

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# A Motor-Imagery BCI System Based on Deep Learning Networks and Its Applications

---

Jzau-Sheng Lin and Ray Shihb

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.75009>

---

## Abstract

Motor imagery brain-computer interface (BCI) by using of deep-learning models is proposed in this paper. In which, we used the electroencephalogram (EEG) signals of motor imagery (MI-EEG) to identify different imagery activities. The brain dynamics of motor imagery are usually measured by EEG as non-stationary time series of low signal-to-noise ratio. However, a variety of methods have been previously developed to classify MI-EEG signals getting not satisfactory results owing to lack of characteristics in time-frequency features. In this paper, discrete wavelet transform (DWT) was applied to transform MIEEG signals and extract their effective coefficients as the time-frequency features. Then two deep learning (DL) models named Long-short term memory (LSTM) and gated recurrent neural networks (GRNN) are used to classify MI-EEG data. LSTM is designed to fight against vanishing gradients. GRNN makes each recurrent unit to capture dependencies of different time scales adaptively. Similar scheme of the LSTM unit, GRNN has gating units that modulate the flow of information inside the unit, but without having a separate memory cells. Experimental results show that GRNN and LSTM yield higher classification accuracies compared to the existing approaches that is helpful for the further research and application of relative RNN in processing of MI-EEG.

**Keywords:** motor imagery, brain-computer interface (BCI), recurrent neural network (RNN), long-short-term memory (LSTM), gated recurrent neural network (GRNN)

---

## 1. Introduction

Brain-computer interface (BCI) system provides one of the most important aspects, which is an alternative way of communication through brain signals. It is just to translate electroencephalogram

(EEG) signals from a reflection of brain activity into user action through system's hardware and software. A BCI system provides a communication channel not based on nerves and muscles that allow users to communicate by electrodes contacting on scalp. It has attracted increasing attention of a variety of research fields including neuroscience, machine learning, pattern recognition, rehabilitation medicine, and so on.

Motor imagery (MI) is an important research topic in the field of BCI that mentally simulates a given action, e.g., imaging the motions of the limbs [1]. It refers to visualization of a limbic activity, or any other movement, without the actual execution of the motion imagined. It leads to various changes in the connectivity between the neurons present in the cortex. This results in either an event-related desynchronization (ERD) or event-related synchronization (ERS) of mu rhythms. These effects are due to the changes in the chemical synapses of the neurons, the change in strength between the interconnections or the change of intrinsic membrane properties of local neurons. Since extracted from scalp EEG, MI-EEG has the characteristics of nonlinear, nonstationary, and time-varying.

In the research field of MI-EEG-based BCI, several researchers have proposed different strategies. Tomida et al. [2] presented an active data selection method for MI-EEG classification in 2015. Rejecting or selecting data from multiple trials of EEG recordings is crucial in the selection method. To aim at brain machine interfaces (BMIs), they proposed a sparsity-aware method to select data from a set of multiple EEG recordings during MI tasks. An extraction approach with transform-based feature for MI tasks classification was proposed by Baali et al. [3]. A signal-dependent orthogonal transform was used, referred to as linear prediction singular value decomposition (LP-SVD), for feature extraction. They used a logistic tree-based model classifier to classify the extracted features into one of four motor imagery movements. In 2016, Wu et al. [4] used the fuzzy integral with particle swarm optimization (PSO), which can regulate subject-specific parameters for the assignment of optimal confidence levels for classifiers. Lin and Lo [5] constructed a MI-based BCI system to control an electric wheelchair. They used discrete wavelet transform (DWT) to transform EEG signals into frequency domain and applied SVM to classify them into different commands. Chatterjee and Bandyopadhyay [6] used SVM and multilayered perceptron (MLP) for MI-EEG classification in 2016. They showed that both SVM and MLP were suitable for such MI classifications with the accuracy of 85 and 85.71%, respectively. The symmetric positive-definite (SPD) covariance matrices of EEG signals carry important discriminative information proposed by Xie et al. [7] for MI BCI system in 2016. Chatterjeel et al. [8] examined the quality of feature sets obtained from wavelet-based energy entropy with variation of scale and wavelet type for MI classification in 2016. They have verified their study with three classifiers—Naive Bayes, MLP and SVM. Jois et al. [9] compared several classification techniques for motor imagery-based BCI in 2015. They indicated that common features, e.g., band power values, present that the single EEG trials can be extracted by suitable methods for classification using SVM, neural networks, or ensemble classifiers. The classifiers yield different efficiencies and are compared to find the optimal technique for same number of features. They believed the neural net techniques were proved to be the most efficient. One obstacle of the traditional neural networks for their broader application is the initial weights need to be chosen carefully. Generally, small values could make the multilayer

network untrainable owing to weight diffusion, while large initial values of the weights could result in poor local minima [10]. In order to resolve this problem and construct high descriptive-ability neural networks, a new model of strategies and algorithms, called deep learning (DL), has been successfully developed and becomes prevailing in several fields [11].

There are many ways in machine learning for data classification. The most popular and proven method in recent decades is “Artificial Neural Network (ANN).” We know how artificial neural networks adjust weights so that the error between output and input becomes smaller. But even so, this is far from the “artificial intelligence” that we want. If the computer can analyze the data to find the features, then it is closer to the artificial intelligence we want, that is to say, the created computer can think. DL allows computers to analyze their own data to find “features,” rather than decided by human beings with features, just as computers can have deep thinking to learn. DL uses not only a multilayer neural network but also an auto-encoder for unsupervised learning.

