Hybrid Convolutional Recurrent Neural Networks Outperform CNN and RNN in Task-state EEG Detection for Parkinson's Disease

Xinjie Shi*†, Tianqi Wang†, Lan Wang†, Hanjun Liu‡ and Nan Yan†
* School of Software Engineering, University of Science and Technology of China, Hefei, Anhui, China
† CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China
‡ Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China
E-mail: lhanjun@mail.sysu.edu.cn, nan.yan@siat.ac.cn

***Abstract*—In hospitals, brain-related disorders such as Parkinson's disease (PD) could be diagnosed by analyzing electroencephalograms (EEG). However, conventional EEG-based diagnosis for PD relies on handcrafted feature extraction, which is laborious and time-consuming. With the emergence of deep learning, automated analysis of EEG signals can be realized by exploring the inherent information in data, and outputting the results of classification from the hidden layer. In the present study, four deep learning algorithm architectures, including two convention deep learning models (convolutional neural network, CNN; and recurrent neural network, RNN) and two hybrid convolutional recurrent neural networks (2D-CNN-RNN and 3D-CNN-RNN), were designed to detect PD based on task-state EEG signals. Our results showed that the hybrid models outperformed conventional ones (fivefold average accuracy: 3D-CNN-RNN 82.89%, 2D-CNN-RNN 81.13%, CNN 80.89%, and RNN 76.00%) as they combine the strong modeling power of CNN in temporal feature extraction, and the advantage of RNN in processing sequential information. This study represents the an attempt to use hybrid convolutional recurrent neural networks in classifying PD and normal take-state EEG signals, which carries important implications to the clinical practice.**

***Keywords*-deep learning; EEG; classification; Parkinson's disease**

I. INTRODUCTION

Parkinson's disease (PD) is a common neurodegenerative disorder [1], affecting more than 1% of the population over 50 years old with increasing prevalence year by year [2]. Typical symptoms associated with PD include tremor, muscle rigidity, postural instability, bradykinesia (slowness of movement), and dysphonia (voice disorders) [3], due to the degeneration of dopaminergic neurons in the substantia nigra pars compacta of the basal ganglia. These clinical motor symptoms, together with non-motor symptoms, are usually the basis of conventional diagnosis. With recent advances in neuroscience, new tools such as electroencephalography (EEG), become potentially applicable for the detection of PD.

EEG-based diagnosis is easy to set up, non-invasive in nature, and provides high temporal resolution. However, traditional diagnosis is based on manual analysis, which precludes its application to a large volume of data. Fortunately, advances in classification renders the automated analysis of EEG signals realizable. Yuvaraj and colleagues [4] extracted higher-order spectra (HOS) features from resting-state EEG signals, which were fed to various classifiers, including decision tree (DT), fuzzy K-nearest neighbor (FKNN), K-nearest neighbor (KNN), naïve Bayes (NB), probabilistic neural network (PNN), and support vector machine (SVM). Despite obtaining an optimal mean accuracy of 99.62% to differentiate PD patients from their healthy counterparts, handcrafted HOS features are time-consuming to extract, and requires highly trained clinicians or neurologists. A more devastating drawback is the loss of spatial and temporal information during feature extraction.

A step forward towards the automated analysis of EEG signals could be achieved by unsupervised feature learning based on the deep learning model. Wen and Zhang [5] constructed the deep convolution network and autoencoders-based model (AE-CDNN) to extract feature representations from unlabeled EEG in epilepsy patients and healthy controls. Features were also obtained by principal component analysis and sparse random projection. Different common classifiers, including KNN, SVM, DT, NB, random forest (RF), multilayer perceptron (MLP), and AdaBoost algorithm (ADB), were applied to verify the effectiveness of features extracted. The results showed that the classification accuracies based on unsupervised feature learning from AE-CDNN were the highest, and were not inferior to the results of other studies.

