Multi-stream 3D Convolution Neural Network with Parameter Sharing for Human State Estimation

*Chin-Teng Lin[1], Jia Liu[1], Chieh-Ning Fang[1], Shih-Ying Hsiao[2], Yu-Cheng Chang[1] and *Yu-Kai Wang[1]

[1] Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia;
[2] Brain Research Center, Institute of Electrical and Control Engineering, National Chiao Tung University, Taiwan

***Abstract*—The large amount data is one challenge in electroencephalogram (EEG) analysis, in which the channel (2D), time (1D) and spectral (1D) are generally considered. Convolutional neural networks (CNNs) have drawn much attention for automatic feature learning in various fields. Meanwhile, many studies have demonstrated integration from multiple sources and decisions could boost performance. However, CNN for EEG analysis usually involves millions of parameters, which easily leads to overfitting. A new model of a multi-stream 3D CNN with parameter sharing is proposed for EEG. Two EEG datasets, the lane-keeping task (LKT) dataset and sleep dataset, are applied. For LKT dataset, the proposed multi-stream 3D CNN with parameter sharing model achieves 0.5486 root mean square error (RMSE), showing improvement by at least 2.77% compared to the other approaches. In the sleep dataset, the error rate of the proposed model was 24.65%, showing at least 10.28% improvement in performance compared to the other methods. The lower RMSE and error rate show that the multi-stream 3D CNN with parameter sharing model efficiently extracts significant features from EEG data. Moreover, the sharing mechanism even reduces the risk of overfitting and number of parameters by comprehending common representations among multiple streams.**

Index Terms—convolutional neural network, parameter sharing, EEG

I. INTRODUCTION

Deep learning has gained great attention in the machine learning area due to its ability to learn important features from multiple layers hierarchically and build high-level representations automatically [1]. Deep learning algorithms have yielded competitive performance in multiple domains, such as image recognition[2, 3], human action recognition [4-6], image captioning [7, 8], speech recognition [9, 10], and brain-computer interfaces (BCIs) [11-15]. A convolutional neural network (CNN) is a typical type of deep learning structure in which a series of convolutional layers and pooling layers are applied to the input images[16]. CNNs can achieve superior performance because the hierarchy of features is built and extracted effectively from lower-level features [16, 17]. Encouraged by the significant capabilities of extracting features, recent studies [5, 18] have demonstrated further improved performance by incorporating several sources or fusing decisions from multiple CNNs. Simonyan and Zisserman published a two-stream architecture that is composed of spatial and temporal networks [5]. Final predictions were computed by averaging the outputs of the spatial and temporal networks. Currently, there are an increasing number of datasets in which the dimension of critical information is more than three. Ji *et al*. first applied a three-dimensional convolutional neural network (3D CNN) to extract motion features from spatial and temporal domains [4]. Compared with 2D networks, 3D CNNs can encode richer spatial information and extract more representative features via a hierarchical architecture trained with 3D samples [19].

As an effective way to monitor changes in human behaviour and brain states [20, 21], electroencephalogram (EEG) data have been greatly promised by deep learning (DL) and machine learning (ML) to learn good feature representations from raw signals. Several studies [13-15, 22-25] have leveraged ML or DL algorithms to analyse EEG signals. For example, Reddy *et al*. proposed fuzzy time-delay common spatial-spectral patterns to model human behaviour through EEG dynamics during drowsiness driving [23]. Wang *et al*. leveraged a support vector machine with a radial basis function kernel to distinguish brain dynamics to track attentional changes during distracted driving [24]. In our previous research, Hung *et al*. applied a 3D CNN with channel-wise convolution to extract EEG features and predict fatigue levels during continuous driving [14]. Chen *et al*. used raw EEG data to train deep learning models that could select specific features for better target detection when performing rapid serial visual presentation tasks [15]. Due to these successful ML- or DL-based EEG analysis, raw EEG signals and extracted features from the frequency, temporal or spatial domains can provide a high possibility to investigate physiological changes and model human performance.

However, an inevitable issue is that while training EEG data with all information from frequency, temporal and spatial domains to achieve a higher performance, the enormous inputs from the combination of multiple networks into one model could generate a mass of related parameters as well, which easily encounters overfitting problems during the training phase. Besides caused by multiple domains network, the overfitting problem is more serious for EEG analysis itself because there are high diversities and variations across the data from multiple individuals. To overcome this problem, transfer learning and multi-task learning have been broadly applied for EEG-based training and analysis [26, 27]. The characteristics of transfer learning [28, 29] and multi-task learning [30] are suitable to overcome overfitting and subject-variant problems. Transfer learning aims to use knowledge from one or more source domains to help the learning in a target domain [31, 32]. Transferring the pre-trained parameters from the source domain to the target domain avoids training complex models from

scratch [33]. Multi-task learning makes models comprehend multiple tasks simultaneously. Features in different networks enable transfer across related tasks to reinforce each other effectively [31, 33, 34]. In summary, both transfer learning and multi-task learning can share representations between related tasks or different domains. These characteristics can benefit overfitting and subject-variant problems in EEG analysis.

As one of the most widely used approach for multi-task learning with neural networks, hard parameter sharing was applied to learn a common representation for all tasks through completely sharing weights/parameters between tasks [35]. This shared feature is used to model the different tasks, usually with additional, task-specific layers that are learned independently for each task. In fact, research summarized that the risk of overfitting the shared parameters is an order N, where N is the number of tasks that is smaller than overfitting the task-specific parameters from the output layers [30, 36]. In other words, during training phrase, more tasks the model learning simultaneously, the more model has to find a representation that captures all of the tasks and the less is the chance of overfitting on original task. Thereby, hard parameter sharing acts as regularization for multiple tasks, which is capable to greatly reduces the risk of overfitting.

Inspired by the above mechanisms and approaches, we propose a model, multi-stream 3D CNN with parameter sharing, in this study to enhance the communication between multiple streams and reduce the number of parameters for EEG-based analysis. To fully exploit multimodal information for important characters in different domains of EEG data, the 3D CNN has been proven to achieve better performance and results in previous studies [4, 14]. In this work, we hypothesized that all information from EEG signals can improve the reliability and stability of human states identification. Note that the dimensions of EEG signals include spatial (2 dimension), temporal (1 dimension) and frequency (1 dimension). We then leverage convolution neural network (CNN) to extract the significant patterns for modelling the relationships between the brain dynamics and human states. As each frequency band has different functional characteristics, we use one CNN model for every frequency bin in this study. Meanwhile, the increased or decreased amplitude is the main index across all frequency bins. The sharing parameters across frequency bins is applied to extract the significant changes of amplitude in the different CNN models. Proven by experimental results, with the sharing mechanism, the model converges fast and avoids overfitting since it has a smaller number of parameters [37] than 3D CNN, CNN and neural network (NN) and recurrent neural network models. Additionally, the transferable features between different sources also benefit the subject-variant issue in EEG analysis and further improve overall BCI performance.

The rest of the study is organized as follows. The structure of a multi-stream 3D CNN with parameter sharing is illustrated in Section 2. Implementation details and experimental results of the LKT and sleep datasets are given in Section 3 and Section 4, respectively. Section 5 discusses the results observed in two experiments. Section 6 summarizes the conclusions of the experimental results.

II. MULTI-STREAM 3D CNN WITH PARAMETER SHARING

The architecture of the multi-stream 3D CNN and the setting of parameter sharing are introduced in this section.