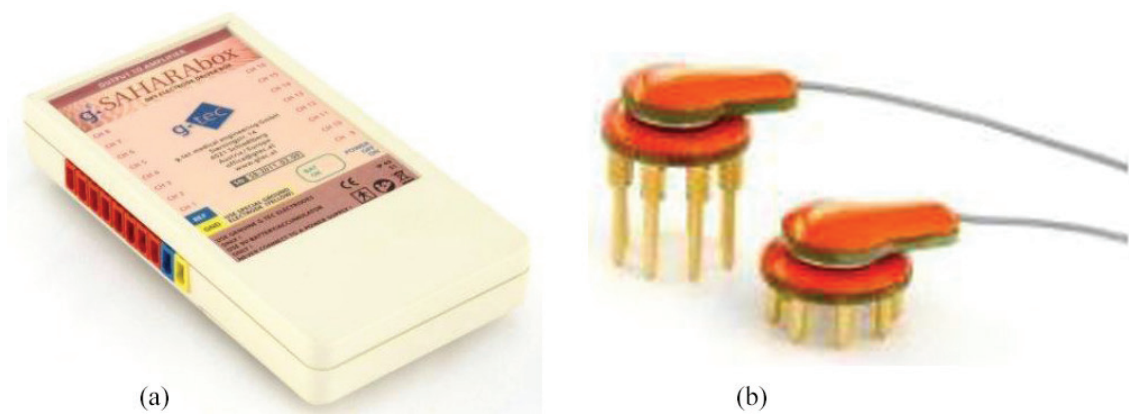
Recurrent neural networks (RNN), one of the models in DL, have proved promising results in many field [12–15] recently, especially when input and/or output are of variable length. In the application of EEG signals classification, Petrosian et al. [16] first applied RNN and wavelet transform to classify EEG signals. RNN is not satisfied in scalp EEG owing to the scalp EEG containing interference resulted from external noises. Besides, the input of the RNN does not have a special signal preprocessing, the RNN network has some problems such as gradient explosion and gradient vanish. Fully using characteristics in time-frequency features of signals, RNN with LSTM [17], have recently emerged as an effective deep learning model in a wide variety of applications that involve sequential data. The LSTM-based RNN can not only solve the problems in RNN but also store the long time information. In 2016, Li et al. [18] proposed an LSTM-based RNN integrated with DWT to classify the EEG signals. The LSTM is designed to fight against vanishing gradients through a gating mechanism. Gated recurrent neural network (GRNN), proposed by Cho et al. [19] in 2014, makes each recurrent unit to capture variable-length sequences adaptively. Similar scheme of the LSTM unit, GRNN has gating units that modulate the flow of information inside the unit, but without having a separate memory cell. In GRNN, the parameters at each level are shared through the whole network.

In this chapter, LSTM and GRNN combined with the DWT to classify the EEG signals were proposed. The average power spectrum of MI-EEG signals was calculated and the effective time segment was also determined. Then, DWT is applied to each channel of MI-EEG to extract the effective time-frequency characteristics. Finally, LSTM and GRNN were used as classifiers to recognize the MI-EEG signals. The experimental results showed that GRNN and LSTM methods can make full use of the time-frequency information of MI-EEG, as well as time sequence information, and can get better recognition performance.

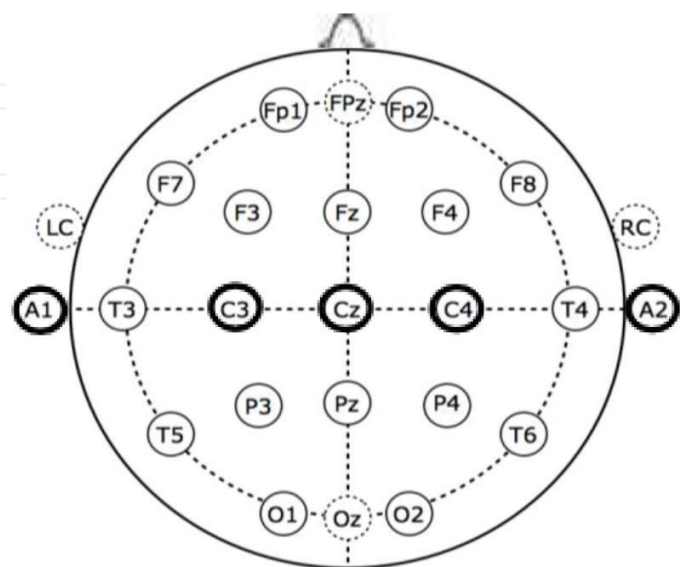
The rest of this chapter is organized as follows: Section 2 describes the system architecture; wavelet transform is described in Section 3; Section 4 presents the LSTM-based recurrent network; the GRNN is discussed in Section 5; Section 6 shows the experimental results; the application to control an electric wheelchair is shown in Section 7; and finally, the discussions are given in Section 8.

## 2. System architecture

The proposed BCI system is integrated as EEG signals extracting subsystem through the Emotiv EPOC chip, g.SAHARAbbox system, and g.SAHARA electrodes. The g.SAHARAbbox system and g.SAHARA electrodes are shown in **Figure 1**. The system’s electrodes are dry manner and nonintrusive conductive system that allows 16 EEG channels to be embedded into the input of EPOC chip at the same time. The electrode locations C3, C4, and Cz based on the international 10–20 system, shown in **Figure 2**, were used to extract EEG signals, while locations A1 and A2 were used as reference points. For the MI-EEG signals, two motion-imagination brain signals were recognized, respectively. One is “imagining right-hand action” and the other is “imagining left-hand action.” In order to establish a sampling model, we captured 9-s EEG signals for every imagining action from every channel. And, the extracted brainwave signal is transformed through DWT to obtain the spectrums in frequency domain. Then, the frequency feature was calculated and classified into different categories by using LSTM and GRNN.



