Deep Transfer Learning for Cross-subject and Cross-experiment Prediction of Image Rapid Serial Visual Presentation Events from EEG Data

Mehdi Hajinoroozi[1(✉)], Zijing Mao[1], Yuan-Pin Lin[2], and Yufei Huang[1]

[1] University of Texas at San Antonio, San Antonio, USA
Mehdi.hajinoroozi@my.utsa.edu, mzjl68@hotmail.com, Yufei.huang@utsa.edu

[2] Institute of Medical Science and Technology, National Sun Yat-sen University, Kaohsiung, Taiwan
yplin@mail.nsysu.edu.tw

Abstract. Transfer learning (TL) has gained significant interests recently in brain computer interface (BCI) as a key approach to design robust predictors for cross-subject and cross-experiment prediction of the brain activities in response to cognitive events. We carried out in this.aper the first comprehensive investigation of the transferability of deep convolutional neural network (CNN) for cross-subject and cross-experiment prediction of image Rapid Serial Visual Presentation (RSVP) events. We show that for both cross-subject and cross-experiment predictions, all convolutional layers and fully connected layers contain both general and subject/experiment-specific features and transfer learning with weights fine-tuning can improve the prediction performance over that without transfer. However, for cross-subject prediction, the convolutional layers capture more subject-specific features, whereas for cross-experiment prediction, the convolutional layers capture more general features across experiment. Our study provides important information that will guide the design of more sophisticated deep transfer learning algorithms for EEG based classifications in BCI applications.

Keywords: Transfer learning · Deep convolutional neural networks · EEG signals

1 Introduction

Rapid Serial Visual Presentation (RSVP) is a widely used EEG-based brain computer interface (BCI) paradigm designed to study human brain response to time-lock rare target stimuli [1]. RSVP has also found many applications including BCI keyboard, smart learning, etc. Like in most BCI systems, designing robust classifier for accurate prediction of RSVP target event from EEG measurements is a crucial component and it has benefited from the advancement in machine learning and signal processing. While the XDAWN filter [2] and Bayesian linear discriminant algorithm (BLDA) [3]

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC 2017, Part I, LNAI 10284, pp. 45–55, 2017.
DOI: 10.1007/978-3-319-58628-1_4

represent two state-of-the-art shallow algorithms for RSVP target event classification, deep learning has also gained much interest for this classification recently. To this end, we have conducted comprehensive investigations of convolutional neural network (CNN) models and showed that the spatial-temporal CNN (STCNN) model can achieve considerable performance improvement over both XDAWN and BLDA in predicting RSVP events, [4] demonstrating the ability of deep learning to learn robust and complex EEG discriminate features.

To achieve this improved performance, deep learning requires a large amount of training data. However, collecting large training data for a single user is expensive and laborious. Prolonged BCI training time can also induce fatigue, thus deteriorating user performance. It is therefore desirable to integrate data from other subjects performing the same or similar BCI experiments. However, it is well known that there is a large variation in individual brain responses to the same stimuli and brute-force combing data from different subjects might degrade rather than improve the performance. Instead, transfer learning [5–7] provides a principle paradigm for identifying and adapting discriminate information in data across different subjects or experiments to help improve subject-specific classification performance. However, developing deep learning based transfer learning algorithms for RSVP event prediction and general EEG-based classification is still an open topic, yet to be investigated.

Because of the nature of deep learning algorithm and architecture, transfer deep learning models can be easily implemented through its fine tune process. However, fine-tuning does not always lead to improved performance and an important investigation of feature transferability of CNN models for image recognition [11] has showed that the transferability decreases with layers, where the lower convolution layers tend to learn general features more transferable and higher fully connected layers are more likely to learn less transferable, task-specific features. This result has inspired new deep transfer learning algorithms such as deep adaptive network that optimize the transferable features in CNN.

