Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **1**, January **2018**

Deep Learning Models for EEG-based Rapid Serial Visual Presentation Event Classification

Fu-Quan Zhang[1,2]

[1]Fujian Provincial Key Laboratory of Information Processing and Intelligent Control
Minjiang University, Fuzhou, China
[2]School of Software
Beijing Institute of Technology
Room A9-301, Jiyuan, Juyuan, Jinshan, Fuzhou, 350008, China
8528750@qq.com

Zi-Jing Mao, Yu-Fei Huang

Department of Electronic Engineering
University of Texas at San Antonio One UTSA Circle, San Antonio
Arrow Oaks 126, San Antonio, Texas, 78249, USA
Zijing.Mao@utsa.edu, Yufei.Huang@utsa.edu

Lin Xu

Innovative Information Industry Research Institute
Fujian Normal University
Fuqing Brunch of Fujian Normal University, Campus County 1,
Longjiang Road, Fuqing, Fuzhou, 350300, China
xulin@fjnu.edu.cn

Gangyi Ding

School of Software
Beijing Institute of Technology
Room 611, Software Building, Beijing Institute of Technology,
5 South Zhongguancun Street, Haidian, Beijing, 100081, China
dgy@bit.edu.cn

Received April, 2017; revised November, 2017

Abstract. *We consider deep learning (DL) for event classification using electroencephalogram (EEG) measurements of brain activities. We proposed HDNN or hierarchical deep neural network, and CNN4EEG, a new convolution neural network (CNN). Both DL models are designed to improve the representation of spatial and local temporal correlations inherent in EEG data. These models were tested for image target prediction in a time-locked rapid serial visual presentation (RSVP) experiment. The performances were compared with the state-of-the-art RSVP classification algorithms including HDCA and XDAWN and with other popular machine learning algorithms. The results show that global spatial local temporal CNN (CNN4EEG) achieved a 13% improvement over the best competing non-DL algorithm and a 9% improvement over canonical CNN for image processing, and a 6% over deep neural network (DNN). Our results suggest that the unique design of temporal and spatial filters in CNN4EEG can improve the representation of EEG data, hence the prediction performance.*

Keywords: EEG; Rapid serial visual presentation; Deep learning; Hierarchical deep neural network ; Convolution neural network; CNN4EEG

1. **Introduction.** EEG measurement of electrical activity of brain has found a wide range of applications in diagnosis of neurological diseases, cognitive science, and brain computer interaction[1, 2]. An important problem concerning these applications is event prediction based on EEG data. Past research has identified a host of event-related potentials (ERPs) such as P300 and N1 that are indicative of different basic sensory, cognitive, and motor events. However, as demonstrated in [3], the ERPs can change in both magnitude and timing with subjects and experiments, making cross-subject prediction based on ERPs less reliable. Additionally, significant improvement in spatial and temporal resolution of EEG has tempted us to predict much more complex cognitive events that can produce a variety of EEG patterns highly convoluted in space, time, and frequency. Accurate prediction of these events cannot be accomplished by using these basic ERPs and this has led to widespread adoptions of state-of-the-art machine learning algorithms. However, as in most machine learning applications, to achieve best prediction performance requires extraction of carefully engineered features in time, frequency, or transferred independent component, whose construction often relies on highly specific domain expertise. Yet, it becomes less clear how to improve beyond using expert-engineered features and in cases when expert prior knowledge is limited, finding discriminant features can be quite problematic.

