Comparative Study of Wet and Dry Systems using EEG-based Convolution Neural Network

Taweesak Emsawas (✉ taweesak@ai.sanken.osaka-u.ac.jp)

   Osaka University   https://orcid.org/0000-0003-0115-9251

Tsukasa Kimura

   Osaka University

Ken-ichi Fukui

   Osaka University

Masayuki Numao

   Osaka University

Research

RESEARCH

Comparative Study of Wet and Dry Systems using EEG-based Convolution Neural Network

Taweesak Emsawas[1]*
, Tsukasa Kimura[2]
, Ken-ichi Fukui[2]
and Masayuki Numao[2]

---

*Correspondence:
taweesak@ai.sanken.osaka-u.ac.jp
[1]Graduate School of Information
Science and Technology, Osaka
University, Osaka, Japan
Full list of author information is
available at the end of the article

Abstract

Brain-Computer Interface (BCI) is a communication tool between humans and systems using electroencephalography (EEG) to predict certain cognitive state aspects, such as attention or emotion. For brainwave recording, there are many types of acquisition devices created for different purposes. The wet system conducts the recording with electrode gel and can obtain high-quality brainwave signals, while the dry system expressly proposes the practical and ease of use. In this paper, we study a comparative study of wet and dry systems using two cognitive tasks: attention and music-emotion. The 3-back task is used as an assessment to measure attention and working memory in attention studies. Comparatively, the music-emotion experiments are used to predict the emotion according to the subject's questionnaires. Our analysis shows the similarities and differences between dry and wet electrodes by calculating the statistical values and frequency bands. Besides, we further study the relative characteristics by conducting the classification experiments. We proposed the end-to-end models of EEG classification, which are constructed by combining EEG-based feature extractors and classification networks. A deep convolution neural network (Deep ConvNet) and a shallow convolution neural network (Shallow ConvNet) were applied as the feature extractor of temporal and spatial filtering from raw EEG signals. The extracted feature is then forwardly conveyed to a long short-term memory (LSTM) to learn the dependencies of convolved features and classify attention states or emotional states. Additionally, transfer learning was utilized to improve the performance of the dry system by using transferred knowledge from the wet system. We applied the model not only on our dataset but also on the existing dataset to verify the model performance compared with the baseline techniques and the-state-of-the-art models. Using our proposed model, the result shows the significant differences between accuracy and chance level in attention classification (92.0%, S.D. 6.8%) and SEED dataset's emotion classification (75.3%, S.D. 9.3 %).

Keywords: Brain-Computer Interface (BCI); Electroencephalography (EEG); Attention Classification; Music-Emotion Classification; Convolution Neural Networks

1 Introduction

Brain-Computer Interface (BCI) is the concept of a communication tool between humans and systems by using a physiological signal, especially brain signals, to translates or interpret the neuronal information into a cognitive state such as at-

tention or emotion [1]. BCI and the BCI-related, such as brain-machine interface (BMI) [2], are on the mainstream of emerging technologies from 2010 to 2020, according to the Gartner hype cycle[1], and possibly achieves the plateau reached in more than 10 years. For brain monitoring, the electrical activity's recording method is an electroencephalography (EEG), which uses electrodes attached to the scalp and measures voltage fluctuations from ionic channels. These signals are amplified to fitting the processing and then digitized to machine-interpretable data [3].

There are many sensors built to record EEG signals and use in various disciplines, both medical and non-medical purposes [4]. The medical propose conducts the EEG recording using a wet system, the EEG recording with placing electrodes on the scalp with a conductive gel [5]. Many researchers typically use wet electrodes in expert disciplines, such as clinical neurology, neurobiology, and medical physics. For example, these signals can analyze the dynamic changes in Alzheimer's disease [6] and Parkinson's disease [7] directly through the statistical values or frequency band. Moreover, it can utilize the classification tasks by using machine learning techniques to diagnose epilepsy patients [8]. A wet system is a useful tool for these analyses, but it is meticulous in using and need time and cost for data recording. Applying the concept of BCI, a wearable requirement is needed. The dry sensors are generally designed based on wearable and mobility concepts and use in practical application and non-medical tasks to investigate, entertain, or monitor brain activity [9][10][11]. Nevertheless, the dry system's recorded signal might be worse than the wet system because it is more sensitive to environmental noise and the artifacts [12][13]. Many studies applying dry sensors instead of wet sensors can achieve appropriately and accurately accomplishment [14], which relates to the tremendous growth of technology. Moreover, machine learning techniques, especially deep learning, can automatically and effectively deal with noise without any technical and expert signal processing. Thus, according to earlier points, the combination of dry sensor and machine learning will be the future of non-invasive BCI with low-cost and high-performance systems.

In EEG-based cognitive recognition, it is required processing to proceed from raw signals to cognitive labels orderly, such as data preprocessing, data cleaning, feature extraction, and classification. The time-frequency domain is the most popular technique from [15] and [16] reviews for feature extraction, such as power spectral density (PSD). Meanwhile, a convolutional neural network (ConvNet) can efficiently perform deep feature learning by using the advantage of the hierarchical pattern, notably successfully applying in computer vision [17]. Many studies apply Deep ConvNets with a variety of different architectures. For EEG feature learning, the Shallow Convnet [18], which is inspired by filter bank common spatial pattern (FBCSP) [19], specifically adapted to decode band power features. The main idea is employing temporal and spatial filtering with convolution layers to extract the time-frequency features from raw signals. The extracted output is generally and directly connected to the classification layer. One of the effective time series classification networks is a long short-term neural (LSTM) network [20]. LSTM architecture can preserve temporal dependencies by using their memory blocks and gates. Combining ConvNet and LSTM can effectively extract the temporal-spatial features, such

---

[1]https://www.gartner.com/

as ConvNet with LSTM in the housing energy consumption [21]. Furthermore, we have found that BCI researchers face the data quantity problem because there are time, cost, and effort limitations. According to the recent review of emotion recognition using EEG signals [22], the number of subjects is varied considerably from 1 to 161 subjects with a median of 15 subjects, and the works with less than 15 subjects are 47 percent from 99 selected papers. Due to the quantity limitation, the concept of transfer learning is used as a learning paradigm to acquire knowledge from one system and apply it to solve the related system [23].

This paper contributes to two studies: (1) the data analysis of wet and dry systems, and (2) the EEG-based classification model. We conducted two cognitive experiments to study a comparison of wet and dry systems. The attention experiment assessed using a 3-back task and learned the attention/non-attention classes of EEG signals. For more difficult BCI-task, we learned the positive and negative emotions of EEG signals by stimulating with the MIDI song. In terms of classification, the proposed method utilizes Deep ConvNet and Shallow ConvNet to perform feature extractor and LSTM networks to perform classification learning. These classification results also show the dependencies among datasets and systems comparatively. To improve dry system performance, we further applied transfer learning by freezing the ConvNet of the wet system and used it as a feature extractor of the dry system. Because our dataset is considerably small, collecting from 5 subjects, so we also verify the model performance by implementing on the existing datasets. DEAP [24] and SEED datasets [25] are top 2 citation datasets on EEG-based emotion recognition [1]. Thus, this study is expected to contribute to developing the combination of non-invasive BCI with a dry electrode and machine learning techniques for a future brainwave translation tool.

In this paper, we first review the related researches and techniques in Section 2. For the proposed model, we applied on self-dataset and existing datasets. The self-dataset consists of 2 cognitive tasks described the data acquisition, preprocessing, and feature extraction in section 3. Section 4 discusses the data analysis and differential comparison of wet and dry systems on both self-datasets. Then, the proposed model, Shallow ConvNet with LSTM network, is explained in section 5. The experiments and results are systematically presented in section 6. Our discussions on this study can be found in section 7. Finally, the conclusion and future work are described in section 8.

2 Related Works

2.1 Frontal Brain Electrical Activity

Frontal brain electrical activity is related to cognitive skills, such as emotional expression, working memory, and working memory capacity. Many studies attempt to achieve cognitive recognition with various stimuli by focusing on frontal brain activity. Trainor [26] found that the pattern of asymmetrical frontal EEG activity distinguished valence of the musical excerpts and higher relative left frontal EEG activity to joy and happy musical excerpts. In the research, overall brain activity in the frontal region shows a relation with musical stimuli significantly, consistent with many of the literature's findings [27]. This study not only deals with the emotional state but also the working memory [28]. The frontal theta activity plays an active

role in working memory. Theta activity increases while doing the working memory task [29].