A. *Multi-stream 3D CNN*

In the following, the proposed multi-stream 3D CNN architecture is described. The proposed architecture developed for EEG signal analysis is shown in Fig. 1. *N* 3D CNN streams are trained in the proposed model, and each stream focuses on a specific frequency bin (1 Hz, 2 Hz to *N* Hz). The different frequency bin has different functional characteristics which can be indexed by the increased and/or decreased amplitude values in each frequency bin. the increased and decreased amplitude values from one frequency bin. are the major indexes There is a consensus of broad approval that high disparities in brain states can exist in different frequency domains [25, 38, 39]. The separation of spectrum information for each stream aims to make each stream focus on a single frequency and obtain more accurate features and knowledge of brain states from specific frequency domains, which enables us to classify different interpretations of the brain state on the spectrum more precisely [40]. Detailed information on each layer is provided in the following sections.

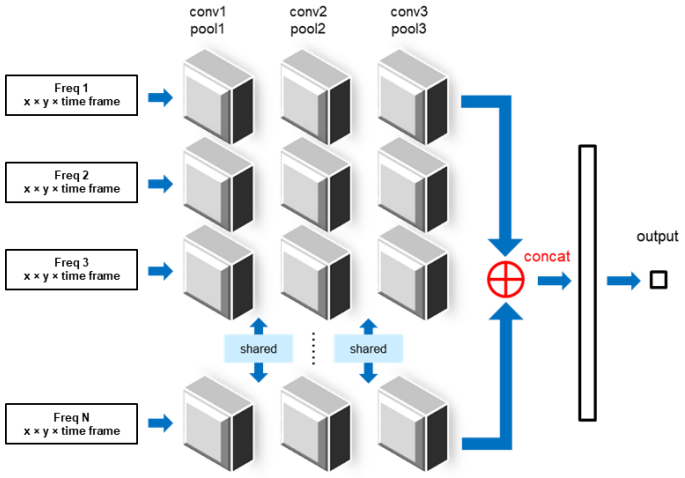

Fig. 1. Architecture of the multi-stream 3D CNN with parameter sharing. The input x and y indicate channel information. *N* 3D CNN streams are trained in respective model, and each stream focuses on a specific frequency bin (1 Hz, 2 Hz to *N* Hz). The corresponding convolutional layers share the same set of parameters. See Section B below for a detailed method of updating the same set of parameters in convolutional layers.

1) Convolution Layer: At the convolutional layer, 3D convolution is performed and defined by

$$v_{ij}^{xyz} = relu\left(b_{ij} + \sum_{m}\sum_{p=0}^{P_i+1}\sum_{q=0}^{2_i+1}\sum_{r=0}^{R_i+1} w_{ijm}^{pqr} v_{(i+1)m}^{(x-p)(y-q)(z-r)}\right), \quad (1)$$

where v_{ij}^{xyz} is the unit at position (x, y, z) in the ith layer and jth feature map, b_{ij} is the bias between the two layers, m is the mth filter in the convolutional layer, P, Q and R are the kernel sizes of the filters, w_{ijm}^{pqr} is the weight at position (p, q, r) of the mth filter, and a rectified linear unit (ReLU) is the activation function.

2) Pooling Layer: Pooling layers are connected to reduce the sample of feature maps from the convolution layers, which sub-samples the volume spatially independently in each depth slice of the input volume. We apply the 3D pooling operation and 2D pooling operation at the last layer to sub-sample the data points,

which reduces the spatial and temporal resolution of every feature map. Both kinds of operations perform sub-sampling on the feature maps without overlapping. They select the maximum unit from the specific region. 2D max-pooling and 3D max-pooling use 2×2 filters and 2×2×2 filters, respectively. With pooling layers, models can reduce the size of representation to decrease the number of parameters and computation.

3) Fully Connected Layer and Output Layer: These two layers present the function of classification/recognition in the whole proposed model architecture. After a sequence of convolutional layers and pooling layers, the size of the output feature maps of each stream becomes 4×4×4. All feature maps are concatenated in the channel dimension and flattened into a one-dimensional vector. Each unit in this vector is fully connected to each of the 128 units. Then, the fully connected layer performs regression to predict with the soft-max activation function.

With the aid of a multi-stream 3D CNN, feature and specific knowledge from different frequency domains can be trained and grabbed precisely. However, the usage of twenty 3D CNN models undoubtedly results in overfitting problems due to the large number of parameters trained in the model. The lack of shared representations causes the whole model to not comprehend the relevant parts insufficiently [41]. Thus, to converge the model fast and avoid overfitting, we applied parameter sharing to transfer similar features among multiple streams to help the model grab relevant parts effectively.

B. Parameter Sharing

Considering the above architecture of a multi-stream 3D CNN, the major challenge is the computational complexity and the time needed to train large networks [42]. In the proposed architecture for human state estimation, there are *N* different modes for the selected frequency bins (1 Hz, 2 Hz to *N* Hz). Each frequency band has different function characteristics which can be investigated through the increased or decreased amplitude [43]. Then one CNN model is used to extract the significant features from each single frequency bin. One essential characteristic of CNNs is that the feature map can reflect the affine transformations in the input image. Therefore, the feature map can provide the outputs that are subjected to the transformed inputs. As the increased or decreased amplitude is the common features across different frequency bins, we then apply the same parameters to all CNN models. In previous studies, parameter sharing demonstrated that the number of training parameters can be dramatically reduced [44, 45]. In this way, the sharing mechanism can extract the significant spectrum changes across different human states.

In Fig. 1, the blue blocks with 'shared' text indicate that the corresponding convolutional layers use the same set of parameters. The method of updating the same set of parameters in convolutional layers is introduced as follows. The weights and biases in the corresponding convolutional layers are represented by $\theta = \{w_*, w_+, \cdots, b_*, b_+, \cdots\}$, and the initial value of θ is θ^0. The gradients of the parameters of the n^t- stream are defined as

$$\nabla L(\theta)_n = \begin{bmatrix} \partial L(\theta)/\partial w_* \\ \partial L(\theta)/\partial w_+ \\ \vdots \\ \partial L(\theta)/\partial b_* \\ \partial L(\theta)/\partial b_+ \\ \vdots \end{bmatrix} \quad (2)$$

where L is loss function. First, we compute the gradients of each stream of 1 Hz to *N* Hz spectral data. Then, we sum the gradients of multiple streams and update the parameters by the total value. As represented in Equations 3 and 4, the model renews the trainable parameters every iteration i until convergence.

$$Compute\ \nabla L(\theta^{\mathrm{i}}) = \mathrm{M}_{n1^*}^{N} \nabla L(\theta^{\mathrm{i}})_n \quad (3)$$

$$\theta^{\mathrm{i}2^*} = \theta^{\mathrm{i}} - \nabla L(\theta^{\mathrm{i}}) \quad (4)$$

The sum of gradients and updating of parameters of different source domain data will help the network learn and keep all essential knowledge and avoid retraining similar parameters, especially for the temporal dimensions, with every iteration.

III. EXPERIMENT ▮: LKT DATASET

In the first experiment, the multi-stream 3D CNN with parameter sharing and other methods (more detail in the next paragraph) on the LKT dataset. We used leave-one-subject-out cross-validation to evaluate regression and classification performance. In addition, the 10% training set was assigned as the validation set, which was used to determine the stopping epoch. Those models were trained by the adaptive moment estimation (Adam) optimizer with a mini-batch size of 32. The rates of start learning and decay are 0.001 and 0.96, respectively. The early stop threshold is 25 epochs.