To address these issues, we investigate deep learning (DL) solutions in this paper. Deep learning is a class of machine learning algorithms that possess a multi-layered architecture. DL draws significant attention in recent year because of its record-beating performance in image recognition, speech recognition, and many other applications[4, 5]. The key to DLs success is its ability to automatically discover discriminant feature representations directly from raw data that are essential for accurate prediction. The reason it adapted to EEG is EEG contained both temporal and spatial information; while CNN has the capability of capture spatial correlation such as image processing [6] and temporal correlation such as speech recognition [7]. Therefore, we proposed to use CNN to capture both spatial and temporal correlation in EEG signals. A few existing work also demonstrated the power of DL, especially convolution neural network (CNN), in EEG classification. In [8], a CNN is applied for predicting epileptic seizure based on intracranial EEG recording. The architecture includes 3 convolution layers, 2 max-pooling layers and 1 fully-connected layer. In this CNN, temporal filters were applied separately on data from individual channels to capture temporal EEG correlations in the convolution layer and the fully connected layer is included to capture the spatial correlation between channels. However, using different temporal filters on different channels could distort spatial correlations, making extraction of correct spatial correlation in the fully connected layer difficult. A similar architecture was also developed in [9]. Another CNN proposed in [10, 11] applies EEG spatial filters such as XDAWN [10, 12] to each of the time samples across all the EEG channels in its convolution layer. The problem with this CNN is that it does not capture local temporal correlations in EEG. The third type of CNN for EEG classification is described in [13, 14]. The CNN model is built for a single channel recording and the inputs to CNN are time-frequency spectrum power values of the channel. Accordingly, a time-frequency convolution filter was applied in the convolution layer.The problem with this CNN is that it completely ignores the spatial correlation.

In this paper, we propose two new DL architectures, namely, hierarchical deep neural network (HDNN), and CNN4EEG to improve the DL representation of spatial and local temporal correlations in EEG data. We tested the performance of these new DL algorithms for target image detection in a rapid serial visual presentation (RSVP) experiment.

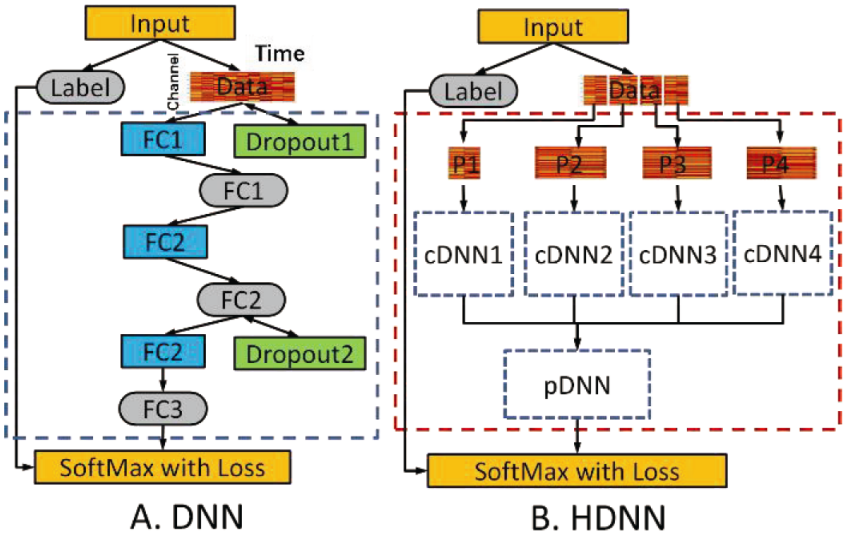

FIGURE 1. Architecture of DNN and HDNN. The blue squares in A are fully connected modules and gray ovals are hidden units. The cDNN and pDNN modules of HDNN in B use the same architecture as shown inside the blue dotted box in A. In HDNN, we separated each 1s epoch into multiple sub-epochs and push them into each individual DNN modules with multiple fully connected layers. Then the output of each DNN modules were concatenated as the input to a new DNN (pDNN) module and the label layer was on top of the pDNN module.

The results show that CNN4EEG achieved a performance improvement of 13 percentage points over the best shallow classifier we tested, 9 percentage points over canonical CNN for image processing, and 6 percentage points over DNN.