To investigate an association between brain activity and working memory, Wayne Kirchner introduced the n-back task [30]. It is a performance task that is commonly used as an assessment in cognitive research. In the n-back task, The subject is requested to memorize a single character's sequence on the screen continuously. The subject task is to indicate if the current character is the same as the previous n-step before.

2.2 Segmentation

In EEG-signals, there are several techniques to investigate the reaction between the raw signal and cognitive states. The event-related potential (ERP) is a powerful method to allow the observation to reflect specific cognitive processes within the defined period [31]. This technique is restricted to the concrete time-related, not on the continuously recording in practical application. Pragmatically, recording the whole raw signals and interpreting the particular responses is more realistic and useful for future non-invasive BCI applications. However, it requires the necessary processing to separate and investigate a significant stage automatically. The most popular technique is the window segmentation that slices the raw signal into a similar window size and slides it with or without overlapping [32]. Accordingly, we can observe the cognitive stages from each sliding window comparatively. Moreover, we can also observe the relationship between the sliding window sequence [11]. Besides, the continuous recognition while doing the task was an effective method for tracking the cognitive reporting oscillations and provides an opportunity to understand human emotional processes better [33]. However, we need to consider the possibility of acquiring continuous emotional states.

Therefore, this paper utilized the window segmentation to acquire the input data for all model comparison. In sequence-to-sequence learning, window recognition performs the model to classify the window sequence and label sequence pairs individually. On the other hand, the sequence-to-label learning accumulates the window-predicted results into the designated class.

2.3 Traditional Approaches

In traditional approaches of both ERP or window learning, we indispensably need efficient techniques for signal processing and feature extraction. In the study of non-stationary signals like EEG, time-frequency methods can bring up additional information by considering dynamical changes. The power spectral density (PSD) is the common applicable time-frequency feature presented in continuous spectra, which utilizes the fast Fourier transform (FFT) to calculate the spectrogram. The PSD integral over a specified frequency band computes the average power in the signal over that frequency band. For example, it is defined as theta, alpha, and beta bands with slightly different ranges.

Then, the extracted features are classified by the machine learning model. Many classification models are used in EEG-based recognition. In this paper, we apply the support vector machine (SVM) and random forest (RF), which is successfully

achieved the state-of-the-art in many time-series classification [34]. SVM is the discriminative classifier that utilizes the labeled data to optimize the decision boundary. RF is an ensemble classifier that operates by assembling the aggregation of decision trees. It can reduce the overfitting problem in normal decision trees and reduce the variance and, therefore, improve accuracy.

2.4 EEG-based Convolution Neural Networks

Deep Convolutional Neural Networks (Deep ConvNets) have reformed the computer vision research field with the learning ability of hierarchical patterns (Krizhevsky et al. 2012). Then, Deep ConvNets have been successfully applied and achieved in many research fields. On the EEG-based convolution network, we generally apply the 3 types of convolution operation through the signals: temporal, spatial, and temporal-spatial filters. The shallow convolutional neural networks (Shallow ConvNet)[18] is inspired by filter bank common spatial patterns (FBCSP). The main idea of FBCSP is transformations by combining the multi-filtering such as frequency filtering, spatial filtering, and feature selection. To imitate the same, Shallow ConvNet constructed band power feature extraction with 2 convolutional layers that perform temporal convolution and spatial filtering. The feature extractor layer is then combined with the pooling layer and classification layer in the final step. This research shows the study of attention using Shallow ConvNet and fully-connected network [35].

2.5 Existing Datasets

We are not only interested in the quality of recorded signals, but the quantity of data in BCI research is also a tremendous problem according to the cost and time manipulation, especially the wet system. Fortunately, the quantity of our research may not be at a reliable level. We further study the well-known and open-access datasets to increase the reliability of our proposed method.

This study introduces DEAP [24] and SEED [25] datasets, which are widely used in emotion recognition using physiological signals. DEAP is a multimodal dataset, including 32 EEG channels and peripheral physiological signals. The raw EEG signals were applied preprocessing, bandpass filtering, EOG artifact removal. Then the signals were downsampled from 512 to 128 Hz. There are 32 subjects recorded as each watched 40 one-minute highlights of music excerpts. After recording, the subjects were asked to rate arousal, valence, and dominance questionnaires under the continuous scales from 1 to 9. On the other hand, the SEED dataset recorded the 32 EEG channels of 15 subjects while watching 15 film-excerpts for 3 sessions. The film-excerpts stimulate positive, negative, and neutral emotion with the same amount.

3 Self-dataset Acquisition

This study collected dataset by ourselves from 5 healthy subjects, which are 4 males and 1 female. They are graduate students of Osaka University, between 20 and 30 years of age. The subjects were asked to finish two different cognitive tasks: attention experiments using the 3-back task and music-emotion experiments using the brAInMelody application. Meanwhile, the EEG signals were recorded while doing these cognitive tasks. The experiments were conducted twice in the corresponding task and setting but properly using different dry and wet systems.

3.1 Sensors

In this study, the experiments were conducted with two systems comparatively. The wet device is Polymate AP1532 (Miyuki Giken, Japan), an electrode cap (Easycap GmbH, Germany) and Ag/AgCl electrodes with a cable connection to AP monitor program (Miyuki Giken, Japan). The sampling rate is set to 1000 Hz. On the other hand, the dry device is an imec sensor named EEG brAInMelody Gen-3 compatible[2] with brAInMelody application and Nyx software. The sampling rate of the sensor is 256 Hz. EEG placement is in accordance with the international 10-20 system. Ten electrodes (Fz, Fpz, T3, F7, F3, Fp1, Fp2, F4, F8, and T4.) are placed in the frontal brain region. The Fpz and Fz are set as ground and reference electrodes. Consequently, The total signals gained are 8 signals.

3.2 Attention Experiment Setting using 3-Back Task

In the experiments, the subject performed the 33 repetitions of a 3-back task, which is memorizing the 3-back character. Firstly, the subject was introduced by instructions and then wore the sensor. Before doing the task, the '+' character appeared in the middle of the screen for 20 seconds. At the same time, the brain signals began to be recorded as a baseline. Then, the 33 repetitions of the 3-back task started sequentially. Each repetition consists of a 0.5-second character and a 2-second blank screen. Additionally, the experiment conducted in a closed room with minimal noise, and the subject was asked to minimize their movement to avoid the noises. To compare 2 systems, all subjects performed the same task 2 times, with a different random set of characters, in dry and wet settings. Fig.1 shows the illustration and detail of this experiment.

This study classifies the recorded signals into 2 classes: attention and non-attention. The recorded signal of the 3-back task is an attention class, and the baseline signal is a non-attention class.

3.3 Music-Emotion Experiment Setting

For the music-emotion experiment, the introductory and environment are the same as the previous attention experiment. Each subject was asked to listen to 5 MIDI songs. Each song was added a 20-second baseline with no sound before the music started. After listening, the subjects immediately filled in the questionnaires, consisting of valence and arousal level between -1 to 1. They performed the same playlist for 2 times in a dry setting and a wet setting, respectively. The subjects also filled in the questionnaires for 2 times because their feelings might slightly change in the second round. Moreover, the subject is required to close the eyes while recording a baseline and listening to music. The Fig.2 shows the illustration and detail of this experiment.

In the music-emotion experiment, we study to investigate the emotional state from the EEG signals. The target labels or emotions obtained from the subject's questionnaires are valence and arousal values from -1 to 1. Because the emotion label is ambiguous, the study merely classifies 2 classes: positive and negative. The positive class represents positive valence and arousal, and the rest is a negative class.

---

[2]https://www.imec-int.com/en/articles/imec-and-holst-centre-introduce-eeg-headset-for-emotion-detection

3.4 Data Preprocessing

In our recorded signals, there are missing values caused by the device connection. So, the signals were manipulated by duplicating the previous 8 samples. Then, we applied a notch filter to remove the 60 Hz power line noise and applied artifact removal by using the EEGLAB [36] toolbox to avoid the severe contamination of EEG signals automatically. The sampling rate of wet sensors was downsampled into 256, which is equivalent to the dry sensors. We continue to apply sliding window techniques with 2-second window size and 1-second overlapping for segmentation and sequential analysis. For the baseline classification, the PSD was calculated from each window sequentially as the extracted features: theta band (4-7 Hz.), alpha band (8-12 Hz.), and beta band (13-32 Hz.). The process is shown at the top of Fig. 3. While ConvNet's feature extraction of each window, the convolution is applied to the raw signals for learning the temporal and spatial filtering automatically and directly without any signal preprocessing. The learning ability of a neural network can also deal with noises or artifacts from unprocessed signals.