Compared with conventional classifiers, deep neural networks have strong modeling power to explore the inherent information in data, and can be directly applied for classification. Because of that, researchers have attempted to use deep learning algorithms for automated analysis of EEG data. Schirrmeister et al. [6] evaluated convolutional neural network (CNN) of different architectures with different design choices against a widely used baseline method, filter bank common spatial patterns (FBCSP) which revealed that CNN could achieve at least as good performance as FBCSP (mean decoding accuracies FBCSP 82.1%, CNN 84.0%). The

same algorithm architecture was also applied to distinguish pathological from normal EEG recordings in the Temple University Hospital (TUH) EEG Abnormal Corpus [7], which reached substantially better accuracies than best reported results for this dataset (about 6% better, 85% vs. 79%). It is worth mentioning that both studies visualized the learned features, demonstrating that CNN indeed learned to use spectral power changes in different frequency bands for the decoding decision. Roy et al. [8], on the other hand, evaluated the performance of a number of algorithms, including 1D-CNN, 2D-CNN, deep 1D convolutional gated recurrent neural network (1D-CNN-RNN), and time-distributed convolutional recurrent neural network (TCNN-RNN) on different representation input (time-series, spectrogram, and Gramian Angular Fields; GAF) for the same dataset. As reported, the best performance was obtained by directly feeding the time-series data to 1D-CNN-RNN, achieving a 3.47% increase compared to previously reported accuracies. Besides, a hybrid convolutional recurrent network architecture has been applied to a brain computer interface (BCI) [9], in which the input EEG signal was fed to RNN and CNN structure for temporal and spatial feature learning in parallel. The learned features were then stacked for feature transformation, and were input to the classifier. The hybrid architecture yielded an accuracy of 95.53%, which is significantly higher than any other state-of-the-art methods.

Deep neural networks have been employed to the detect PD automatically (e.g., Oh et al., [10] employed a thirteen-layer CNN architecture). However, most studies used EEG signals collected in resting state. Given that PD is characterized by the gradual degradation of motor function, and the loss of dopaminergic neurons in basal ganglia will compromise its control over speech. Considerable studies, including one conducted by our lab [11], have found that patients with PD are associated with deficits in auditory-motor integration for vocal pitch regulation. Specifically, patients with PD exhibited abnormal compensation (i.e., larger vocal response peak) when they heard their vocal pitch unexpectedly shifted upward or downward. At the cortical level, larger P2 responses during pitch perturbation were induced by enhanced activity in the superior and inferior frontal gyrus, premotor cortex, inferior parietal lobule, and superior temporal gyrus. Since there are significant differences in the cortical responses between PD patients and healthy controls, we speculate that implementing deep neural networks to the task-state EEG data would better differentiate pathological from normal population.

In the present study, four deep learning algorithm architectures were designed to tackle the task of EEG detection for PD. The first two models are single CNN and RNN, which are compared to the hybrid convolutional recurrent neural network. Combining advantages of both CNN and RNN, we hypothesize that the hybrid model will yield higher accuracies for classification.

II. Deep Learning Architectures

Four deep learning architectures were designed in this study to tackle the task of EEG decoding. In this section, we first explain the basic ideas of CNN and RNN, followed by a description of the four models used in this study. It is worth noting that all the four models used raw data as input, and the first one (i.e., CNN [6]) served as a baseline model. At the end of this section, we describe how to represent the EEG input for these models, as well as the training strategy.

A. Convolutional Neural Network

Three basic layers are stacked together to build a CNN: convolution layer, pooling layer, and fully connected (dense) layer [12, 13]. First of all, the input signal is connected to the convolution layer to perform convolution operation using a kernel (window) [14]. Results of the operation are generated as a feature map for the next layer. Between two convolution layers, a pooling layer is used to reduce the size of the feature map, and hence enables faster computation. Every neuron of the pooling layer is connected to every neuron in the fully connected layer, where high-level features were used to classify the input signal into various classes [15, 16].

B. Recurrent Neural Network

Compared with conventional feedforward network architectures, RNN has inherently strong modeling power to learn sequential information, as neurons send feedback signals to the other neurons in the same hidden layer (i.e., the input of the hidden layer considers the output from the preceding time steps), which provides RNN with the memory from the states in the history [17]. Our model takes the time series input signals from the previous temporal information using gated recurrent unit (GRU) [18].

C. Deep Neural Network Architecture

Table I shows the internal details of all layers in each architecture.

