TOPICAL REVIEW • OPEN ACCESS

Deep learning for electroencephalogram (EEG) classification tasks: a review

To cite this article: Alexander Craik *et al* 2019 *J. Neural Eng.* **16** 031001

View the article online for updates and enhancements.

Topical Review

Deep learning for electroencephalogram (EEG) classification tasks: a review

Alexander Craik[1] , Yongtian He and Jose L Contreras-Vidal

Department of Electrical and Computer Engineering, Noninvasive Brain–Machine Interface
System Laboratory, University of Houston, Houston, TX 77204, United States of America

E-mail: alexcraik@gmail.com (A Craik)

CrossMark

Abstract

Objective. Electroencephalography (EEG) analysis has been an important tool in neuroscience
with applications in neuroscience, neural engineering (e.g. Brain–computer interfaces, BCI's),
and even commercial applications. Many of the analytical tools used in EEG studies have used
machine learning to uncover relevant information for neural classification and neuroimaging.
Recently, the availability of large EEG data sets and advances in machine learning have
both led to the deployment of deep learning architectures, especially in the analysis of EEG
signals and in understanding the information it may contain for brain functionality. The
robust automatic classification of these signals is an important step towards making the
use of EEG more practical in many applications and less reliant on trained professionals.
Towards this goal, a systematic review of the literature on deep learning applications to EEG
classification was performed to address the following critical questions: (1) Which EEG
classification tasks have been explored with deep learning? (2) What input formulations
have been used for training the deep networks? (3) Are there specific deep learning network
structures suitable for specific types of tasks? *Approach*. A systematic literature review of EEG
classification using deep learning was performed on Web of Science and PubMed databases,
resulting in 90 identified studies. Those studies were analyzed based on type of task, EEG
preprocessing methods, input type, and deep learning architecture. *Main results*. For EEG
classification tasks, convolutional neural networks, recurrent neural networks, deep belief
networks outperform stacked auto-encoders and multi-layer perceptron neural networks in
classification accuracy. The tasks that used deep learning fell into five general groups: emotion
recognition, motor imagery, mental workload, seizure detection, event related potential
detection, and sleep scoring. For each type of task, we describe the specific input formulation,
major characteristics, and end classifier recommendations found through this review.
Significance. This review summarizes the current practices and performance outcomes in the
use of deep learning for EEG classification. Practical suggestions on the selection of many
hyperparameters are provided in the hope that they will promote or guide the deployment of
deep learning to EEG datasets in future research.

[1] Author to whom any correspondence should be addressed.

(Some figures may appear in colour only in the online journal)

Deep learning strategy

AE	Auto-encoder
CNN	Convolutional neural network
Conv	Convolutional layer
DBN	Deep belief network
FC	Fully connected
Hid.	Hidden layers
Ind.	Independent
LSTM	Long short term memory
MLPNN	Multi-layer perceptron neural network
N/A	Not addressed
OC	Output classes
RBM	Restricted Boltzmann machine
RNN	Recurrent neural network
SAE	Stacked auto-encoders
Strat	Strategy
SVM	Support vector machine

Activation

ReLU	Rectified linear unit
ELU	Exponential linear unit
L-ReLU	Leaky rectified linear unit
SELU	Scaled exponential linear unit
PReLU	Parametric ReLU

Input type

AA	Average adjusted
CW	Channel-wise approach
f	Features
FFM	Fourier feature map
FFT	Fast Fourier transform
MAD	Mean absolute difference
N	Normalized
PSD	Power spectral density
Spect.	Spectrogram
Stat	Statistical
STFT	Short time Fourier transform
SV	Signal values
SVD	Singular value decomposition
SWD	Swarm decomposition
WD	Wavelet decomposition

Artifact strategy

AR	Algorithmically removed artifacts
ALI	Artifacts left In for analysis
MR	Manually removed artifacts

Task information

AD	Alzheimer's
BI	Bullying indices
ER	Emotion recognition
GC	Gender classification
Hr	Hours
MI	Motor imagery
Mult.	Multiple datasets
MW	Mental workload
ND	Neurological disorders
O	Other
SD	Seizure detection
SS	Sleep stage scoring
Subj	Subjects
TE	Treatment effectiveness

1. Introduction

Electroencephalography (EEG) is widely used in research involving neural engineering, neuroscience, and biomedical engineering (e.g. brain computer interfaces, BCI) [1]; sleep analysis [2]; and seizure detection [3]) because of its high temporal resolution, non-invasiveness, and relatively low financial cost. The automatic classification of these signals is an important step towards making the use of EEG more practical in application and less reliant on trained professionals. The typical EEG classification pipeline includes artifact removal, feature extraction, and classification. On the most basic level, an EEG dataset consists of a 2D (time and channel) matrix of real values that represent brain-generated potentials recorded on the scalp associated with specific task conditions [4]. This highly structured form makes EEG data suitable for machine learning. A great number of traditional machine learning and pattern recognition algorithms have been applied on the EEG data. For example, independent component analysis (ICA) is commonly used for artifact removal [5]; principle component

analysis (PCA) and local Fisher's discriminant analysis (LFDA) are typically used to reduce dimensionality of the features [5]; classic supervised learning methods such as linear discriminant analysis (LDA), support vector machines (SVM), and decision trees are common in neural classification [6, 7]; and canonical correlation analysis (CCA) is frequently used to identify steady-state visual evoked potentials (SSVEPs).

Neural networks did not immediately receive the high attention seen today in neural classification applications because of practical issues, such as very long computation time and problems with the vanishing/exploding gradients [8]. Fortunately, the availability of large datasets and the recent development of graphic processing units (GPU's) brought neural network researchers an inexpensive and powerful solution to their hardware bottleneck [9], allowing them to investigate deep learning architectures (neural network architectures containing at least two hidden layers). These innovations have led to an exponential increase in interest and applications of deep learning in the past decade. Indeed, it significantly improved performance in a wide range of traditionally challenging domains, such as images [10], videos [11], speech [12], and text [13]. Because neural networks iteratively and automatically optimize its parameters, they are generally believed to require less prior expert knowledge about the dataset to perform well [9]. This advantage led to early adaptations in the realm of medical imaging [14] which usually involves large datasets that are otherwise difficult to be interpreted, even by experts. Recently, due to the increasing availability of large EEG datasets, deep learning frameworks have been applied to the decoding and classification of EEG signals, which usually are associated with low signal to noise ratios (SNRs) and high dimensionality of the data.