In the attention dataset, the data sequence individually consists of 84 pairs of input data and labels obtaining from 19 baseline windows and 65 doing-task windows. Each system entirely contains 420 pairs of window input and label output. Thus, there are 25 pairs of train and test samples for each system, but the input data was sliced into sliding windows, and all of them were paired with the same emotion label. Each system of music-emotion dataset totally consists of 870 pairs of input windows and labels.

In DEAP and SEED dataset, the input was initially processed by the provider, as mentioned in section 2.5. We then applied a 2-second window sliding with a 1-second overlap to slice the raw data into window input for classification tasks. To simplify the DEAP's continuous scales, the rating was classified into two classes of emotion: the positive class is positive arousal and valence, and the rest is the negative class. While in the SEED dataset, we selected only the positive and negative classes. Thus, all of our experiments are a binary classification with different chance levels according to the dataset.

4 Data Analysis

This section shows the difference between dry and wet electrodes by calculating the statistical values and frequency bands. Firstly, we investigate the raw signal through the range, mean, and standard deviation, which shows the statistical report of voltages and features calculated from all subjects data in Table 1. These can easily notice the diversity of both systems.

Table 1 shows the significant wet-dry relationship and also the relationship between attention and music-emotion tasks. The wide range of dry system on both tasks shows signal instability and susceptibility of the sensor, but the wet system is more stable and sensible. While considering the task relevances, the attention experiments, the subject is required to do the 3-back task intently with prompt and continuous response. It caused the signal stability, which can observe from the shallow range. In contrast, in the music-emotion experiment's recording session, the subject did not move while listening to music, but the higher range is acquired. These mentions that music listening varies the raw data or gain high frequency than the attention task.

Considering the example of attention experiment, Fig. 4 shows the topoplots and histograms of subject 4. As the related research about the frontal brain activity in section 2, our attention experiment shows the relative result. These findings are consistent with the results of frontal brain activity reported by many studies. The attention class is associated with increasing low-frequency (theta) and mid-frequency bands (alpha), especially left-right electrodes: T3, F7, F8, and T4. Whereas, the high-frequency band, the beta band, decreases when doing the task. These frequencies reflect attentional requests such as alertness and memorization. Unfortunately, there is an inconsistent part of the increasing and decreasing between baseline and task. These might be affected by the baseline focusing, which the subject might highly pay attention to the '+' character. On the other hand, considering the system comparison, we could see both systems' tendencies fortunately. Notably, the theta and beta band of the left-right frontal brain identifies the increase while doing the 3-back task in both wet and dry systems. For the beta band, it also has decrease tendencies comparatively.

On the other hand, in the example of the music-listening task, which shows in Fig. 5, the PSD features of subject 2 is plotted. The subject questionnaire is positive valence and arousal, which is a positive class. The related research associates our observation that the left frontal EEG activity is greater relative to happy and joyful music. We observe that theta and alpha bands have a similarity in increasing while the beta band is contrast. The high-frequency features might be sensitive and contaminated by the environment or unrelated noises. From this observation and investigation, we decided to continue the classification on these datasets in section 6.

5 Proposed models

The proposed model utilizes ConvNet and LSTM network for binary classification of cognitive tasks. For feature extraction, we implemented the convolution network for the brain-signal decoding method. Deep ConvNet and Shallow ConvNet were applied in this experiment to show the decoding. The original Deep ConvNet and Shallow ConvNet performs Fully Connected Networks (FCN) as a classifier in the final layer to discriminate the classes [18]. In this work, we applied the LSTM network instead to perform the convolution features. Firstly, we reshape the raw signals into the 2D shape, consisting of time and electrode channels, as the input for feature learning layers.

The setting of Deep Convnet is shown in Fig. 6. The input is convolved with 24 temporal filters, and then 24 spatial filters are sequentially performed. The deep features of temporal-spatial filtering are extracted by 3 blocks of convolution and pooling layers with twice the number of filters in each block (48, 96, 192). We apply the batch normalization and activation with an exponential linear unit (ELU) for each convolution and pooling layer. On the other hand, the setting of Shallow ConvNet and LSTM networks is shown in Fig. 7. This network is more compatible and less trainable parameters than the previous one and can effectively learn the EEG features. The model contains 40 temporal and spatial filtering constructed by convolution networks continuously. Then, the network applies the batch normalization layer and ELU activation after convolving the features. Finally, we average the values by the pooling layer before sending it to the classification layers. Consequently,

the extracted features of Deep ConvNet or Shallow ConvNet are conveyed to two layers of 50 blocks LSTM networks for learning dependencies of feature sequence and summarized into predicted class. In the final classification layer, the output is activated with a sigmoid function. Their kernel size and pool size are adjusted based on each dataset's window size and sampling rate. Fig. 6 and Fig. 7 show the setting of our dataset which is 512 window size, 256 sampling rate, and 8 electrode channels.

For transfer learning, this study transfers the knowledge of feature extraction from the wet system to the dry system. The implementation is the freezing of feature extractor or ConvNet in the wet system and then apply it to the dry system. Then, we retrain only the LSTM layers to classify the attention or emotion classes.

6 Experiments and Results

This section presents the binary classification learning implemented on self-datasets and existing-datasets. The baseline models are RF and RBF-SVM classifiers combined with PSD features. The state-of-the-art models from [18] are Deep ConvNet and Shallow ConvNet with FCNs. Moreover, the proposed models are Deep ConvNet and Shallow ConvNet with LSTM networks shown in section 5. However, our small datasets may cause an overfitting problem, and it can not transcend the reliability of data quantity. We further applied all techniques on the SEED and DEAP dataset to prove to model performance. All models implemented the leave-one-out (LOO) cross-validation for all subjects to investigate the subject-independent performance. We performed self-implementation of all techniques in the same environment and technical setting. For baseline settings, the grid search was applied for parameter tuning to select the setting in these experiments. All neural network architectures were fitted using Adam optimizer to minimize the binary cross-entropy loss function with 100 epochs and 16 batch size for self-datasets but 64 batch size for existing datasets. The evaluation metric is the accuracy of binary classification even there is the problem of an unbalanced class. To investigate the learning performance in this problem, we present the chance level, which is the majority class's accuracy when predicting constantly. All the results are shown as the accuracy and standard deviation across all subjects in Table 2 and the significant comparison by histograms in Fig. 8.

6.1 Self-Dataset Classification

Based on the overall results in 5 subjects cross-validation, the attention classification is easier classified than music-emotion classification, which is observed by the average accuracy and standard deviation. A high standard deviation can show the subject variation and complexity.

In the attention results shown in Fig. 8a, there are the same chance levels of 77.5% , S.D. 0.0% accuracy in both wet and dry systems. All classification models achieved significantly higher results than the chance level (p<0.01), except for Shallow ConvNet with FCN and Deep ConvNet with LSTM. Our proposed models, Shallow ConvNet with LSTM network, achieved the highest accuracy in both dry and wet attention experiments: 90.0%, S.D. 4.5% and 92.0%, S.D. 6.8%, respectively. We observed that the accuracies of Shallow ConvNet with LSTM are higher

than the chance level (p<0.01) and baseline techniques (p<0.05). Moreover, we can see the improvement of replacing FCN with LSTM network on attention classification, especially combining of Shallow ConvNet and LSTM network. Considering the system comparison, we found that overall performances can learn with higher consistency observed in the models' accuracies and standard deviations.

While in music-emotion experiments, all results, except Deep ConvNet with LSTM in dry system, can not significantly outperform the chance levels (p>0.05), which are of 70.2%, S.D. 16.2% of dry system and 71.5%, S.D. 20.4% of wet system. Classification results are lower than the attention results explicitly. All results are shown in 8b, where we found no significant difference in performance across all models (p>0.05). The highest performance is the Deep ConvNet with LSTM network in dry system (71.7%, S.D. 15.5%) and RF in wet system (70.3%, S.D. 22.5%).