- **CNN:** Following [6], the first architecture was a CNN with convolution-max-pooling structure, in which the first two convolutional layers were designed to handle EEG input, followed by two pooling-convolutional layers, two fully connected layers and a softmax classification layer (Fig. 1A). Specifically, the kernel in the first layer performs a convolution over time, and the kernel in the second layer performs a spatial filtering over electrodes.
- **RNN:** Our study [11] has revealed that PD patients elicit larger P2 responses during pitch perturbation. To classify input signals on the basis of temporal information, RNN was designed by splitting the raw data to 700 steps, with each step containing 64 units (i.e., step length). The outputs of the last 20 steps' states from the RNN module were fed to fully connected layers, and then softmax layer for classification (Fig. 1B).
- **2D-CNN-RNN:** We took advantage of the CNN structure for spatial information extraction and employed the RNN structure for its strong modeling

power to explore temporal relevance in time-series data by designing a hybrid convolutional recurrent neural network. The CNN structure includes one layer, in which the kernel performs a spatial filtering over electrodes. The output feature map was then fed to the RNN structure, which includes 350 steps, with each step containing 40 units (i.e., step length). The outputs of the last 50 steps' states from the RNN structure were fed to fully connected layers, and finally softmax classification layer (Fig. 1C).

- **3D-CNN-RNN:** To better process spatial information, a more sophisticated hybrid convolutional recurrent neural network was applied. Compared with 2D-CNN-RNN, the CNN structure was designed with two convolutional layers. In the first layer, the kernel performs a convolution over time, and in the second layer, the convolution operation is performed on the 3D feature map, which could better extract spatial feature compared with 2D-CNN-RNN. All the other structures are exactly the same as in 2D-CNN-RNN (Fig. 1D).

D. Input Representation

EEG can be represented as a time series of topographically organized images (i.e., voltage distributions across the scalp surface). Thus, a number of studies have taken the power spectra of EEG decoding as an input of CNN (e.g., [19]). The rationale of applying spatial filters to the global pattern is based on the assumption that EEG signals approximate a linear superposition of spatially global voltage patterns caused by multiple dipolar current sources in the brain [20]. In this view, however, the hierarchical compositionality of local and global EEG modulations in space could hardly be observed. Considering that EEG also preserves temporal hierarchies of local and global modulations (e.g., in nested oscillations [21-23]), the CNN can be designed to learn spatially global unmixing filters in the entrance layers, and temporal hierarchies in the deeper layers. Following [6], we represented the input as a 2-dimensional array with the number of time steps (700) as the width and the number of electrodes (64) as the height. Further advantage of this approach is that it reduces the input dimensionality compared with treating EEG as a time series of image

E. Training Strategy

Training and testing were carried out for each model. During training, fivefold cross-validation (CV) was used, in which the EEG data obtained from 40 PD patients (4000 trials; details of the data are introduced in the Section "EEG Data Acquisition and Preprocessing") and 30 healthy controls (HC; 3000 trials) were split into five uniform portions. Out of the five portions, four were used to train the model, and the rest were kept for testing. This procedure was iterated five times, producing datasets denoted CV1-CV5. Trials obtained from the same participant were involved in either training or testing phase, and would not be used in both phases. For each dataset, the designed model underwent five training rounds, and the result of each training round were used to calculate the accuracy and standard deviation (SD).

TABLE I. PARAMETERS OF THE FOUR MODELS

Model	Hyper-parameter	Value
CNN	1st convolution	Filter [1,26], stride [1,1], depth 40
	2nd convolution	Filter [64,40], stride [1,1], depth 40
	1st max-pooling	Filter [1,3], stride [1,3]
	3rd convolution	Filter [40,26], stride [1,1], depth 80
	2nd max-pooling	Filter [1,4], stride [1,4]
	4th convolution	Filter [80,11], stride [1,11], depth 100
	Fully connected	1000+500
RNN	Steps × step length	700×64
	RNN neural size	6×64
	Fully connected	1280+300
2D-CNN-RNN	1st convolution	Filter [64,2], stride [1,2], depth 40
	Steps × step length	350×40
	RNN neural size	6×40
	Fully connected	2000+300
3D-CNN-RNN	1st convolution	Filter [1,26], stride [1,1], depth 40
	2nd convolution	Filter [64,40], stride [1,1], depth 40
	1st max-pooling	Filter [1,3], stride [1,3]
	Steps × step length	225×40
	RNN neural size	6×40
	Fully connected	2000+300
Shared Parameters	Batch size	100
	Input size	64×700
	Learning rate	0.001
	Activation function	ReLU
	Classification function	Softmax
	Optimization function	Adam optimization