A. LKT Dataset and Pre-processing

LKT is a lane-keeping task whose aim is predicting the driver's driving performance based on fatigue level. This study obtained the approval of the local ethics committee (Institutional Review Board of Taipei Veterans General Hospital, Taiwan), and were recorded at National Chiao Tung University. The experiment was conducted in a 360 degree virtual reality lab [46, 47] with a motion platform, which simulates the driving environment in reality. The entire dataset includes 71 one-hour sessions from 37 normal participants. The performance of the drivers is measured by the reaction times (RTs). RT is the latency between the onset of deviation and the onset of response. During the period, the participants were aware of the deviation and drove the car back to the cruising lane with a steering wheel, as shown in Fig. 2. A high RT indicates that the participants were relatively drowsy in this virtual reality highway driving scenario; in contrast, a lower RT indicates that the participants were relatively alert. The proposed model with a sharing mechanism aims to predict RT by extracting crucial information on human brain dynamics.

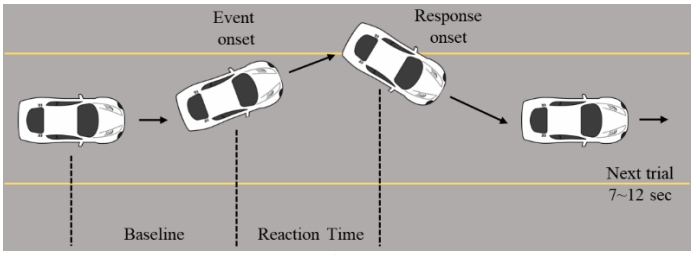

Fig. 2. The experimental scenario of the LKT dataset. The cruising car was randomly drifted to the left or right, and the participants were asked to control the car back to the cruising lane. We use a 5.8-second EEG signal before the event onset (baseline) as the input data for the multi-stream 3D CNN, and the period between the event onset and the response onset is defined as the reaction time. The period between two continuous trials is approximately 7 to 12 seconds.

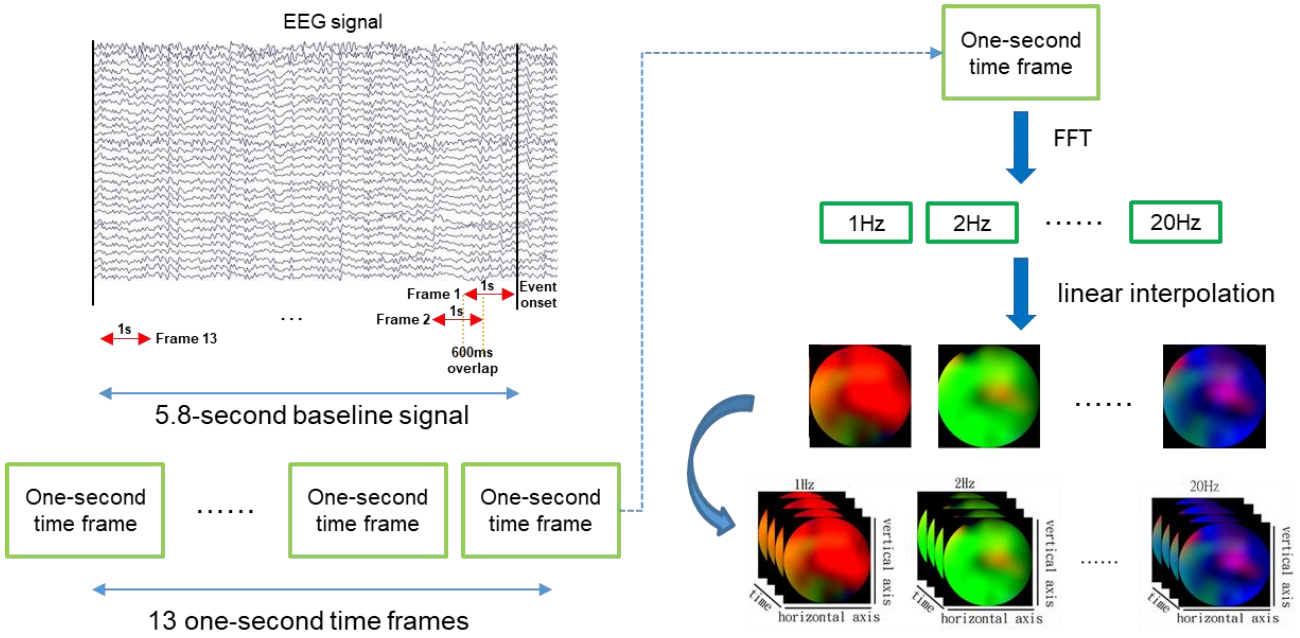

Fig. 3. Structure of LKT data pre-processing and feature extraction in the multi-stream 3D CNN algorithm. For each trial, EEG signals were segmented into a 5.8-second baseline and divided into 13 one-second time frames with 600 ms overlap between two frames. FFT was then applied to translate the frequency information. There were 20 frequency bins generated (between 1 Hz to 20 Hz). Each frequency bin was composed of 13 sequential 2D EEG channel maps. Then the 5.8-second EEG signals were transformed into 4D data format, which are 2D EEG channel map × time frame × frequency bin.

Fig. 3 shows the data structure of our multi-stream 3D input data. We applied EEGLAB for EEG data preprocessing [48]. The raw EEG data was first downsampled into 250-Hz, followed by a high-pass filter (1 Hz), and a low-pass filter (50 Hz). Then the EEG signals were segmented into 5.8-second baseline signals, which began 5.8 seconds before event onset (refer Fig. 2). The baseline signal was divided into 13 one-second time frames with a 600 ms overlap between two frames. Fast Fourier transform (FFT) was performed to transform a time frame into a 2D EEG channel map. Twenty frequency bins were generated from 1 Hz to 20 Hz with a 1 Hz difference. Each frequency bin was composed of 13 sequential EEG channel maps. Additionally, the 2D EEG channel maps included the frequency power of all EEG channels in one single time frame. Hence, the 5.8-second EEG signals were transformed into 4D data format, which are the 2D EEG channel maps, time frame and frequency bin. In terms of normalization, the training and testing datasets were normalized by subtracting the mean of the training dataset and dividing by the standard deviation of the training dataset. Additionally, the RTs in each trial in a session were divided by the tenth percentage shortest RT in their session. After data pre-processing and normalization, we applied the multi-stream 3D CNN with parameter sharing to analyse the temporal and spatial information effectively. Each stream focused on addressing one frequency bin of EEG signals. Therefore, the input shape for each 3D CNN stream was 2D EEG channel map × time frame.

TABLE I
DETAILS OF THE MULTI-STREAM 3D CNN ARCHITECTURE FOR THE LKT DATASET

Layer	conv1	pool1	conv2	pool2	conv3	pool3	fc1	fc2
kernel size	3×3×3	2×2×2	3×3×3	2×2×2	3×3×3	2×2×1	-	-
stride	1	2	1	2	1	2	-	-
number of maps	16	16	32	32	64	64	128	1
height × width	32×32	16×16	16×16	8×8	8×8	4×4		
depth	13	7	7	4	4	4		
Activation function	ReLU		ReLU		ReLU		ReLU	Softmax

B. *Regression Performance on the LKT Dataset*

The structure of each 3D CNN stream is shown in Table I. We adopt 3×3×3 (height, width, depth) filters in all three convolutional layers, in which there were 16, 32 and 64 filters, respectively. In the pooling layers, we utilize max-pooling with 2×2×2 filters in the first and second convolutional layers. The third

TABLE II
INPUT CONFIGURATION AND COMPARISON RESULTS OF PROPOSED MODELS AND OTHER ALGORITHMS FOR THE LKT DATASET

Method	Multi-stream 3D CNN with parameter sharing	Multi-stream 3D CNN	4D CNN	3D CNN	CNN	NN
Input data points (*shape*)	266240 (*2D EEG channel map × time frame × frequency bin*)			7800 (1D *EEG channel × time frame × frequency bin*)	600 (1D *EEG channel × frequency bin*)	
Parameters amount of convolution layer	69K	1393K	208K	69K	23K	-
RMSE	0.5486±0.1257	0.5549±0.1344	0.5638 ±0.1391	0.5862 ±0.1359	0.5992 ±0.1489	0.6188 ±0.1374
Improvement of Proposed model	-	1.14%	2.77%	6.85%	9.22%	12.79%

convolutional layer utilizes a 2×2×1 filter. After performing convolution and reducing the dimensionality, the output of the last pooling layer in each stream has 64 feature maps with a size of 4×4×4. They are concatenated in the channel dimension into a size of 4×4×4×1280 and flattened to a one-dimensional vector. Then, the fully connected layers perform regression to predict the RTs.

