# Sustained Attention Driving Task Analysis based on Recurrent Residual Neural Network using EEG Data

Yurui Ming[1], Yu-Kai Wang[1], Mukesh Prasad[1], Dongrui Wu[2], Chin-Teng Lin[1]
[1]Centre for Artificial Intelligence, School of Software, FEIT, University of Technology Sydney, Australia
[2]DataNova LLC, NY, USA
[1]{Yurui.Ming@student, Yukai.Wang@, Mukesh.Prasad@, Chin-Teng.Lin@}uts.edu.au, [2]drwu09@gmail.com

***Abstract*— this paper proposes applying recurrent residual network (RRN) for analyzing electroencephalogram (EEG) data captured during a simulated sustained attention driving task. We first address the suitableness of utilizing residual structure as well as adopting recurrent structure for EEG signal processing. Then based on these descriptions a recurrent residual network is tailored and depicted in detail. Thirdly we use an EEG dataset obtained from a sustained-attention experiment for our model justification. By applying the RRN model to the experimental data and via the competitive result achieved, we demonstrate the elegance of the proposed model. At last, we discuss the characteristics of the learned filters and their interpretations from EEG frequency band perspectives.**

***Keywords—Deep Learning, Recurrent Residual Network (RRN), EEG***

## I. INTRODUCTION

People's interests in the structure and functionality of our brain and the accompanying cognitive process can be dated back to ancient Egypt [1]. With the recent boosts of machine learning, such a trend is only enhanced due to the need of inspiration for more powerful neural network models. There are two ways to conduct brain research. One is from micro perspective, a.k.a. from molecular or cellular level; the other is from macro perspective, a.k.a. from functional level [2]. It is evident that collective behavior of neurons wins its significance over a single neuron since some conclusion cannot be drawn from the behavior of the latter. Electroencephalogram (EEG) which reflects the rhythm of neuron clusters provides a mirror reflecting the brain activity and its indications, hence gain widespread usage in brain research.

However, there are two challenges for research conducted at this regard. The first lies in the captured EEG data itself. Considering the non-invasive way of EEG acquisition, one hand it provides the convenience throughout data capturing, on the other hand since the electrical signals need propagating through various head structural layers with different physical properties, distortions are unavoidable during the process. In addition to the device imperfection, like impedance variation and line noise, artifacts are always present. Second, due to the current understanding and knowledge of physiological and cognitive process, the still-existing limits put constraints on the perspectives from which EEG data can be investigated or interpreted. Alternatively speaking, if some models are taken to analyze the data, it tends to be a difficult task in deciding what kind of features are available or feasible for such models. Plus the pending noise, the result usually deteriorates most of the time.

Recent years have witnessed the achievements and breakthroughs of deep neural networks in various applications [3-5]. One advantage of deep neural networks over other models is automatic feature learning or extraction. It means during the end-to-end learning process the most suitable features will emerge themselves instead of being discovered or crafted manually. So facing the challenge of EEG data analysis, it is appealing to investigate deep neural networks to have some inspirations. In this paper, we will consider two structures which are the most common organic fabrics among deep learning applications, namely, residual structure [6] and recurrent structure [7].

In the following, our work and contributions are lying on several folds as below: (1) we highlight the relations between traditional EEG analysis and the corresponding deep network structures. Based on these we propose our recurrent residual network. (2) we make use of an EEG dataset captured during an experiment concerning with people's vigilance to analyze and show the very competitive result achieved by our model. (3) we interpret the learned filters from frequency band perspectives and correlated the current approach with the traditional EEG signal processing methods.

## II. NETWORK STRUCTURE

To propose neural network structures suitable for EEG data analysis, we first highlight some tactics during EEG analyzing as inspirations. Recall that a typical treatment adopted is normalization or its variant. One example is shown as in Fig. 1, a mean value for each channel in the baseline range is calculated, and values in analyzing segment are centered around this mean value. In detail, for a given channel $p$, let $m_p = \frac{1}{|T_b|}\sum_{i\in T_b} v_p[i]$, then $\bar{v}_p[j] \doteq v_p[j] - m_p, \forall j \in T_a$. Such average normalization could be done in the time domain or frequency domain depending on different situations. Either case, the motivation is to counter-bias the different power levels of signals among different subjects participating the same experiment. Due to physiological traits or impendence caused by contact variations between electrodes and scalp, the average voltage level in the signal varies. However, it is wise to learn the essential alteration pattern of the signal, but not the direct component of the waveform data.