III. EEG DATA ACQUISITION AND PREPROCESSING

A. Participants

Forty participants with PD (15 females and 25 males; mean age = 63.53, SD = 4.95) and 30 age-matched HC (12 females and 18 males; mean age = 64.72, SD = 5.74) were recruited for the experiment. Each participant passed a bilateral screening test to verify the hearing status. Informed consent was obtained from all participants and the study was approved by the institutional board for human research of The First Affiliated Hospital at Sun Yat-sen University in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

B. Experimental Setup

The experiment was carried out in a sound-attenuated booth. Participants heard their self-voice feedback with a gain of 10 dB sound pressure level (SPL) relative to their vocal output. Participants were instructed to produce the vowel /u/ for approximately 5-6 seconds at their conversational pitch and loudness level, while listening to their voice unexpectedly pitch-shifted upwards 200 cents (100 cents = 1 semitone).

During each vocalization, 5 pitch shifts (200 ms duration) were presented with an inter-stimulus interval (ISI) of 700-900 ms and the first one occurred 500-1000 ms after the vocal onset. Participants produced 40 consecutive vocalizations, resulting in 100 +200-cent trials.

Participant's voice was picked up by a dynamic microphone (model DM2200, Takstar Inc.), amplified by a MOTU Ultralite Mk3 firewire audio interface, and pitch-shifted by the Harmonizer controlled by a MIDI program (Max/MSP, v.5.0 by Cycling 74). The pitch-shifted signals were finally amplified by an ICON NeoAmp headphone amplifier and fed back to participants through insert earphones (ER1-14 A, Etymotic Research Inc.). The transistor-transistor logic (TTL) pulses were generated to mark the onset of each pitch perturbation, and sent to the EEG recording system via a synch DIN cable. The original and pitch-shifted voice signals as well as the TTL pulses were digitized with a sampling frequency of 10 kHz by a PowerLab A/D converter (model ML880, AD Instruments), and recorded using LabChart software (v.7.0 by AD Instruments).

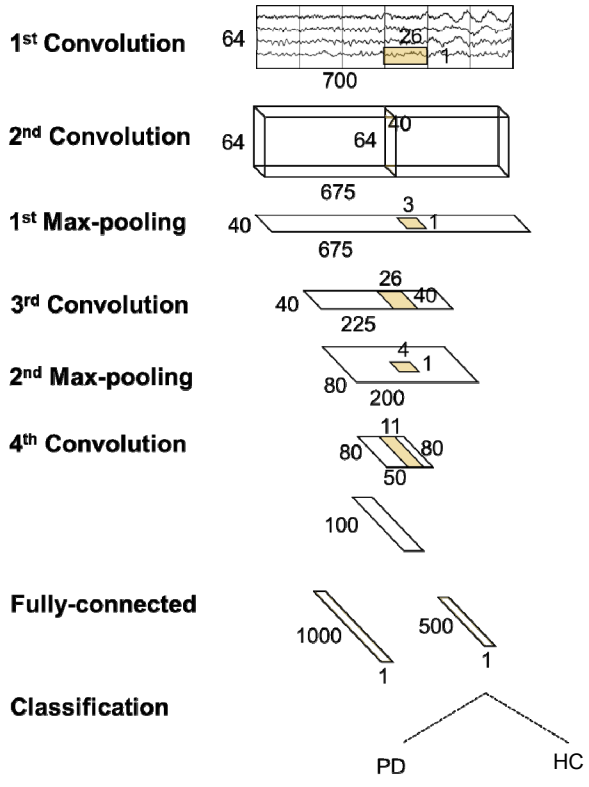
C. EEG Data Online and Offline Processing

The EEG data were recorded using a 64-electrode Geodesic Sensor Net (Electrical Geodesics Inc.) with 1 kHz sampling rate and referenced against the vertex (Cz). The signals were amplified by a Net Amps 300 amplifier (Electrical Geodesics Inc.) and recorded onto a Macintosh computer. During the online recording, impedances of individual sensors were kept below 50 kΩ.