This systematic review of the literature on deep learning applications to EEG classification attempts to address critical questions: which EEG classification tasks have been explored with deep learning (section 3.3)? What input formulations have been used for training the deep networks (section 3.4)? Are there specific deep learning network structures suitable for specific types of tasks (section 4.1)? To bridge this gap, we compiled all peer-reviewed published EEG classification strategies using deep learning and reviewed their EEG preprocessing methods, network structure and performance. By analyzing the overall trends and with architectural comparisons made in individual studies, different EEG classification tasks were found to be classified more effectively with specific architecture design choices. This information was compiled into a recommendation workflow diagram, which can serve as a starting point for the initial architecture design phase in future applications of deep learning to EEG classification.

2. Methods

2.1. Search methods for identification of studies

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), a systematic review and meta-analysis procedure [15], was used to identify studies and narrow down the collection for this review of deep learning applications to EEG signal classification, as shown in figure 1. This search was conducted on 22 December 2018 within both Web of Science and PubMed databases using the following group of keywords: ('*Deep Neural Network*' OR '*Deep Learning*' OR '*Deep Belief Network*' OR '*Deep Machine Learning*' OR '*Deep Convolutional*' OR '*Representation Learning*' OR '*Boltzmann Machine*' OR '*Deep Recurrent*' OR '*Deep LSTM*') AND ('*EEG*' OR '*Electroencephalography*'). Duplicates between the two databases and studies not within the inclusion criteria (outlined below) were excluded. Full texts of the remaining studies were then screened.

The following criteria were used to exclude unqualified studies:

- Electroencephalography only—To reduce variability in the studies, studies with multi-model dataset, e.g. EEG analysis combined with other physiological signals (electrooculography, electromyography) or videos were excluded.
- Task classification—This review focused solely on the classification of tasks performed by humans using their EEG signals. Other studies, such as power analysis, non-human studies, and feature selection with no end classification, were excluded.
- Deep learning—In this review, deep learning is defined as neural networks with at least two hidden layers
- Time—Given the fast progress of research in this topic, only studies published within the past five years were included in this review.

2.2. Data extraction and presentation

The following data categories were collected (see appendix for details):

a. *Task information*

- Task type
- Number of subjects
- Total length of analyzed data

b. *Artifact removal strategy*

- Manual
- Automatic
- No cleaning or removal

c. *Frequency range used for analysis*

d. *Input formulation*

- EEG signal features
- Channel selection methods

e. *Deep learning strategy, main characteristic, number of classifier layers, and output classes*

- Convolutional Neural Network (CNN), number of convolutional layers, activation
- Deep belief network (DBN) and number of restricted boltzmann machines (RBM's)
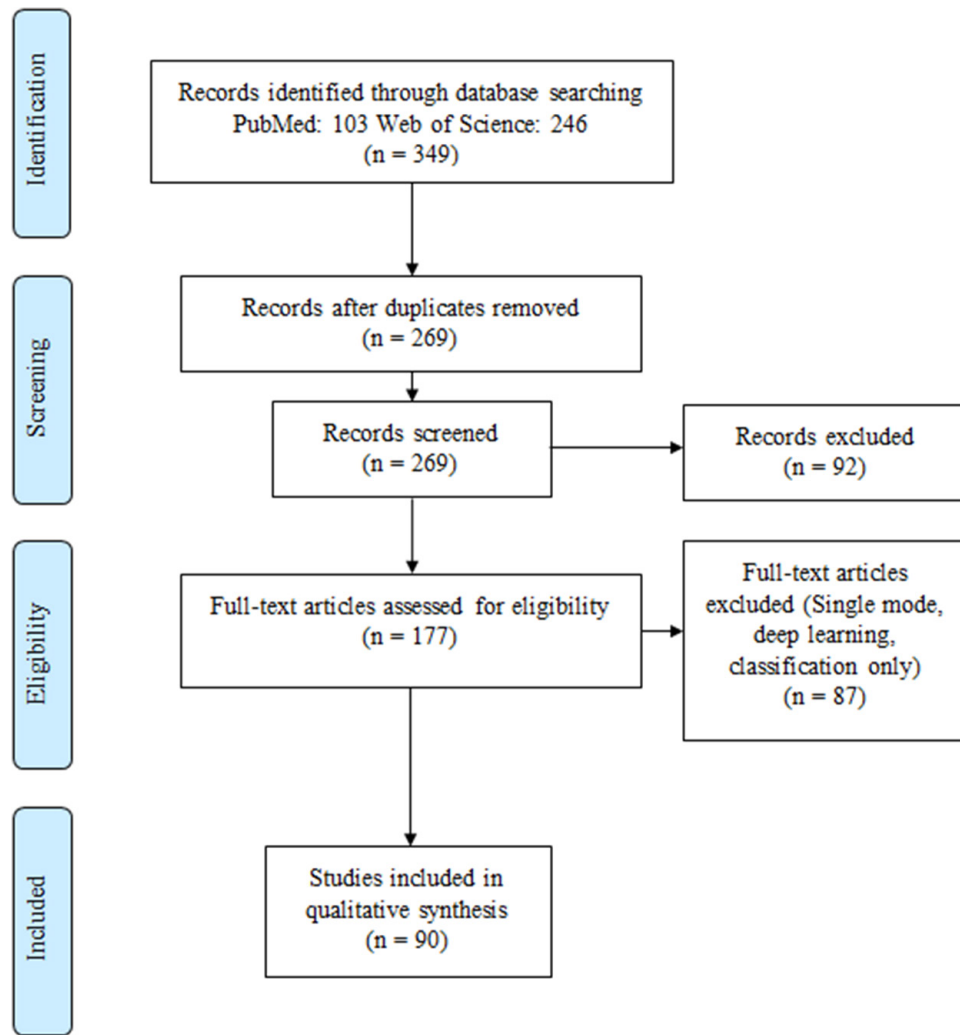
Figure 1. PRISMA study selection diagram. The diagram moves through the four stages of PRISMA study selection: identification, screening, eligibility, and included. This process led this review from an original count of 349 studies to a final count of 90 studies.