It can be observed to some extent, by reflection of the work in [6], the above treatment of EEG data coincides with the origination of residual structure as in Fig. 2, of which the motivation is to learn just the essence. For instance, suppose an identity mapping was optimal under some extreme case, it would be easier and better to learn nothing instead of learning an identity mapping by a stack of nonlinear layers. For a comparable interpretation with the EEG case, note in Fig. 2, there are no weights associated with the bypass $X$ branch. The learned part is $\mathcal{F}(X)$. Let $\mathcal{H}(X) = \mathcal{F}(X) + X$, so the equivalent learned part is $\mathcal{F}(X) = \mathcal{H}(X) - X$, or some centered or normalized version learned. The above explanation encourages residual structure being adopted as the fundamental building blocks for designing a neural network for EEG data processing.

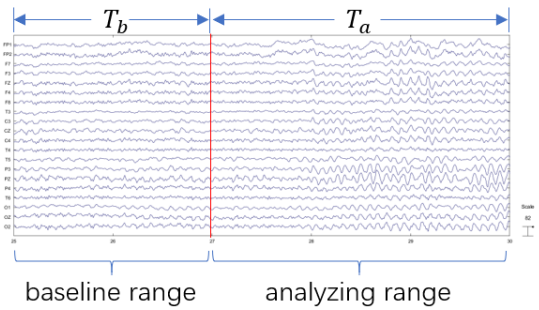



Fig. 1. Averaging normalization

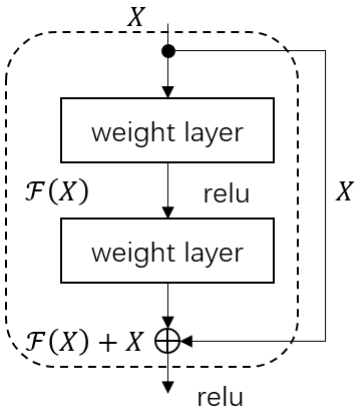



Fig. 2. Diagram of residual block.

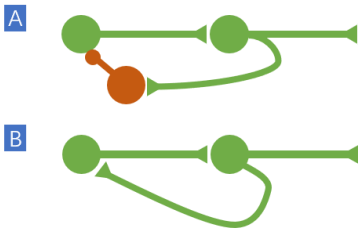



Fig. 3. Biological recurrent structures (A) Inhibition (B) Excitation

There is already work in utilizing recurrent neural network (RNN) for EEG analysis [8], where features extracted from convolutional sub-network are pipelined to recurrent sub-network for inference. Considering artificial neural networks are inspired by the biological structures of mammalian brains, which are ubiquitously endowed with types of recurrent structures as in Fig. 3 [9], it makes sense to begin with the recurrent structures just from initial layers instead of postponing to the final layers. Based on this argument and contrary to work in [8], the proposed network is thoroughly built upon recurrent structures.

Now we depict the recurrent residual blocks (RRB) which constitute the fabric of the network. Just as in Fig. 4, the dashed box confines the boundary of the residual block shown in a time-unfolded manner. Let $X_{l-1}$ denote the output from the bottom layer, $S_l^{t-1}$ denotes the state of the block at a previous time step, then the update of the block at layer $l$ and time $t$ is governed by the following formulas:

$$S_l^t = \text{relu}\left(W_l^i \cdot X_{l-1} + W_l^s \cdot S_l^{t-1} + b_l\right) \quad (1)$$

$$O_l^t = S_l^t + X_{l-1} \quad (2)$$

However, to explore the relationship between multiple channels, the linear transformation in (1) is taken place by convolution. For efficiency, $X_{l-1}$ and $S_l^{t-1}$ are concatenated together and applied convolution as in (3):

$$S_l^t = \text{relu}(\text{conv}(X_{l-1} \odot S_l^{t-1}, W_l) + b_l) \quad (3)$$