6.2 Wet-Dry Transfer Learning

Applying the transfer learning strategy in section 5, this experiment focused on improving the dry system using the transfer learning from the wet system. In the attention experiment, the new accuracies of Deep ConvNet and Shallow ConvNet with retained LSTM networks are 88.2%, S.D. 3.7% (+6.6% from 81.6%, S.D. 14.0) and 90.0%, S.D. 4.5% (+0.0% from 90.0%, S.D. 4.5%), respectively. While in the music-emotion experiment, the new accuracies are 67.1%, S.D. 18.0% (-4.6% from 71.7%, S.D. 15.5%) and 69.2%, S.D. 19.7% (+1.2% from 68.0%, S.D. 19.3%), respectively. From the results, we can see the small improvement of Deep ConvNet in the attention experiment and Shallow ConvNet in the music-emotion experiment without significance level (p>0.05).

6.3 Existing-Dataset Classification

In the DEAP dataset results, the study classified the binary classes with 64.2%, S.D. 11.2% chance level among 32 subjects. Fortunately, all classification models could not outperform the chance level significantly (p>0.05). However, Shallow ConvNet and RF similarly achieved the highest accuracy with different standard deviations at 60.4%, S.D. 10.1% and 60.4%, S.D. 12.4%, respectively.

For the SEED dataset, the chance level is 50.0%, S.D. 0.0%, which is calculated from the balanced-label stimulus on 15 subjects. All models were superior to the chance level significantly (p<0.01). Again, the Shallow ConvNet achieved the highest accuracy with 75.3%, S.D. 9.3% accuracy. Replaced FCN with LSTM network of ConvNet-based models can improve the performance of both datasets, especially the Shallow ConvNet with LSTM network outperformed on both existing datasets.

7 Discussion

In this section, we discuss on 3 aspects consisting of system comparison, dataset comparison, and model comparison.

7.1 Wet and Dry systems

To acquire and analyze brain activity, we need a device to record EEG brain-signal. In this research, we conducted our experiments with the dry and wet systems on the same tasks to investigate the electrode comparison. According to the result, we can

not deny that wet systems are better than dry systems obviously. For classification results, the wet systems outperformed dry systems, which were observed by the average accuracy. Even though there is a data quantity problem in this study, the statistical analysis in Section 4 shows the relationship between wet and dry systems. Fortunately, the classification results also show related tendencies and characteristics following the corresponding tasks. From the wet-dry transfer learning experiment, we inconsiderably discovered the possibility of using a wearable sensor in a real-world application by transferring the knowledge from laboratory sensors.

7.2 Dataset Comparison

Considering the dataset comparison, this study recorded the EEG signals with two different cognitive tasks and two different systems from 5 subjects. After performing classification with machine learning techniques, we realized that our dataset is too small for machine learning, and It is not reliable to verify the model performance. We further applied our proposed model to the existing datasets, which are more subjects and more stimuli. Based on cognitive tasks, we divided all datasets into attention and emotion datasets. The attention dataset is only the 3-back task, which classifies the attention and non-attention classes, while the rest are emotion classification with different stimuli. In our dataset, the subject is stimulated by music listening. DEAP's stimuli are music video excerpts, and SEED's stimuli are movie excerpts. We found that subjects' emotions were elicited from different sources, and then their labels were obtained from the questionnaires after performing the task. The emotional scale might be different depending on the individual subject's feelings. The variety can be occurred by emotion, subject, or questionnaires individually. However, SEED provided the labels derived specifically from the movie excitation, not from the actual subject's emotion. These may generalize the complexity and affect the classification performance. Our results show that the music-emotion datasets are not explicit separable and more difficult tasks than the attention experiments. Meanwhile, the DEAP dataset is more complicated than the SEED dataset.

7.3 EEG Classification Model

Analyzing the baseline techniques, SVM can not perform properly because of the sensitivity to the noise, while the ensemble modeling, like RF, performs good results in all datasets with multiple decision tree models. It can learn to reduce the generalization error of the prediction, and also can deal with subject-independence of the EEG classification. Additionally, the raw signals might include environmental noise, muscle artifact, or eye-movement artifact. PSD features were directly extracted from these raw signals with predefined and unlearnable preprocessing. It means that the EEG signals with more noise directly and easily affect the learning performance, and it is not appropriate to BCI practical applications. In contrast, considering the ConvNet models, training with neural networks can automatically learn additional preprocessing without predefined knowledge. The architecture of the ConvNets is used as a feature extractor or EEG decoder, and then it needs the added layer for discriminative learning. Instead of using the simple fully connected layer, which individually and independently learns the feature perception on

convolved features, the LSTM network can interpret the dependencies across time steps efficiently. Based on the classification results, it is helpful for temporal-spatial learning like multivariate time-series classification, such as brainwave signals. Considering the results, it might not see the significant differences between these two ConvNet, but the Shallow ConvNet applied less parameter than the Deep ConvNet. Moreover, when adding the sample, we can retrain the model with new additional data with pre-trained knowledge from the existing datasets or subjects.

8 Conclusion

This study presents the comparative study of wet and dry systems by performing 2 cognitive tasks and proposes the EEG-based classification model by constructing the convolution network with LSTM networks. According to the data analysis, we can see the tendency relationship between dry and wet systems, not only on the statistical reports but also on classification results. Besides, the frontal brain activity occurs while performing the attention and music-emotion trials, similar to many literature evidence. We further proposed the model to classify the binary classes for cognitive recognition of emotion and attention. Overall results show that ConvNet with LSTM networks outperformed the baseline with PSD and ConvNet with FCNs on all datasets significantly. Fortunately, our studies are limited to small imbalanced-dataset problems. It directly caused reliability and also the overfitting problem when classified by machine learning techniques. Additionally, we performed experiments on the existing dataset for verifying the proposed model performance to relieve this problem. The result and analysis show that our proposed models mostly outperformed the baseline and the state-of-the art models. Finally, in the future BCI application, we strongly suppose that a non-invasive electrode with a dry system combined with machine learning techniques may become a future brainwave translation tool.

Author details
[1]Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. [2] The Institute of Scientific and Industrial Research (ISIR), Osaka University, Osaka, Japan.

References
1. Al-Nafjan, A., Hosny, M., Al-Ohali, Y., Al-Wabil, A.: Review and classification of emotion recognition based on EEG brain-computer interface system research: A systematic review. Applied Sciences (Switzerland) **7**(12) (2017). doi:10.3390/app7121239
2. Carmena, J.M., Lebedev, M.A., Crist, R.E., O'Doherty, J.E., Santucci, D.M., Dimitrov, D.F., Patil, P.G., Henriquez, C.S., Nicolelis, M.A.L.: Learning to control a brain–machine interface for reaching and grasping by primates. PLOS Biology **1**(2) (2003). doi:10.1371/journal.pbio.0000042
3. Shih, J.J., Krusienski, D.J., Wolpaw, J.R.: Brain-computer interfaces in medicine. Mayo Clinic Proceedings **87**(3), 268–279 (2012). doi:10.1016/j.mayocp.2011.12.008

4. Hochberg, L.R., Donoghue, J.P.: Sensors for brain-computer interfaces. IEEE Engineering in Medicine and Biology Magazine **25**(5), 32–38 (2006)

5. Volosyak, I., Valbuena, D., Malechka, T., Peuscher, J., Gräser, A.: Brain–computer interface using water-based electrodes. Journal of Neural Engineering **7**(6), 066007 (2010). doi:10.1088/1741-2560/7/6/066007

6. Jeong, J.: Eeg dynamics in patients with alzheimer's disease. Clinical Neurophysiology **115**(7), 1490–1505 (2004). doi:10.1016/j.clinph.2004.01.001

7. Klassen, B.T., Hentz, J.G., Shill, H.A., Driver-Dunckley, E., Evidente, V.G.H., Sabbagh, M.N., Adler, C.H., Caviness, J.N.: Quantitative eeg as a predictive biomarker for parkinson disease dementia. Neurology **77**(2), 118–124 (2011). doi:10.1212/WNL.0b013e318224af8d. https://n.neurology.org/content/77/2/118.full.pdf

8. Smith, S.: Eeg in the diagnosis, classification, and management of patients with epilepsy. Journal of neurology, neurosurgery, and psychiatry **76 Suppl 2**, 2–7 (2005). doi:10.1136/jnnp.2005.069245