- Recurrent neural network (RNN), number of RNN layers, type of RNN unit
- Stacked auto-encoders (SAE), number of hidden layers, activation
- Multi-layer perceptron neural network (MLPNN), number of hidden layers, activation
- Hybrid architectures, types of algorithms, corresponding main characteristics, activation

f. *Highest achieved accuracy or other performance metric*

3. Results

This section first details the preprocessing methods found within this review. Then, the general types of tasks, the input formulations, and architecture trends are analyzed. The results section finishes with a case study over a public dataset, which allows for a comparison on specific deep learning design choices.

3.1. Which EEG classification tasks have been explored with deep learning?

The tasks presented within these studies fell into six groups: emotion recognition (16%), motor imagery (22%), mental workload (16%), seizure detection (14%), and sleep stage scoring (9%), event related potential detection (10%), and other studies (13%), which include Alzheimer's classification [16–18], bullying indices detection [19], depression [20], gait pattern classification [21], gender classification [22], detection of abnormal EEG [23–25], and transcranial stimulus treatment effectiveness [26]. The following provides a description for the general protocols for these tasks.

3.1.1. Emotion recognition tasks. Emotion recognition tasks typically involve having subjects watch video clips which have been assigned specific emotions by experts prior to the viewing [27]. EEG was measured during these viewings and an emotion self-assessment typically followed. The self-assessment and original emotion class was then converted into a pair of valence and arousal values, a widely used system to describe emotions

[28]. The primary drive for emotion recognition studies is the eventual application in brain–machine interfaces as understanding a patient's emotion will help the underlying algorithm decide whether a selected movement was the intended movement. More generally, emotion recognition studies help computers better understand the current emotional state of the user.

3.1.2. Motor imagery tasks. Motor imagery tasks involve having the subject imagine certain muscle movements on limbs and/or the tongue [29]. Their applications are mostly BMI-related in that eventual BMI applications will need to accurately classify a user's intended movement.

3.1.3. Mental workload tasks. Mental workload tasks involve measuring EEG data while the subject was under varying degrees of mental task complexity. There were many methods used to categorize the levels of mental workload, including driving simulation studies [30], live pilot studies [31], and responsibility tasks [32]. For driver and pilot studies, mental workload was classed based on statistics of the subject's behavior, such as reaction time and path deviation. For responsibility studies, workload was classed based on the increasing number of actions the subject was responsible for. This kind of task may be applied in two general areas: cognitive stress monitoring or BMI performance monitoring.

3.1.4. Seizure detection tasks. For seizure detection studies, EEG signals of epileptic patients are recorded during periods of seizure and during seizure-free periods [33]. EEG signals were also recorded from non-epileptic patients for some datasets as a control class. These studies were designed for the eventual application for detecting upcoming seizures and preemptive notification of the epileptic patient.

3.1.5. Sleep stage scoring tasks. Sleep stage scoring studies, which was the task type with the fewest number of studies, record the EEG signals of subjects overnight. These signals were scored by experts and classified into sleep stages 1, 2, 3, 4, and the rapid eye movement stage. The eventual application for research of this kind focuses on reducing the reliance on trained personnel in the analysis and understanding of patient sleep stages.

3.1.6. Event related potential tasks. Studies that focused on the detection and classification of event related potentials typically record EEG from subjects who are undergoing a visual presentation task. In these tasks, a subject watches a rapid sequence of pictures or letters with the aim of focusing attention on specific indicators. Once a specific letter or image appears, a stereotypical response is seen in the EEG data, typically in the form of a P300 response. These tasks are useful in research due to the relatively clean signal (minimization of artifacts) and high signal-to-noise ratio [34], which are characteristics not typically seen in EEG data. Research into event related potential tasks will help lead to improved non-verbal communication systems using EEG.

3.2. Preprocessing methods

EEG data is inherently noisy because EEG electrodes also pick up unwanted electrical physiological signals, such as the electromyogram (EMG) from eye blinks and muscles on the neck. There are also concerns about the motion artifacts occurring from cable movement and electrode displacement when the subject moves. The identification and removal of EEG artifacts have been extensively studied in the past literature [35–38], which will not be repeated in this review. Outside of the 41% of studies that did not address any specific artifact removal process, studies approached the artifact removal process in one of the following three strategies (shown in figure 2): (1) manual removal (29%), (2) automatic removal (8%), and (3) no cleaning/removal (22%).

Surprisingly, more than a quarter of the studies (26 of 90 studies) removed artifacts manually. It is indeed easy to visually identify abrupt outliers, for example when signals are lost or when intense EMG artifacts are present. However, it is difficult to identify persistent noisy channels or noise that is sparsely presented in multi-channel recordings. More importantly, manual data processing is highly subjective, rendering it difficult for other researchers to reproduce the procedures. Together with the 22% of studies that did not take any actions to remove artifacts, 63% of the studies reviewed did not systematically remove EEG artifacts. The most frequent artifact-removal algorithms used in the remaining 8% of the studies reviewed were independent component analysis (ICA) and discrete wavelet transformation (DWT).

Most studies used frequency domain filters to limit the bandwidth of the EEG to be analyzed. This is useful when there is a certain frequency range of interest so that the rest can be safely discarded. About half of the studies low pass filtered the signal at or below 40 Hz, which is in or below the typical low gamma band. The filtered frequency ranges, organized by task type, and artifact removal strategies are shown in figure 2, which shows that most studies employed some form of artifact removal strategy in addition to reducing the frequency ranges to be analyzed.

An interesting question yet to be answered is whether it is still necessary to filter the data if the deep learning is capable of pulling such information from unfiltered data. To the authors' best knowledge, there are no studies that specifically analyze whether deep learning can achieve comparable results without any artifact cleaning or removal process.