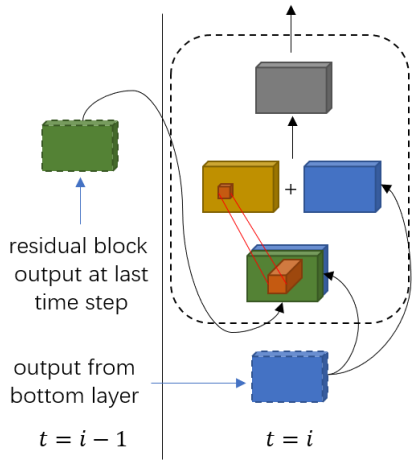



Fig. 4. Recurrent residual block: temporal recurrency and spatial residue

Fig. 5 shows the whole network represented in a succinct form. Since EEG data is one dimensional sampled waveform data constituted by multiple channels, convolutions involved are all in one-dimensional (1D) form. It is worthy of mentioning that channels for input data are the effective channels (sometimes based on selection) and not necessary must be the channels during data acquisition. EEG experiment tends to be endurance ones, so there is case where some channel data may be dominated by noise and becomes useless. As in Fig. 5, the prepared segmented multi-channel EEG data is fed into the network consecutively.

For hidden layers, according to (2) regarding RRB update, $S_l^t$ and $X_{l-1}$ must have the same dimensions to be added together. If adjacent chained RRBs have different feature maps,

another standard convolutional block must be introduced to adjust the number of feature maps. In some cases, it is preferable to reduce the resolution of feature maps from the previous layer, and this can be fulfilled by setting the stride parameter greater than 1 in this convolutional layer. The inserted adjusting convolutional block (ACB) together with RBB form a more massive block. It relies on the repetition of the block in the dotted box in Fig. 5 to enhance the learning capability of the network. Also, the number of repetitions can be customized along with the dimension of the targeted problem.

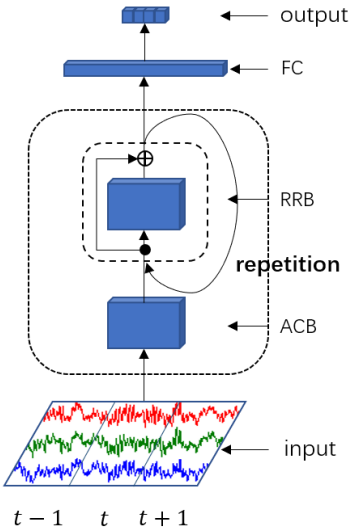



Fig. 5. Recurrent residual network: learning capability deponds on repetion of buidling blocks in dotted box

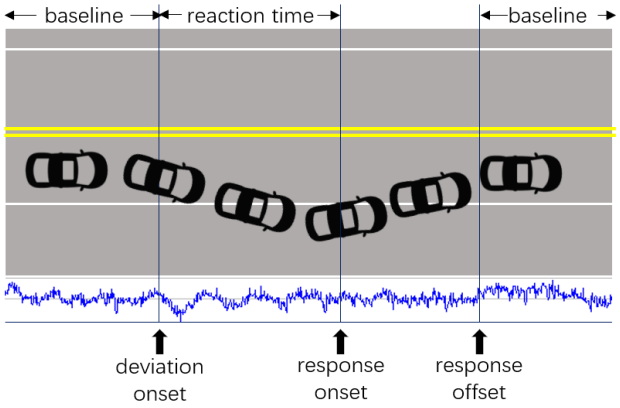
The output of the topmost RRB is flattened and fed into fully-connected (FC) layers for final classifying or regressing purpose. The result section of this paper gives an instantiation or a specific network configuration of the above general network model.

## III. DATASET DESCRIPTION

To justify the effectiveness of the proposed model, we make use of an EEG dataset captured from a sustained-attention driving task [10]. The experiment aimed to investigate people's driving performance with relation to the mental state or vigilance. According to [11], people's fatigue is the most vital factor causing vehicle accidents in commuting and transportation. Research at this regard is always of meaningfulness.