The comparison results and input structures of the multi-stream 3D CNN with and without parameter sharing, as well as other state-of-the-art algorithms, namely, 4D CNN [49], 3D CNN [14], CNN and NN on the LKT dataset are described in Table II. According to the results, the sharing mechanism made the convergence rate faster and eased overfitting due to a smaller number of parameters. The RMSE of the multi-stream 3D CNN with parameter sharing was 0.5486±0.1257. The RMSE improved over the performance of the 4D CNN, 3D CNN, CNN, NN by 2.77%, 6.85%, 9.22% and 12.79%, respectively. The results indicated that our proposed model with a sharing mechanism has the capability to predict RT by extracting crucial information on human brain dynamics.

Fig. 4 shows the average RMSE and standard deviation for each method. To evaluate the regression performance statistically, a t-test with Bonferroni correction was performed for pair comparisons. The multi-stream 3D CNN with parameter sharing has the highest regression performance with a lower standard deviation, and the performance is significantly higher than 3D CNN, CNN and NN ($p < 0.005$ with Bonferroni correction).

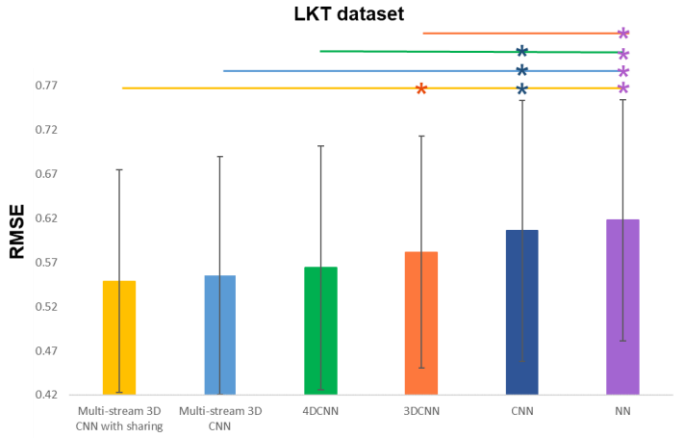

Fig. 4. Average RMSE and standard deviation of LKT dataset. The colour bars represent the comparing algorithms respectively, and the error bars indicate standard deviations. * indicates significant difference among pair methods with Bonferroni correction ($p < 0.005$).

C. LKT Dataset Visualization

In this section, we visualized topo-images and feature maps to explore the relationships between brain dynamics in specific bands and driving performance.

First, we show the input data in the theta and alpha bands under alert and drowsy conditions, as shown in Fig. 5. Previous EEG studies showed that frequency power in theta (4-7 Hz) and alpha (8-13 Hz) bands was a significant indicator for monitoring brain dynamics during a driving task [38, 50, 51]. The increased theta and alpha activities are related to driver drowsiness. Furthermore, the theta power of posterior brain regions dramatically increases if the driver is at a high fatigue level [52, 53]. A high RT indicates that the participant is relatively drowsy; in contrast, a low RT indicates that the participant is relatively alert. To understand the influence of parameter sharing, we visualize the feature maps distributed in all convolutional layers. The feature maps of the multi-stream 3D CNN with and without parameter sharing are shown in Fig. 6. In the alpha band, both models caught the increased power of the drowsy condition. With sharing, the trend of the brain dynamic state over time was extracted more correctly. Without a sharing mechanism, the model confused drowsiness with alerts in the theta band.

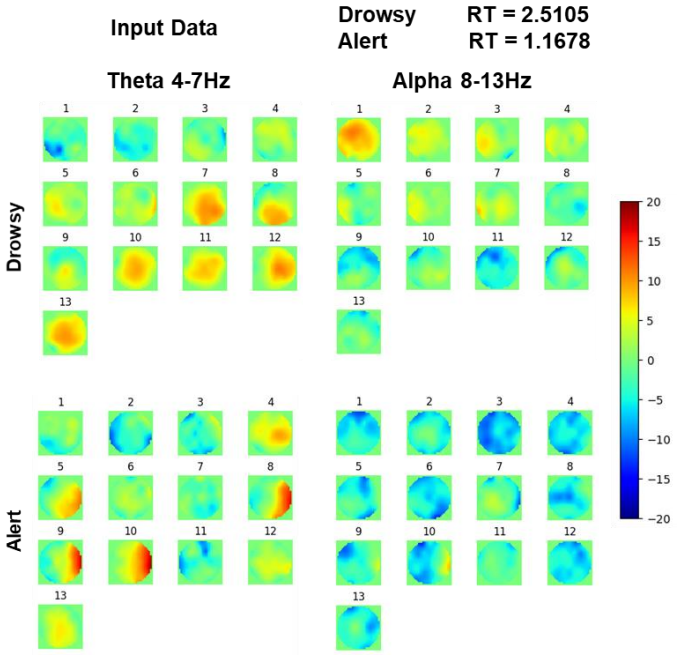

Fig. 5. The input data of multi-stream 3D CNN. The baseline signals are divided into 13 one-second time frames with a 600 ms overlap between two frames. The labels of each image refer to the sequence of time frames, e.g., label 1 means 0-1000 ms; label 2 means 400-1400 ms and label 13 means 4800-5800 ms. The top half of topo-images relates to drowsy conditions, and the bottom half of topo-images relates to alert conditions. The right half part and left half part are topo-images of theta and alpha bands, respectively.

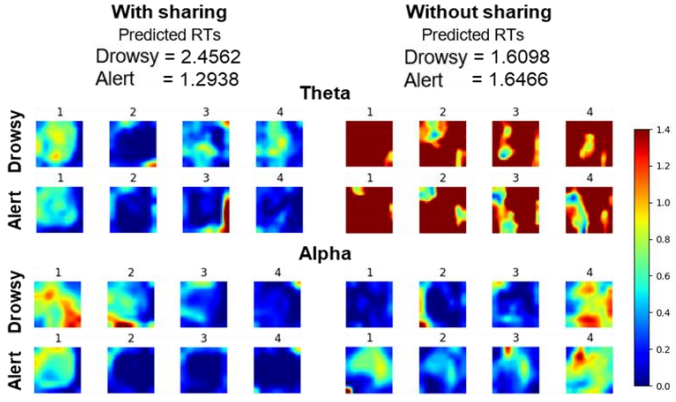

Fig. 6. Feature maps of third convolutional layers in the theta and alpha bands. The left half is the predicted RTs and feature maps of the multi-stream 3D CNN with parameter sharing. The right half part shows the predicted RTs and feature maps of the multi-stream 3D CNN without parameter sharing.