9. Chi, Y.M., Wang, Y., Wang, Y., Maier, C., Jung, T., Cauwenberghs, G.: Dry and noncontact eeg sensors for mobile brain–computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering **20**(2), 228–235 (2012). doi:10.1109/TNSRE.2011.2174652

10. Liu, N.-H., Chiang, C.-Y., Chu, H.-C.: Recognizing the degree of human attention using eeg signals from mobile sensors. Sensors (Basel, Switzerland) **13**, 10273–86 (2013). doi:10.3390/s130810273

11. Emsawas, T., Fukui, K.-i., Numao, M.: Feasible affect recognition in advertising based on physiological responses from wearable sensors. In: Ohsawa, Y., Yada, K., Ito, T., Takama, Y., Sato-Shimokawara, E., Abe, A., Mori, J., Matsumura, N. (eds.) Advances in Artificial Intelligence, pp. 27–36. Springer, Cham (2020)

12. Duvinage, M., Castermans, T., Dutoit, T., Petieau, M., Hoellinger, T., Saedeleer, C., Seetharaman, K., Cheron, G.: A p300-based quantitative comparison between the emotiv epoc headset and a medical eeg device, vol. 765 (2012). doi:10.2316/P.2012.764-071

13. Guger, C., Krausz, G., Allison, B., Edlinger, G.: Comparison of dry and gel based electrodes for p300 brain–computer interfaces. Frontiers in Neuroscience **6**, 60 (2012). doi:10.3389/fnins.2012.00060

14. Ratti, E., Waninger, S., Berka, C., Ruffini, G., Verma, A.: Comparison of medical and consumer wireless eeg systems for use in clinical trials. Frontiers in Human Neuroscience **11**, 398 (2017). doi:10.3389/fnhum.2017.00398

15. Kim, M.K., Kim, M., Oh, E., Kim, S.P.: A review on the computational methods for emotional state estimation from the human EEG. Computational and Mathematical Methods in Medicine **2013** (2013). doi:10.1155/2013/573734

16. Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., Yang, X.: A review of emotion recognition using physiological signals. Sensors (Switzerland) **18**(7) (2018). doi:10.3390/s18072074

17. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. Neural Information Processing Systems **25** (2012). doi:10.1145/3065386

18. Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T.: Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping **38**(11), 5391–5420 (2017). doi:10.1002/hbm.23730. 1703.05051

19. Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, Cuntai Guan: Filter bank common spatial pattern (fbcsp) in brain-computer interface. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 2390–2397 (2008)

20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**, 1735–80 (1997). doi:10.1162/neco.1997.9.8.1735

21. Kim, T.-Y., Cho, S.-B.: Predicting residential energy consumption using cnn-lstm neural networks. Energy **182**, 72–81 (2019). doi:10.1016/j.energy.2019.05.230

22. Alarcão, S.M., Fonseca, M.J.: Emotions recognition using eeg signals: A survey. IEEE Transactions on Affective Computing **10**(3), 374–393 (2019)

23. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering **22**(10), 1345–1359 (2010). doi:10.1109/TKDE.2009.191

24. Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing **3**, 18–31 (2011). doi:10.1109/T-AFFC.2011.15

25. Zheng, W.-L., Lu, B.-L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Transactions on Autonomous Mental Development **7**(3), 162–175 (2015). doi:10.1109/TAMD.2015.2431497

26. Trainor, L.: Frontal brain electrical activity (eeg) distinguishes valence and intensity of musical emotions. Cognition and Emotion - COGNITION EMOTION **15**, 487–500 (2001). doi:10.1080/02699930126048

27. Lin, Y., Wang, C., Jung, T., Wu, T., Jeng, S., Duann, J., Chen, J.: Eeg-based emotion recognition in music listening. IEEE Transactions on Biomedical Engineering **57**(7), 1798–1806 (2010). doi:10.1109/TBME.2010.2048568

28. Gevins, A.S., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., Rush, G.: Monitoring working memory load during computer-based tasks with eeg pattern recognition methods. Human factors **40** 1, 79–91 (1998)

29. Jensen, O., Tesche, C.: Frontal theta activity in humans increases with memory load in a working memory task. The European journal of neuroscience **15**, 1395–9 (2002). doi:10.1046/j.1460-9568.2002.01975.x

30. Kirchner, W.K.: Age differences in short-term retention of rapidly changing information. Journal of Experimental Psychology **55**(4), 352 (1958). doi:10.1037/h0043688

31. Luck, S.: An Introduction to The Event-Related Potential Technique, (2005)

32. Jenke, R., Peer, A., Buss, M.: Feature extraction and selection for emotion recognition from eeg. IEEE Transactions on Affective Computing **5**(3), 327–339 (2014)

33. THAMMASAN, N., MORIYAMA, K., FUKUI, K.-i., NUMAO, M.: Continuous music-emotion recognition based on electroencephalogram. IEICE Transactions on Information and Systems **E99.D**(4), 1234–1241 (2016). doi:10.1587/transinf.2015EDP7251

34. Karim, F., Majumdar, S., Darabi, H., Harford, S.: Multivariate lstm-fcns for time series classification. Neural

Networks **116** (2018). doi:10.1016/j.neunet.2019.04.014

35. Fahimi, F., Zhang, Z., Goh, W., Lee, T.-S., Ang, K., Guan, C.: Inter-subject transfer learning with end-to-end deep convolutional neural network for eeg-based bci. Journal of Neural Engineering **16** (2018). doi:10.1088/1741-2552/aaf3f6

36. Delorme, A., Makeig, S.: Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. Journal of Neuroscience Methods **134**, 9–12 (2004)
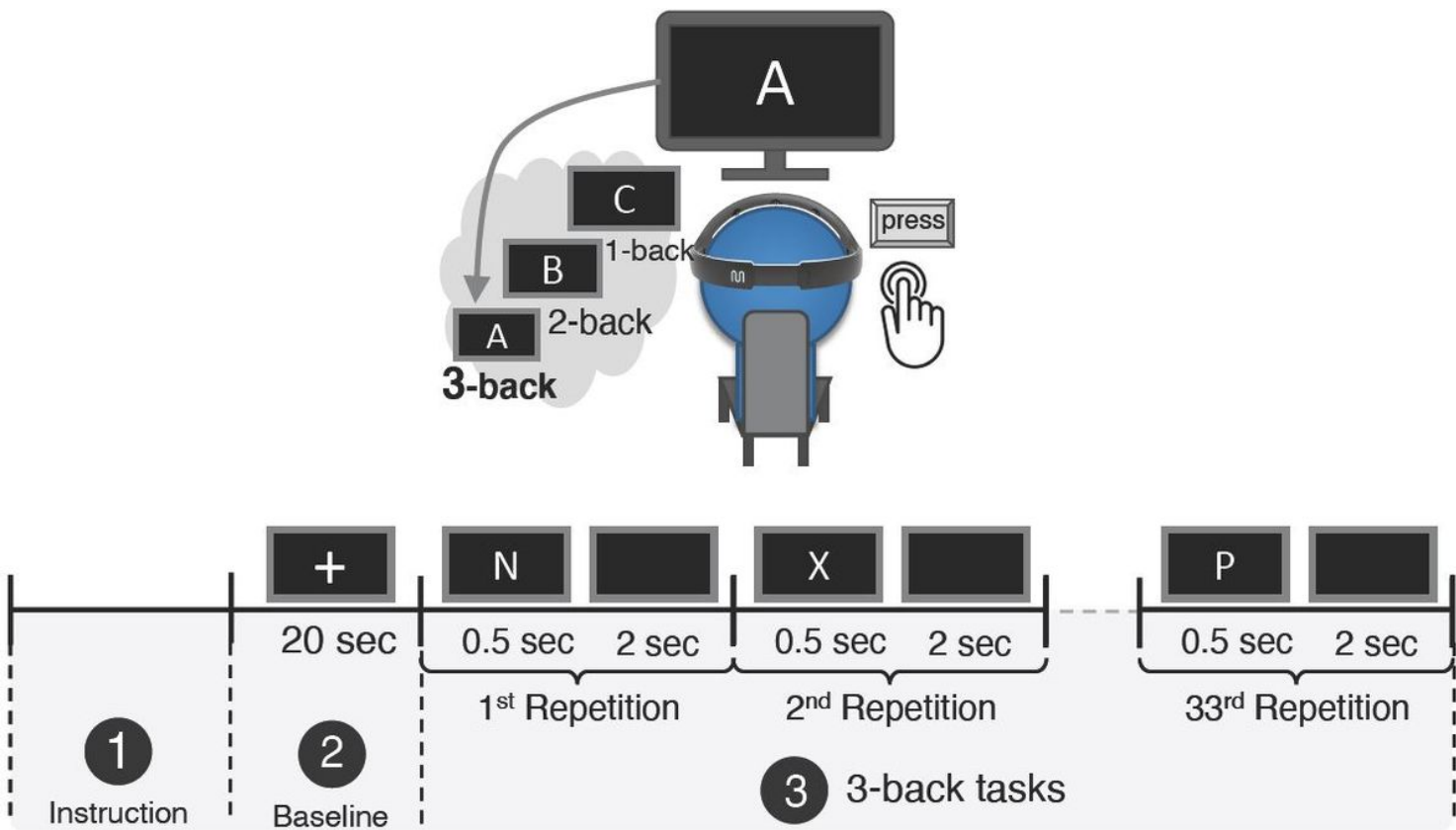
Figures