For the experiment scenario, subjects who are certified drivers wearing EEG cap were participating in an endurance driving on a virtual four-lane high way. The experiment was conducted under laboratory environment, with simulated highway projected on multiple giant screens to mimic highway scene and a converted car installed on a maneuverable platform to simulate car conditions.

Fig. 6 shows the paradigm of the experimental procedure. On usual the car is on cruising along one lane. Then a deviation to the car is deliberated introduced (deviation onset) to have the car randomly drift leftward or rightward. The subjects need to steer the wheel to have the car back to the original cruising lane. The moment for subject's taking reaction is called response onset. The duration lasting from deviation onset to response onset is called reaction time, treated as an indicator for subjects' vigilance. Once the car backs to the original cruising state, it is regarded as response offset, and another new transaction begins. One iteration from deviation onset, response onset to response offset is called a trial, and the collection of trials for a certain subject during one experiment is called a session. The number of sessions for subject $i$ is denoted by $N_s^{(i)}$. And the number of




trials accumulated for all $N_s^{(i)}$ sessions is denoted by $N_T^{(i)}$.

Fig. 6. Paradigm of sustain-attention driving test: test subjects' behaving and simutanously recorded EEG

EEG data was captured simultaneously and continuously during the whole process using Scan SynAmps2 Express system. For EEG data, it usually undergoes several preprocessing for later analysis by various models. In this research to take advantage of an automatic feature extraction and learning capabilities of the network model, only very limited preprocessing is applied. In detail, a manual channel selection is first applied to the whole captured data to have perverted channel data removed. Then it is down-sampled to 250Hz, following by a band-pass FIR filter with 0.5Hz to 50Hz. At last, EEG data of specified duration (6 seconds in this paper) in the baseline segment just before deviation onset is extracted into individual epochs for analysis. Fig. 7 shows the details of each preprocessing procedures. The individual epoch data is denoted by $X$.

We are unable to "magically" read the vigilance or mind state of test subjects, so it usually requires a delicate design to indirect measure it. The above experiment achieves this by associating mind states with reaction time (RT). We manually map RTs into different vigilance states according to certain thresholds. Here we only consider a supervised dichotomous classification problem, so only two states, alertness and fatigue are put into consideration. Actually, during the experiment, the subjects only needed to operate the steering wheel in reacting to lane-perturbation events and free from the accelerator and brake pedal controlling, so the primary factor impacting performance was vigilance. We can safely exclude additional factors like attention distraction. Here RT longer than 2.1 seconds is treated as fatigue

label, while RT shorter than 0.7 seconds is treated as alertness label. $Y$ denotes labels for facilitating descriptions.

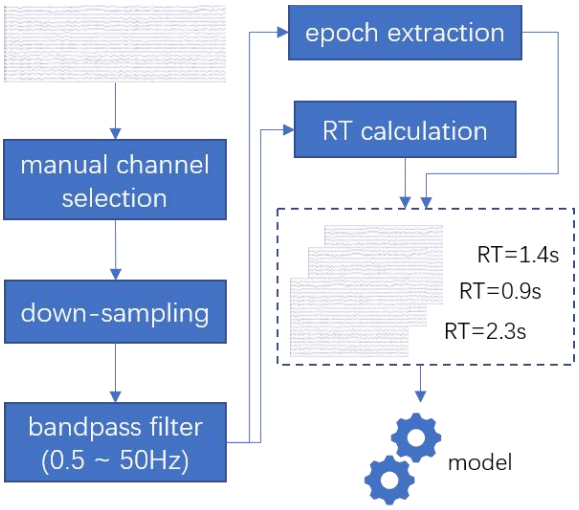



Fig. 7. EEG waveform data preparation flow

## IV. Experimental Result

For deep neural network model, it is preferred large dataset for training. One challenge of EEG experiment is that samples are sometimes costly to acquire. To meet such requirement, we did not take some leave-out scheme. Instead, we use as many samples as feasible to evaluate the performance of our model. In detail, samples from different subjects are blended together and then divided according to the ratio 0.8:0.1:0.1 for training, validation, and testing. Considering the imbalanced samples between different subjects and to counter a biased learned model, we introduce two criteria $N_S^{(i)} \geq 2$ and $N_T^{(i)} \geq 300$. Based on those criteria samples from 7 subjects are selected out for analysis, resulting around 2072 samples for training, 259 samples for cross-validation and testing respectively. We performed 10-fold model validation by using the divided samples.