However, the extent to which the STCNN (as a deep convolutional neural network) layers can be transferred and if the transferability result for image recognition still holds for RSVP event prediction and general EEG-based classification are unclear. To answer this very important question, in this work, we investigate how transferable the layers of STCNN are. Specifically, we determine if the features learned in each layer of the STCNN are general to different subjects or experiments or subject-/experiment specific in the case of RSVP event prediction. We investigated both cross-subject and cross-experiment predictions and interestingly, we showed that the fully connected layer features are specific features and cannot be transferred. On the other hand, the convolution layer features are extracting some general features but are not completely general. In addition, transferring the features from source domain to target domain and performing fine-tuning result in the best classification in target domain.

The rest of this paper is organized as follows. In Sect. 2, we introduce the datasets used for this investigation. In Sect. 3, we explain the STCNN architecture for RSVP event classification. In Sect. 4, we discuss the procedures of our investigation of feature transferability in different layers of the STCNN and demonstrate the results for both cross-subject and cross-experiment predictions. Concluding remarks are provided in Sect. 5.

2 Description of Data

In this work, we used EEG data from three RSVP experiments to study STCNN feature transferability for both cross-subject and cross-experiment event prediction. In an RSVP experiment, subjects are asked to identify a target image from a continuous burst of image clips presented at a high rate. The target image can be predefined or decided by certain rules. Subjects EEG signals are recorded during this process. The patterns in EEG signals are different when the subject is presented with target or non-target images. Three different RSVP data sets used in this work are CT2WS [19], Static Motion [20] and Expertise RSVP [24]. In the CT2WS RSVP experiment, short grayscale video clips as target and non-target stimuli (targets are moving people or vehicles, and non-targets are plants or buildings) are presented at 2 Hz (every 500 ms). The experiment included 15 subjects, where each subject participated in a 15-min. session, where EEG data were recorded by Biosemi device with 64 channels, at sampling rate of 512 Hz. In the Static motion RSVP experiment, target and non-target images static images presented at speed of 2 Hz. 16 subjects have taken part, where each subject participated in a 15-min. session, and EEG data were collected with a Biosemi headset with 64 electrodes at a sampling rate 512 Hz. The Expertise RSVP experiment consists of a 5-Hz presentation of color images of indoor and outdoor scenes, where the target images come from one of following five categories: stair, container, poster, chair, and door [25]. The experiment consists of 10 subjects, where each subject participated in 5 sessions of 60-min. presentation. EEG data were collected with Biosemi EEG headsets with 256 electrodes at a sampling rate of 512 Hz. The data from all three datasets were first band-pass filtered with a bandwidth of 0.1–55 Hz to remove DC and electrical noise and then down-sampled to 128 Hz to reduce feature dimension and cover the whole frequency band after filtering. For Expertise RSVP, only 64 channels (based on the 10–20 system) were selected. Following the procedure described in [17], one-second epochs of the EEG samples time-locked to each target/non-target onset were extracted for all subjects, where the size of each EEG epoch is 64×128. For cross-subject prediction, we used Expertise RSVP. Specifically, we called samples from subject 1 to 5 including 65831 epochs as dataset A and those from subject 6 to 10 including 62553 epochs as dataset B. For cross-experiment prediction, we combined the EEG epochs from CT2WS and Static Motion data sets which contain 21680 EEG epochs and we call this C data set.

3 Spatial-Temporal CNN for RSVP Event Prediction

In this section we provide the explanation for STCNN architecture and also how the transfer learning can be performed by STCNN.

3.1 Architecture of Spatial-Temporal CNN

We discuss next the architecture of spatial-temporal convolutional neural network (STCNN), a deep learning model for classification of the RSVP EEG data sets [7–13].