2. The RSVP experiment and data preprocessing.

2.1. The Cognitive Technology Threat Warning System. In this study, we considered an RSVP experiment called the Cognitive Technology Threat Warning System (CT2WS) [15]. In CT2WS, all stimuli were short grayscale video clips, with targets including moving people or vehicles, and non-targets including plants or buildings. Subjects were required to find target images and press the button once they saw a target. Stimuli were presented at 2 Hz (one stimulus every 500 ms) and brain signals were recorded with 64-channel Biosemi EEG systems at a sampling rate of 512 Hz. There were 15 subjects and each performed a 15-min session. The voluntary, fully informed consent of the subjects was obtained as required by federal and Army regulations [16, 17].The investigator has adhered to Army policies for the protection of human subjects [17]. The goal is to predict if the subject sees a target clip using EEG recordings.

2.2. Data Preprocessing and Prediction Objective. The raw EEG samples were first bandpass filtered with a bandwidth of 0.1-55 Hz. The lower cutoff frequency at 0.1Hz aims to reduce the distortion to data in low frequencies (10Hz) [18] and that at 55 Hz aims to remove the electrical artifacts. Down-sampling was performed to reduce the sampling rate to 128 Hz, which is the maximum down-sampled frequency that does not produce aliasing at the high-passed frequency. According to the procedure described in [3], 1-sec EEG epochs after each target/non-target onset were extracted for all subjects to be used for training and prediction. However, 80% of epochs are non-target. To remove the

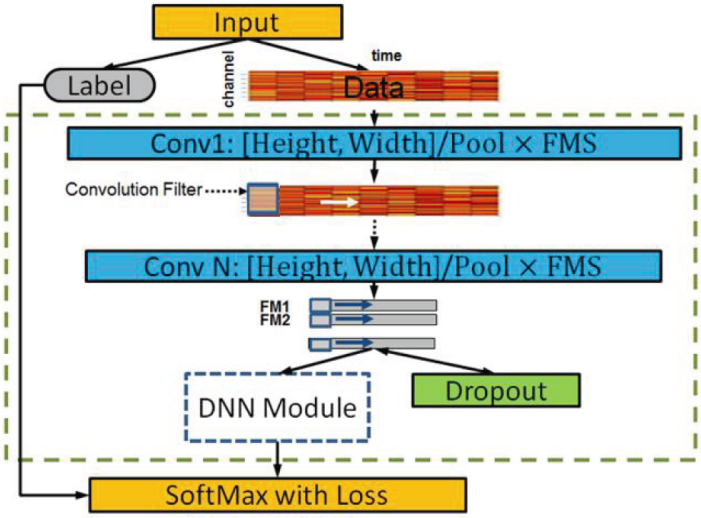

FIGURE 2. CNN4EEG architecture. There are two convolution layers and blue boxes are convolution operations, where the texts inside represent [kernel shape]/ MP width * Feature Map Size. "FM" denotes feature map

potential bias in the trained classifier towards predicting non-target events, we balanced the target and non-target epochs by randomly sampling equal numbers of target and non-target. In all, about 10,400 epochs (700 epochs per subject) were obtained. Each epoch is of dimension (64*128) and was normalized by subtracting the mean of the corresponding channel and then dividing by the standard deviation.

3. The proposed deep learning models.

3.1. **Introduction to Deep Neural Network.** We first discuss the DNN architecture and its basic modules for this classification. These modules will be the basic building blocks of the proposed new architectures. An overview of DNN modules is shown in Fig. 1-A. The input to DNN is a vectorized EEG epoch (1*(64*128)) in this case) and the classification (target or non-target) is produced in the output layer, where the SoftMax function is used. Multiple hidden layers are included in- between the input and the output layer. Each hidden layer includes a fully connected (FC) architecture linking input and output hidden units. The Rectified Linear Unit (ReLU) [19] activation function is added in each hidden layer because the ReLu function learns much faster and does not have neuron saturation like the sigmoid function. Dropout modules [20] are also included either on top of the input layer (Dropout=0) or below the last hidden layer (Dropout=1) as a sparse constrain to prevent over-fitting [21]. Notice that DNN does not intentionally seek to represent spatial and temporal correlations in EEG. Therefore, it would take larger training samples and longer training iterations to learn the underlying spatial and temporal information.