For comparison, we take support vector machines (SVM) as an example representing traditional machine learning methods. Generally speaking, traditional approaches rely on intensive pre-processing of EEG data to draw independent features for later analysis [18-20]. Sometimes they are impractical from application perspective. However in this paper one purpose is to justify the feasibility of directly utilizing EEG data in waveform, so the input to SVM can be treated as a flatten vector with length $n_c \cdot 6 \cdot n_s$ of the preprocessed trial data $X$. About other configuration of SVM, we relax the constraint to the most margin violations with parameter $C = 1$.

For deep learning models, we adopt multilayer perceptron (MLP) and convolutional neural network (CNN) as benchmarks. These two networks are also capable of feature extraction; however, the capabilities may vary due to structural properties.

For proposed RRN, referring to Fig. 5, all filters for ACB have the same length of 3, and their feature maps are pre-defined by the adjacent RRBs. All the filters for RRB have the same length of $l = 21$, and their feature maps began from 8 and doubled every two repeated blocks. The repetition of the block (1 ACB plus 1 RRB) is 12. So the numbers of feature maps are $8 * 2^{\lfloor \frac{n}{2} \rfloor}, n = 0, 1, \cdots 11$.

For the input into RRN model, 6-second epoch data X extracted from baseline is divided into six non-overlapping segments, with each lasting for 1 second, denoted by $X(i), i = 1, \cdots 6$. We use the corresponding label $y$ as learning signal for training, and criteria for testing. So the RRN is a many-to-one paradigm, and the span of unfold time steps is 6. Further, if we denote the sampling rate ($n_s$) by 250, the number of channels ($n_c$) by 14, the input at each time step is a matrix of dimension $n_s \times n_c$, the total input is of dimension $6 \times n_s \times n_c$.

For MLP, it is constituted by five fully-connected layers. The number of nodes in each layer begins from 4096 and decrease by half. So the number of nodes for the topmost hidden layer is 256. For inputs into the model, the whole 6 seconds duration data are concatenated channel-wise, which turns to be a vector with length $(6 \cdot n_s) \cdot n_c$. For CNN, we use a diminished version of the proposed RRN. We remove the recurrent structure and retained all other configurations. An alternative view is recurrent step is restricted to one, to have it behave like a feed-forward network. The input segments are stacked together to feed into the network all at once, which is a matrix of dimension $(6 \cdot n_c) \times n_s$, to cater to this change.

With the above data preparation and model configuration, we train all the models with the batch number equal to 1/8 of total training samples for 1000 iterations, equivalent to 125 epochs. Then we use all the testing samples for evaluating the performance of individual models. The results were as follows:

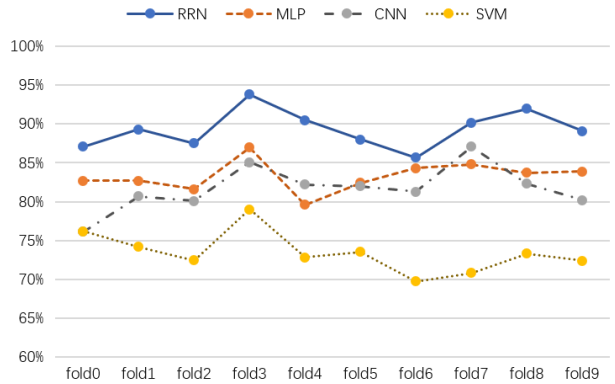



Fig. 8. Test accuracies of different models for each fold

Table I. Average test accuracies of different models

| Model | SVM | MLP | CNN | RRN |
|---|---|---|---|---|
| **Accuracy** | 0.734 | 0.832 | 0.817 | **0.893** |

It can be observed from Fig. 8 that the test accuracy of our proposed model is supreme over all other models for each fold. As indicated, SVM relying on manual-crafted features had difficulty in performing well here. For MLP and CNN, their feature extraction capabilities assist the performance to surpass SVM. However, when compared with the proposed RRN, which

is endowed with all the characteristics of MLP and CNN, the performance is no doubt the best among all the models just as indicated in Table I.