STCNN is a CNN model, specially designed in order to extract spatial correlations and local temporal correlations of the EEG signals. STCNN similar to regular CNNs structure includes convolutional layers as feature extractors and fully connected layers (FC) on top of the neural network as classifier. Let $\mathbf{v} \in \mathcal{R}^{M\times 1}$ denote an input vector 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 cth and tth weight of the pth feature map for hidden layer k and qth feature map for hidden layer $k-1$, where $c = 1,\ldots,c'$, $t = 1,\ldots,t'$ with $c' \times t'$ as the kernel size and p = 1,…,P, q = 1,…,Q, where P, Q are feature map sizes (FMS) as hyper-parameters to be learned. Then, the p^{th} FM at the output of the convolutional layer is:

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

where $\mathbf{v}_{ct}$ is input element corresponding to the EEG measurement from channel c at time t, ReLU represents the rectified linear function [19, 20] $f(x) = max(0,x)$. 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/non-target events. In current design which is specific for EEG signals STCNN contains two convolution layers to capture both spatial and temporal correlations in EEG signals. In the first convolutional layer, kernels of size 64 × Conv1W ($c' = 64, t' =$ Conv1W) is applied to sub-epochs, where each kernel slides in the whole epoch from the start to the end to generate a 1 × (128-Conv1W + 1) feature map [21–23]. Figure 1 shows the structure of the STCNN.

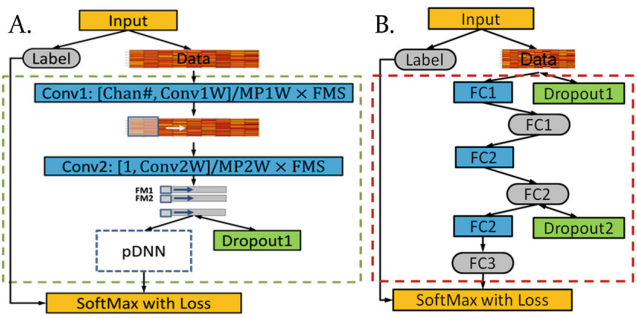

Fig. 1. STCNN architecture. **A.** The designed CNN architecture. There are N convolution layers and blue boxes are convolution operations, where the texts inside represent [kernel shape]/MP width × feature map size. "FM" denotes feature map. **B.** The detailed architecture of the DNN Module in A. The gray ovals are hidden units. (Color figure online)

3.2 Transfer Learning with Spatial-Temporal CNN

Suppose that we have two datasets, generated from the same experiment but for different subjects or from two similar experiments (e.g. two RSVP experiments). Particularly, we call them as source and target domain datasets separately. We further assume that a STCNN has been trained by using the source domain dataset. The goal of transfer learning is to train another STCNN by using the target domain dataset and by transferring the architecture and common features from source domain STCNN. Common features between source and target domain refers to the features learned by STCNNs that are general across these two domains. To perform transfer learning with STCNN, we consider in the paper simple weight transfer and fine-tuning, i.e., we copy the weights of the source domain STCNN to the target domain STCNN and then perform fine tuning. The weight transfer can be carried out by layers. The investigation of transferability of a layer is to study if the source domain weights in this layer contain general or task-specific features. We investigate two approaches. In the first approach, we transfer the weights of a certain layer and fix them in target domain model, which means that the transferred weights will not change and no fine-tuning will be performed on them when the target domain STCNN is trained. We call the first approach "Frozen (fixed) transferred layer approach". For the second approach, we can transfer the weights of a layer from source domain STCNN to target domain model and then that transferred layer gets fine-tuned while the target domain STCNN is being trained. We called the second approach the "fine-tuned transferred layer approach".

4 Results

In this section, we show the results on the transferability of STCNN for both cross-subject and cross-experiment predictions. We used area under the curve (AUC) as a measurement of the prediction performance. To obtain an AUC for an algorithm, a 10-fold cross validation (CV) was performed, where for each CV, the data were randomly separated into 10 equal sized parts with one part used for validation and the remaining 9 parts used for the trained the model. This is done 10 times and the average performance is considered as classification performance of the model. In following sections, we first show the baseline performance of STCNN and other state-of-the-art shallow algorithms and then present the results on the transferability of STCNN for both cross-subject and cross-experiment predictions.