IV. EXPERIMENT II: SLEEP DATASET

In this experiment, another EEG dataset, sleep dataset is used to compare the performance between the proposed model (ref. Fig. 1) and other algorithms. All same algorithms with the same training strategies for the LKT dataset were applied.

A. Sleep Dataset and Pre-processing

The sleep dataset included nine participants (5 males and 4 females; at the age of 22±2 years). All participants were asked to keep a regular daily routine and filled out the sleep schedule for one week before the experiment. Participants were required to sleep for one night in the sleep laboratory at National Chiao Tung University, Taiwan. All experimental procedures were approved by the local ethics committee (Institutional Review Board of Taipei Veterans General Hospital, Taiwan). The sleep experimental environment was especially designed based on a home-style bedroom to avoid the impact of an unfamiliar environment. In addition, there were cameras for monitoring and emergency phones that could be used if the participants felt uncomfortable or had physiological problems. The temperature of the experimental room was controlled at 24 to 26 degrees Celsius, which is a suitable range for sleep.

The experimental duration was from 22:30 to 07:30. At approximately 23:30, the experiment started with simultaneous EEG, electrooculography (EOG), and electrocardiography (EMG) recordings until the participants woke up. An experienced sleep technologist from Hospital follows the scoring rules in the American Academy of Sleep Medicine (AASM) to score all recorded EEG signals. In general, there are five sleep stages including wake stage (W), non-rapid eye movement (NREM) stage 1 (N1), NREM stage 2 (N2), NREM stage 3 (N3) and rapid eye movement (REM) stage. The N1, N2 and REM stages are regarded as light sleep in which people more easily wake up [54]. N3 is referred to as slow-wave sleep (SWS) or deep sleep [54]. During the whole night of sleep, the stage changes with cycles of REM and NREM, normally in the order of N1 → N2 → N3 → N2 – REM. Each cycle lasts for 90 to 110 minutes, and the cycle repeats four to six times a night. The proposed model with a sharing mechanism aims to classify diverse sleep stages by extracting crucial information on human brain dynamics.

The protocol of pre-processing and feature extraction for sleep dataset is same as LKT data (refer Fig. 3 for LKT data), except different signal length for baseline and overlap was extracted to fit sleep dataset. We only extracted the third and fourth cycles (180th ~ 360th minute) and segmented the EEG signals into 30-second trials. The data in one trail are divided into 29 two-second time frames with a one second overlap between two frames. FFT was then applied to every time frame to translate the frequency/spectral information from 1 Hz to 20 Hz. Each frequency bin was composed of 29 sequential EEG channel maps. Additionally, the 2D EEG channel maps include the frequency power of all EEG channels in one single time frame. Hence, the 30-sec EEG signals were transformed into 4D data format, which are the 2D EEG channel map, time frame and frequency bin. In terms of normalization, the training and testing datasets were normalized by subtracting the mean of the training dataset and dividing by the standard deviation of the training dataset.

B. Classification Performance on the Sleep Dataset

The structure of each 3D CNN stream with parameter sharing is shown in Table Ⅲ. We adopted 3×3×3 (height, width, depth) filters in all three convolutional layers, in which there were 16, 32 and 64 filters, respectively. In the pooling layers, we utilized max-pooling with 2×2×2 filters. After performing convolution and reducing the dimensionality, the output of the last pooling layer in each stream had 64 feature maps with a size of 4×4×4. They were concatenated in channel dimensions into sizes of 4×4×4×1280 and flattened to a one-dimensional vector. Then, the fully connected layers classified the EEG signals into five categories.

TABLE III
DETAILS OF THE MULTI-STREAM 3D CNN ARCHITECTURE FOR THE SLEEP DATASET

Layer	conv1	pool1	conv2	pool2	conv3	pool3	Fc1	fc2
kernel size	3×3×3	2×2×2	3×3×3	2×2×2	3×3×3	2×2×2	-	-
stride	1	2	1	2	1	2	-	-
number of maps	16	16	32	32	64	64	128	5
height × width	32×32	16×16	16×16	8×8	8×8	4×4		
depth	29	15	15	8	8	4		
Activation function	ReLU		ReLU		ReLU		ReLU	Softmax

The comparison results as well as input configurations of the multi-stream 3D CNN with parameter sharing and other methods are summarized in Table Ⅳ. With the sharing mechanism, the model could avoid overfitting owing to fewer parameters. Additionally, we compared the system performance with other methods, including 3D CNN, CNN and NN.

The error rate of our proposed model was 0.2465±0.07, which represented improvements over the performance of 3D CNN, CNN and NN by 10.28%, 14.80% and 16.06%, respectively. The model with sharing mechanism used 69K parameters (same as 3D CNN in TABLE IV), but the error rate improved by 10.28%. The results demonstrate that the proposed model can successfully capture significant features and relationships from brain signals.

Fig. 7 shows the average error rate and standard deviation for each method. The t-test with Bonferroni correction indicated that there was no statistically significant difference in classification performance among the different methods on the sleep dataset ($p < 0.005$ with Bonferroni correction).

TABLE IV
INPUT CONFIGURATION AND COMPARISON RESULTS OF PROPOSED MODELS AND OTHER ALGORITHMS FOR THE SLEEP DATASET

Method	Multi-stream 3D CNN with parameter sharing	Multi-stream 3D CNN	3D CNN	CNN	NN
Input data points (*shape*)	593,920 (*2D EEG channel map* × *time frame* × *frequency bin*)		11,600 (1D *EEG channel* × *time frame* × *frequency bin*)	400 (1D *EEG channel* × *frequency bin*)	
Parameters amount of convolution layer	69K	1,393K	69K	23K	-
Error rate	0.2465±0.07	0.2635±0.09	0.2719±0.08	0.2830±0.08	0.2861±0.13
Improvement of Proposed model	-	6.91%	10.28%	14.80%	16.06%

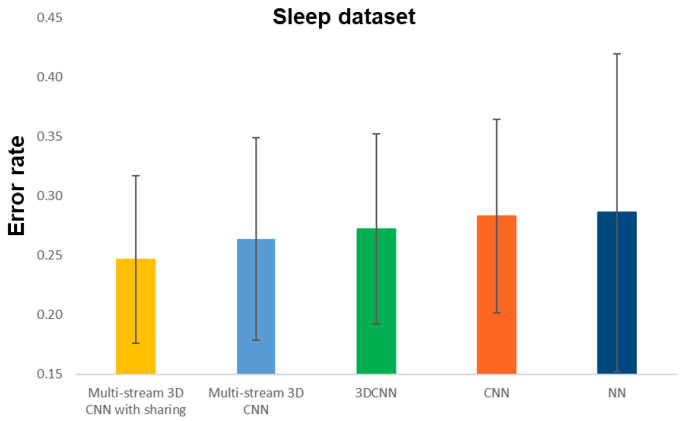

Fig. 7. Average error rate and standard deviation of the sleep dataset. The colour bars represent the comparing algorithms respectively, and the error bars indicate standard deviations. The proposed multi-stream 3D-CNN with parameter sharing reach the lowest error rate, but there is no significant difference as comparing with the other state-of-the-art algorithms.

C. *Sleep Dataset Visualization*

In the AASM guidelines [54], the characteristics of each sleep stage were recorded in detail. Fig. 8 shows the input data and feature maps across four sleeping states (N1, N2, N3 and REM). In the delta bands, the frequency power gradually increases from the N1 stage to the N3 stage (ref. Fig. 8(a)). This indicates that the participant will deeply sleep [55]. The difference between the N1 and N2 stages is also in the theta band (ref. Fig. 8(b)). As a person moves from N1 to N2, brain activity might increase in the theta band. When the participant is in the REM stage, lower frequency power in the alpha band occurs (ref. Fig. 8(c)). The participant might have intense dreams during REM sleep since the brain is more active.