## V. Discussion

It is worth investigating the learned filters to inspire EEG data analysis in traditional ways. However, to directly investigate them in time domain is a little difficult. Sometimes convolution can still be interpreted as local template matching. In two-dimensional (2D) case, the pattern is usually manifest to check, however, in 1D it is a little abstract to inspect in a straight manner. For 1D, if we plot the filters out, we can only spot the peaks and troughs, in addition to delays when compared with some other signals. It is hard to draw definite conclusions or have apparent intuitions. In another way, convolution is closely linked to filtering. So by investigating the power spectra of the learned filters in the frequency domain, we can take an indirect way to estimate their impacts or modulation on original EEG data. However, we first clarify where to concentrate.

The frequency Ω of unit radian/second for the EEG data in continuous case is associated with $\omega$ of unit radian/sample in a discrete case by $\omega = \Omega T_s$, where $T_s$ is the sampling frequency which is 250Hz. Since the original EEG data has undergone a bandpass filter with the highest frequency equal to 50Hz, we denote the highest frequency by $\Omega_h$ which is $2\pi * 50$. Now $\omega_h = \Omega_h T_s = 2\pi * 50/250 = 0.4\pi$. Namely, when utilizing DFT to convert the learned filters from a time domain to frequency domain via formula $H(e^{j\omega}) = \sum_{-\infty}^{+\infty} h(n)e^{-j\omega n}$, we only need to focus on the range $[0, 0.4\pi]$. Because in theory there should be severely attenuated or even no signal beyond 50Hz.

Considering the initial layers are inclined to learn some concrete features, so we extract the learned filters from the second hidden layer. Even for this layer, it has 512 filters, which is impractical to inspect one by one. So we do clustering to the filters and investigate their collective behaviors. The filters are clustered into eight categories using Euclidean metric. The frequency amplitudes of centroids of each category and the corresponding number of elements for each category are drawn in Fig. 9. Both angular frequencies in different units are labeled on the $x$-axis.

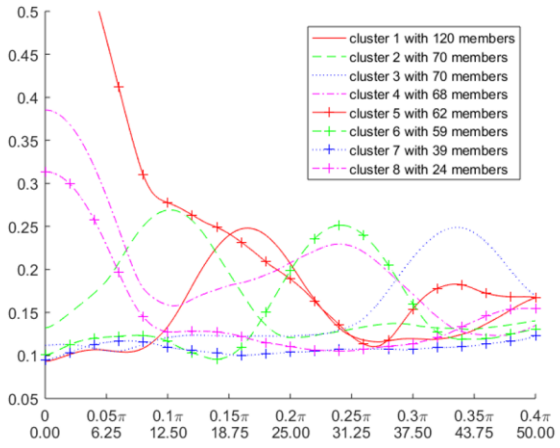



Fig. 9. Clustering of frequency amplitudes of learnt filters

As from Fig. 9, the solid line represents the cluster with the most elements, which spanned across the upper alpha band and almost beta band. The dashed line represents the cluster with the second most elements, which spanned across the alpha band and slightly partial beta band. These are coincided with our previous research [10, 12], peers' research [13] and general neuroscience recognitions [14]. For the beta band, the postulate tends to emphasize its correlation with mind state like vigilance [15]. We frequently link alpha bands to visuomotor conditions [16, 17]. According to the design of the experiment, these two bands should play critical roles in unveiling test subjects' mental states and behaviors. So it is interesting to see if what was automatically learnt agrees with what to be expected.

The category with the least members is represented by a dash-dot-plus line, roughly speaking it is acting as low-pass filtering. However, is argued that the lower frequencies, for example, delta band, are not that useful in most scenario as some research [8] just put it out of consideration. It is also interesting to see least auto-learned filters favor such approach. Although we have difficulty in explaining the dot-plus line representing the category with second least members, we are glad to see to some extent all the filters tend to be evenly distributed covering all the available bands.

