific features and combined them with EEG raw data to capture more spatial and temporal correlations of target or non-target event and improve the EEG-based RSVP target detection performance. We tested on X2 Expertise RSVP dataset to show the experiment results. We conducted detailed performance evaluations among different features and feature combinations with traditional classification models and different CNN models for within-subject and cross-subject test. Compared with state-of-the-art traditional Bagging Tree (BT) and Bayesian Linear Discriminant Analysis (BLDA) classifiers, our proposed combined features with CNN models achieved 1.1% better performance in within-subject test and 2% better performance in cross-subject test. This shed light on the ability for the combined features to be an efficient tool in RSVP target detection with deep learning models and thus improved the performance of RSVP target detection.

key words: *RSVP, EEG, feature selection, deep learning, CNN*

1. Introduction

Brain computer interfaces (BCI) rely on machine learning (ML) algorithms to decode the brain's electrical activity into decisions. One application of BCI is to use electroencephalograph (EEG) signals to detect rare target images within a large collection of non-target images using a rapid serial visual presentation (RSVP) paradigm [1]. There are multiple detection and classification approaches that have been applied to the RSVP tasks, including bagging tree (BT) [2], linear discriminant analysis (LDA) [3], Bayesian Linear Discriminant Analysis (BLDA) [4], support vector machines (SVM) [5], hierarchical discriminant component analysis (HDCA) [6], [7], xDAWN [8], and deep learning (DL) methods [9]. Deep learning is a class of machine learning algorithms with a multi-layered architecture which has recently achieved outstanding performance in a variety of applications including images, videos, speech, and text recognition [10]–[15]. The key to DL's success is its ability to automatically extract discriminant feature representations directly from big raw data. Although DL application in EEG data classification is still in its infancy, as the dataset size increases, the advantage of DL over traditional machine-learning techniques is becoming more apparent.

A few existing works has demonstrated the power of DL in EEG classification, especially convolutional neural network (CNN). In [16], a CNN was applied for predicting epileptic seizure based on intracranial EEG raw recording. [17] transformed EEG activities into a sequence of topology-preserving multi-spectral images and trained a deep recurrent convolutional network for learning feature representations that are invariant to inter and intra subject differences. [18] explored a deep CNN architecture for generalized single-trial EEG raw data classification across subjects. Another CNN proposed in [19], [20] applied XDAWN [20] to each of the time samples across all the EEG channels in its convolution layer. The CNN model described in [21], [22] was built for a single channel EEG recording and the inputs to CNN were time-frequency spectrum power values of the channel. Our previous work [23], [24] investigated the impact of different convolutional filters in DL and showed different techniques of training DL models to improve the EEG classification performance in RSVP target detection tasks.

Above recent work all used raw EEG data to train DL models, with the assumption that DL models could learn discriminative feature representations by itself [25]. However, this is based on enough training samples are available. It is not yet clear what kind of specific features can be selected and combined with EEG raw data working better in capturing spatial and temporal correlations in EEG stimulated by RSVP task and improve the performance of DL models. From another aspects, feature selection is non-trivial because useful features for classification tasks will definitely benefit the performance greatly [26], [27], specifically for RSVP task which normally has visual evoked potential at P300, and also the visual evoked potential at the image showing frequency rate, such as the 5.5 Hz pattern showed in X2 Expertise RSVP dataset from BCIT [28]. In this paper, we tried to extract RSVP specific features and combined them with EEG raw data to improve the RSVP

Manuscript received March 15, 2018.
Manuscript revised August 2, 2018.
Manuscript publicized January 22, 2019.

[†]The authors are with the Department of Electrical and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China.
[††]The authors are with the Department of Electrical and Computer Engineering, University of Texas at San Antonio, One UTSA Circle San Antonio, TX 78249–0669, USA.
[†††]The author is with the Faculty of Information Science and Technology, Aichi Prefectural University, Aichi, 480–1198 Japan.
a) E-mail: chenjx_sust@foxmail.com
b) E-mail: Zijing.Mao@utsa.edu
c) E-mail: zr_ellen@163.com
d) E-mail: yufei.huang@utsa.edu
e) E-mail: helifeng@ist.aichi-pu.ac.jp
DOI: 10.1587/transinf.2018EDP7095

target classification performance. We tested on X2 Expertise RSVP dataset to show the experiment results. We firstly proposed to find the feature combinations that could achieve the best performance for within-subject RSVP tasks. Then we applied the combined features on different machine learning and deep learning models to show the robustness of such feature combinations on cross-subject RSVP tasks. Compared with state-of-the-art traditional Bagging Tree (BT) and Bayesian Linear Discriminant Analysis (BLDA) classifiers, our proposed combined features with CNN models achieved 1.1% better performance in within-subject test and 2% better performance in cross-subject test.

The remainder of the paper was organized as follows. Section 2 described material and methods including the RSVP dataset, the acquisition of EEG data, data preprocessing, feature extraction and selection and the proposed DL models. Section 3 reported the testing experiments and the results. Section 4 discussed the results. Section 5 offered some concluding observations and suggestions for future work.