There are two convolution layers and blue boxes are convolution operations, where the texts inside represent [kernel shape]/ MP width * Feature Map Size. FM denotes feature map.

3.2. **The proposed hierarchical DNN (HDNN).** HDNN is designed to capture the local temporal correlations in EEG data and is inspired by HDCA[22], a popular algorithm for RSVP target detection. For HDNN input, the raw EEG data is first divided along time into non-overlapping $64 \times W$ sub-epochs, where is the length of the sub-epoch.These

sub-epochs are fed into independent child-DNNs (cDNN). Each cDNN is a DNN without the output layer. The output hidden units of all cDNNs are then concatenated and fed into a fully-connected, regular DNN called parent-DNN (pDNN). Fig. 1-B shows the proposed HDNN architecture. Compared with DNN, HDNN uses cDNNs to learn local temporal EEG features. Although these cDNNs can have their unique weights to be able to learn unique and shared temporal patterns in EEG sub-epochs, we request cDNNs to have shared weights because the training samples are limited in this case. The weight sharing acts as a regularization strategy to emphasize similar local features across different sub-epochs. In addition, it also reduces size of the model and subsequently increases the speed of training.

3.3. **The proposed convolution neural network for EEG (CNN4EEG).** with $M = C \times T$, where both C (channel size) and T (time samples) are 64 in this case. Also, let W_{ct}^{pq} represent the c^{th} and t^{th} weight of the p^{th} feature map for hidden layer k and q^{th} feature map for hidden layer k-1, where c=1,,c',t=1,,t' with $c' \times t'$ as the kernel size and p = 1, , P, q = 1, , Q, where P, Q are feature map (FM) sizes as hyper-parameters to be learned. Then, the p^{th} FM at the output of the convolutional layer is:

$$\text{convolution}\left(\mathbf{v}\right)_{ct}^{p} = \text{ReLU}\left(\sum_{q=1}^{Q} W_{ct}^{pq} * \mathbf{v}_{ct} + b_p\right) \tag{1}$$

where $\mathbf{v}_{ct}$ is input element corresponding to the EEG measurement from channel c at time t, ReLU represents the rectified linear function $f\left(x\right) = max\left(0, x\right)$. Asterisk sign is convolution operation as $W_{ct}^{pq} * \mathbf{v}_{ct} = \Sigma_{u=-c'}^{c'} \Sigma_{v=-t'}^{t'} W_{ct}^{pq} \mathbf{v}_{c-u,t-v}$ and b_p is the bias parameter for p^{th} feature map. We can see from (1) that the kernel filters for all channels at time t form a spatial filter. After the convolutional layer, an MLP is added to combine all FMs for prediction of target/nontarget events. The number of FMs and the number of hidden units are CNN hyper-parameters to be tuned during training. is a convolution neural network but designed specifically to capture spatial and local temporal correlations using multiple filters or kernels. As regular CNN, CNN4EEG also contains a combination of convolution layers and fully connected layers. Particularly, CNN4EEG has two convolution layers specifically designed to capture both spatial and temporal correlations in EEG.

In the first convolution layer, a 64 × Conv1W kernel is applied to sub-epochs and particularly, the kernel slides from the start to the end of the epoch to generate a $1 \times (128 - \text{Conv1W} + 1)$ feature map (Fig. 2). Multiple kernels can be applied and each generate a different feature map. The feature map size (FMS) is a model parameter to be determined. After the first convolution layer, a Max-Pooling (MP) layer with width (MP1W) is applied to the feature maps. MP passes the maximal marginalized information to the next layer; it not only decreases the dimension of a feature map but also factors out the nuisance variations over time. After MP, the second convolution layer is constructed to extract deeper and more abstract temporal patterns. This layer applies 1-D kernels each with width Conv2W to the outputs of MP. Another MP layer (with a MP width MP2W) is then followed. For both convolution layers, we always use a filtering stride of 1 to achieve a maximum overlapping between consecutive embedding; this enables CNN4EEG to capture detailed temporal dynamics in EEG. Note that additional convolution layers with temporal filters can be added to extract more abstract correlations. In current CNN4EEG, only two layers are used. On top of the convolution layers, a fully connected DNN module is added. Fig. 2 illustrates the architecture of CNN4EEG. Compared with the conventional CNN for image recognition such as ImageNet [23], the