**Figure 1.** The subsystems in the proposed BCI: (a) g.SAHARAbbox system and (b) g.SAHARA electrodes.



**Figure 2.** Locations C3, C4, and Cz are used in the 10–20 system.



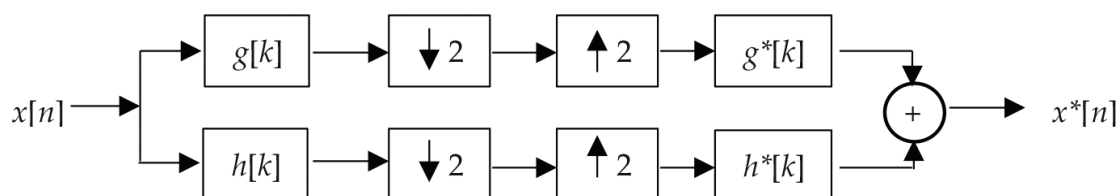
In order to speed up the processing of DWT and update the classification performance in the deep learning algorithms, the NVIDIA Jetson TK1 is used in the proposed system. In the platform, NVIDIA Tegra K1 SoC is embedded with a super computing core NVIDIA Kepler. So that it is a high-speed computing system for rapid development and deployment in computer vision, robotics, medical applications, and more. Additionally, an FPGA module named Xilinx Virtex4 XC4VFX12 is also applied to control external system such as electric wheelchair.

### 3. Discrete wavelet transform

The concept of wavelet was proposed by Jean Morlet in 1981. In this chapter, The Daubechies wavelet, proposed by Dr Daubechies in 1988 [20], was used to extract the features from EEG signals. It is often used in signal compression, digital signal analysis and noise filtering, and so on. In Daubechies wavelet, several series db wavelets can get better performance in signal analysis. In this chapter, db4 wavelets were used to extract main features from EEG signals. Multiresolution analysis in the WT algorithm was proposed by Mallat [21] in 1989. When a signal resolution has a high-degree variation in a proper area, it is difficult to get detailed features while the multiresolution strategy can decompose the lower layer signal to get more information. Therefore, the decomposed low-frequency signal can be decomposed continuously to display more features. However, the decomposed iterations of the signal are so many to make the number of samples so few that results in less obvious characteristics of the signal.

Therefore, the number of signal decomposition layer is limited. In the wavelet decomposition, the original signal is input to a low-pass filter  $g[k]$  and a high-pass filter  $h[k]$ , respectively. The low-pass filter retains the consistency of the original signal, and the high-pass filter reserves the variability of the original data. Discrete wavelet transform can be combined with wavelet function and scale function. In the low-frequency part, it has a high frequency resolution and low temporal resolution, while there was a lower frequency resolution and a higher time resolution in the high-frequency part. The discrete wavelet transform decomposition and recombination is shown in **Figure 3** and the multiresolution analysis in the WT is shown in **Figure 4**.

The left half is wavelet decomposition, after the high-pass and low-pass decomposition and then downsampling to get two groups of detailed signal and the approximate signal. The right half in **Figure 3**, the decomposition of the series for the rise of sampling, and then through the high-frequency synthesis filter and low-frequency synthesis filter can be reconstructed.



**Figure 3.** Discrete wavelet decomposition and reconstruction.

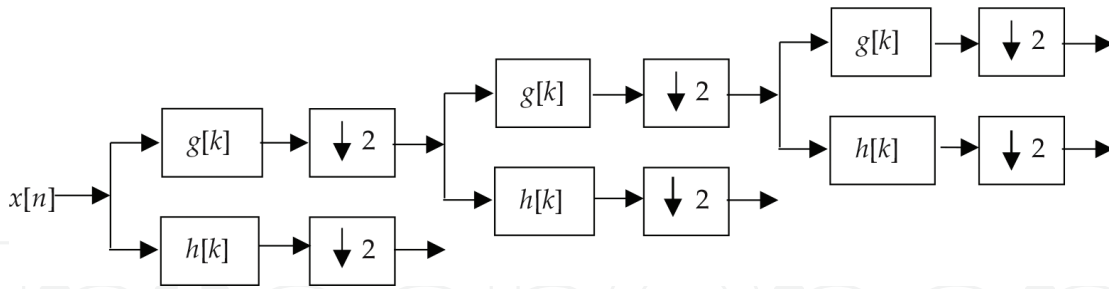


Figure 4. Discrete wavelet multiresolution decomposition.

#### 4. LSTM-based recurrent network

RNNs are popular networks that have shown great promise in many sequential tasks. RNNs are called recurrent because they perform the same task for every element of a sequence, with the output being depended on the previous states. Recently, several researchers have developed more sophisticated types of RNNs to deal with some of the shortcomings of the vanilla RNN model. Training an RNN is similar to training a traditional neural network (TNN). Because RNNs trained by TNN's style have difficulties in learning long-term dependencies due to the vanishing and exploding gradient problem. LSTMs do not have a fundamentally different architecture from RNNs, but they use a different function to calculate the states in hidden layer. The memory in LSTMs is called cells and can be thought as black boxes that take as input the previous state and current input. Internally, these cells decide what to be kept in (and what to be erased from) memory. They then combine the previous state, the current memory, and the input. It turns out that these types of units are very efficient at capturing long-term dependencies. In this chapter, a peephole-connection LSTM, proposed by Gers and Schmidhuber [22], is applied and shown in **Figure 5**. In **Figure 5**, the state of forget gate  $f_t$ , shown as in Eq. (1), is decided by a sigmoid function from the previous cell state  $C_{t-1}$ , the previous hidden layer state  $h_{t-1}$  and input data  $x_t$ .