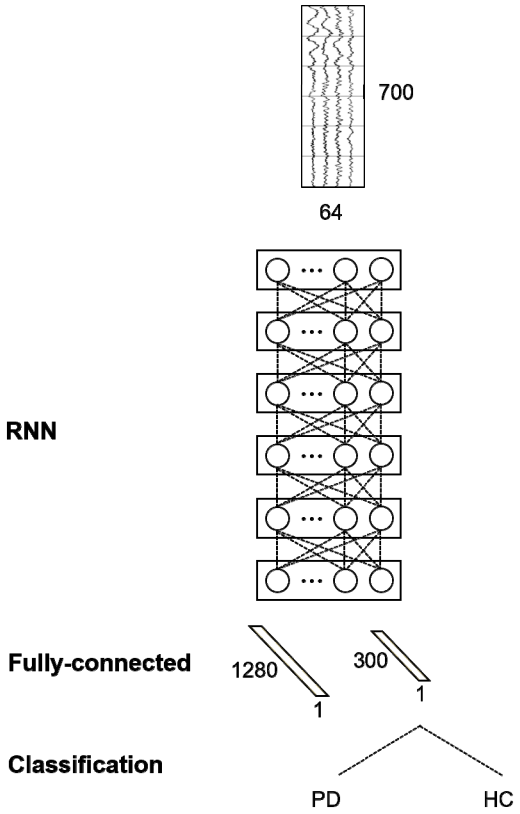
Offline signal processing was carried out using NetStation software. Raw data were band-passed filtered (1-20 Hz) and segmented with a window of -200 ms before and 500 ms after the onset of the pitch shift. Data were then re-referenced to the average of the electrodes on each mastoid, and baseline-corrected. Recorded trials with excessive muscular activity, eye blinks, or other activities beyond the range of -50 to 50 μv were rejected. We used this relatively simple method as a large number of trials were available.

IV. Results and Discussion

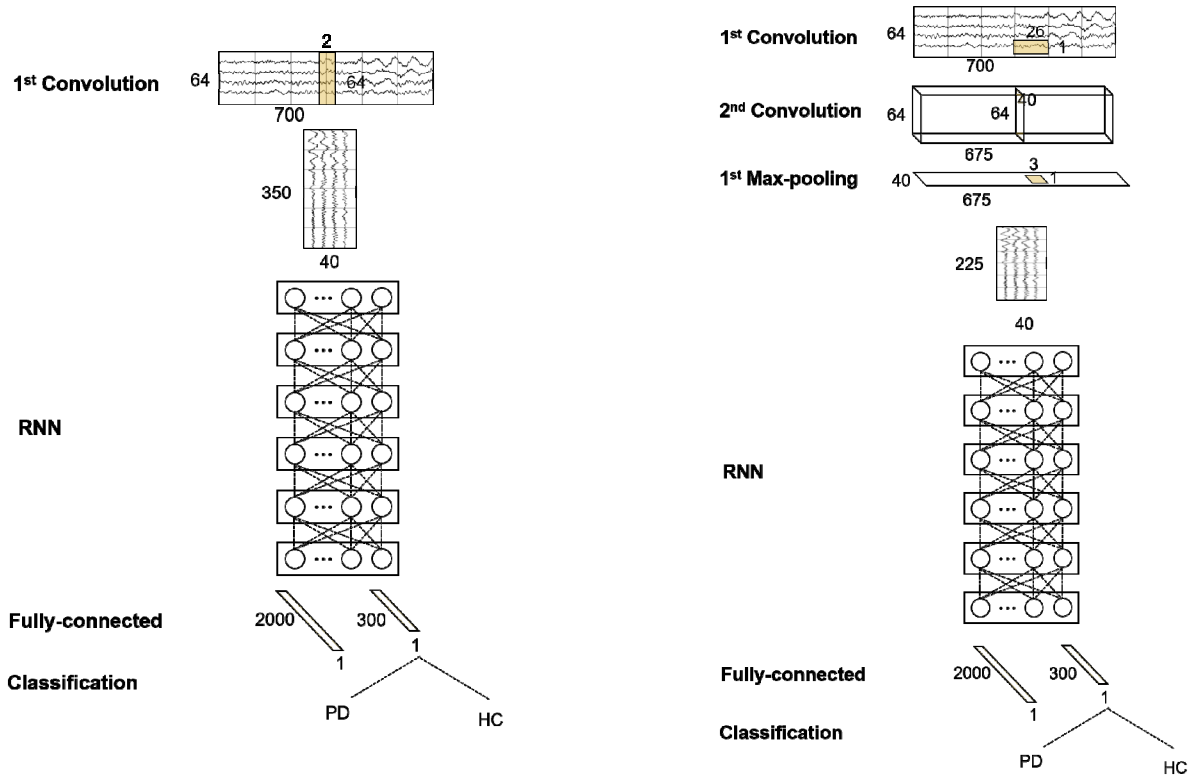
Table II shows the performance of each model (accuracy and SD) in each dataset (i.e., CV1-CV5) respectively. As can be seen, 3D-CNN-RNN yielded the highest accuracy (fivefold average 82.89%) among the four models, followed by 2D-CNN-RNN (81.13%), CNN (80.89%), and RNN (76.00%). Our results suggested that hybrid models (3D-CNN-RNN and 2D-CNN-RNN) were superior than single CNN and RNN models, as the former employed a vertical learning method, which combined the strong modeling power of CNN in temporal feature extraction, and the advantage of RNN in processing sequential information. Specifically, the more sophisticated hybrid model (3D-CNN-RNN) performed better than the one with a simpler CNN structure (2D-CNN-RNN), probably because the two convolutional layers could better learn spatial features of EEG input signals. Our RNN model, though not as good as the other models, reached an accuracy above 75%, indicating that temporal information could be used to differentiate PD from normal EEG signals. The result also provided further evidence to our previous findings in [11]. Finally, it is worth mentioning that compared with single CNN, the other three models employed much fewer hyperparameters, which substantially saved computational resources.