Figure 1 Attention experiment setting with 3-back task

Figure 2 Music-emotion experiment setting

Figure 3 Data preprocessing and feature extraction

Figure 4 Topoplots and histograms to show a comparison between dry and wet systems on the 3-back task of subject 4. The rows denote the frequency bands: theta, alpha, and beta respectively. The columns denote the baseline and 3-back task sessions, and the last column is different between these sessions.

Figure 5 Topoplots and histograms to show a comparison between dry and wet systems on the music-listening task of subject 2. The rows denote the frequency bands: theta, alpha, and beta respectively. The columns denote the baseline and music-listening sessions, and the last column is different between these sessions.

Figure 6 The network structure of Deep ConvNet and LSTM networks

Figure 7 The network structure of Shallow ConvNet and LSTM networks

Figure 8 Classification results; error bars denote the standard deviation for all subjects and stars indicate a significant improvement compared to chance levels.

Tables

Table 1 The statistical report of voltages and 3 frequency band features in our 4 collected datasets

Task	Type	Raw data (μV)			PSD features ($\times 10^{-11}$)					
					Theta		Alpha		Beta	
		Range	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Emotion-Music	dry	21764.39	-34.65	68.65	9.65	6.21	5.22	3.48	5.61	2.32
	wet	1783.34	0.01	11.06	0.09	0.05	0.08	0.05	0.13	0.07
3-Back Task	dry	727.47	-26.25	39.07	8.64	3.46	4.51	1.40	5.54	1.21
	wet	486.42	-0.04	12.54	0.11	0.06	0.07	0.04	0.18	0.05

Table 2 Classification results determine with accuracy and standard deviation of all subjects. The bolds denote the highest accuracy among the corresponding dataset.

		Baseline with PSD		ConvNet with FCN		ConvNet wi
Dataset	Chance Lvl	RBF-SVM	RF	Deep	Shallow	Deep
Attention (Dry)	77.4, S.D. 0.0	81.8, S.D. 2.1	81.8, S.D. 2.1	82.5, S.D. 12.2	84.6, S.D. 7.7	81.6, S.D. 14.0
Attention (Wet)	77.4, S.D. 0.0	84.2, S.D. 1.0	84.2, S.D. 1.0	88.5, S.D. 3.4	90.5, S.D. 4.4	90.0, S.D. 6.0
Music-Emotion (Dry)	70.2, S.D. 16.2	68.1, S.D. 23.4	69.1, S.D. 19.5	62.4, S.D. 17.3	59.25, S.D. 24.8	**71.7, S.D. 15.5**
Music-Emotion (Wet)	71.5, S.D. 20.4	69.5, S.D. 23.0	**70.3, S.D. 22.5**	67.4, S.D. 21.2	68.7, S.D. 24.7	67.7, S.D. 21.2
DEAP [24]	64.2, S.D. 11.2	53.4, S.D. 7.2	**60.4, S.D. 10.1**	56.4, S.D. 10.0	58.8, S.D. 13.7	59.8, S.D.12.1
SEED [25]	50.0, S.D. 0.0	51.0, S.D. 0.6	74.0, S.D. 8.2	72.0, S.D. 7.8	74.6, S.D. 9.3	73.0, S.D. 7.8

Figures

Figure 1

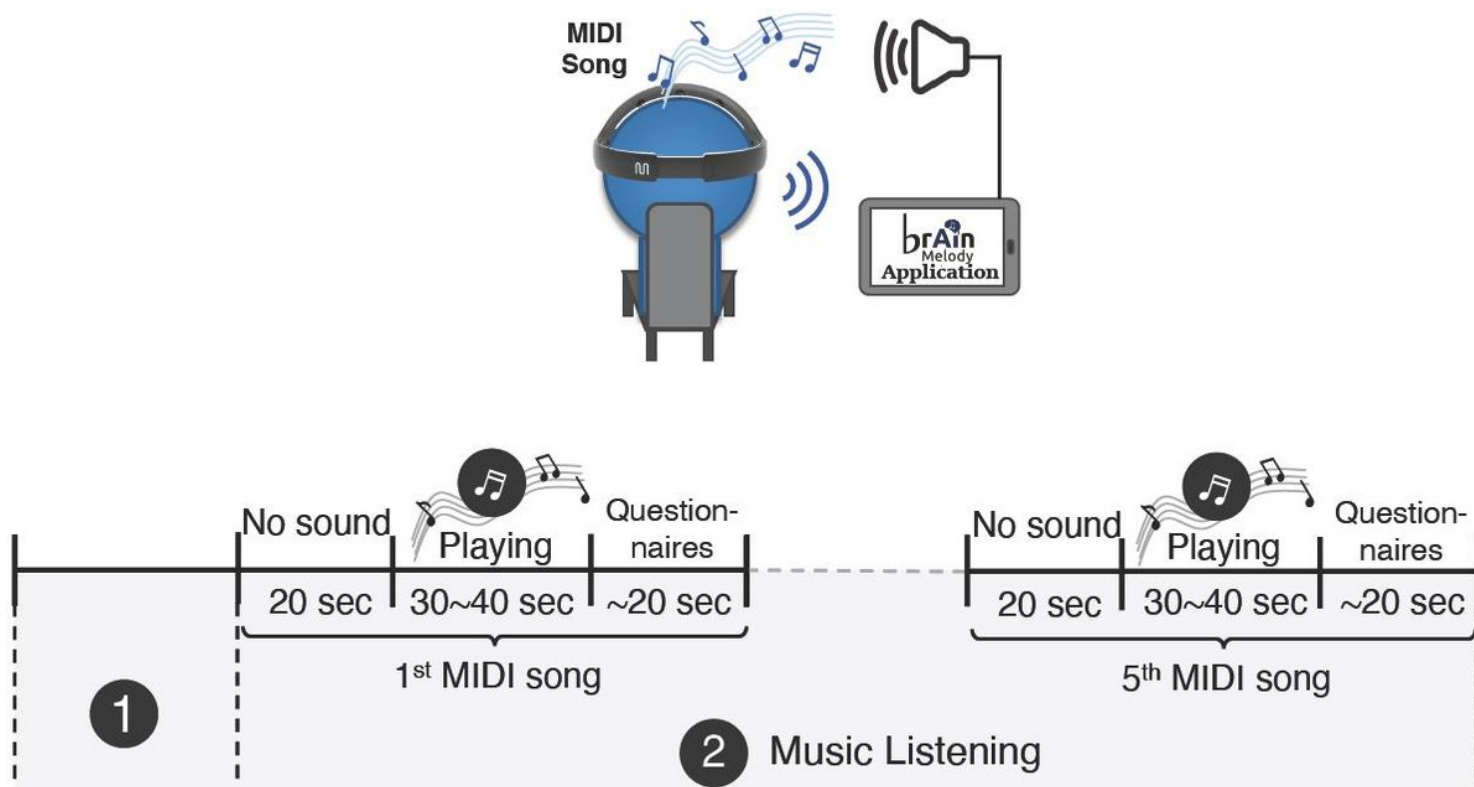Attention experiment setting with 3-back task