$$f_t = \sigma(w_{cf}C_{t-1} + w_{xf}x_t + w_{hf}h_{t-1}) + b_f \quad (1)$$

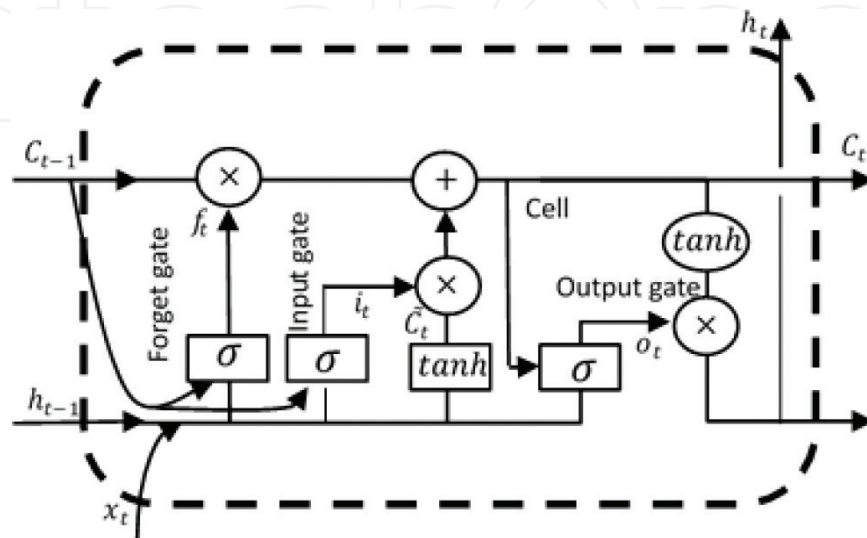
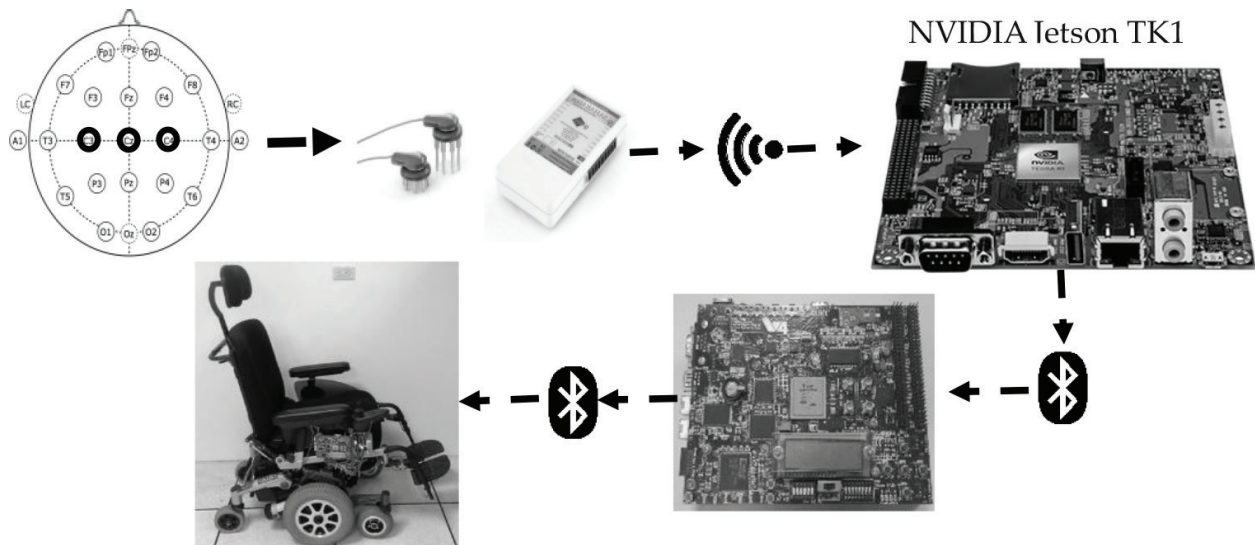


Figure 5. The block diagram of LSTM.



**Figure 6.** The proposed BCI control system.

From **Figure 6**, we can find the cell state shown as Eq. (2), calculated with the previous cell state  $C_{t-1}$ , forget-gate state  $f_t$ , and  $i_t \cdot \tilde{C}_t$ .

$$C_t = f_t * C_{t-1} + i_t \cdot \tilde{C}_t \quad (2)$$

where

$$i_t = \sigma(W_{c,i} * C_{t-1} + w_{x,i} * x_t + w_{h,i} * h_{t-1} + b_i) \quad (3)$$

and

$$\tilde{C}_t = \tanh(w_{x,c} * x_t + w_{h,c} * h_{t-1} + b_c) \quad (4)$$

Finally, the output-gate state  $O_t$  and hidden-layer state  $h_t$  are computed by Eq. (5) and Eq. (6), respectively.

$$o_t = \sigma(w_{c,o} * C_t + w_{x,o} * x_t + w_{h,o} * h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \tanh(C_t) \quad (6)$$

## 5. Gated recurrent neural network (GRNN)

The GRNN was proposed by Cho et al. [19] in order to make each recurrent unit to extract dependencies of different timescales adaptively. The GRNN, shown in **Figure 7**, has gating units that modulate the flow of information inside the unit like the LSTM unit but without having a separate memory cell. The parameters in the GRNN are updated as follows:



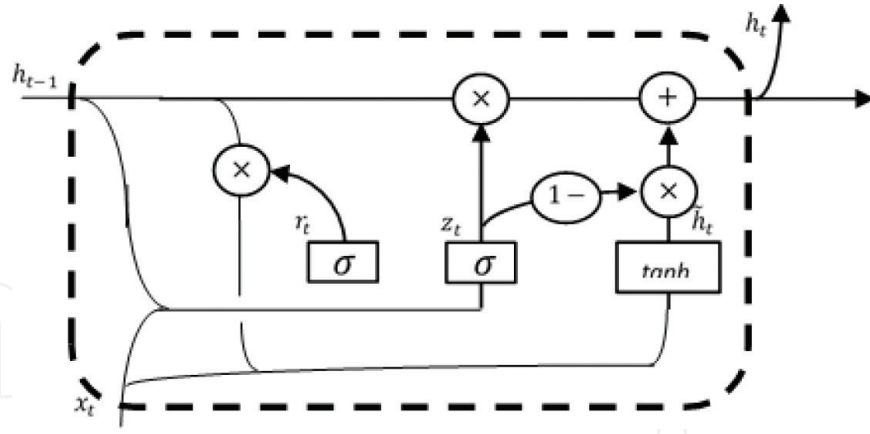


Figure 7. The block diagram of GRNN.

$$z_t = \sigma(w_{x,z} * x_t + w_{h,z} * h_{t-1} + b_z) \quad (7)$$

$$r_t = \sigma(w_{x,r} * x_t + w_{h,r} * h_{t-1} + b_r) \quad (8)$$

$$\tilde{h}_t = \tanh(w_{x,h} * x_t + w_{h,r} * (r_t * h_{t-1}) + b_h) \quad (9)$$

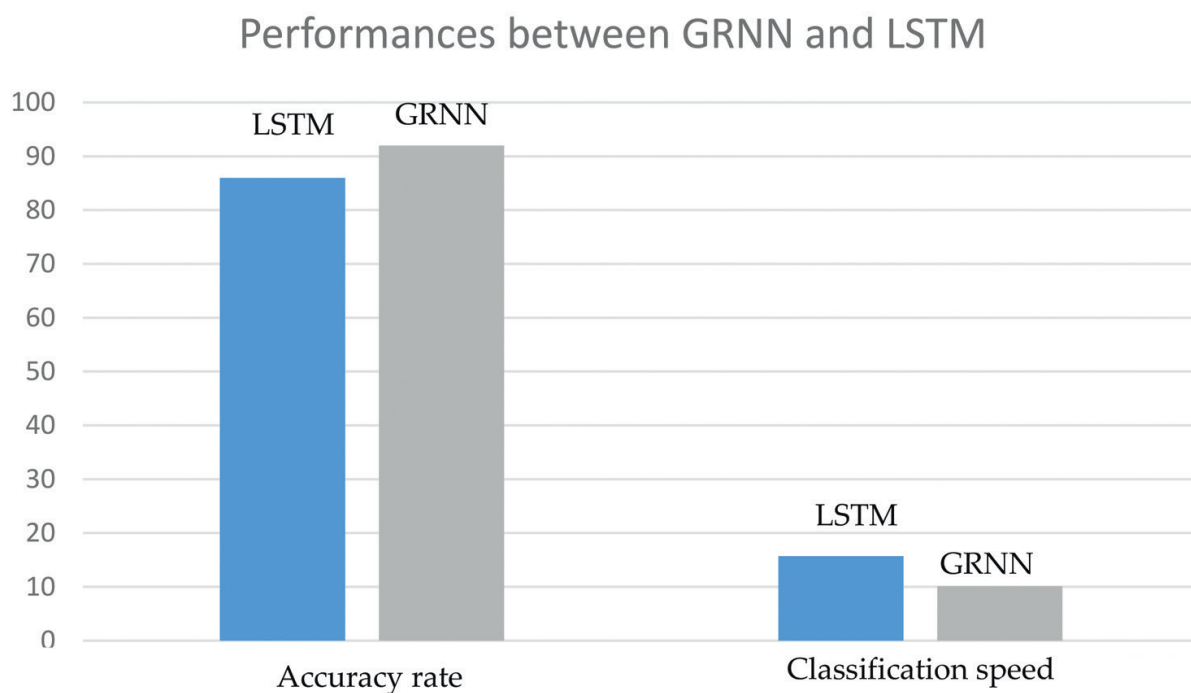
$$h_t = z_t * h_{t-1} + (1 - z_t) * \tilde{h}_t \quad (10)$$

where  $x_t$  is the input vector,  $h_t$  is the output vector in hidden layer,  $z_t$  is the vector of update gate, and  $r_t$  is the vector of reset gate, respectively.

## 6. Experimental results

In this chapter, C3, Cz, and C4 are used to capture brainwave signals. Each subject wore an Ultracortex helmet connected with g.tec dry electrode and Emotiv EPOC chip to record MI-EEG signals including to imagine right-hand and left-hand movements. Each imaginary action was consumed 9 s for a data set. The EEG signals were extracted 28 times and transformed by wavelet transform to obtain their features. Therefore, we can obtain 140 sets for 5 subjects and these data sets were divided into 112 groups for training and 28 groups for testing. The experimental data acquisition process is down to obtain a data set every 9 s with an interval of 2 min. The waiting time is set on the first 2 s, then a stimulus signal was sound indicating that the testing process is started and a cross sign “+” is displayed for 1 s. Then, the left or right arrow is displayed to hint a subject imagining the moving of left or right hand. The sampling rate is 128 Hz for the acquisition process.