2. Material and Methods

2.1 Dataset Introduction

In this paper, we used BCIT X2 Expertise RSVP EEG dataset [28], [29] to carry out our experiments. This dataset consisted of a rapid presentation of color photographs (512× 662 pixels) of indoor and outdoor scenes. The images were presented at the frequency of 5 Hz (~200 ms per image) and subtended a visual angle of approximately 9°. Every 10s a blank screen with the word "blink" was presented to give participants a chance to blink without missing stimuli. There were 10 subjects and each subject participated 5-session experiments. Each session of RSVP task consisted of 6 blocks of 10 min each. All scenes contained only inanimate objects and were manually scaled and cropped. Some scenes contained target objects and others did not. Before each block participants were instructed as to the class of target objects for that block. There were 5 classes of target objects: containers, chairs, doors, stairs, and posters. Before the beginning of each block, the "ready screen" would indicate the target class for that block. During the RSVP, participants were instructed to press a button only when they saw an object from the current target class. The order of the target classes was randomly chosen for each participant (blocks 1–5); however, the last block (block 6) always had the same target class as the first block. In addition to target class, target probability varied across each block. Six target probability values (0.01, 0.03, 0.05, 0.07, 0.09, and 0.11), one for each block, were randomly assigned at the beginning of the task [28], [29]. EEG recordings were digitally sampled at 512 Hz from 256 scalp electrodes over the entire cortex of each subject using a BioSemi Active Two system (Amsterdam, Netherlands).

2.2 Data Preprocessing

X2 Expertise RSVP raw dataset was preprocessed with the PREP pipeline [30] which performed automated noise removal, bad channel detection, and referencing in a way that allowed users to specialize the data to particular applications without having to work with the raw data. With PREP pipeline, the raw EEG data was firstly band-pass filtered with a bandwidth of 0.1-32 Hz to remove its line noise. Then the true signal mean was estimated and the signal referenced by this mean was used to find the bad channels. A robust z-score algorithm was used to detect noisy channels that contain any NaN (not-a-number) data or that have significant periods with constant values or very small values. The spherical option of EEGLAB *eeg_interp* function was used for channel interpolation. The original channel size of this dataset was 256, but we selected a subset of 64 channels based on the 10-20 system 64-channel locations. Its original sampling rate was 512 Hz and then down sampled to 64 Hz for the convenience of calculation.

Following the procedures described in [31], epochs were extracted across time and all contained channels. In [32], one second after an image onset event was often used to capture the dynamics that user see a target at about 300ms and analysis it at about 600ms. Others in [33] also recommend to extract epoch between 200ms before and 800ms after image onset. In this paper, for each of 10 subjects, we extracted one-second epochs of his initial EEG data time-locked to each target/non-target image onset as target/non-target epochs, resulting in the size of each epoch is 64 (channels) by 64 (time points). To balance the training data, we randomly selected the same number of non-target epochs as target epochs, resulting in a total of 107,592 epochs for ten subjects and ~10,000 epochs including ~5,000 target epochs and ~5,000 non-target epochs for per subject. We then organized the raw EEG feature for each subject by randomly selecting 10% target and non-target epochs as testing samples, the left 90% as training samples including random 10% as validating samples.

2.3 Feature Extraction and Selection

We extracted EEG features from time domain and frequency domain separately. In time domain, we took the amplitudes of the extracted 1-s EEG epochs as raw EEG feature (Raw) which can directly reflect the statistical characteristics of EEG signals in a period of time. Due to great difference of raw feature magnitude over different epochs will leads to instable numerical calculation and poor testing performance, we also extracted the normalized feature (Norm) of each epoch of raw EEG feature by firstly normalizing the training and validating data across epochs by z-score normalization [24] and the normalized parameters Mu and sigma were then used to normalize the testing data across epochs. Event related potentials (ERP) such as P300 [30] are also widely used time-domain features for RSVP tasks. So, we extracted

the ERP which allocated at 200-400ms of each 1-s epoch as P300 feature.

ERP can also be achieved in the frequency domain. In sensorimotor control, the amplitudes of mu (8-12Hz) and beta (18-25Hz) rhythm were used to map thoughts to one or two-dimensional movement [34]. We extracted frequency ERP features by Auto Correlation and Fast Fourier Transform to convert the raw EEG feature into power spectral data at specified frequency range of 5-6Hz, 1-10Hz and the whole frequency range of 0.1-32Hz which were called Freq5-6Hz, Freq1-10Hz and Freq feature separately. We also extracted the power spectral density (PSD) of the whole frequency range of the raw EEG feature to intuitively observe the distribution and variation of EEG rhythm. We applied 0.25-second Hanning window without overlapping on each channel of 1-s epochs to extract 16 PSD features per channel. As a result, we obtained 7 different types of single EEG features including Raw, Norm, P300, Freq5-6Hz, Freq1-10Hz, Freq and PSD for each subject.

Feature selection was often applied after feature extraction to eliminate the unrelated and redundant features, so as to learn the discriminative feature subset, reduce the number of features and improve the accuracy of the model. The feature selection algorithms include filter and wrapper approaches, where filter approaches select subsets by their information content without directly evaluating their classification performance and wrapper methods use a classifier to search subsets by their predictive accuracy on test data. Filter approaches are much faster, while the wrapper methods can produce a set that achieves better classification performance. In this paper, we did not apply any filter and wrapper method to make feature selection. As combination features took into account both local features and global features of EEG data, we just selected some of the above single features and make them combined. We investigated RSVP target classification performance of shallow classifiers and deep learning models on these seven different simple features and their selected combinations respectively to find out the best one exhibiting the highest performance and make sure which kind of features are most efficient for deep learning models.