TABLE 1. Model parameters for DNN

DNN Parameters	Search Space
NHL	1, 2, …, 6
HUS	100, 200, 400, …, 3200
Dropout	0, 1

TABLE 2. Model parameters for HDNN

HDNN Parameters	Search Space
NHL for cDNN & pDNN	1, 2, 3, 4
HUS for cDNN & pDNN	100, 200, 400, 800
Dropout for pDNN	0, 1
Sub-epoch Size	2, 4, 8

main differences in CNN4EEG are the specific designs of the filters in each layers. In CNN, each convolution layer implements a bank of 2-D square FIR filters to extract local features at different scales in images [8]. However, although an EEG epoch is arranged as a 2-D matrix, it cannot be treated directly as an image because the arrangement of channels (rows) in EEG data epochs does not reflect the spatial relationship of the channels. Therefore, the square kernels cannot correctly capture the spatial correlations in EEG.

4. **Results.**

4.1. **Construction of training data and process of training.** We partitioned the epochs into the validation, training, and testing set for hyper-parameter training with validation dataset and performance evaluation for testing dataset. We selected one individual subject (subject 1) as our validation set for model selection. Model selection is an important step of deep learning training; it determines the model hyper parameters including the number of hidden layers (NHL), the hidden unit size (HUS), position of dropout modules, max-pooling width, and number of feature maps (NFM). Since these deep learning models have a large number of model combinations, the model parameters need to be trained carefully and efficiently to be able to achieve good performance [24]. Specific model parameters to be trained and the search space for DNN, HDNN, and CNN4EEG are detailed in Table I-III. The search space was defined to balance the trade-off between a deeper architecture and limited training samples. 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 [25] 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 weight decay was 0.0005. Here, we applied random search [26] 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.

For the remaining 14 subjects, a leave-one-subject-out cross-validation (CV) was performed on the best model to test the cross-subject prediction performance. In each round of CV, samples from 13 subjects were used for training, while those from the other subject were for testing. On average, there were 9000 training samples and 700 testing samples

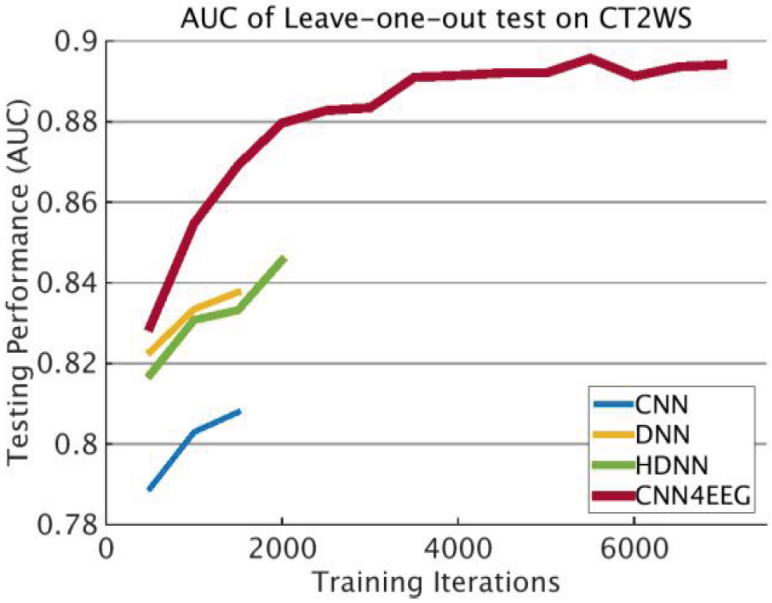

FIGURE 3. Learning curves of the 4 DL models. Stopping TI is selected by early stopping