According to those characteristics, we then focused on exploring whether the feature maps included the relationships between the changes in sleep stages. We want to understand what different information is learned between multi-stream 3D CNNs with and without parameter sharing, which improves the performance. The feature maps of the third convolutional layers in the delta band, theta band and alpha band are shown in Fig. 8(a-c), respectively. In the delta band, both models extracted the frequency power, which gradually increased from the N1 stage to the N3 stage. In the theta band, the feature maps extracted by the models with and without parameter sharing were significantly different. With the sharing mechanism, the model with the highest power occurs in the N3 stage. In contrast, the trend of human brain activity from the N1 to N3 stages could not be extracted successfully without parameter sharing. In the alpha band, the model with parameter sharing obtained decreased frequency power in the REM stage. Without a sharing mechanism, the variation among the four sleep stages was not extracted.

V. Discussion

To enhance communication between multiple streams, we utilized a method to combine parameter sharing and a multi-stream 3D CNN for EEG processing in this study. Each stream has a specific spectrum input. With parameter sharing, the model had the capability to extract crucial information to distinguish physiological conditions and capture the trend of brain dynamic states more exactly. Additionally, the sharing mechanism also improves the convergence rate and avoids overfitting by decreasing the number of parameters. In the following sections, we discuss the results observed in both the LKT and sleep datasets.

A. *Brain dynamics under fatigue driving*

In the LKT dataset, the multi-stream 3D CNN outperforms the NN, CNN, 3D CNN and 4D CNN (Ref. Table II) and extracts the increased theta and alpha activities in the drowsy condition (Ref. Fig. 5 and Fig. 6). On the contrary, the model without a sharing mechanism might not extract significant features between drowsy and alert conditions because of lack of the shared information across similar frequencies. As illustrated in Fig. 5, without sharing, the model cannot distinguish drowsiness between alert states by similar patterns in the theta band. With sharing, the trend of the brain dynamic state is extracted more correctly. This may be because features can be transferred across related tasks to reinforce each other [30] and effectively distinguish the similarity with the sharing mechanism. In addition to greatly decreasing the number of parameters and avoiding the overfitting problem, the proposed model has good performance in processing EEG signals with high similarity.

B. *Brain dynamics across different sleeping stages*

The proposed multi-stream 3D CNN reaches lower error rate for sleeping stage identification as shown in Table IV and Fig. 7. According to feature maps (Ref. Fig. 8), the sharing mechanism could obtain important patterns of brain dynamics for distinguishing sleeping stages (N1, N2, N3 and REM). As shown in Fig. 8, the increased activities over all three major frequency ranges can be extracted for sleeping stage identification. In contrast, a multi-stream 3D CNN without a sharing mechanism could not efficiently extracted the significant patterns of brain activities across the different sleeping stages. Particularly, the patterns in alpha band across all sleeping stages are pretty similar (Ref. Fig. 8(c)). However, the proposed models have the capability to extract significant patterns of brain dynamics to distinguish sleeping stages.

VI. Conclusion

In this study, we include all channel, time and frequency domains to pursue lower error rate for prediction problem. Two high dimension datasets including LKT and sleep are used for validation purpose. To handle the huge amount of data, the overfitting and parameters are the challenges. The proposed multi-stream 3D CNN with parameter sharing mechanism successfully overcome the above challenges. The experimental results demonstrate that the proposed model outperforms the other approaches.

In the LKT dataset, the multi-stream 3D CNN with parameter sharing reaches 0.5486 RMSE which is significantly improved over the RMSEs of the 3D CNN, CNN and NN by 6.85%, 9.22% and 12.79%, respectively. In the sleep dataset, the error rate is 24.65%. Compared with the 3D CNN, CNN and NN, the performance of the proposed method is improved by 10.28, 14.80% and 16.06%, respectively. Meanwhile, the feature maps demonstrate that the sharing mechanism effectively extracts the significant brain dynamics in both LKT and sleep experiments. The increased theta and alpha activities in the drowsy state can be explored from the LKT dataset. According to feature maps of

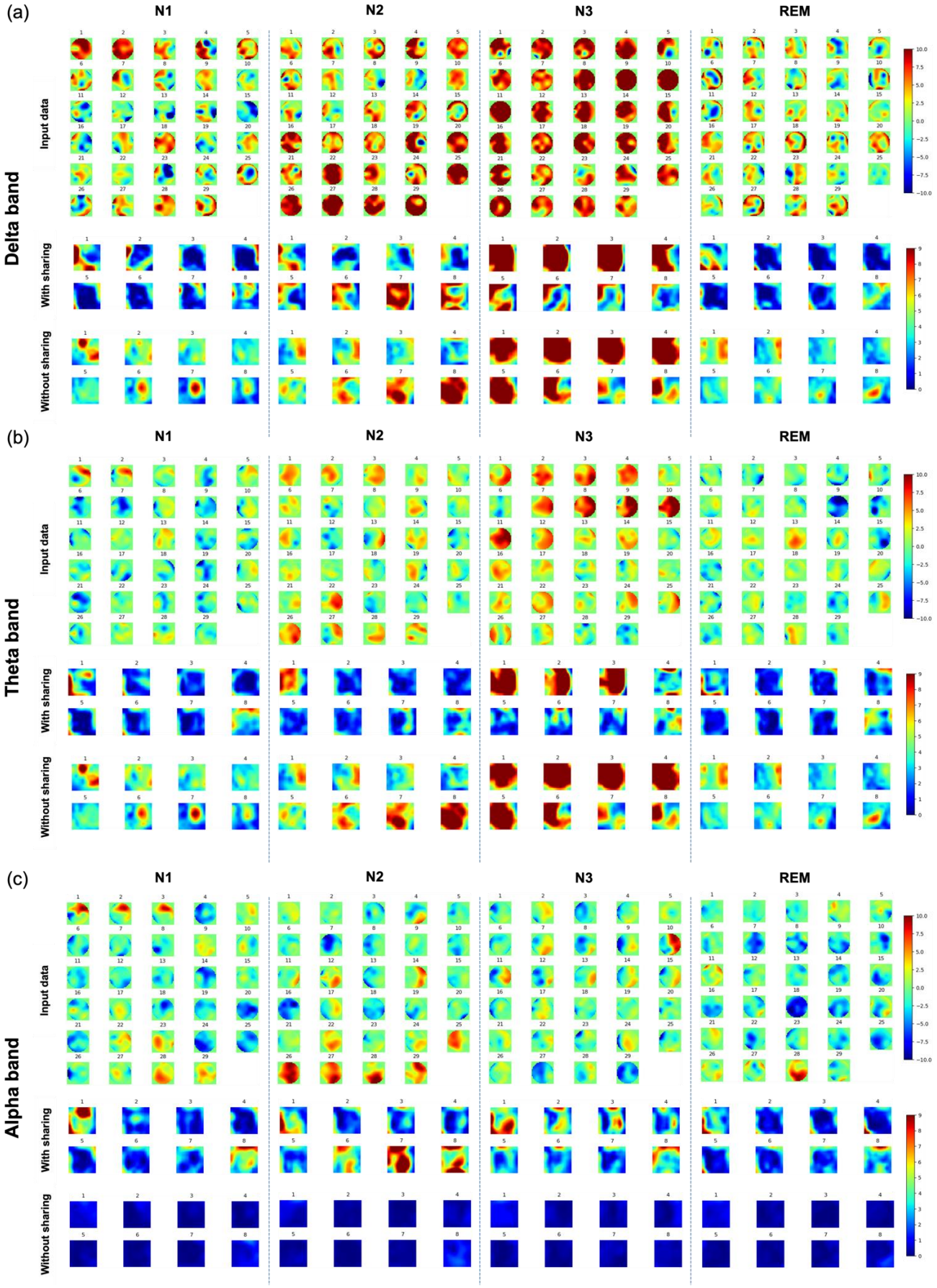

Fig. 8. Input data and feature maps of third convolutional layers. (a-c) are feature maps in the delta band, theta band and alpha band, respectively. The upper part of the sub-figure is the input data. The middle part of the sub-figure shows feature maps of the model with parameter sharing. The bottom part of the sub-figure shows feature maps of the model without parameter sharing.