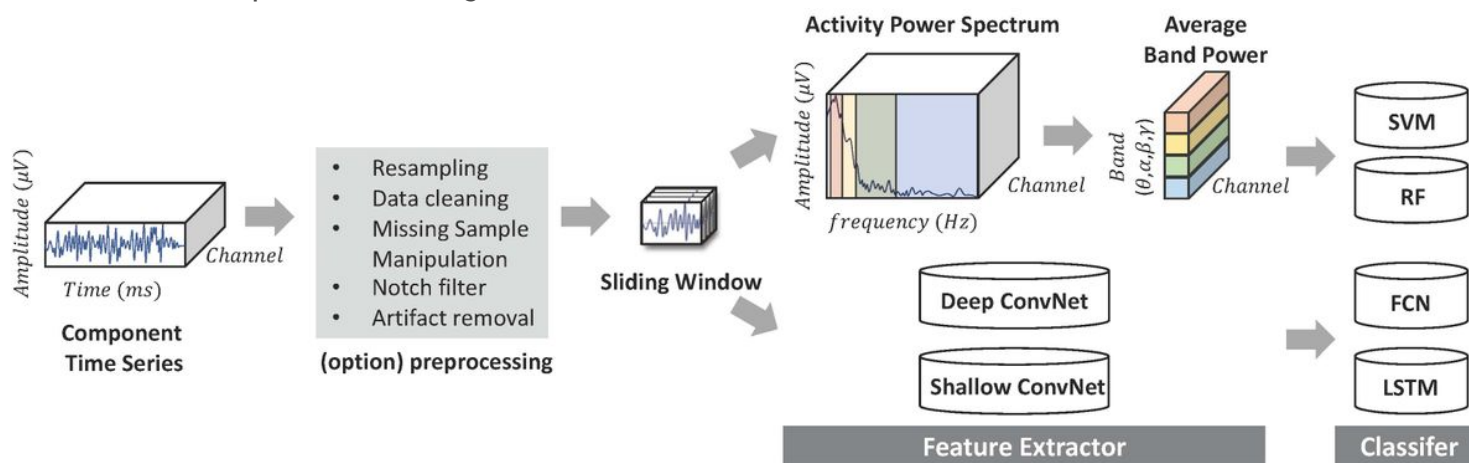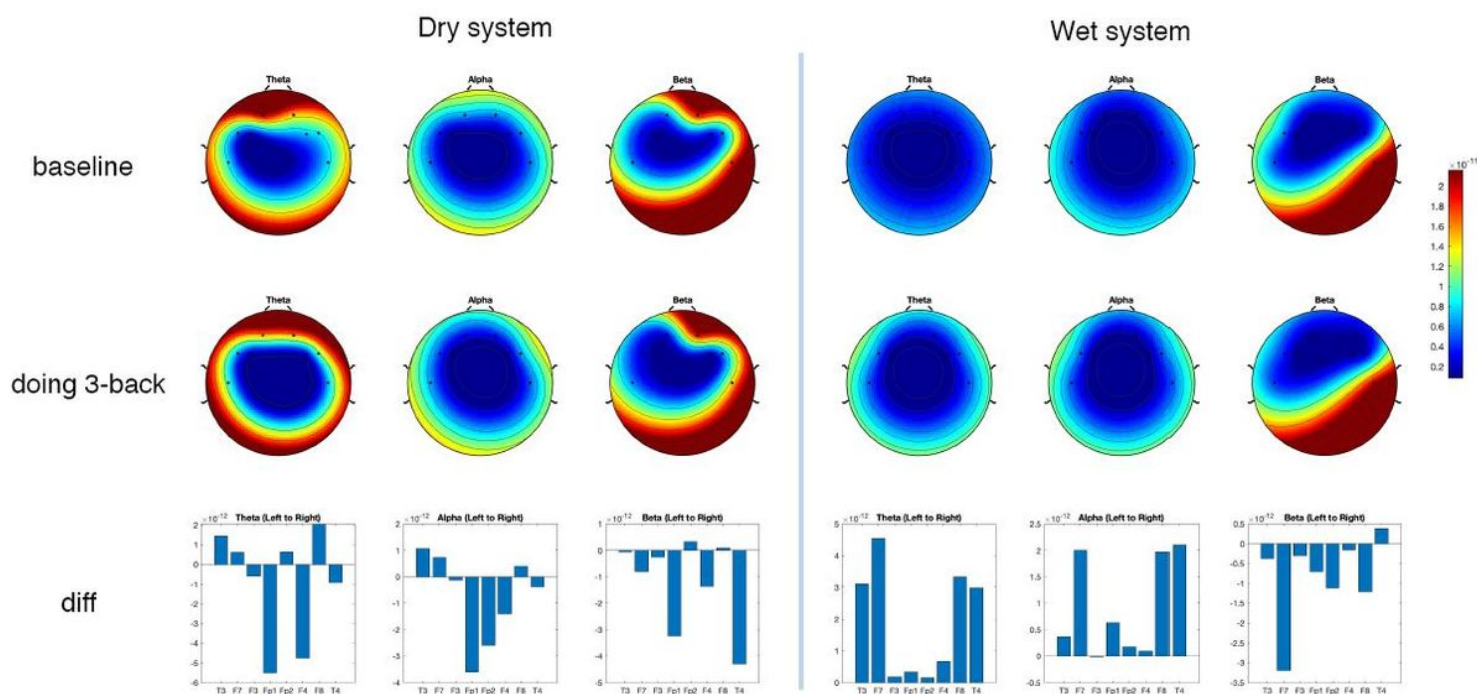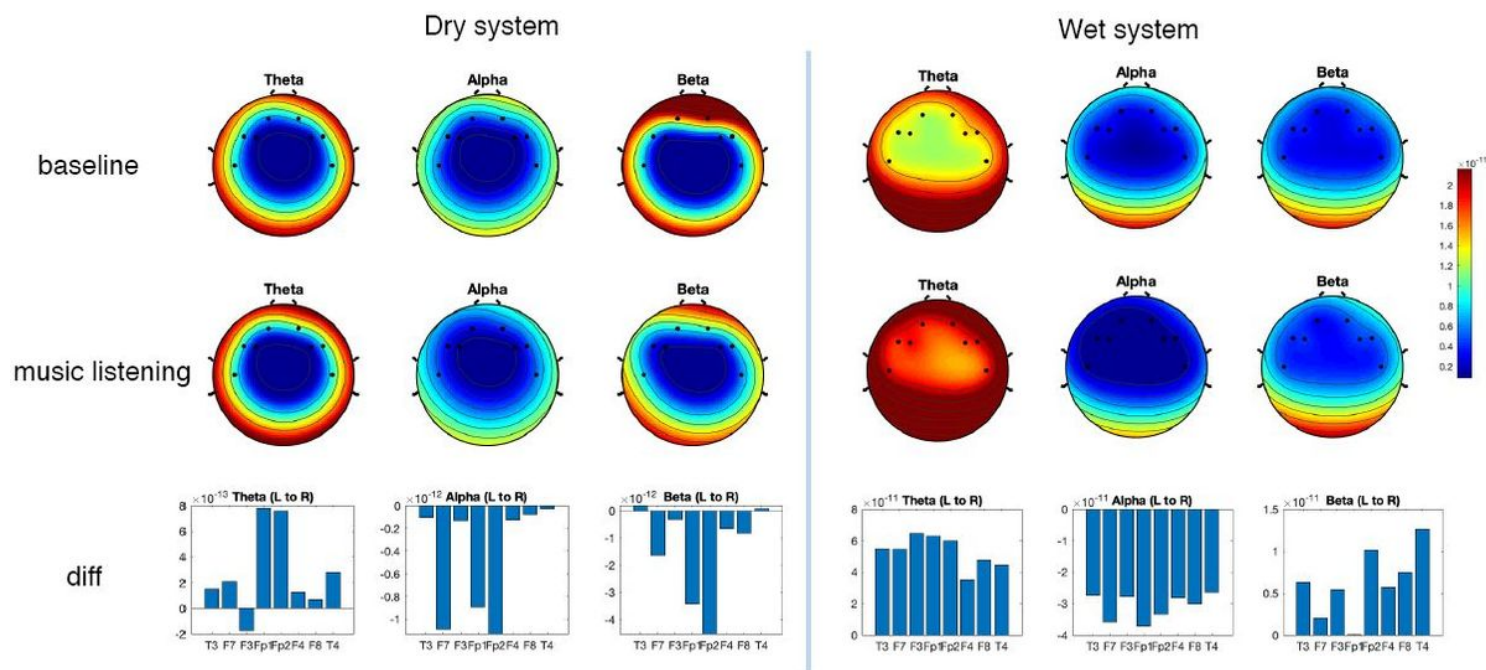**Figure 2**

Music-emotion experiment setting

Figure 3

Data preprocessing and feature extraction

Figure 4

Topoplots and histograms to show a comparison between dry and wet systems on the 3-back task of subject 4. The rows denote the frequency bands: theta, alpha, and beta respectively. The columns denote the baseline and 3-back task sessions, and the last column is different between these sessions.