2.4 Deep Learning Models

In this paper, we explored the deep learning models especially CNN models for target prediction in RSVP tasks. The CNN architecture contained multiple convolutional layers followed by fully connected layers (Fig. 1 (b)). In each convolutional layer, multiple filters or kernels were convolved with the input data (vectorized EEG epochs). These filters were designed to capture different local spatial features. The output of a convolution layer from one kernel was called a feature map (FM). The fully connected layers then combined all the feature maps at the output of the last convolution layer. Here we used the CNN architecture described in [7], [35]. Let W_c^k represent the c^{th} weight of the k^{th} kernel for $c = 1, \ldots, 64$ channels and $k = 1, \ldots K$, where K was

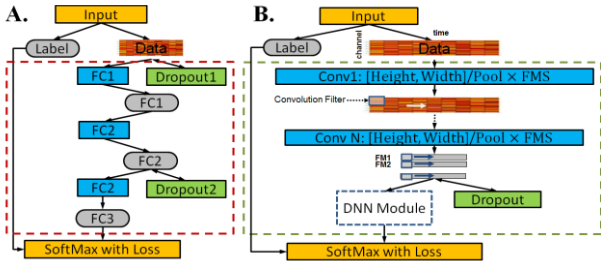


Fig. 1 (a) The designed DNN architecture with fully connected modules and hidden units expressed in gray ovals; (b) the designed CNN architecture. There were N convolution layers and blue boxes represented convolution operations, where the texts inside represented [kernel shape]/MP width × Feature Map Size. "FM" denoted feature map.

a hyper-parameter to be learned. Let $\boldsymbol{v} \in \mathcal{R}^{M\times 1}$ denote an input vector with $M = 64\times 64$ in this case. Then, the k^{th} FM at the output of the convolutional layer was:

$$convolution(\boldsymbol{v})_{kt} = ReLU\left(\sum_{c=1}^{64} W_c^k \boldsymbol{v}_{ct}\right)$$
$$for\ t = 1, \ldots, 64 \tag{1}$$

where v_{ct} was the input element corresponding to the EEG measurement from channel c at time t, ReLU [4] represented the rectified linear function f(x) = max(0, x). We saw from the Eq. (1) that the kernel filters for all channels at time t formed a spatial filter. After the convolutional layer, a multilayer perceptron (MLP) was applied to combine all FMs for prediction of target/non-target events. The number of FMs and the number of hidden units were CNN hyper-parameters [9] to be tuned during training.

For convolution filters, we considered 8 combinations of local (L)/global (G) with spatial (S)/temporal (T) filters. A spatial (S) (or temporal (T)) filter referred to the one that focused on EEG channels (or time samples), because EEG data had both spatial and temporal correlations. All together, these 8 combinations of local (L)/global (G) vs spatial (S)/temporal (T) filters were: LS, LT, GS, GT, LSLT, LSGT, GSLT, GSGT. Our previous experiment results [23], [24] had shown that GT and GSLT convolution filters had better performance than others. For testing simplicity, we applied GTCNN and GSLTCNN models on our selected features to test the RSVP target classification performance and considered a local filter with kernel size of 5 across time samples and a global filter with kernel size across all channels which number were 64 in our dataset.

Model selection was an important step in CNN training, which determined the model hyper-parameters including the number of hidden layers, the hidden unit size, the position of dropout modules, the max-pooling width, and the number of feature maps. Specific model parameters needed to be trained and their search space for each of the 8 designed CNNs were convolution layer size (1 to 3 layers) and feature map size (8 to 128 feature maps in log2 scale) for each convolution layer. The search space was defined to balance the trade-off between a deeper architecture and limited training samples. For simplicity, we fixed the local filter

kernel size as 5×5. If pooling was applied, the pooling kernel size would be 2×2. The top fully-connected DNN modules (Fig. 1 (a)) had 2 hidden layers with 128 hidden units in each layer. Dropout layer was after the first fully-connected layer and all layers used the ReLU activation function for faster approximation. All models were trained using stochastic gradient descent (SGD) on a mini-batch size of 32 epochs with an exponential decay for the learning rate and momentum. The strategy of early stopping [36] was applied to determine the training iterations, where the maximum training iteration was set to be 10,000. The initial learning rate was 0.01 with a decay rate about 1.01, the momentum was 0.9, and decay weight was 0.0005. Here, we applied random search [37] to find the best model combination, where we tested 32 randomly picked models and the best model was selected as the one that produced the largest Area Under the Curve (AUC) on the validation dataset.

3. Experiments and Results

3.1 ERP of RSVP Experiment

For X2 Expertise RSVP dataset, we performed normalization before we put target and non-target epochs into training. This process tried to eliminate the strong image showing pattern which is about 5 Hz of each evoked potential region corresponding with the image showing frequency in the experiment. The top of Fig. 2 indicates the non-target and target epoch ERPs of raw data in X2 Expertise Experiment, and there is no obvious difference can be observed between ~200ms and ~300ms with the image showing pattern smearing on target ERP and there is slightly more obvious pattern smearing on target ERP at ~400ms. These two patterns correspond to the visual (~300ms) and analysis (~400ms) events. While after normalization, patterns at these two locations have been emphasized and showed in the bottom of Fig. 2. The transition of activation in target ERP is from occipital region to parietal and frontal lobe also indicates the visual process.

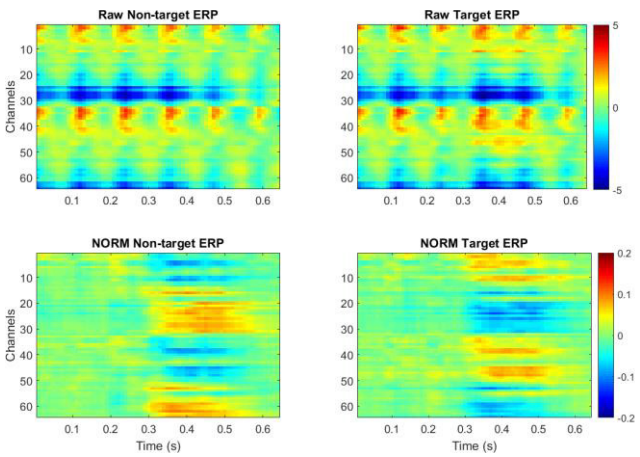


Fig. 2 X2 Expertise ERP of raw data (top) and normalized data (bottom) of target and non-target epochs