sleep data, the proposed model obtains important clues for distinguishing sleep stages. From sleeping stage N1 to N3, the increased delta and theta activities have been extracted. The decreased activities in both alpha and beta bands are in the REM stage. Overall, the reliable and high performance of the proposed 3D CNN multi-stream model has been demonstrated by both LKT and sleep datasets. In the near future, we will extend the proposed 3D CNN multi-stream model with parameter sharing to more EEG datasets and other research fields to ease the training complexity and investigate the insight of datasets with high regression performance.

ACKNOWLEDGEMENT

This work was supported in part by the Australian Research Council (ARC) under discovery grant DP210101093 and DP220100803. Research was also sponsored in part by the Australia Defence Innovation Hub under Contract No. P18-650825, US Office of Naval Research Global under Cooperative Agreement Number ONRG - NICOP - N62909-19-1-2058, and AFOSR – DST Australian Autonomy Initiative agreement ID10134. We also thank the NSW Defence Innovation Network and NSW State Government of Australia for financial support in part of this research through grant DINPP2019 S1-03/09 and PP21-22.03.02.

REFERENCES

[1] R. Salakhutdinov, J. B. Tenenbaum, and A. Torralba, "Learning with hierarchical-deep models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1958-1971, 2013.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[3] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, 2015.

[4] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221-231, 2013.

[5] K. Simonyan, and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14)*, Cambridge, MA, USA, 2014, pp. 568-576.

[6] J. Tu, M. Liu, and H. Liu, "Skeleton-Based Human Action Recognition Using Spatial Temporal 3D Convolutional Neural Networks," *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 2018, pp. 1-6.

[7] Z. Liu, and Y. Wang, "TV News Story Segmentation Using Deep Neural Network," *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1-4, 2018.

[8] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proceedings of the 32nd International Conference on Machine Learningpp*, vol. 37, 2015, pp. 2048-2057.

[9] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. N. Sainath, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, 2012.

[10] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, 2013, pp. 6645-6649.

[11] H. Cecotti, and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 433-445, 2011.

[12] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Joint Classification and Regression via Deep Multi-Task Multi-Channel Learning for Alzheimer's Disease Diagnosis," *IEEE transactions on bio-medical engineering*, vol. 66, no. 5, pp. 1195-1206, 2019.

[13] L. Jingwei, C. Yin, and Z. Weidong, "Deep learning EEG response representation for brain computer interface," *2015 34th Chinese Control Conference (CCC)*, 2015, pp. 3518-3523.

[14] Y. C. Hung, Y. K. Wang, M. Prasad, and C. T. Lin, "Brain dynamic states analysis based on 3D convolutional neural network," *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 222-227.

[15] Y. C. Chen, Y. K. Wang, N. R. Pal, and C. T. Lin, "Exploring Covert States of Brain Dynamics via Fuzzy Inference Encoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2464-2473, 2021.

[16] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, and J. Cai, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354-377, 2018.

[17] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," *IEEE Transactions on Bio-Medical Engineering*, vol. 63, no. 3, pp. 664-675, 2016.

[18] Z. Wu, Y. G. Jiang, X. Wang, H. Ye, and X. Xue, "Multi-stream multi-class fusion of deep networks for video classification," *24th ACM international conference on Multimedia*, 2016, pp. 791-800.

[19] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection," *IEEE transactions on bio-medical engineering*, vol. 64, no. 7, pp. 1558-1567, 2017.

[20] C.-T. Lin, M. Nascimben, J.-T. King, and Y.-K. Wang, "Task-related EEG and HRV entropy factors under different real-world fatigue scenarios," *Neurocomputing*, vol. 311, pp. 24-31, 2018.

[21] C.-T. Lin, Y.-T. Liu, S.-L. Wu, Z. Cao, Y.-K. Wang, C.-S. Huang, J.-T. King, S.-A. Chen, S.-W. Lu, and C.-H. Chuang, "EEG-based brain-computer interfaces: a novel neurotechnology and computational intelligence method," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 3, no. 4, pp. 16-26, 2017.

[22] M. Hajinoroozi, Z. Mao, and Y. Huang, "Prediction of driver's drowsy and alert states from EEG signals with deep learning," *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2015, pp. 493-496.

[23] T. K. Reddy, V. Arora, L. Behera, Y. Wang, and C. Lin, "Multiclass Fuzzy Time-Delay Common Spatio-Spectral Patterns With Fuzzy Information Theoretic Optimization for EEG-Based Regression Problems in Brain–Computer