(A) CNN



(B) RNN



(C) 2D-CNN-RNN (D) 3D-CNN-RNN

Figure 1. Four deep learning architecture (A) CNN, (B) RNN, (C) 2D-CNN-RNN, (D) 3D-CNN-RNN

TABLE II. CLASSIFICATION ACCURACY AND STANDARD DEVIATION OF THE FOUR DEEP LEARNING MODELS (%)

Accuracy						
Model	**CV1**	**CV2**	**CV3**	**CV4**	**CV5**	**Fivefold Average**
CNN	89.46	84.10	89.60	68.08	73.20	80.89
RNN	81.63	76.31	81.87	64.73	75.44	76.00
2D-CNN-RNN	87.93	83.76	89.54	69.67	74.73	81.13
3D-CNN-RNN	89.29	85.63	93.17	71.17	75.19	82.89
Standard Deviation						
Model	**CV1**	**CV2**	**CV3**	**CV4**	**CV5**	**Fivefold Sum**
CNN	1.34	2.56	3.23	4.97	2.29	14.40
RNN	2.26	6.01	3.15	6.69	2.03	20.14
2D-CNN-RNN	2.73	5.50	5.11	2.75	2.99	19.08
3D-CNN-RNN	1.66	1.59	2.43	2.27	1.64	9.60

We did notice the variance of accuracies across fivefold datasets, which may stem from individual differences in cortical responses instead of the group-specific traits. We rechecked the data carefully, and found that trails of some participants in the test set have significantly different cortical responses compared with those from the same group, which may confuse the classifier. Another limitation of the present study is the imbalance and relatively small amount of data in PD and HC group, as deep learning is sensitive to data disparity. Therefore, the four models may not unlock their potentials to be deep learning algorithms with a large number of hyperparameters.

Future research will be conducted based on larger and more balanced datasets between PD and HC group. Besides, with the technique of visualization and the attention mechanism, we aim to unfold the topography of the causal contributions of features learned by the algorithm, which provides useful tools for task-related brain mapping in the temporal and spatial domain.

V. CONCLUSION

In the present study, we designed four deep learning algorithm architectures—two convention deep learning

models (CNN and RNN) and two hybrid convolutional recurrent neural networks (2D-CNN-RNN and 3D-CNN-RNN)—to tackle the task of PD detection based on EEG signal. The results showed that our models outperform the conventional deep learning models. The present study highlights the potential of deep learning algorithms for automated classification of raw EEG data without handcrafted features, which introduces a new perspective to facilitate clinical decision making.