3.2 Baseline Results on Within-Subject Experiment with Selected Features

The motivation of this preliminary test was to search the stable feature combinations that provided the best performance based on two state-of-the-art RSVP algorithms: BT and BLDA [22]. Bagging tree is one of efficient machine learning algorithms based on bootstrap aggregating method, which can ensemble meta-algorithm designed to improve the stability and accuracy of statistical classification and regression. It also reduces variance and helps to avoid overfitting [2]. BLDA is a three-level hierarchical Bayesian model which yields a fast algorithm resulting in reasonable comparative performance in terms of test set likelihood [4]. In our previous studies [22], these two algorithms had been proved to perform better than the other traditional ones on EEG classification. Therefore, we chose them as baseline algorithms comparing with DL models. Here, we used X2 Expertise RSVP dataset to test the within-subject prediction performance on selected features. We preprocessed the initial EEG data and made feature extraction and selection as the methods described in Sects. 2.2 and 2.3. As a result, we got 7 types of single features including Raw, Norm, Freq, Freq5-6Hz, Freq1-10Hz, P300, PSD for each subject. For each type of these features, we carried out 10-fold (from S01 to S10) cross validation experiments. For each fold, 10% target and non-target epochs of its subject was randomly selected as testing samples, the left 90% as training samples including random 10% as validating samples. We made the average performance of these 10 folds as the final result of each type of feature. Figure 3 showed our provided baseline results for these 7 types of single features. From Fig. 3, we observed that in frequency domain both BT and BLDA models had its best performance (0.66 and 0.67) on Freq features which covered the whole frequency range and in time domain BLDA had the best performance of 0.7 on Raw and Norm feature as well as the BT model. The next step was to find out feature combinations with better performance.

We chose to use either Raw or Norm feature combined with other 5 types of features because Raw and Norm fea-

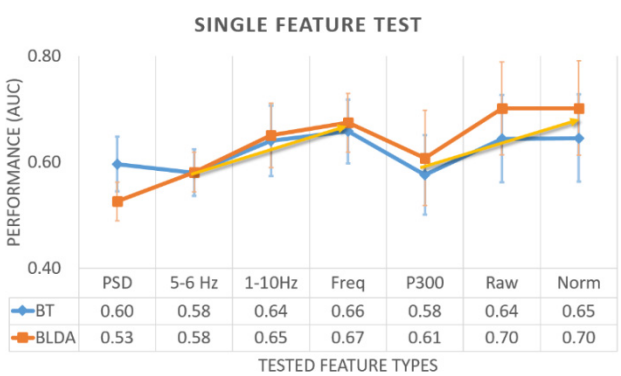


Fig. 3 Tested baseline results for 7 types of single features

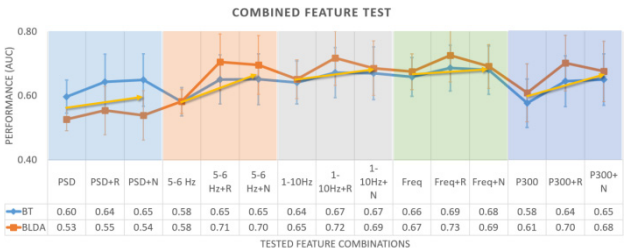


Fig. 4 Tested baseline results for combined features

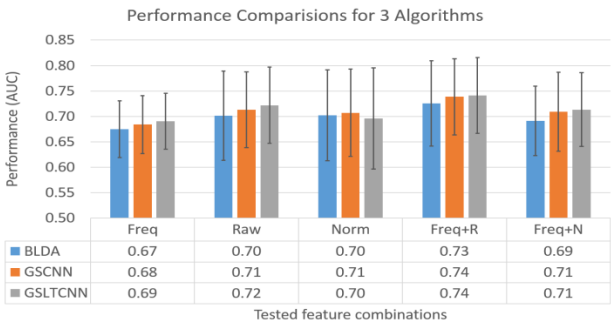


Fig. 5 The performance of three models for within-subject X2 Expertise RSVP dataset

tures achieved the best performance according to the above results. Therefore, we had 10 different combinations. Figure 4 showed the testing results on these combined features in five blocks, among which each block included three results from the original feature, the combined features with raw and normalized feature. We observed both BLDA and BT models have the best performance of 0.73 and 0.69 on Freq feature combined with the Raw feature (Freq+R) and have the suboptimal performance of 0.69 and 0.68 on Freq feature combined with the Norm feature (Freq+N). So, we selected to use the three best single features including Freq, Raw and Norm and their two types of combined features including Freq+N and Freq+R to carry out the within-subject and cross-subject experiments on DL models.

3.3 Within-Subject Performance on DL models

We tested the above five selected features on our designed CNN model with global spatial filters (GSCNN) and CNN model with global spatial and local temporal filters (GSLTCNN). Figure 5 showed the within-subject performance of three models on 10-fold X2 Expertise RSVP dataset, which indicated the performance increased when combining Raw feature with Freq feature. GLSTCNN and GSCNN models both achieved the best performance on Freq+R feature. What's more, DL models all presented higher performance than BLDA model on all features except Norm feature. GSLTCNN also presented the same performance tendency as GSCNN. Figure 6 showed the training, validation and testing learning curve for GSLTCNN model. From this figure, we observed the training accuracy always increased with training iteration, while testing and