## VI. Conclusion

In this paper we proposed a recurrent residual neural network for processing EEG data, especially from time domain with very limited preprocessing. We detailed the motivation of adopting such structures and demonstrated the competitive results achieved by comparing it with other benchmark models. We also examined the learned filters and highlighted their characteristics and emphasized the correlations with EEG signal processing in traditional ways.

## VII. Acknowledgement

This work was supported in part by the Australian Research Council (ARC) under discovery grant DP180100670 and DP180100656. This research was also sponsored in part by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022 and W911NF-10-D-0002/TO 0023. The views and the conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S Government. The U.S Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## References

[1] Mohamed W., "The Edwin Smith Surgical Papyrus: Neuroscience in Ancient Egypt," IBRO History of Neuroscience, July 2014.

[2] Zull J., The art of changing the brain: Enriching the practice of teaching by exploring the biology of learning, Sterling, Virginia: Stylus Publishing, LLC, 2002.

[3] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp. 1-9, 2015.

[4] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le, "Sequence to Sequence Learning with Neural Networks," arXiv:1409.3215v3, 2014.

[5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," Nature 521, pp.436–444, 2015.

[6] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," CVPR, Vol.2016-, pp.770-778, 2016.
[7] Karpathy Andrej, Johnson Justin, and Li Fei-Fei, "Visualizing and Understanding Recurrent Networks," Cornell Univ. Lab. 2015.
[8] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," ICLR, 2016.
[9] Purves D, Augustine GJ, Fitzpatrick D, et al., editors. Neuroscience. 2nd edition. Sunderland (MA): Sinauer Associates; 2001.
[10] Chun-Hsiang Chuang, Li-Wei Ko, Tzyy-Ping Jung, and Chin-Teng Lin, "Kinesthesia in a Sustained-attention Driving Task," NeuroImage, vol. 91, pp.187-202, 2014.
[11] http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/chapter3.pdf?ua=1
[12] Liu Y.T., Chuang C.H., Wang JM, and Lin C.T., "Changes in Alertness and the EEG Effective Connectivity in a Sustained-Attention Driving Task," Journal of Neuroscience and Neuroengineering, 2016.
[13] Foong R., Ang K.K., et el, "An analysis on driver drowsiness based on reaction time and EEG band power," Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Vol.2015-, pp.7982-7985, 2015.
[14] Pfurtscheller G., Lopes da Silva, and F. H., "Event-related EEG/MEG synchronization and desynchronization: Basic principles," Clinical Neurophysiology 110, pp.1842–57, 1999.
[15] Gola Mateusz, Magnuski Mikoaaj, etc, "EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects," International Journal of Psychophysiology, 2013.
[16] Roberts Daniel M., Fedota John R, et el, "Prestimulus Oscillations in the Alpha Band of the EEG Are Modulated by the Difficulty of Feature Discrimination and Predict Activation of a Sensory Discrimination Process," Journal of cognitive neuroscience, Vol. 26, no. 8, pp. 1615-1628, 2014.
[17] Rihs Tonia A., Michel Christoph M., and Thut Gregor, "Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization," European Journal of Neuroscience, Vol. 25. no. 2, pp. 603-610, 2007.
[18] Z. Cao, M. Prasad and C. T. Lin, "Estimation of SSVEP-based EEG complexity using inherent fuzzy entropy," 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, 2017, pp. 1-5.
[19] Wu, Dongrui & King, Jung-Tai & Chuang, Chun-Hsiang & Lin, Chin-Teng & Jung, Tzyy-Ping. (2017). Spatial Filtering for EEG-Based Regression Problems in Brain-Computer Interface (BCI). IEEE Transactions on Fuzzy Systems. PP. 10.1109/TFUZZ.2017.2688423.
[20] J. Andreu-Perez, F. Cao, H. Hagras and G. Z. Yang, "A Self-Adaptive Online Brain‑Machine Interface of a Humanoid Robot Through a General Type-2 Fuzzy Inference System," in IEEE Transactions on Fuzzy Systems, vol. 26, no. 1, pp. 101-116, Feb. 2018..