Figure 5

Topoplots and histograms to show a comparison between dry and wet systems on the music-listening task of subject 2. The rows denote the frequency bands: theta, alpha, and beta respectively. The columns

denote the baseline and music-listening sessions, and the last column is different between these sessions.
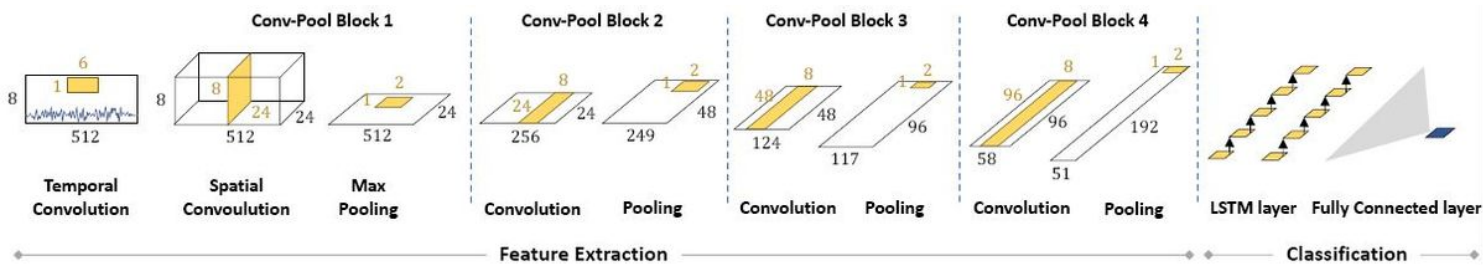

Figure 6

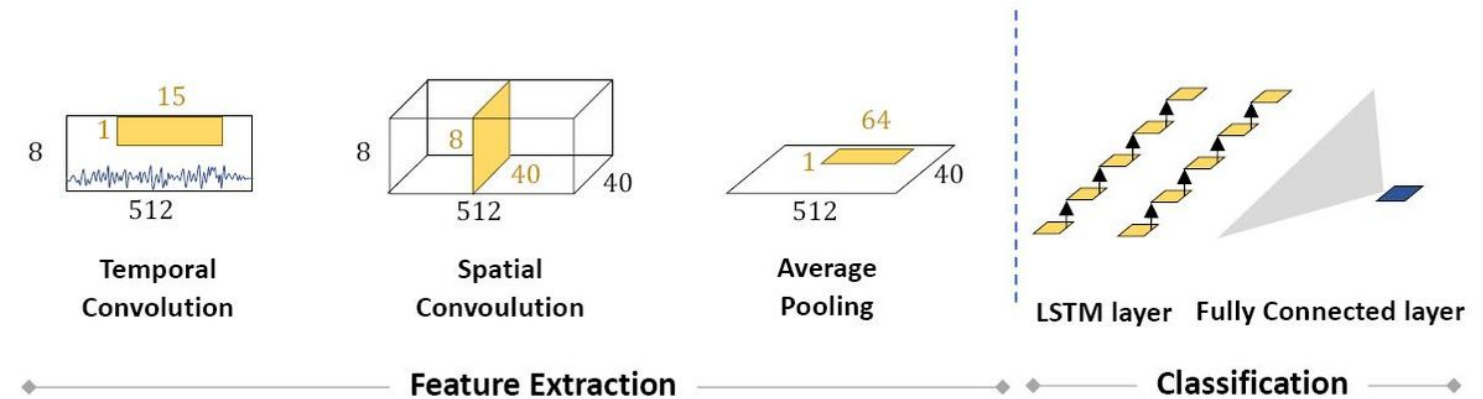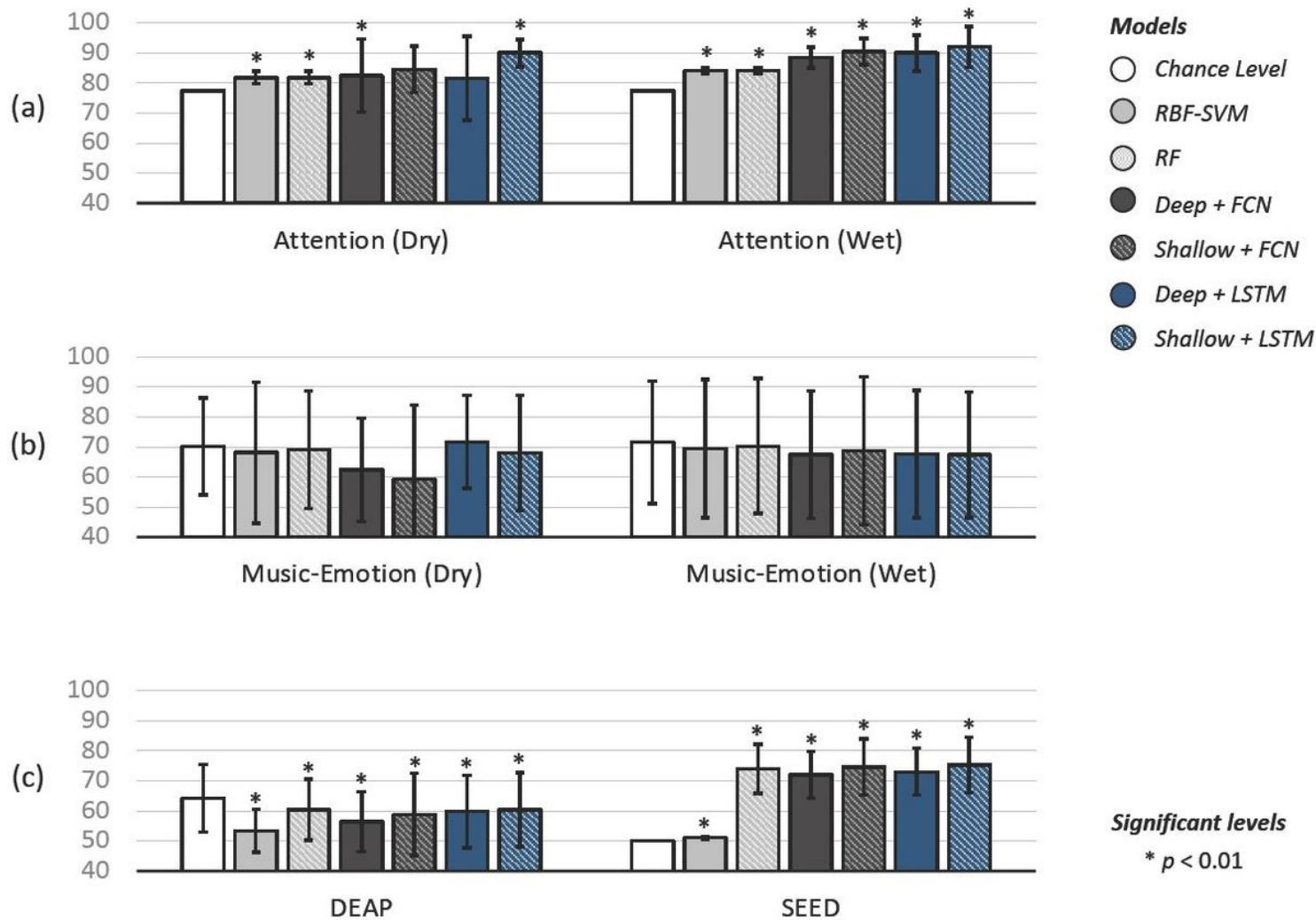The network structure of Deep ConvNet and LSTM networks

Figure 7

The network structure of Shallow ConvNet and LSTM networks

Figure 8

Classification results; error bars denote the standard deviation for all subjects and stars indicate a significant improvement compared to chance levels.