Interface (BCI)," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 10, pp. 1943-1951, 2019.
[24] Y.-K. Wang, T.-P. Jung, and C.-T. Lin, "EEG-based attention tracking during distracted driving," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 23, no. 6, pp. 1085-1094, 2015.
[25] K.-C. Huang, et. al., "An EEG-based fatigue detection and mitigation system," *International journal of neural systems*, vol. 26, no. 04, pp. 1650018, 2016.
[26] Y.-P. Lin, and T.-P. Jung, "Improving EEG-Based Emotion Classification Using Conditional Transfer Learning," *Frontiers in Human Neuroscience*, vol. 11, no. 334, 2017.
[27] C. Tan, F. Sun, and W. Zhang, "Deep transfer learning for EEG-based brain computer interface," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 916-920.
[28] S. J. Pan, and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.
[29] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1717-1724.
[30] S. Ruder, "An overview of multi-task learning in deep neural networks," *National Science Review*, vol. 5, no. 1, pp. 30-43, 2018.
[31] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 9, 2016.
[32] B. Cheng, D. Zhang, and D. Shen, "Domain transfer learning for MCI conversion prediction," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 15, 2012, pp. 82-90.
[33] M. E. Taylor, and P. Stone, "Cross-domain transfer for reinforcement learning," *2018 IEEE International Conference on Robotics and Automation (ICRA)* , 2018, pp. 879-886.
[34] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14)*, 2014.
[35] R. Caruana, "Multitask Learning: A Knowledge-Based Source of Inductive Bias." in *Proceedings of the Tenth International Conference on Machine Learning*, 1993
[36] J. Baxter, "A Bayesian/Information Theoretic Model of Learning to Learn via Multiple Task Sampling," *Machine Learning*, vol. 28, no. 1, pp. 7-39, 1997.
[37] J. K. Gupta, M. Egorov, and M. Kochenderfer, "Cooperative multi-agent control using deep reinforcement learning," *International Conference on Autonomous Agents and Multiagent Systems*, 2017, pp. 66–83.
[38] R. S. Daniel, "Alpha and theta EEG in vigilance," *Perceptual and Motor Skills*, vol. 25, no. 3, pp. 697-703, 1967.
[39] Z. Cao, C. Chun-Hsiang, K. Jung-Kai, and L. Chin-Teng, "Multi-channel EEG recordings during a sustained-attention driving task," *Scientific Data*, vol. 6, no. 1, pp. 1-8, Dec 2019
[40] O. Reyes, and S. Ventura, "Performing multi-target regression via a parameter sharing-based deep network," *International journal of neural systems*, vol. 29, no. 09, pp. 1950014, 2019.
[41] G. Y. Park, et. al., "Neural correlates of spatial and nonspatial attention determined using intracranial electroencephalographic signals in humans," *Human Brain Mapping*, vol. 37, no. 8, pp. 3041-3054, 2016.
[42] V. Sze, C. Yu-Hsin, Y. Tien-Ju, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," in *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295-2329, 2017.
[43] J. J. Newson, and T. C. Thiagarajan, "EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies," *Frontiers in Human Neuroscience*, vol. 12, no. 521, 2019-January-09, 2019.
[44] Y. Xiong, N. Liu, Z. Xu, and Y. Zhang, "A parameter partial-sharing CNN architecture for cross-domain clothing retrieval," *2016 Visual Communications and Image Processing (VCIP)*, 2016, pp. 1-4.
[45] D. Dai, L. Yu, and H. Wei, "Parameters Sharing in Residual Neural Networks," *Neural Processing Letters*, vol. 51, no. 2, pp. 1393-1410, 2020.
[46] R.-C. Wu, C.-T. Lin, S.-F. Liang, T.-Y. Huang, Y.-C. Chen, and T.-P. Jung, "Estimating driving performance based on EEG spectrum and fuzzy neural network," *2004 IEEE International Joint Conference on Neural Networks*, 2004, pp. 585-590.
[47] C.-T. Lin, I.-F. Chung, L.-W. Ko, Y.-C. Chen, S.-F. Liang, and J.-R. Duann, "EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 7, pp. 1349-1352, 2007.
[48] A. Delorme, and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9-21, 2004.
[49] C.-T. Lin, C.-H. Chuang, Y.-C. Hung, C.-N. Fang, D. Wu, and Y.-K. Wang, "A driving performance forecasting system based on brain dynamic state analysis using 4-D convolutional neural networks," *IEEE transactions on cybernetics*, vol. 51, no. 10, pp. 4959-4967, 2021.
[50] L. N. Boyle, J. Tippin, A. Paul, and M. Rizzo, "Driver performance in the moments surrounding a microsleep," *Transportation research part F: traffic psychology and behaviour*, vol. 11, no. 2, pp. 126-136, 2008.
[51] C. Dissanayaka, D. Cvetkovic, C. R. Patti, S. S. Shahrbabaki, B. Ahmed, C. Schilling, and M. Schredl, "Sleep onset detection with multiple EEG alpha-band features: comparison between healthy, insomniac and schizophrenic patients," *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2015, pp. 1-4.
[52] M. Simon, E. A. Schmidt, W. E. Kincses, M. Fritzsche, A. Bruns, C. Aufmuth, M. Bogdan, W. Rosenstiel, and M. Schrauf, "EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions," *Clinical Neurophysiology*, vol. 122, no. 6, pp. 1168-1178, 2011.
[53] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2352-2359, 2009.
[54] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. Vaughn, "The AASM manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 2012.
[55] G. Keklund, and T. ÅKerstedt, "Objective components of individual differences in subjective sleep quality," *Journal of sleep research*, vol. 6, no. 4, pp. 217-220, 1997.

Chin-Teng Lin received a Bachelor of Science from National Chiao-Tung University (NCTU), Taiwan in 1986, and holds Master's and PhD degrees in Electrical Engineering from Purdue University, USA, received in 1989 and 1992, respectively. He is currently a distinguished professor and Co-Director of the Australian Artificial Intelligence Institute within the Faculty of Engineering and Information Technology at the University of Technology Sydney, Australia. He is also an Honorary Chair Professor of Electrical and Computer Engineering at NCTU. For his contributions to biologically inspired information systems, Prof Lin was awarded Fellowship with the IEEE in 2005, and with the International Fuzzy Systems Association (IFSA) in 2012. He received the IEEE Fuzzy Systems Pioneer Award in 2017. He has held notable positions as editor-in-chief of IEEE Transactions on Fuzzy Systems from 2011 to 2016; seats on Board of Governors for the IEEE Circuits and Systems (CAS) Society (2005-2008), IEEE Systems, Man, Cybernetics (SMC) Society (2003-2005), IEEE Computational Intelligence Society (2008-2010); Chair of the IEEE Taipei Section (2009-2010); Distinguished Lecturer with the IEEE CAS Society (2003-2005) and the CIS Society (2015-2017); Chair of the IEEE CIS Distinguished Lecturer Program Committee (2018-2019); Deputy Editor-in-Chief of IEEE Transactions on Circuits and Systems-II (2006-2008); Program Chair of the IEEE International Conference on Systems, Man, and Cybernetics (2005); and General Chair of the 2011 IEEE International Conference on Fuzzy Systems. Prof Lin is the co-author of Neural Fuzzy Systems (Prentice-Hall) and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning (World Scientific). He has published more than 390 journal papers including over 180 IEEE journal papers in the areas of neural networks, fuzzy systems, brain-computer interface, multimedia information processing, cognitive neuro-engineering, and human-machine teaming, that have been cited more than 28,600 times. Currently, his h-index is 78, and his i10-index is 354.

Jia Liu received the B.S. (Honours) degree in Telecommunications Engineering from Beijing Information Science & Technology University (BISTU), China, in 2015, and the Master of Telecommunications Engineering from The University of Melbourne, Australia, in 2017. She is currently a Ph.D. student in University of Technology Sydney, Australia. Her research interests include brain–computer interface, mobile brain/body imaging (MoBI), cognitive neuro-engineering.

Chieh-Ning Fang received the B.S. degree in electrical and computer engineering and the M.S. degree in electrical and control engineering from National Chiao-Tung University in 2014 and 2016, respectively. Her research interests cover deep learning, neural network, image recognition, and pattern recognition. Future research aims at developing a brain–computer interface associated with deep learning algorithm to improve its performance.

Shih-Ying Hsiao received the B.S degree in Electrical Engineering from Tunghai University, Taichung, Taiwan in 2016, and the M.S. degree in electrical and control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2019. Her current research interests include brain–computer interface, deep learning, image recognition, and pattern recognition.

Yu-Cheng Chang received the B.S. degree in vehicle engineering from the National Taipei University of Technology, Taipei, Taiwan, in 2008, the M.S. degree with a specialization in system and control from the Department of Electrical Engineering, National Chung Hsing University (NCHU), Taichung, Taiwan, in 2010. He is currently working toward the Ph.D. degree in computer science with the University of Technology Sydney, Ultimo, NSW, Australia. From 2014 to 2016, he was a Research Assistant with the Department of Electrical Engineering, NCHU. His current research interests include fuzzy systems, human–machine autonomous, and brain state analysis.

Yu-Kai Wang received the M.S. degree in biomedical engineering and the Ph.D. degree in computer science in 2009 and 2015, respectively, both from the National Chiao Tung University, Taiwan. He is currently a Senior Lecturer and core member of the Australia Artificial Intelligence Institute within the Faculty of Engineering and Information Technology at University of Technology Sydney, Australia. He has organised the Computational Intelligence and Brain Computer Interface symposium in IEEE Symposium Series on Computational Intelligence (2020-2022). His current research interests include computational neuroscience, human performance modelling, Brain-Computer Interface and human-agent interaction.