4.1 Baseline Performance of STCNN

We first evaluated the baseline performances of STCNN in dataset A, B and C, respectively, and compared with the state-of-the-art shallow learning algorithms including Bagging, XDAWN-LDA (XLDA) and LDA. Figures 2, 3 and 4 show the classification AUC performances for dataset A, B and C, respectively. They show that STCNN outperforms all three tested shallow machine learning algorithms in all three datasets. STCNN has the highest gain in dataset B, where it achieved $\sim 8\%$ improvement.

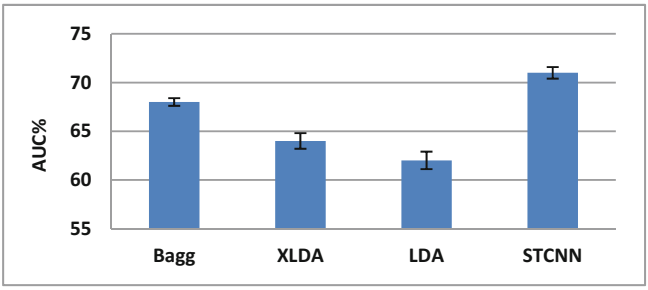

Fig. 2. AUC classification performances for dataset A.

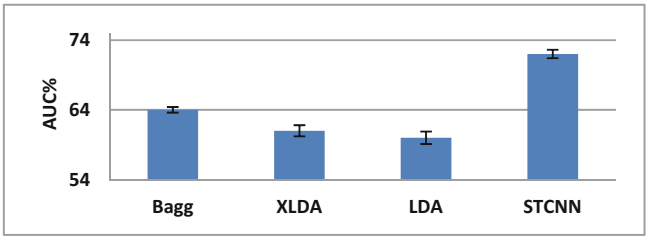

Fig. 3. AUC classification performances for dataset B.

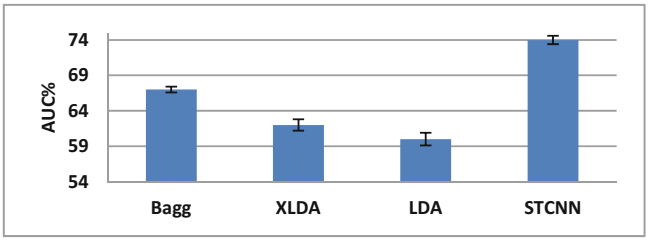

Fig. 4. AUC classification performances for dataset C.

4.2 Investigation of STCNN Transferability for Cross-subject Prediction

In this section, we investigate how transferable are the weights learned in different layers of the STCNN for cross-subject prediction. In this case, we first train a STCNN model using a source domain dataset and our goal is to transfer this model to a target data set. Apparently in cross-subject prediction the source domain contains the EEG epochs of the subjects, which are not seen in target domain and source domain and target domain contain completely different subjects. In order to study the transfer learning for cross subject prediction, we alternate dataset A and B as source and target datasets.

In the following, we used AnB(+) and BnA(+) to represent how the transfer learning is being performed, where AnB means that A is the source domain, B is the target domain, there are n transferred layers and If "+" is also included, then the transferred layers are also fined-tuned. Figure 5 depicts the results for transferring from A to B. The green dot (Base B) is the baseline classification performance of the STCNN trained only using dataset B. The blue dots which are named AnB show the performance of the frozen transferred layer approach and the red dots AnB+ show the performances of the fine-tuned transferred layer approach. We can see that AnB performance drops continuously from the convolutional layers to the fully connected layer, comparing with the baseline performance, where the drops in convolutional layers are higher than in fully connected layers. This suggests that all the layers contain subject-specific features, where convolutional layers seem to capture most of these subject-specific features as performances of frozen fully connected layers does not induce too much drop anymore. The fact that the largest performance drop is about 5% also suggests that all layers also contain a significant amount of general information. This is confirmed by the results of the fine-tuned transferred layer approach (AnB+), where fine-tuning after weights transfer significantly improves the performance and the improvement is pronounced particularly for the convolutional layers. Moreover, when all the layers are transferred and fine-tuned, the highest classification performance 73.28% is achieved, which is 2.69% higher than the baseline performance.