3.3. What input formulations have been used for training the deep networks?

EEG signals are intrinsically noisy and suffer from channel crosstalk [27]. In typical scalp EEG recording settings, each EEG electrode picks up signals from the area nearby, making the spatial resolution coarse (several centimeters). Unmixing the signals is not trivial because of the anisotropic volume conduction characteristics in human brain tissues, skull, scalp, and hair. Therefore, one of the standing challenges in EEG data analysis is how to formulate inputs. So, what input
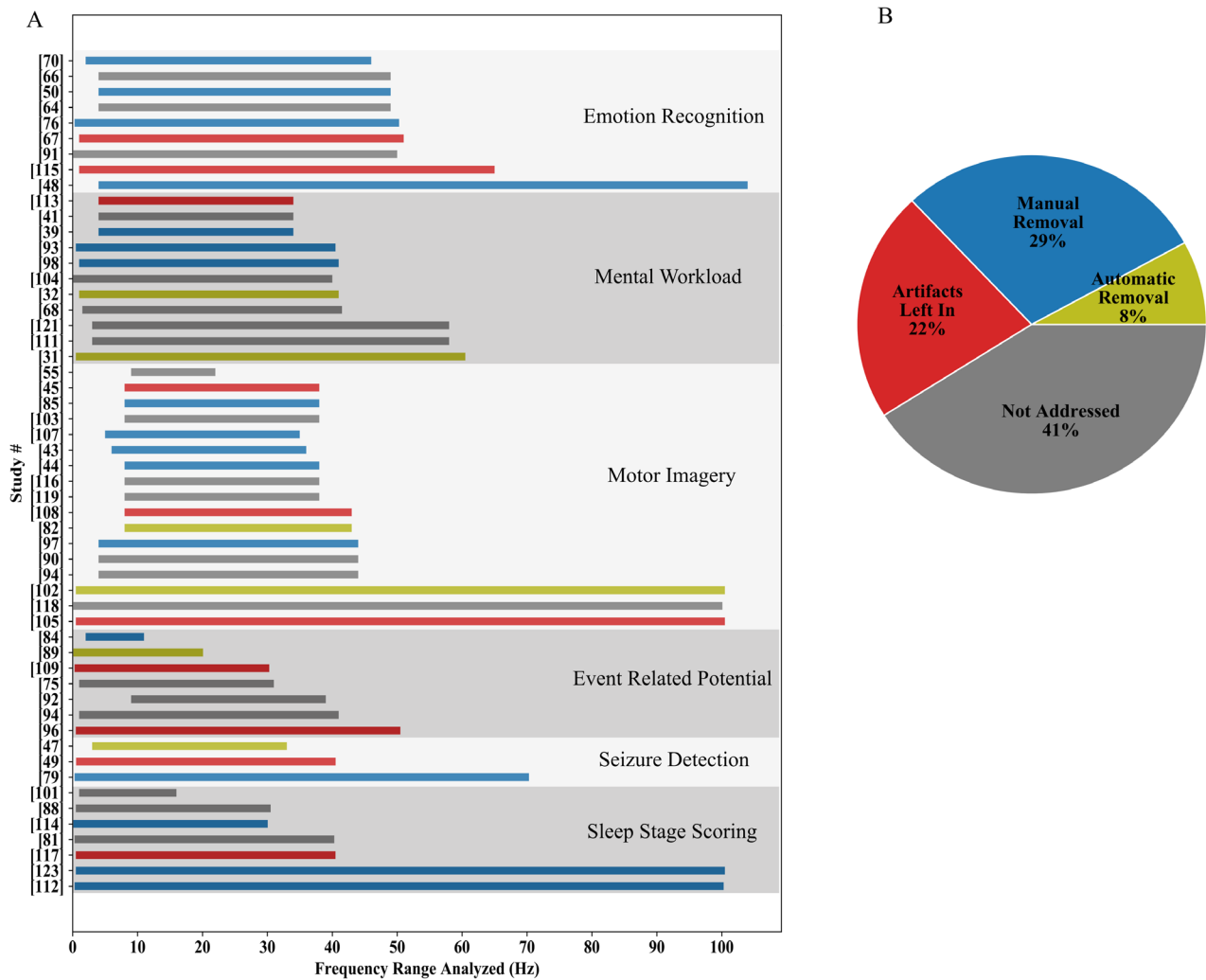
Figure 2. Artifact removal and filtering strategies. (A) Frequency range used in EEG analysis for each identified study, organized by task type. The colors of the bars correspond to different artifact removal strategies. Red bars indicate studies that discussed the conscious decision to leave the artifacts as contaminants in the data, whereas dark grey bars represent studies that did not address any artifact removal procedure at all. Studies are grouped according to the application type. (B) The percentage of different artifact removal strategies across all studies.

formulations have been used and how effective have they been shown to be in the context of deep learning?

Studies fell into three types of input formulation categories: calculated features (41%), images (20%), and the signal values (39%) (figure 3(A)). The selection of input formulation relied heavily on the task and deep learning architecture, so these decisions will also be described in the respective sections below.

Unsurprisingly, a wide range of methods have been developed to extract features from the EEG, accounting for 41% of all identified studies (figure 3(A)). EEG data is commonly analyzed in frequency domain because they are often found to be associated with behavioral patterns [4]. Power spectral density (PSD), wavelet decomposition, and statistical measures of the signal (i.e. mean, standard deviation) are the three most common input formulations used in the reviewed studies.

Notably, many neural networks, especially CNN's, use spectrograms generated from the EEG data as inputs [23, 39–41]. Spectrograms are traditionally used as a post-processing tool to visualize the data [42]. However, CNN's unprecedented ability to learn images has enabled spectrograms to be used as an input to the classifier. Other image formulations include creating Fourier feature maps [43–45] and designing 2D or 3D grids [46–49].