In this chapter, LSTM and GRNN are used as the EEG classifiers. MI-EEG features were extracted for C3, Cz, and C4 and classified into two groups. Therefore, the neurons of input and output layers of LSTM and GRNN were set three and two, respectively. In order to obtain



**Figure 8.** The performance competition between GRNN and LSTM.

Authors	Features	Classifiers	Accuracy rates
Christin Schafer [24]	Wavelet	Bayes	89.29%
GAO Xiaorong [24]	ERD	LDA	86.43%
Akash Narayana [24]	AR	LDA	84.29%
The proposed LSTM	DWT	LSTM	92.83%
The proposed GRNN	DWT	GRNN	94.50%

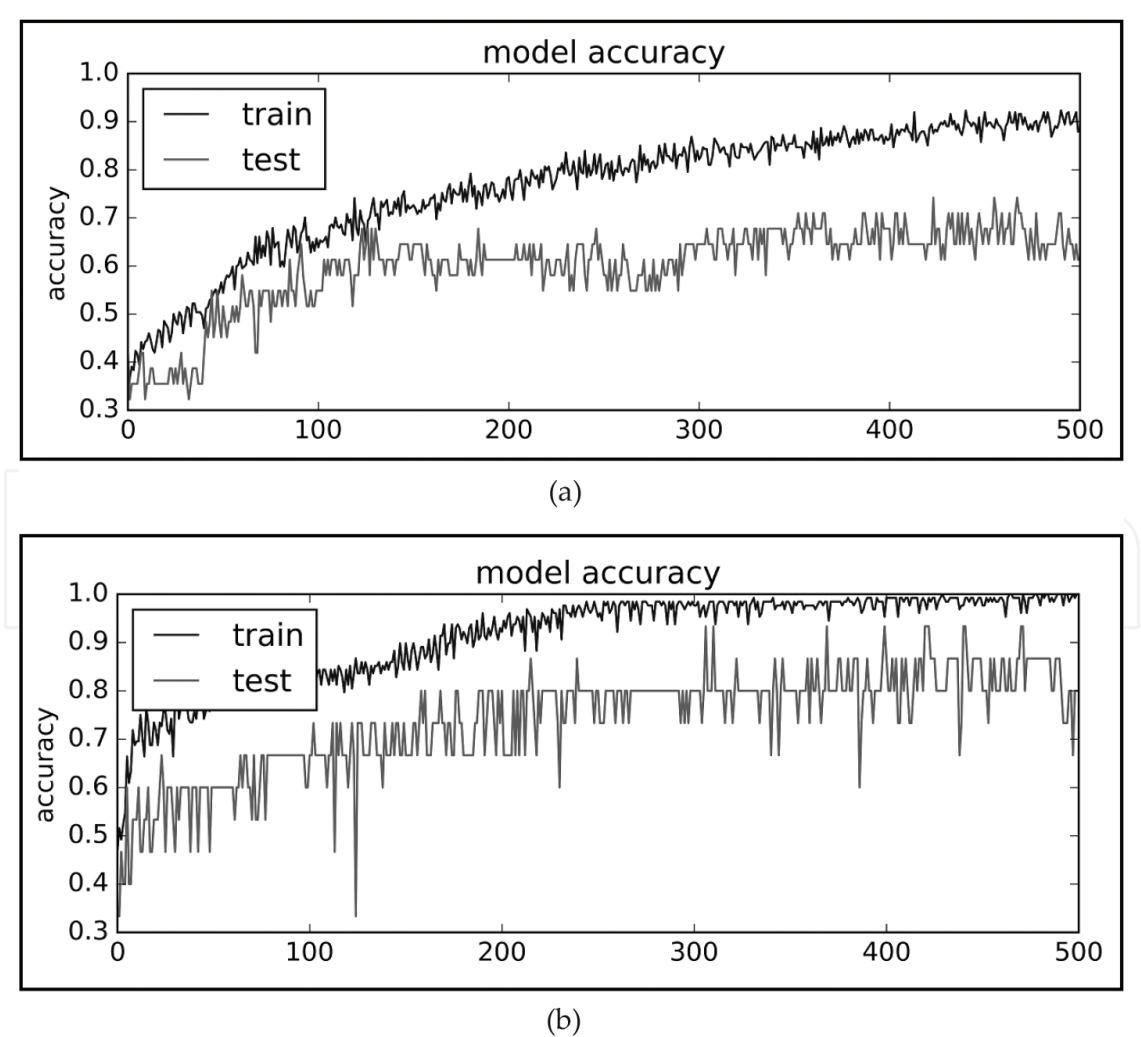
**Table 1.** The accuracy rates of different strategies for BCI Competition 2003.

better performance for classification, the hidden layer is set into 7 neurons, and therefore, we can obtain the length of MI-EEG characteristic sequence being 15, while the channel number of MI-EEG-based BCI is 3. In order to evaluate the classification results and obtain a reliable and stable model, this model performs 500 cross validation to calculate the classification accuracy. In 2009, Smith [23] indicated that the nervous system is significantly important to integration of information and to the range of behaviors in which the system can stably engage and among which the system can flexibly switch. However, the nervous system, the body, and the environment each possess their own complex intrinsic dynamics, and these are always in continuous interaction with each other. Human intelligence reveals both remarkable stability and nimble flexibility. Stability emerges from the incorporation of the past into the present. Flexibility, requires an abandonment of (or selection among) past ways, a shifting of responses to meet new circumstances. For the consideration of stability and flexibility, the proposed methods are compared to other strategies based on “BCI Competition 2003” [24]. The experimental results

are shown in **Table 1**. From **Table 1**, we can find that the proposed method can get better performance than others. Additionally, the GRNN is better than the LSTM with 2.67% and 5 ms in the performances of accuracy and classification speed that is shown in **Figure 8**.