Figure 6 shows the results of transferring from B to A. Very similar results can be seen in this case. In addition, fine-tuning of all transferred layers results in the highest classification performance 72.42%. Taken together, the results show that for cross-subject predictions, both convolutional and fully connected layers in STCNN contain both general features that can be transferred and subject specific information that cannot be transferred. It is notable that fine-tuning of all the transferred layers using the target domain data achieves the best performance and improves the baseline performance.

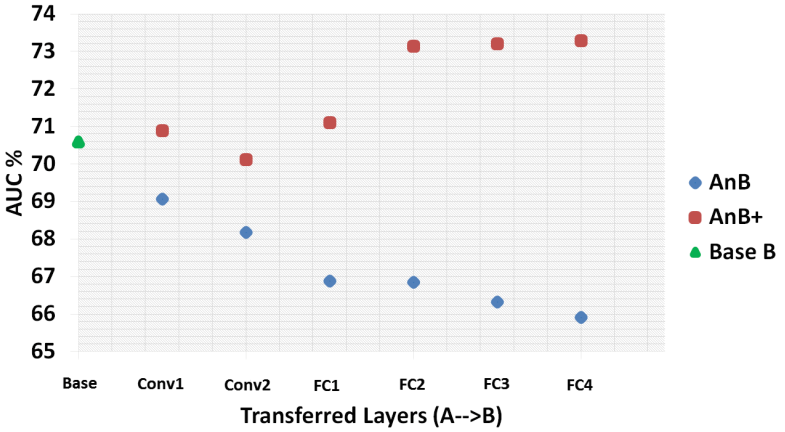

Fig. 5. AUC classification performance of the target domain **B** when the features transferred from source domain **A**. (Color figure online)

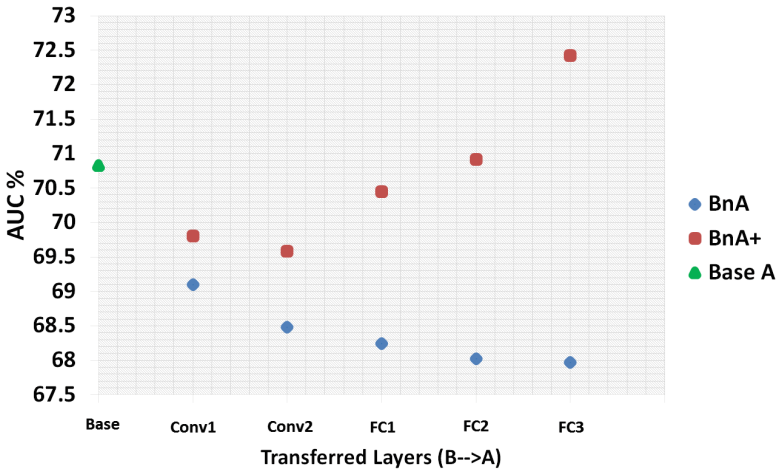

Fig. 6. AUC classification performance of the target domain **A** when the features transferred from source domain **B**.

4.3 Experimental Results for Cross-experiment Transferability Study

In this section, we study the cross-experiment transferability with STCNN. In this case, we consider dataset C as the source domain and A and B are considered two individual target domain datasets. Figures 7 and 8 depict the transferability from C to A and B, respectively. Once again, the green dot is baseline performance trained with only the target domain data. The blue dots named CnA and CnB show the results for the frozen transferred layer approach and the red dots in CnA+ and CnB+ show the results for the fine-tuned transferred layer approach. From CnA and CnB we observe again that the