Signal strength in the time domain is also used directly as inputs to the neural networks (39%). Traditionally, this approach is usually associated with particular hand-engineered time domain features, such as the power spectral density features [50]. Neural networks promise to automatically learn complicated features from large amounts of data, prompting the idea of end-to-end learning [51]. This machine learning philosophy encourages researchers to feed raw signal
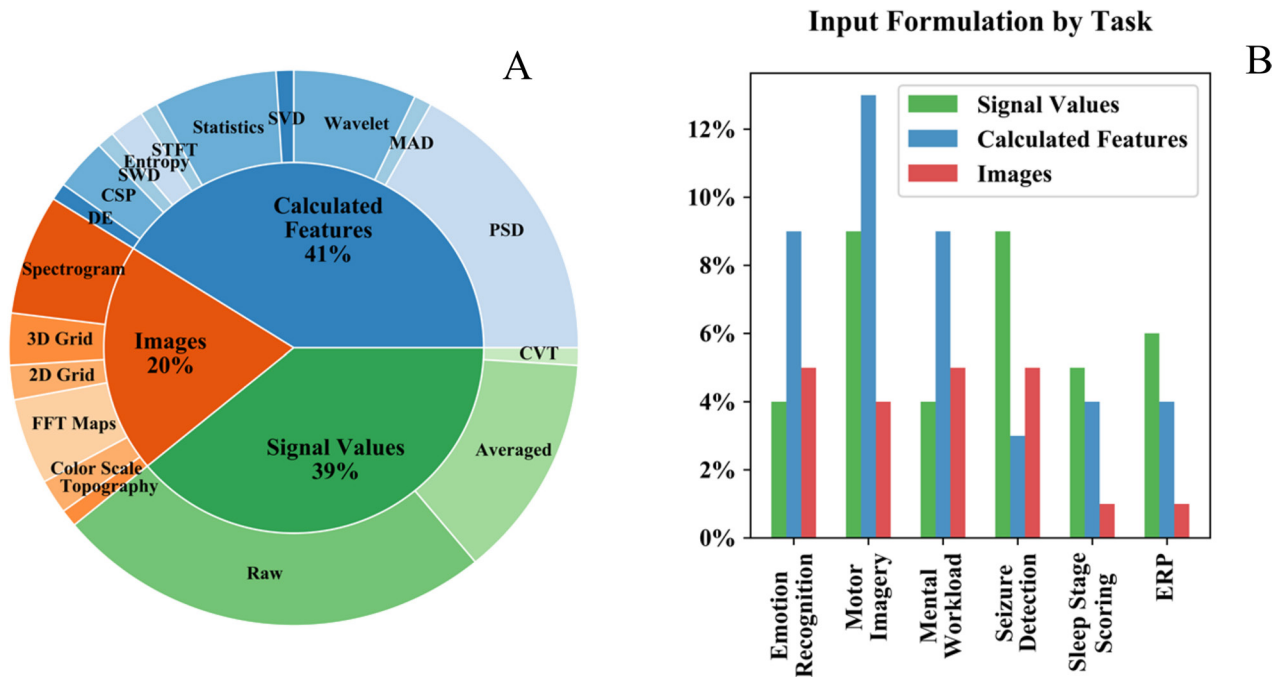
Figure 3. (A) Input formulations across all studies. The inner circle shows the general input formulation while the outer circle shows more specific choices. (B) General input formulation compared across different tasks. Most tasks had calculated features as inputs, with seizure detection studies instead having a much higher proportion of signal values. Key—CVT: complex value transformation, CSP: common spatial pattern, DE: dynamic energy, FFT: fast Fourier transform, MAD: mean absolute difference, PSD: power spectral density, STFT: short time Fourier transformation, SVD: singular value decomposition, SWD: swarm decomposition

values directly into the neural network without hand-designed features, which may contribute to the practice of directly analyzing raw EEG data with deep learning.

When it comes to task-specific input formulations, three tasks (emotion recognition, mental workload, and motor imagery tasks) heavily chose calculated features. The most prevalent input formulation tactic for the remaining three types of tasks (seizure detection, sleep stage scoring, and event related potential analysis) was to use the signal values as inputs. Seizure detection studies showed both the highest proportion of using signal values as inputs as well as the smallest proportion of studies using calculated features, which lends to the idea that calculating features is not suitable for that task. Generating spectrogram images as inputs has seen success in all tasks, but both event related potential and sleep stage scoring tasks only had a single study attempted to use images as inputs, so further research into effective inputs for these tasks is needed. The proportions of input formulations by task are shown in figure 3(B).

3.4. Deep learning architecture trends

3.4.1. Architecture design choices. This section of the review focuses on understanding the trends in the formation of specific deep learning architectures, namely, the primary characteristic and end classifier. This aggregated information is displayed in figure 4.

The most prevalent architecture design framework, CNN's (43%), involve alternating layers of convolution with pooling layers (typically maximum pooling layers). The primary

design features for CNN's was the number of convolutional layers and the type of end classifier. DBN's followed CNN's at 18% as the second most prevalent choice. DBN's are composed of a number of stacked restricted Boltzmann machines followed by an end classifier, which is typically a number of fully-connected layers. Hybrid architectures made up the next group at 12% of the total number of the studies, divided, as shown in figure 4, into two divisions: hybrid CNN's and Hybrid MLP's. Hybrid CNN's, in addition to convolutional and pooling layers, include an additional architecture type, such as the inclusion of a number of recurrent layers or restricted Boltzmann machines. Hybrid MLPs are composed of a number of dense layers and the inclusion of another type of deep learning algorithm. The next group of architectures by proportion of the total study count was RNN's (10%), which are composed of a number of recurrent layers (each layer containing a study-specific number of recurrent units) followed by a number of fully-connected layers. MLPNN's, which had a primary design feature of the number of hidden layers as its only reviewed characteristic, followed with 9% of the study count. The final group of studies used an SAE (8%), which has the total number of fully connected layers followed by, in all cases, a single fully connected layer.

3.4.2. Activation functions. Activation functions were collected for applicable deep learning architectures across all studies. For convolutional layers, 70% of studies employing convolutional layers for the deep learning architectures used rectified linear unit (ReLU) as the layer's activation function. No other activation function breached 10% of the
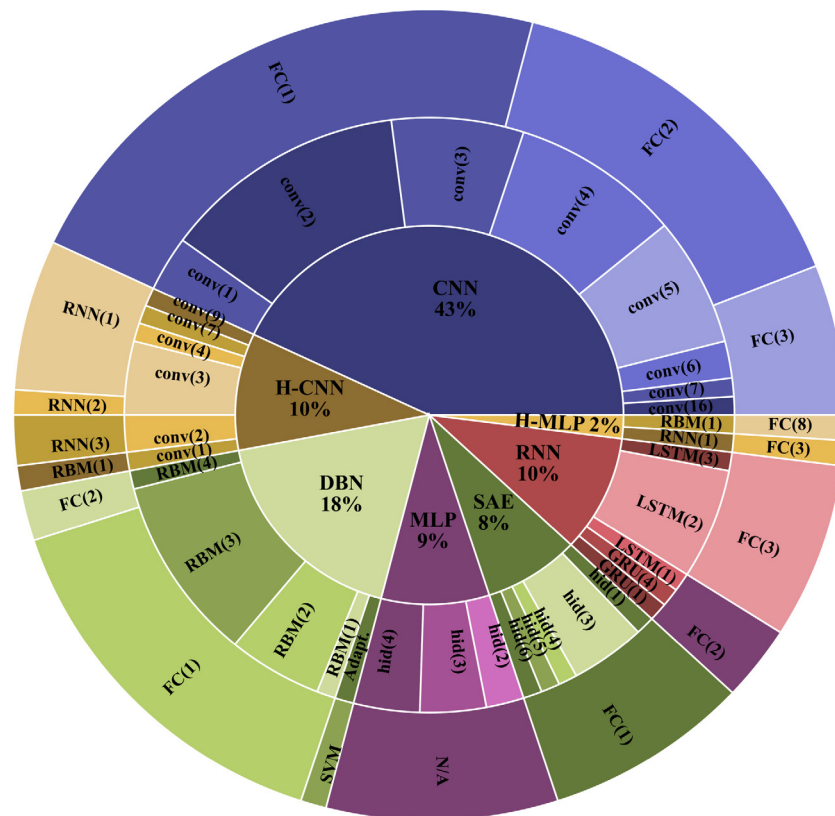
Figure 4. Deep learning architectures across all studies. The inner circle shows the general deep learning strategy, the middle circle shows the primary design feature for the specified deep learning strategy, and the outer circle shows the secondary design characteristic. Key—Adapt.: adaptive number of RBM's conv(#): number of convolutional layers, CNN: convolutional neural network, DBN: deep belief network, FC(#): number of fully connected layers, hid(#): number of hidden layers, H-CNN: hybrid CNN, H-MLP: hybrid MLP, MLP: multi-layer perceptron, RBM(#): number of restricted Boltzmann machines, RNN: recurrent neural network, RNN(#): number of recurrent layers, SAE: stacked auto-encoders.