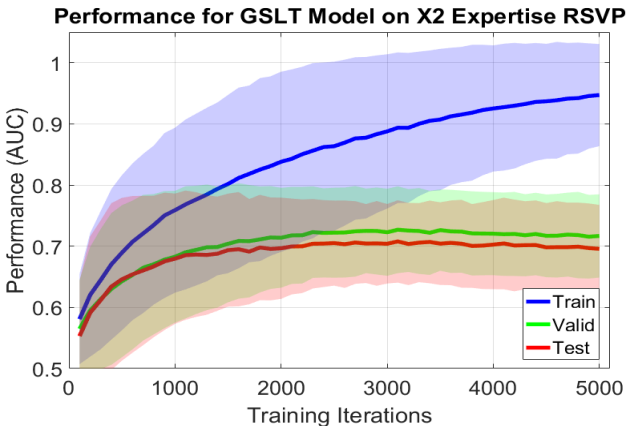


Fig. 6 Learning curve of GSLTCNN model trained on within-subject cross validation test for X2 Expertise RSVP experiment. Three lines represents the 10-fold mean learning curve, the three half transparent areas under the three average learning curves are the standard deviation for the 10-fold results.

Table 1 Structure of 5-fold cross-subject validation datasets for X2 Expertise RSVP task

Fold Name	Test Size	Train Size	Test data	Train data
RSVP_X2_F01	2180	9000	test data of S02	train data of S01
RSVP_X2_F02	1622	10000	test data of S04	train data of S03
RSVP_X2_F03	1832	10000	test data of S06	train data of S05
RSVP_X2_F04	1632	10000	test data of S08	train data of S07
RSVP_X2_F05	2788	8000	test data of S10	train data of S09

validation accuracy stopped increasing (reach the plateau) in the middle of training. This indicated the overfitting of training data after about 3K iterations.

3.4 Cross-Subject Performance on DL Models

We proposed to perform the cross-subject test on X2 Expertise RSVP dataset in the next step and aimed to show the Freq+R feature with DL models would achieve the best performance. The initial EEG data was preprocessed and epochs were extracted as the methods introduced in Sects. 2.2 and 2.3. In previous within-subject experiments, we had obtained five types of features including Freq, Raw, Norm, Freq+N and Freq+R for each subject. Here, for each type of these features, we built 5 folds of cross-subject validation datasets and each fold contained the training set structured by the training data of this type of feature of each odd-numbered subject (S01, S03, S05, S07 or S09) and the testing set structured by the testing data of this type of feature of each even-numbered subject (S02, S04, S06, S08 or S10). The detail structure of this 5-fold cross-subject validation dataset was shown as Table 1. We made the average performance of these 5 folds as the final result of each type of feature.

We firstly tested the baseline performance of BLDA and BT models on 5-fold cross-subject dataset of these five types of features, which result was shown as Fig. 7.

We then tested on our designed GSCNN and

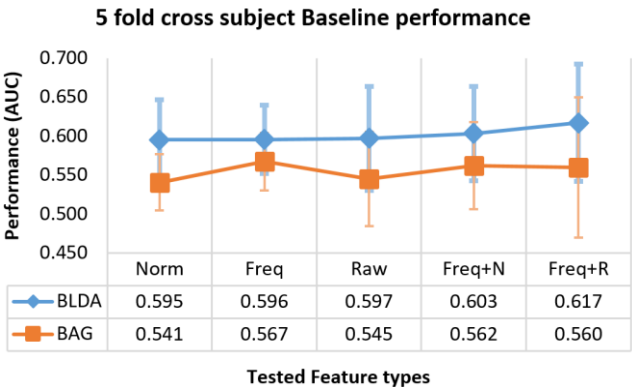


Fig. 7 Baseline performance for selected features on cross-subject X2 RSVP dataset

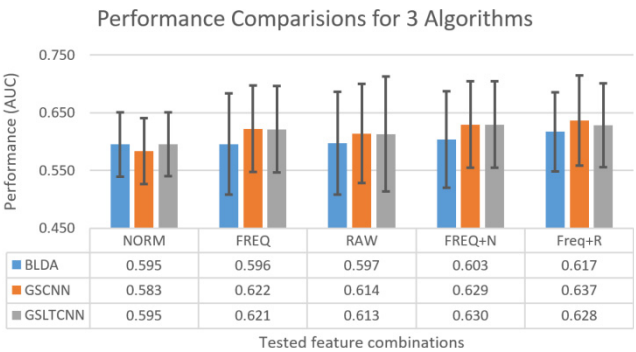


Fig. 8 The performance of cross-subject test for X2 Expertise RSVP dataset

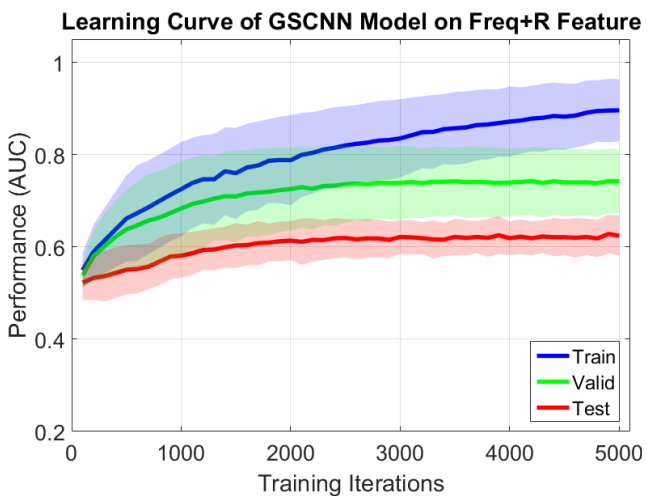


Fig. 9 Learning curve of GSCNN model trained on 5-folder cross-subject X2 Expertise RSVP experiment. Three lines represents the 5-fold mean learning curve, the three half transparent areas under the three average learning curves are the standard deviation for the 5-fold results.