TABLE 3. Model parameters for CNN4EEG

Learning Parameters	HDNN Parameter Choice
Conv1W & Conv2W	1, 2, 4, …, 16
NFM	10, 20, 30, …, 60
MP1W & MP2W	2, 4
NHL for DNN	1, 2, 3, 4
HUS for DNN	100, 200, 400, 800
Dropout for DNN	0, 1

TABLE 4. Training iteration (TI) for 4 DL models

DL Model	CNN	DNN	HDNN	CNN4EEG
TI	1,500	1,500	2,000	7,000

in each CV round. For comparison, we also implemented four popular shallow learning classifiers: SVM, LDA, Bagging tree and multilayer perception (MLP) on both raw EEG data and data from feature extraction. 20 features proposed in [27] for EEG classification were extracted for each channel, spanning 5 principle frequency bands (delta, theta, alpha, beta and gamma bands) [28] and they include four types: dominant frequency, average power of dominant peak, center of gravity frequency and frequency variability. Therefore, for each epoch, the feature selected EEG data (FS-EEG) has a dimension of 64×20. Finally, we also tested the state-of-the-art methods for RSVP classification including HDCA and XDAWN.

4.2. **Cross-subject Classification performance of DL models.** We first examined the performance of four DL models: DNN, CNN, HDNN, and CNN4EEG. The training

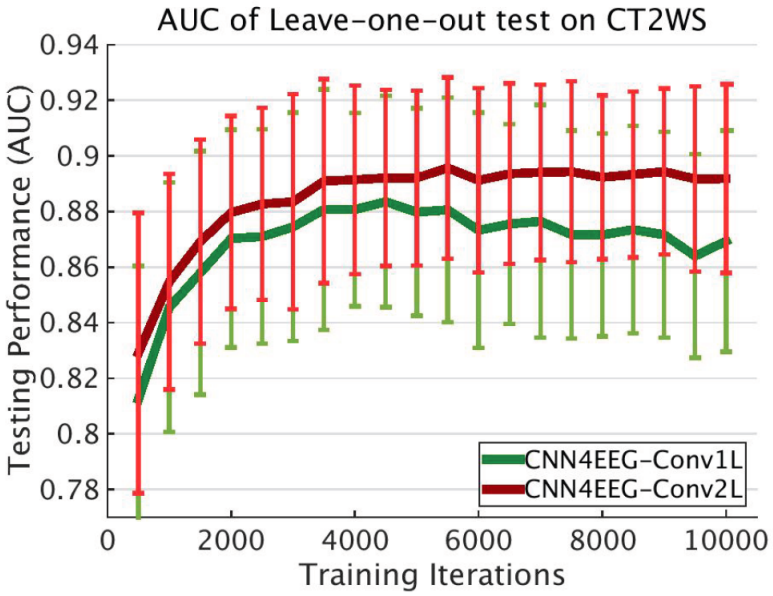

FIGURE 4. Learning curves of one vs. two convolution layers. Each vertical line bar indicates the standard deviation

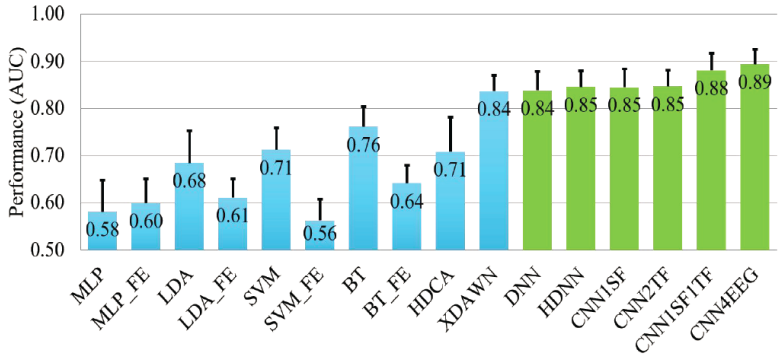

FIGURE 5. Cross-subject prediction AUCs of all 14 tested DL and non-DL algorithms