7. Applications to control an electric wheelchair

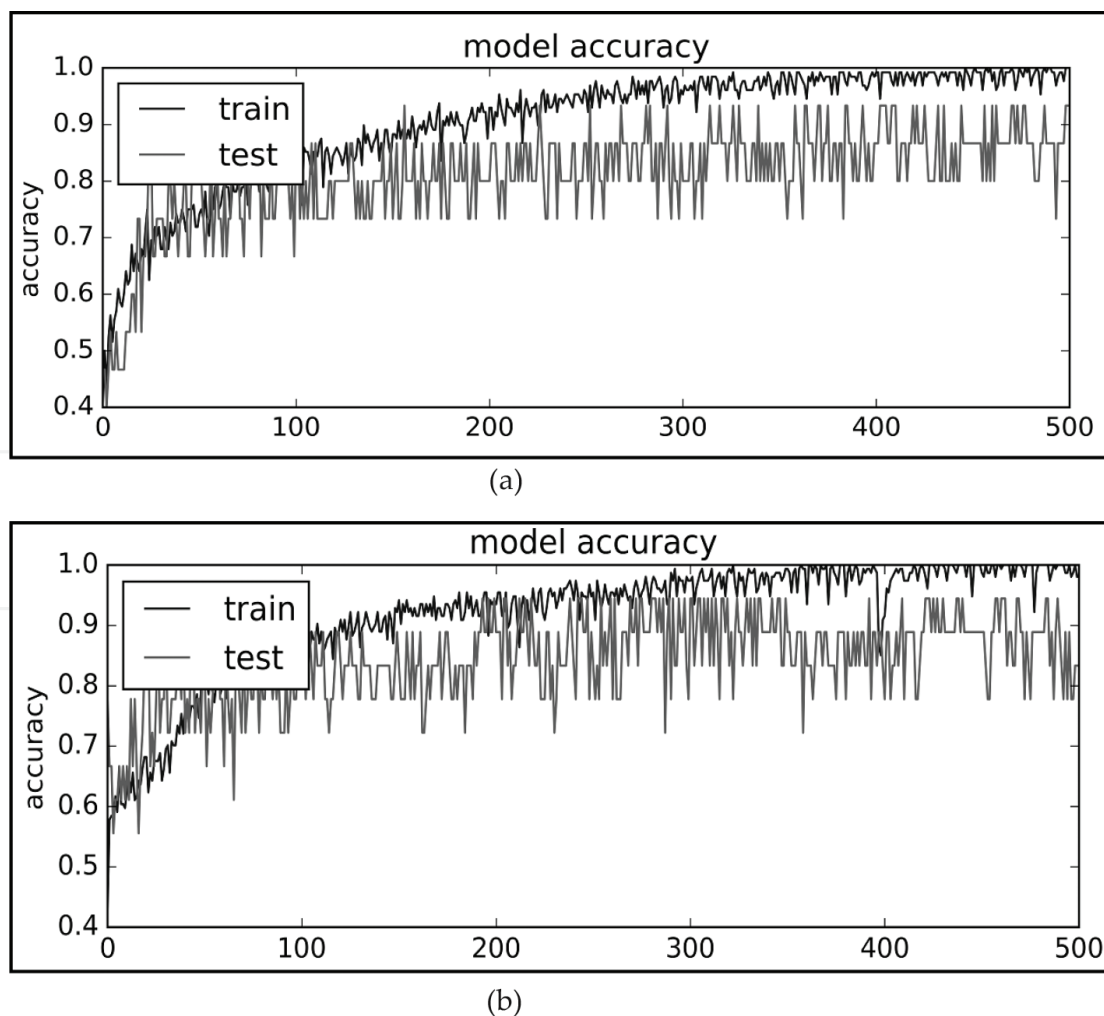
In this section, the proposed BCI system was applied to control an electric wheelchair. During the online experiment, each subject wore the EEG acquisition system with integrated g.SAHARAsys and EPOC chip in the proposed BCI system. Additionally, the EEG signal for eye blinking was added in order to easily control an electric wheelchair to go ahead or emergency stop. For MI-EEG signals, imagining left hand and right hand are translated into turning wheelchair left and right as well as the eye blinking signal is converted into going ahead/emergency stopping. For the purpose of speeding up the extraction and processing EEG signals, the sapling interval was adjusted to 1 s. But these modifications result in losing a few features. Therefore, the db4 wavelet is adjusted to two levels as well as additional one layer is added into hidden layer of LSTM and GRNN networks.



**Figure 9.** The accuracy rates in LSTM and GRNN with db4 wavelets and seven hidden layers. (a) LSTM. (b) GRNN.

Increasing the level number of DWT can directly reduce length of the EEG signals. If the db4 DWT is still used, the extracted signals will lose some features. Thus, reducing the DWT levels can retain more features in the original EEG signals. Increasing the number of hidden layers is due to the increased complexity of the input EEG signals. The more hidden layers are conducive to processing the data with higher complexity. However, too many hidden layers will cause the network to be difficult to converge during the learning process. In this section, additional one layer is added into hidden layer for obtaining better convergence properties. The classification accuracy rates for db4 wavelets by LSTM and GRNN networks with seven layers in hidden layer are shown in **Figure 9**, while the classification accuracy rates for db2 wavelets by LSTM and GRNN networks with eight layers in hidden layer are shown in **Figure 10**. From **Figures 9** and **10**, we can find that the accuracy rates of test data are obviously increased and nearby the accuracy rates of training data for both LSTM and GRNN networks.