Figure 5. The proportions of deep learning architectures by task type. While there was not a clear consensus when looking at all studies together, seizure detection studies showed a clear preference for either CNN's or RNN's, whereas ERP studies showed a preference for CNN's. Sleep stage scoring tasks and mental workload tasks both saw more instances of hybrid architectures applied to the classification task as compared to other architecture types.

total number of studies reporting activation functions. Less prevalent activation functions include exponential linear unit (ELU) (8%), leaky rectified linear unit (leaky ReLU) (8%), and hyperbolic tangent (tanh) (5%). Additionally, there were single studies incorporating the following types of activation functions: parametric ReLU (PReLU), scaled exponential linear unit (SELU), and split tanh. Further analysis of convolutional activation functions is detailed in the discussion section.

Activation functions for fully-connected layers can be grouped into two divisions: non-classifier fully-connected layers and classifier fully-connected layers. The vast majority of classifier fully-connected layers employed a softmax activation function, whereas non-classifier fully-connected layers used the sigmoid activation function. There were only three SAE studies that discussed activation functions and these did not form a consensus, with [52, 53] employing sigmoid activation functions for non-classifier AE layers, while [54] instead using ReLU. More research is needed in order to better understand the most effective activation function for SAE architectures.

3.4.3. Task specific deep learning trends. Emotion recognition, motor imagery, and sleep stage scoring tasks did not show any consensus on the selection of deep learning algorithms. Seizure detection studies were essentially split between using either CNN's or RNN's, showing the highest percentage of studies using RNN's as compared to other tasks. For these studies, only a single study chose to use either an SAE or MLPNN and no seizure detection study used DBN's. Sleep stage scoring tasks had the highest percentage of studies using hybrid formulations, which were evenly represented when compared to studies using CNN's. ERP studies had a clear preference for CNN's (the highest percentage of CNN studies compared to all other tasks). Task-specific deep learning strategy decisions are shown in the figure 5.

3.4.4. Input formulation by deep learning architecture. The specific input formulation strategies varied significantly as a function of the type of deep learning architecture (figure 6). There was no instance among DBN, MLPNN, or SAE studies that used images as inputs, which is unsurprising as image processing is considered to be in the domain of CNN's. For CNN studies that used images as inputs, the average accuracy was essentially the same as CNN studies that used calculated features as inputs, with both input formulation strategies achieving an average accuracy of 84%. This is compared to the average accuracy of 87% for CNN studies that used signal values as inputs. This goes against the intuition that says the more effort applied to the pre-processing stages, the more accurate the classification will be. This points to the surprising conclusion that future research, rather than being inhibited by the desire to spend less time pre-processing, may actually improve results by sending signal values directly into the deep learning framework.

DBN's showed a trend similar to that of CNN studies. DBN studies were split between using signal values and calculated features, with calculated features being the most prevalent

decision. Studies that used calculated features achieved an average accuracy of 85% while signal value studies achieved an average accuracy of 86%. There were no studies that compared the classification accuracies between using signal values and calculated features, thus indicating more research is needed.

RNN's, on the other hand, did not share the trend found with CNN's and DBN's. RNN studies had a relatively even number of instances from all three types of image formulations. RNN studies that used signal values as inputs achieved an average accuracy of 85%, which is less than the accuracies for studies that used calculated features (89%) and images (100%). However, there are significantly less number of studies that used RNN's, so more research is needed for the most effective input formulation strategy for RNN's.

Feature selection seemed to be the only practical choice for MLPNN and SAE studies since only a single study for each architecture chose instead to use signal values. The sole MLPNN study that chose to use signal values [55] achieved an accuracy of 75%. The single SAE study using the signal values as inputs [53] only had a single channel for analysis, but was able to achieve a 96% accuracy.

3.5. Case studies on a shared dataset

The majority of studies covered by this review used EEG datasets that were not publically available. A simple accuracy comparison between studies analyzing different datasets is not valid as they have widely different tasks, subjects, and data procurement procedures. On the other hand, studies that examined identical datasets using different approaches provide a more meaningful comparison. The datasets that are used in more than one study within this review include an emotion recognition dataset ([56], various subsets used in ten studies), a mental workload dataset ([57], used in four studies), two motor imagery datasets from the same BCI competition ([58], one dataset used in six studies and the second dataset used in two studies), a seizure detection dataset ([33], used in three studies), and two sleep stage scoring datasets ([59, 60], which were used in groups of two studies and three studies, respectively).

The *common emotion recognition dataset* (DEAP) [56], the dataset analyzed by the highest number of studies, is a collection of EEG and peripheral signals from 32 subjects participating in a human affective state task. Each subject watched 40 one-minute music videos. After each one minute viewing, the subject performed a self-assessment to classify their levels of arousal, valence, dominance, and liking. Based on the self-assessment and the original classification of the music video, each one-minute segment was labeled with numeric values for valence and arousal level. EOG artifacts were manually removed by the original researchers.