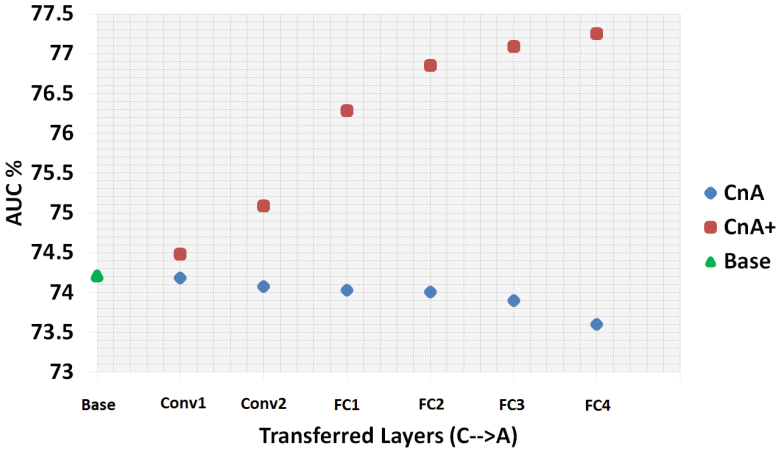

Fig. 7. AUC classification performance of the target domain A when the features transferred from source domain C. (Color figure online)

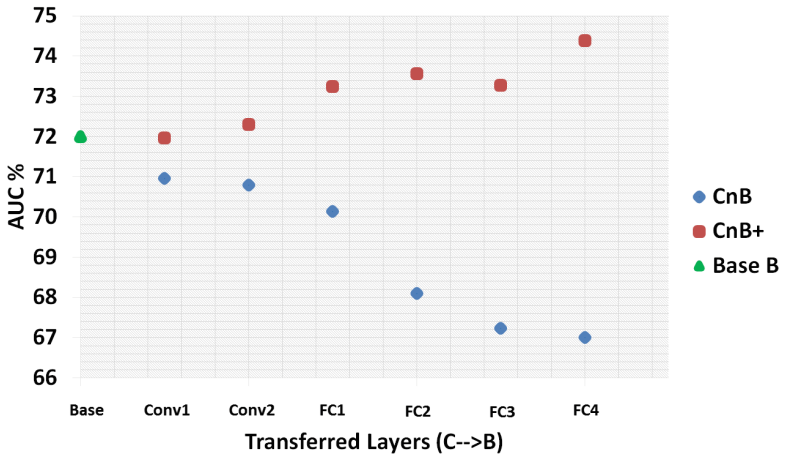

Fig. 8. AUC classification performance of the target domain B when the features transferred from source domain C. (Color figure online)

performance drops with layers, suggesting that all layers learn experimental-specific feature. However, this time there is a lot less amount of drop in convolutional layer; there is almost no significant drop in convolutional layers for C to A. This suggests that the convolutional layers capture a significant portion of features that are general across two different RSVP experiments. In contrast, the fully connected layers contain more experiment-specific features. Since both convolutional and fully connected layers contain experimental specific features, as expected fine-tuning improves the performance and once again fine-tuning of all layers obtains the highest performance 77.24% and 74.38% for transferring from C to A and B respectively.

5 Conclusion

In this work, we studied the transferability of STCNN layers in performing for cross-subject and cross-experiment classification of RSVP target and non-target events using EEG data. We showed that for both cases, all convolutional layers and fully connected layers contain both general and subject/experiment-specific features. For cross-subject prediction, the convolutional layers capture more subject-specific features, whereas for cross-experiment prediction, the convolutional layers capture more general features across experiment. This suggests that the convolutional layers are more likely transferable for cross-experiment predictions. Previously, it has been shown for image recognition that convolutional layers contain general features that can be transferred. Apparently, for EEG based BCI classification, the characteristics of transferability is more complicated. Nevertheless, we show that fine-tuning can improve the baseline performance, which suggests that transfer learning with STCNN has the ability to transfer general features from source domain to improve the performance in the target domain for EEG based classification. This study represents the

first comprehensive investigation of CNN transferability for EEG based classification and our results provide important information that will guide the design of more sophisticated deep transfer learning algorithms for EEG based classifications in BCI applications.