Then, two BCI systems have respectively embedded LSTM and GRNN with db2 wavelets and eight hidden layers are applied to control an electric wheelchair. They can smoothly control an electric wheelchair and the GRNN model can always get better performance than the LSTM.



**Figure 10.** The accuracy rates in LSTM and GRNN with db2 wavelets and eight hidden layers. (a) LSTM. (b) GRNN.

## 8. Conclusions and future prospects

In this chapter, two deep-learning models named LSTM and GRNN were applied to be embedded into a BCI system for MI-EEG signal classification to identify two imagery movements such as imagining right-hand and left-hand actions. In the proposed BCI system, the Emotiv EPOC IC with tg.SAHARAbbox system and g.SAHARA electrodes are used to capture MI-EEG signals on C3, Cz, and C4. In this chapter, we use the Daubechies wavelet to get feature values on db4 and db2 coefficients. The GRNN can make each recurrent unit to capture variable-length sequences adaptively. Modified from LSTM, the GRNN has gating units that modulate the flow of information inside the unit, but without having a separate memory cell. In the GRNN, the parameters at each level are shared through the whole network. From the experimental results, the GRNN can get better performance than other strategies. Additionally, the GRNN can always obtain better performance than the LSTM in the application to control an electric wheelchair.

## Acknowledgements

In this chapter, the research was sponsored by the Ministry of Science and Technology of Taiwan under the Grant 106-2221-E-167-031.

## Author details

Jzau-Sheng Lin \* and Ray Shihb

\*Address all correspondence to: jslin@ncut.edu.tw

Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taiwan

## References

- [1] Decety J, Ingvar DH. Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation. *Acta Psychologica*. 1990;**73**:13-34
- [2] Tomida N, Tanaka T, Ono S, Yamagishi M. Active data selection for motor imagery EEG classification. *IEEE Transactions on Biomedical Engineering*. 2015;**62**:458-467
- [3] Baali H, Khorshidtalab A, Mesbah M, Salami MJE. Transform-based feature extraction approach for motor imagery tasks classification. *Rehabilitation Devices and Systems*. 2015;**3**:1-8. DOI 10.1109/JTEHM.2015.2485261



- [4] Wu S-L, Liu Y-T, Hsieh T-Y, Lin Y-Y, Chen C-Y, Chuang C-H, Lin C-T. Fuzzy integral with particle swarm optimization for a motor-imagery-based brain computer interface. *IEEE Transactions on Fuzzy Systems*. 2016;**2016**. DOI: 10.1109/TFUZZ.2016.2598362
- [5] Lin J-S, Lo C-H. Mental commands recognition on motor imagery-based brain computer Interface. *International Journal of Computing, Consumer and Control*. 2016;**25**:18-25
- [6] Chatterjee R, Bandyopadhyay T. EEG based motor imagery classification using SVM and MLP. In: *Proceeding of International Conference on Computational Intelligence and Networks*; 2016. pp. 84-89
- [7] Xie X, Yu ZL, Lu H, Gu Z, Li Y. Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2016;**25**:504-516. DOI: 10.1109/TNSRE.2016.2587939
- [8] Chatterjeel R, Bandyopadhyal T, Sanyal DK. Effects of wavelets on quality of features in motor imagery EEG signal classification. In: *Proceedings of IEEE WiSPNET 2016 Conference*; 2016. pp. 1346-1350
- [9] Jois K, Garg R, Singh V, Darji A. Comparative analysis of classification techniques for motor imagery based BCI. *IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*; 2015. DOI: 10.1109/WCI.2015.7495507
- [10] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;**313**(5786):504-507
- [11] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013;**35**:1798-1828
- [12] Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. *Studies in Computational Intelligence*. Cham Switzerland: Springer; 2012
- [13] Graves A, Mohamed A-R, Hinton G. Speech recognition with deep recurrent neural networks. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*; May, 2013
- [14] Li N, Chen J, Cao H, Zhang B, Natarajan P. Applications of recurrent neural network language model in offline handwriting recognition and word spotting. In: *Proceeding of 14th International Conference on Frontiers in Handwriting Recognition*; 2014. pp. 134-139
- [15] Moghadam SM, Seyyedsalehi SA. Nonlinear analysis of video images using deep recurrent auto-associative neural networks for facial understanding. In: *Proceedings of 3rd International Conference on Pattern Recognition and Image Analysis*; 2017. pp. 20-25
- [16] Petrosian A, Prokhorov D, Homan R, Dasheiff R, Wunsch DII. Recurrent neural network based prediction of epileptic seizures in intra-and extracranial EEG. *Neurocomputing*. 2000;**30**(1):201-218
- [17] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*. 1997;**9**: 1735-1780

- [18] Li M, Zhang M, Luo X, Yang J. Combined long short-term memory based network employing wavelet coefficients for MI-EEG recognition. In: Proceedings of IEEE International Conference on Mechatronics and Automation; Aug., Harbin, China; 2016. pp. 1971-1976
- [19] Cho K, van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. arXiv Preprint. 2014;arXiv:1409.1259
- [20] Daubechies I. Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics. 1988;41:909-996
- [21] Mallat S. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989;11:674-693
- [22] Gers FA, Schmidhuber J. Recurrent nets that time and count. In: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks; 2000. pp. 189-194
- [23] Smith L. Dynamic Systems, Sensorimotor Processes, and the Origins of Stability and Flexibility. In: Spencer J, Thomas MSC, McClelland JL, editors. Toward a Unified Theory of Development. Oxford: Oxford University Press; 2009
- [24] Blankertz B, Müller KR, Curio G. The BCI Competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. IEEE Transactions on Biomedical Engineering. 2004;51:1044-1051