iterations (TIs) for different DL models are shown in Table IV. Fig. 3 shows the learning curves for AUC performance vs. the training iterations for the DL models. First of all, there was a significant differences in TIs across these 4 models (p¡7e-8, Friedmans test,). Even though CNN4EEG takes the longest training iteration to reach convergence, its computation time is on par with CNN. When comparing the performance, we observed that the best model is CNN4EEG (AUC $= 0.895 \pm 0.032$) , followed by HDNN (AUC $= 0.846 \pm 0.034$)], DNN (AUC $= 0.838 \pm 0.041$)] and CNN (AUC $= 0.808 \pm 0.039$)]. CNN4EEG achieves 9% improvement over CNN (p¡8e-7, t-test,) and 6% of improvement over DNN (p¡4e-4, t-test,). Comparing between HDNN and DNN, we observed slightly improvement of 1% (p=0.204, Wilcoxon sign rank test) but we noted that compared with DNN the training speed of HDNN increased by 10x and its memory decreased by 10x. The fact that CNN has the worst performance underscores the difficulty of square filters to capture correct spatial correlation. Taken together, these results clearly show that the EEG-specific temporal filters in CNN4EEG and HDNN can improve the representation of EEG data, thus leading to better performances. The use of multiple kernels in CNN4EEG entails additional performance improvement.

4.3. **One vs. two convolution layers in CNN4EEG.** We next examined if the deeper temporal features captured by the second convolution layer in CNN4EEG can add discriminant values. To this end, we compared the AUCs of having two vs. one convolution layer. Fig 4 shows the learning curves over the entire 10,000 TI, where the two-layer stops at 7, 000 iterations and the one-layer at 5,000 iterations (given by early stopping). The AUC performance of one layer is 0.8770.035 and therefore the second layer adds about 2% improvement (p¡3e-4, Wilcoxon sign rank test). We also noted that having two layers produced a more stable learning curve, hence less over-fitting. Overall, this result supports the use of additional layers and we also believe more layers could be beneficial if there were more training samples.

4.4. **Overall classification performance.** Performances of DL algorithms were next compared with other non-DL algorithms and the results are shown in Fig. 5. There is a significant performance difference between DL and non-DL algorithms (p¡1e-26, Friedmans test) and overall, DL algorithms outperform the shallow learning algorithms. Particularly, CNN4EEG (AUC=0.8950.032) improved over Bagging Tree (AUC=0.7610.051), the best shallow learning algorithm by 13% (p¡8e-9, t-test). We also performed shallow learning in MLP, LDA and the best performance is achieved by XDAWN in the group of shallow learning algorithms. This might because XDAWN is speicially designed for processing P300 features in EEG signal, which is also a key feature in RSVP experiments; while other shallow learning alogrithms are generic and might not be suitable for EEG data structure.

5. **Conclusion and future work.** We presented in this paper HDNN and CNN4EEG, two new DL models for EEG based event classification. They are designed to improve the representation of spatial and local temporal correlations in EEG. The test results on cross-subject target image detection using CT2WS RSVP data demonstrated their better performance over DNN and CNN and a significant improvement by CNN4EEG over other shallow learning algorithms and the state-of-the-art algorithms for RSVP classification. Since the architecture of CNN4EEG is applicable for general EEG classification, additional effort in the future to investigate its performance for different BCI tasks is desirable.

Acknowledgment. This work was supported by the Army Research Laboratory Cognition and Neuroergonomics Collaborative Technology Alliance (CANCTA) under Cooperative Agreement Number W911NF-10-2-0022 and supported by the Research Program Foundation of Minjiang University under Grants No. MYK17021 and also supported by the Major Project of Sichuan Province Key Laboratory of Digital Media Art under Grants No. 17DMAKL01 and also supported by Fujian Province Guiding Project under Grants No. 2018H0028. We also acknowledge the computational support from Computational Biology Core at the University of Texas at San Antonio, funded by the National Institute on Minority Health and Health Disparities (G12MD007591).