References

1. Bigdely-Shamlo, N., Vankov, A., Ramirez, R.R., Makeig, S.: Brain activity-based image classification from rapid serial visual presentation. IEEE Trans. Neural Syst. Rehabil. Eng. **16**, 432–441 (2008)
2. Cecotti, H., Eckstein, M.P., Giesbrecht, B.: Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering. IEEE Trans. Neural Netw. Learn. Syst. **25**, 2030–2042 (2014)
3. Lei, X., Yang, P., Yao, D.: An empirical Bayesian framework for brain–computer interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. **17**, 521–529 (2009)
4. Rivet, B., Souloumiac, A., Attina, V., Gibert, G.: xDAWN algorithm to enhance evoked potentials: application to brain–computer interface. IEEE Trans. Biomed. Eng. **56**, 2035–2043 (2009)
5. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
6. Cook, D., Feuz, K.D., Krishnan, N.C.: Transfer learning for activity recognition: a survey. Knowl. Inf. Syst. **36**(3), 537–556 (2013)
7. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.-R., Jaitly, N., et al.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Mag. **29**, 82–97 (2012)
8. Shao, L., Zhu, F., Li, X.: Transfer learning for visual categorization: a survey (2014)
9. Razavian, A.S., et al.: CNN features off-the-shelf: an astounding baseline for recognition. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE (2014)
10. Donahue, J., et al.: Decaf: a deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531 (2013)
11. Yosinski, J., et al.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems (2014)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (2012)
13. Patel, A.B., Nguyen, T., Baraniuk, R.G.: A probabilistic theory of deep learning. arXiv preprint arXiv:1504.00641 (2015)
14. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
15. Mirowski, P.W., et al.: Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In: IEEE Workshop on Machine Learning for Signal Processing, MLSP 2008. IEEE (2008)
16. Aytar, Y., Zisserman, A.: Tabula rasa: model transfer for object category detection. In: 2011 IEEE International Conference on Computer Vision (ICCV). IEEE (2011)
17. Li, X.: Regularized adaptation: theory, algorithms and applications. Ph.D. thesis, University of Washington, USA (2007)

18. Yang, J., Yan, R., Hauptmann, A.: Adapting SVM classifiers to data with shifted distributions. In: ICDM Workshops 2007 (2007)
19. U.S Department of Defense Office of the Secretary of Defense: Code of federal regulations, protection of human subjects. 32 CFR 219 (1999)
20. U.S. Department of the Army. Use of volunteers as subjects of research. AR 70-25. Government Printing Office, Washington, DC (1990)
21. Hajinoroozi, M., et al.: Feature extraction with deep belief networks for driver's cognitive states prediction from EEG data. In: 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP). IEEE (2015)
22. Hajinoroozi, M., Mao, Z., Huang, Y.: Prediction of driver's drowsy and alert states from EEG signals with deep learning. In: IEEE 6th International Workshop on Computational Advances in Multi-sensor Adaptive Processing (CAMSAP). IEEE (2015)
23. Hajinoroozi, M., Mao, Z., Jung, T.P., Lin, C.T., Huang, Y.: EEG-based prediction of driver's cognitive performance by deep convolutional neural network. Signal Process. Image Commun. (2016)
24. Touryan, J., Apker, G., Kerick, S., Lance, B., Ries, A.J., McDowell, K.: Translation of EEG-based performance prediction models to rapid serial visual presentation tasks. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) AC 2013. LNCS (LNAI), vol. 8027, pp. 521–530. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39454-6_56
25. Touryan, J., Apker, G., Lance, B.J., Kerick, S. E., Ries, A. J. McDowell, K.: Estimating endogenous changes in task performance from EEG. In: Using Neurophysiological Signals That Reflect Cognitive or Affective State, p. 268 (2015)