GSLTCNN models and the results of 5-folder cross-subject test is shown as Fig. 8 which demonstrated the performance increased when combining Raw feature and Freq feature. Figure 9 showed the training, validation and testing learning curve for the best model GSCNN on the best feature. From this figure, we can see the training accuracy always increases with training iteration, while testing and validation accuracy stop increasing and reach the plateau in the middle of training. This indicates the overfitting of training data after about 3K iterations.

4. Discussion

Through experiments, we firstly analyzed the target and nontarget event related ERP of the raw EEG data and their normalized data collected in X2 Expertise RSVP task through experiment. The results (Fig. 2) showed that the EEG signals induced in this RSVP task had the time-locked property which meant when the time of the target/non-target image onset was known, the delay of the ERP stimulated by the target/non-target image was always invariant, usually 0.3-0.6s after the stimulus onset which corresponds to the visual (~300ms) and analysis (~400ms) events. This proved our 1-s epochs extracted from the initial EEG data time-locked to each target/non-target image onset was effective and discriminative for RSVP target event classification and the normalization helped to demonstrate the visual difference between two ERP patterns related to target and non-target images.

When testing the within-subject baseline results on 7 types of single features, we could see from Fig. 3 that the trend of performance for frequency feature as frequency range increased was ascending and achieved the highest performance considering the full range of frequency from 0.1 to 32 Hz (by BLDA, 0.675, ANOVA p = 0.0013), which indicated more complete range of frequency domain information benefit more on classification. However, the classification on PSD feature did not achieve a good performance (8% degradation comparing with Freq features, t-test p = 2e-5), even considering full range of frequency components. This might because the time samples which was 64 points was too small, and the power given both time and frequency might not be accurate. Similarly, from the temporal domain, we also considered P300 features which allocate at 200ms to 400ms. We observed better performance (by BLDA, 9% improvement with Raw feature, t-test p = 2e-5) when using all the temporal information comparing with P300 features. There was no significant difference observed when applying raw features and normalized features (t-test p = 0.62).

In order to find the feature combinations with best performance, we tested the baseline performance on different 10 combined features. From the results in Fig. 4, we observed the Freq feature covering the whole frequency range and combined with the Raw feature covering the whole-time range achieved the best performance of 0.73 by BLDA model, even 3% (t-test p = 9.5e-4) better than that of using single type of Raw feature. The tendency of performance as including more frequency information (from frequency range 5-6 Hz to 0.1-32 Hz) also increased. The combination of Raw/Norm feature with PSD or P300 even though improved the performance comparing to simply using PSD/P300 features, they did not show better

performance than using Raw or Norm features. In addition, considering the two baseline models, BLDA was observed gaining slightly better performance (1%, ANOVA p = 0.164) than BT. So, we selected to use the three best single features including Freq, Raw and Norm and their two types of combined features including Freq+Norm and Freq+Raw to carry out within-subject and cross-subject experiments on DL models.

As to within-subject test on DL models, from Fig. 5, we observed statistical results of Norm feature comparing to Raw feature was not significant (t-test, p = 0.72), while Raw feature combined with Freq feature showed significance improvement comparing with Norm feature (t-test, p = 0.0076), Raw feature (t-test, p = 0.0033) and Freq feature (t-test, p = 0.0017). Freq+R data with GSLTCNN model achieved the best performance of 74.1% which was 1.9% better comparing with the best performance of 72.2% using single Raw feature, and when comparing with Raw data versus Norm data, 2.6% (t-test p = 0.043) better performance was observed when using Raw data, while no significant difference was observed when comparing Freq+N data with Norm data results (p = 0.205). Additionally, when the results of two DL models were compared, only slightly improvement for GSLTCNN was observed (0.2%, ANOVA p = 0.885). While comparing with the best baseline algorithm BLDA, we observed 1.1% (t-test p = 0.023) better performance in GSLTCNN model. In sum, we observed significant improvement when using Freq+R feature comparing with other features and DL models had better performance than traditional best BLDA model in within-subject X2 Expertise RSVP test.

As to cross-subject test on BLDA and BT models, we observed from Fig. 7 that the Freq+R feature achieved the best performance of 0.617 by BLDA, even 2% (t-test, p = 0.0489) better than the best performance of using single type of Raw feature. In addition, considering the two baseline algorithms, BLDA had 5.7% better performance than BT (ANOVA p = 2.9e-05) on Freq+R feature.

As to cross-subject test on DL models, we observed from Fig. 8 that GSCNN achieved the best performance of 0.637 on Freq+R feature, which was 2% (ANOVA, p = 0.135) better than that of BLDA model on the same feature, 2.3% better (t-test p = 0.0338) than that of using Raw feature and 5.4% (t-test p = 0.0248) better than that of using Norm feature. GSCNN model also gained 3.1% (t-test p = 0.065) better performance when using Raw feature than using Norm feature, 4.6% (t-test p = 0.057) better performance when using Freq+N feature than using Norm feature, while there was no significant difference observed when comparing Raw feature with Freq feature results (0.8%, t-test p = 0.506). For GSLTCNN model, the results also showed the performance consistency applied with different features and its results on Freq feature comparing to Raw feature was also not significant (0.8%, t-test p = 0.405), while Freq+R feature showed significance improvement comparing with Norm features (3.3%, p = 0.0228) and Raw features (1.5%, p = 0.0853). In addition, when comparing with two DL models, improvement for GSCNN was observed (0.9%, ANOVA, p = 0.977). This analysis result indicated the performance of DL models on either Raw or Norm features in both cross-subject and within-subject tests was significantly improved after combining with Freq features and the Freq+R feature had the best performance comparing with other features. As well, DL models had better performance than traditional best BLDA model in both cross-subject and within-subject tests on X2 Expertise RSVP dataset.