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, pp. 432-41, Oct 2008.

[2] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, et al., Brain-computer interface technology: a review of the first international meeting, *IEEE Trans Rehabil Eng* , vol. 8, pp. 164-73, Jun 2000.

[3] P. Sajda, E. Pohlmeyer, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, et al., In a blink of an eye and a switch of a transistor: cortically coupled computer vision, *Proceedings of the IEEE*, vol. 98, pp. 462-478, 2010.

[4] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, et al., Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, *Signal Processing Magazine, IEEE*, vol. 29, pp. 82-97, 2012.

[5] G. E. Dahl, D. Yu, L. Deng, and A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, *Audio, Speech, and Language Processing, IEEE Transactions on*,vol. 20, pp. 30-42, 2012.

[6] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*,2016, pp. 770-778.

[7] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, et al., The Microsoft 2016 conversational speech recognition system, *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*,2017, pp. 5255-5259.

[8] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG, *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*,2008, pp. 244-249.

[9] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, Time series classification using multi-channels deep convolutional neural networks, *Web-Age Information Management, ed: Springer*, 2014, pp. 298-310.

[10] 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, *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 25, pp. 2030-2042, 2014.

[11] H. Cecotti and A. Grser, Convolutional neural networks for P300 detection with application to brain-computer interfaces, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*,vol. 33, pp. 433-445, 2011.

[12] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, xDAWN algorithm to enhance evoked potentials: application to braincomputer interface, *Biomedical Engineering, IEEE Transactions on*, vol. 56, pp. 2035-2043, 2009.

[13] S. Stober, D. J. Cameron, and J. A. Grahn, Classifying EEG recordings of rhythm perception, *15th International Society for Music Information Retrieval Conference (ISMIR14)*, 2014, pp. 649-654.

[14] 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*, 2014, pp. 1449-1457.

[15] A. J. Ries and G. B. Larkin, Stimulus and Response-Locked P3 Activity in a Dynamic Rapid Serial Visual Presentation (RSVP) Task, *US Army Research Laboratory2013*, 2013.

[16] J. Meng, L. M. Merio, K. Robbins, and Y. Huang, Classification of Imperfectly Time-Locked Image RSVP Events with EEG Device, *Neuroinformatics*,vol. 12, pp. 261-275, 2014.

[17] J. Meng, L. M. Merio, 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, p. e44464, 2012.

[18] 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.

[19] X. Glorot, A. Bordes, and Y. Bengio, Deep sparse rectifier neural networks, *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315-323.

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.

[21] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, Deepface: Closing the gap to human-level performance in face verification, *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*,2014, pp. 1701-1708.

[22] A. D. Gerson, L. C. Parra, and P. Sajda, Cortically coupled computer vision for rapid image search, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, pp. 174-179, 2006.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 2012, pp. 1097-1105

[24] Y. Bengio, Learning deep architectures for AI, *Foundations and trends in Machine Learning*, vol. 2, pp. 1-127, 2009.

[25] L. Prechelt, Automatic early stopping using cross validation: quantifying the criteria, *Neural Networks*,vol. 11, pp. 761-767, 1998.

[26] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, On the importance of initialization and momentum in deep learning, *Proceedings of the 30th international conference on machine learning (ICML-13)*, 2013, pp. 1139-1147.

[27] M. V. Yeo, X. Li, K. Shen, and E. P. Wilder-Smith, Can SVM be used for automatic EEG detection of drowsiness during car driving?, *Safety Science*, vol. 47, pp. 115-124, 2009.

[28] W. O. Tatum, Ellen R. Grass Lecture: Extraordinary EEG, *The Neurodiagnostic Journal*, vol. 54, pp. 3-21, 2014.