ACKNOWLEDGMENT

This study was jointly supported by a grant from National Natural Science Foundation of China (61771461 and U1736202) and Shenzhen Fundamental Research Program (JCYJ20170413161611534 and JCYJ20150330102401089).

REFERENCES

[1] J. Valls-Sole and F. Valldeoriola, "Neurophysiological correlate of clinical signs in Parkinson's disease," Clinical Neurophysiology, vol. 113, no. 6, pp. 792-805, Jun 2002.

[2] "Neurological Disorders Public Health Challenges," World Health Organization2006, Available: http://www.who.int/mental_health/neurology/neurological_disorders_report_web.pdf.

[3] C. X. Han, J. Wang, G. S. Yi, and Y. Q. Che, "Investigation of EEG abnormalities in the early stage of Parkinson's disease," Cognitive Neurodynamics, vol. 7, no. 4, pp. 351-359, Aug 2013.

[4] R. Yuvaraj, U. R. Acharya, and Y. Hagiwara, "A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals," Neural Computing & Applications, vol. 30, no. 4, pp. 1225-1235, Aug 2018.

[5] T. X. Wen and Z. N. Zhang, "Deep Convolution Neural Network and Autoencoders-Based Unsupervised Feature Learning of EEG Signals," IEEE Access, vol. 6, pp. 25399-25410, 2018.

[6] R. T. Schirrmeister et al., "Deep Learning With Convolutional Neural Networks for EEG Decoding and Visualization," Human Brain Mapping, vol. 38, no. 11, pp. 5391-5420, Nov 2017.

[7] R. Schirrmeister, L. Gemein, K. Eggensperger, F. Hutter, and T. Ball, "Deep Learning with Convolutional Neural Networks for Decoding and Visualization of EEG Pathology," 2017 IEEE Signal Processing in Medicine and Biology Symposium, 2017.

[8] S. Roy, I. Kiral-Kornek, and S. Harrer, "Deep Learning Enabled Automatic Abnormal EEG Identification," presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Hawaii, USA, 2018.

[9] X. Zhang, L. Yao, Q. Sheng, S. S. Kanhere, T. Gu, and D. Zhang, "Converting Your Thoughts to Texts: Enabling Brain Typing via Deep Feature Learning of EEG Signals," 2017.

[10] S. L. Oh et al., "A deep learning approach for Parkinson's disease diagnosis from EEG signals," Neural Computing and Applications, 2018.

[11] X. Y. Huang et al., "The impact of parkinson's disease on the cortical mechanisms that support auditory-motor integration for voice control," Human Brain Mapping, vol. 37, no. 12, pp. 4248-4261, Dec 2016.

[12] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," Computer Methods and Programs in Biomedicine, vol. 161, pp. 1-13, Jul 2018.

[13] U. R. Acharya et al., "Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network," Future Generation Computer Systems-the International Journal of Escience, vol. 79, pp. 952-959, Feb 2018.

[14] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," Nature Neuroscience, vol. 2, no. 11, pp. 1019-1025, Nov 1999.

[15] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, Proceedings, pp. 994-1000, 2005.

[16] D. Scherer, A. Muller, and S. Behnke, "Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition," Artificial Neural Networks (Icann 2010), Pt Iii, vol. 6354, pp. 92-101, 2010.

[17] T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, and S. Khudanpur, "Recurrent neural network based language model," presented at the Interspeech, 2010.

[18] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches," 2014.

[19] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks," 2016.

[20] P. L. Nunez and R. Srinivasan, Electric Fields of the Brain: The Neurophysics of EEG. USA: Oxford University Press, 2006.

[21] R. T. Canolty et al., "High gamma power is phase-locked to theta oscillations in human neocortex," Science, vol. 313, no. 5793, pp. 1626-1628, Sep 15 2006.

[22] S. Monto, S. Palva, J. Voipio, and J. M. Palva, "Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans," Journal of Neuroscience, vol. 28, no. 33, pp. 8268-8272, Aug 13 2008.

[23] S. Vanhatalo, J. M. Palva, M. D. Holmes, J. W. Miller, J. Voipio, and K. Kaila, "Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep," Proceedings of the National Academy of Sciences of the United States of America, vol. 101, no. 14, pp. 5053-5057, Apr 6 2004.