Considering the universality of our proposed method, we had to admit that the present results were only for target event classification of EEG data elicited by RSVP paradigm with stimulus parameters as X2 Expertise RSVP data set. We tested the universality of our proposed method on traditional classifiers and DL models in within-subject and cross-subject RSVP experiments, but we were not sure whether our proposed method worked better on classification of EEG data elicited by other paradigms with different stimulus parameters. We will explore the universality of our proposed combined features with DL models on other EEG data sets in our future work, which is important when one discusses the feature extraction and selection method.

5. Conclusion

This study established the combined features for deep learning models on RSVP EEG data and shed light on the ability for combined Freq+R and Freq+N features to be an efficient tool in RSVP classification tasks with deep learning models, and thus improved the accuracy of the RSVP target classification. Apart from RSVP, or more generally, EEG feature is individual specific feature which is also of great interest to us for not only diminishing such feature when processing BCI classification tasks but utilizing it to improve the classification performance. The basic idea is to design algorithms that allow learning models to separate and refine the RSVP feature from individual specific feature. We have extracted epochs across all the BCIT data involving more than 200 subjects and designed deep learning models for training individual specific EEG features, which is also one of our work in the future.

Authors' Contributions

JC and RZ carried out the feature selection experiments for deep learning models to find out the feature combination that can achieve the best performance for within-subject and cross-subject RSVP classification tasks and drafted the manuscript. ZM designed deep learning models and participated in drafted the manuscript. YH and LH conceived of the study and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 61806118 and No. 61806144).

References

[1] N. Bigdely-Shamlo, A. Vankov, R.R. Ramirez, and S. Makeig, "Brain activity-based image classification from rapid serial visual presentation," IEEE Trans. Neural Syst. Rehabil. Eng., vol.16, no.5, pp.432–441, 2008.

[2] S.-W. Chuang, L.-W. Ko, Y.-P. Lin, R.-S. Huang, T.-P. Jung, and C.-T. Lin, "Co-modulatory spectral changes in independent brain processes are correlated with task performance," Neuroimage, vol.62, no.3, pp.1469–1477, 2012.

[3] B. Scholkopft and K.R. Mullert, "Fisher discriminant analysis with kernels," Neural Networks for Signal Processing IX 1, 1999.

[4] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient p300-based brain–computer interface for disabled subjects," Journal of Neuroscience methods, vol.167, no.1, pp.115–125, 2008.

[5] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol.20, no.3, pp.273–297, 1995.

[6] J. Meng, L.M. Meriño, K. Robbins, and Y. Huang, "Classification of imperfectly time-locked image rsvp events with eeg device," Neuroinformatics, vol.12, no.2, pp.261–275, 2014.

[7] J. Meng, L.M. Meriño, N.B. Shamlo, S. Makeig, K. Robbins, and Y. Huang, "Characterization and robust classification of eeg signal from image rsvp events with independent time-frequency features," PloS one, vol.7, e44464.2, 2012.

[8] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xdawn algorithm to enhance evoked potentials: application to brain–computer interface," IEEE Trans. Biomed. Eng., vol.56, no.8, pp.2035–2043, 2009.

[9] H. Cecotti and A. Gräser, "Neural network pruning for feature selection-application to a p300 brain-computer interface," ESANN, Citeseer, 2009.

[10] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, pp.1097–1105, 2012.

[11] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," 2013 IEEE international conference on acoustics, speech and signal processing, IEEE, pp.6645–6649, 2013.

[12] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," Proc. IEEE conference on Computer Vision and Pattern Recognition, pp.1725–1732, 2014.

[13] X. Zhang and Y. LeCun, "Text understanding from scratch," arXiv preprint arXiv:1502.01710, 2015.

[14] K.M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," Advances in Neural Information Processing Systems, pp.1693–1701, 2015.

[15] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.4694–4702, 2015.

[16] P.W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, "Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg," 2008 IEEE Workshop on Machine Learning for Signal Processing, IEEE, pp.244–249, 2008.

[17] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," arXiv preprint arXiv:1511.06448, 2015.

[18] J. Shamwell, H. Lee, H. Kwon, A.R. Marathe, V. Lawhern, and W. Noth Wang, "Single-trial EEG RSVP classification using convolutional neural networks," SPIE Defense+ Security, International Society for Optics and Photonics, pp.983622–983622.2, 2016.

[19] H. Cecotti and A. Graser, "Convolutional neural networks for p300 detection with application to brain-computer interfaces," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.3, pp.433–445, 2011.

[20] H. Cecotti, M.P. Eckstein, and B. Giesbrecht, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering," IEEE Trans. Neural Netw. Learn. Syst., vol.25, no.11, pp.2030–2042, 2014.

[21] S. Stober, D.J. Cameron, and J.A. Grahn, "Using convolutional neural networks to recognize rhythm? stimuli from electroencephalography recordings," Advances in Neural Information Processing Systems, pp.1449–1457, 2014.

[22] S. Stober, D.J. Cameron, and J.A. Grahn, "Classifying eeg recordings of rhythm perception," ISMIR, pp.649–654, 2014.

[23] S. Ahmed, L.M. Merino, Z. Mao, J. Meng, K. Robbins, and Y. Huang, "A deep learning method for classification of images rsvp events with eeg data," Global Conference on Signal and Information Processing (Global SIP), 2013 IEEE, IEEE, pp.33–36, 2013.