Table 1 shows the prevailing input and architecture choices along with the highest achieved accuracy by study. The above collection shows a wide range of specific architecture choices applied to the common emotion recognition task. SAE's did not seem to be a good choice for this task, as [48] reported

Input Formulation by Architecture Type

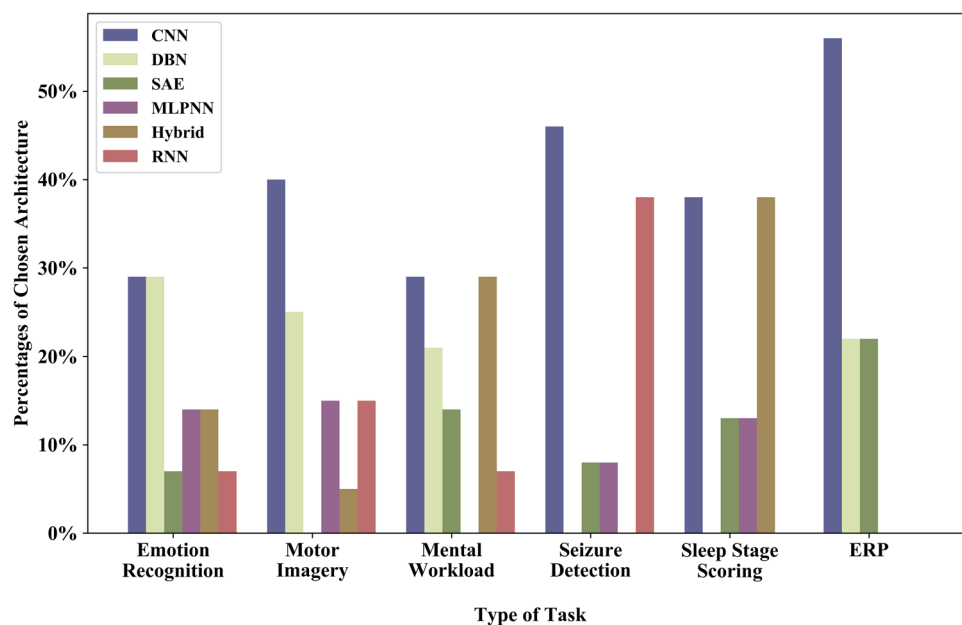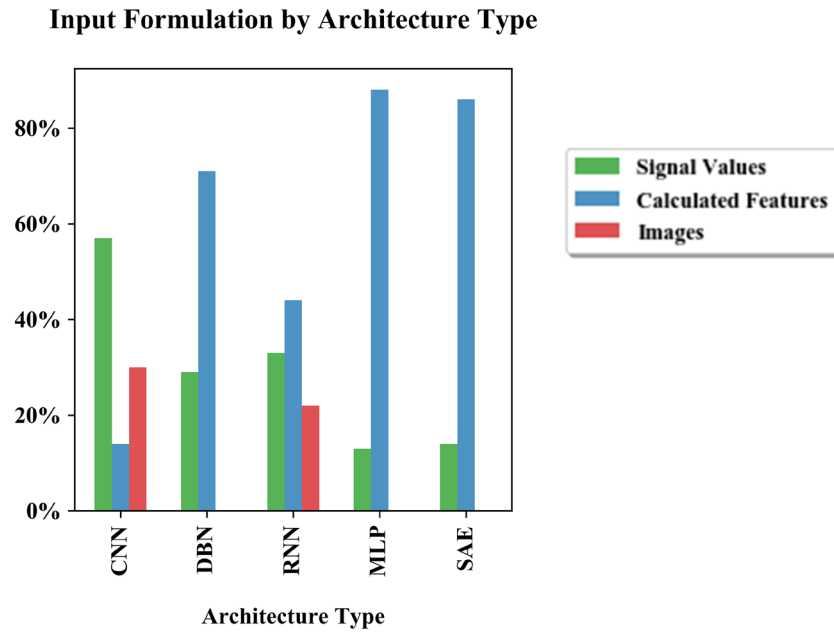**Figure 6.** The proportions of input formulations by deep learning architecture types. While there was not a clear consensus when looking at all studies together, studies that employed either MLP's or SAE's showed a clear preference towards using calculated features. DBN's were solely split between using signal values and calculated features, whereas both CNN and RNN studies saw instances of all three input formulations.

Table 1. A comparison of architecture and input choices across studies using the publicly available DEAP [56] dataset. The table shows study effectiveness as accuracy increases moving down the table.

	Architecture	Input formulation	Channel strategy	Accuracy (%)
[48]	SAE 3 hidden layers 1 dense layer	Principal component analysis, covariate shift adaptation, PSD features	All channels	54
[61]	Hybrid 1 conv layer to 1 RNN layer 2 dense layers	Wavelet derived features	Channel-wise	73
[62]	Hybrid 2 conv layer to 1 dense layer to 1 RNN layer 1 dense layer	Image—color scale	Channels combined in image	75
[63]	CNN 5 conv layers 1 dense layer	Signal values	All channels	81
[64]	MLPNN 4 hidden layers	PSD features	All channels	82
[65]	RNN 2 LSTM layers 1 dense layer	Signal values	All channels	87
[66]	CNN 2 conv layers) 2 dense layers	Image—Fourier feature maps	Channels combined in image	87
[67]	CNN 2 conv layers 1 dense layer	Image—3D grid	Channels combined in image	88
[50]	DBN (3 RBM's) 1 dense layer	PSD features	All channels	89

an accuracy of 54% after exploring various EEG features and neural network designs. Jirayucharoensak *et al* [48] tried three different combinations of calculated features, but found that even the combination of PCE, CSA, and PSD features was not able to rectify the shortcomings of SAE's for this dataset. Xu and Plataniotis [50] compared accuracies between an SAE and several DBN's and found that a DBN with 3 restricted Boltzmann machines performed significantly better than SAE's and DBN's with different numbers of RBM's.