[24] M. Hajinoroozi, Z. Mao, and Y. Huang, "Deep transfer learning for cross-experiment prediction of rapid serial visual presentation events," 2013 IEEE Brain Computer Interface Meeting, pp.55–60, 2013.

[25] Y. Bengio, "Learning deep architectures for ai," Foundations and trends® in Machine Learning, vol.2, no.1, pp.1–127, 2009.

[26] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, How Transferable Are Features in Deep Neural Networks?, Advances in Neural Information Processing Systems, pp.3320–3328, 2014.

[27] I. Guyon and A. Elisseeff, "An introduction to feature extraction," Feature extraction, pp.1–25, Springer, 2006.

[28] J. Touryan, G. Apker, S. Kerick, B. Lance, A.J. Ries, and K. McDowell, "Translation of EEG-Based Performance Prediction Models to Rapid Serial Visual Presentation Tasks," Foundations of Augmented Cognition, pp.521–530, Springer Berlin Heidelberg, 2013.

[29] J. Touryan, G. Apker, B.J. Lance, S.E. Kerick, A.J. Ries, and K. McDowell, "Estimating endogenous changes in task performance from eeg," Using Neurophysiological Signals that Reflect Cognitive or Affective State, 268, 2015.

[30] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K.A. Robbins, "The prep pipeline: standardized preprocessing for large-scale eeg analysis," Frontiers in Neuroinformatics, vol.9, 2015.

[31] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," Journal of neuroscience methods, vol.134, no.1, pp.9–21, 2004.

[32] P. Sajda, E. Pohlmeyer, J. Wang, L.C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M.K. Singh, and S.-F. Chang, "In a blink of an eye and a switch of a transistor: cortically coupled computer vision," Proc. IEEE, vol.98, no.3, pp.462–478, 2010.

[33] S. Kunwar, "Subject independent p300 erp detection and classification," NER2015, 2015.

[34] T. Lan, "Feature extraction feature selection and dimensionality reduction techniques for brain computer interface," OHSU Digital Collections, https://doi.org/10.6083/M4RX9920, 2011.

[35] P. Sajda, A.D. Gerson, M.G. Philiastides, and L.C. Parra, "Single-trial analysis of eeg during rapid visual discrimination: Enabling cortically-coupled computer vision," Towards brain-computer interfacing, 423–444, 2007.

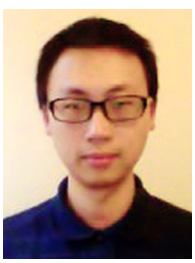
[36] L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," Neural Networks, vol.11, no.4, pp.761–767, 1998.

[37] I. Sutskever, J. Martens, G.E. Dahl, and G.E. Hinton, "On the importance of initialization and momentum in deep learning," ICML (3), vol.28, pp.1139–1147, 2013.

Appendix: Abbreviations

DL: Deep Learning; CNN: convolutional neural network; RSVP: rapid serial visual presentation; EEG: electroencephalograph; BT: Bagging tree; BLDA: Bayesian Linear Discriminant Analysis; BCI: Brain computer interfaces; ML: machine learning; BT: bagging tree; LDA: linear discriminant analysis; BLDA: Bayesian Linear Discriminant Analysis; SVM: support vector machines; HDCA: hierarchical discriminant component analysis; SNR: signal to noise ratio; PSD: power spectral density; MLP: multilayer perceptron; SGD: stochastic gradient descent; AUC: Area Under the Curve; GSCNN: an CNN with global spatial filters; GSLTCNN: an CNN model with global spatial and local temporal filters.

Jingxia Chen received the B.S. and M.S. degrees from Department of electrical and information engineering in Shaanxi University of Science and Technology in China in 2002 and 2005, respectively. During 2013-now, studied for Ph.D. degree in School of Computer Science and Engineering, Northwestern Polytechnical University in China. Now she also works as an associate professor in Department of Electrical and Information Engineering, Shaanxi University of Science and Technology in China. Her research interest is focus on Machine learning and pattern recognition, EEG signal processing and event detection, and deep learning.



Zijing Mao received the M.S. and Ph.D. degrees in Department of Electrical and Computer Engineering in University of Texas, San Antonio in 2013 and 2016, respectively. From 2017, he worked as general manager of Tianjin Yinianbo Technology Co. Ltd and he also simultaneously studied for post-doctoral degree at UTSA. His research interest is focus on Biocomputing, EEG signal processing and event detection, deep learning and brain-computer interface applications.

Ru Zheng received the B.S. and became a graduate student in Department of electrical and information engineering in Shaanxi University of Science and Technology in 2016. Her research interest is focus on EEG signal processing, classification and machine learning.

Yufei Huang received the M.S. and Ph.D. degrees from Department of Electrical Engineering in State University of New York at Stony Brook in U.S.A during 1996-2001. During 2002-now, he taught in Department of Electrical and Computer Engineering, University of Texas San Antonio in U.S.A., and here he was promoted to be a professor in 2012. He gained the Career Award of National Science Foundation in 2005, the Best Paper Award in Artificial Neural Networks in Engineering Conference in 2006, and the Best Paper Award in IEEE Signal Processing Magazine in 2007. His research Interest is focus on computational systems biology, brain-computer interface and deep learning.

Lifeng He received the M.S. and Ph.D. degrees in Nagoya University of industry, Japan in 1994 and 1997, respectively. During 1997-now, he worked in Faculty of Information Science and Technology, Aichi Prefectural University of Japan, and in 2012 he was promoted to be a professor. He was also employed as the Dean of the Academy in Department of Electrical and Information Engineering, Shaanxi University of Science and Technology in 2012. His research interest is focus on image processing, pattern recognition and artificial intelligence.