Convolutional layers were employed by five of these studies. Two of the five were hybrid architectures with the CNN outputting into LSTM RNN modules, but neither attempt breached 75% accuracy. The difference in accuracy in the three standard CNN studies is likely due to the differences in input formulation. Yanagimoto and Sugimoto [63] used signal values as inputs into the neural network, while [66, 67] instead converted the data into Fourier feature maps and 3D grids and achieved accuracies of 87% and 88%, respectively. These
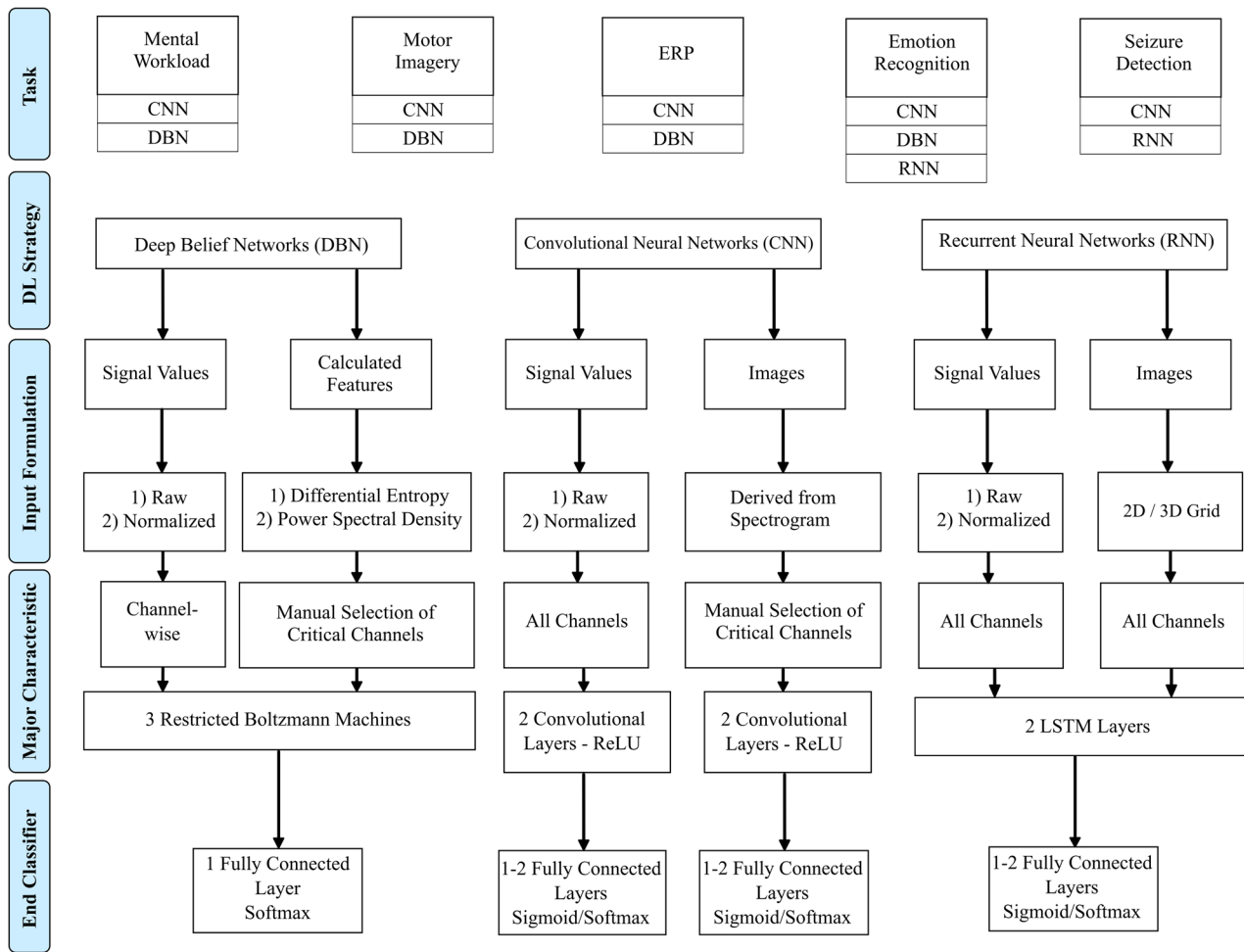
Figure 7. Task-specific deep learning recommendation diagram. The workflow begins with task type (with connected boxes indicating the task's general deep learning architecture recommendation) and leads into deep learning architecture characteristic recommendations, which can serve as the starting point for designing deep learning architectures in future research.

CNN architectures were composed of two convolutional layers each with either a one or two dense layers.

The sole MLPNN architecture applied to this dataset [63] achieved an 82% accuracy, which is comparable to the standard CNN that used signal values, but the input formulation may have been the largest factor in the accuracy difference. Whereas the deep CNN used signal values, which requires significantly less pre-processing, the MLPNN required extensive effort in input pre-processing with PSD features and prefrontal asymmetry channel selection.

The sole architecture that employed an RNN alone (no convolutional layers) within this group [65] was composed of two LSTM layers and a single dense layer. This study was able to achieve an accuracy of 87% while only using signal values as the input. This was unexpected as intuition assumes architectures with greater complexity, such as the two hybrid convolutional recurrent architectures, will lead to higher classification rates. In the case of this dataset, a deep learning recurrent architecture with no convolutional layers was able to outperform the architectures that included both recurrent and convolutional aspects.

For this dataset the most effective architectures reported were DBNs, CNNs, and RNN's. The choice comes down to input formulation in that input images were more suited for CNN's, signal values for RNN's, while calculated features, specifically PSD features, worked better for DBN's. The high accuracy and relative shallowness of the algorithms that dealt with Fourier feature maps and 3D grids may point to the necessity of future research on EEG image generation methods.

4. Discussion

In this section, recommendations for design choices on non-hybrid deep learning architectures based on the type of task are given. These recommendations are based on the prevalent general design choices across the entire dataset with support from studies that varied specific design features. Recommendations are given by task: mental workload, emotion recognition, motor imagery, event related potential, and seizure detection. Sleep scoring applications are not included as the number of studies was too low and there was no consensus on architecture

design choices. The recommendation section will be followed by a discussion on hybrid architecture types.

4.1. Are there specific deep learning network structures suitable for specific types of tasks?

Outside of the common datasets used in some of the studies (see section 3.5), it's difficult to compare classification accuracies achieved by different architecture design choices across tasks and EEG datasets. Moreover, most studies varied different aspects of the algorithm design or input processing. Nevertheless the analysis provided before can help to direct future research using deep learning networks for the range of tasks reviewed herein. Figure 7 depicts a workflow diagram that represents the recommendations formulated from this review.

4.1.1. Mental load tasks. Studies specifically motivated the recommendation that either DBN's or CNN's be used for this type of task. Yin *et al* published three studies all based on one mental workload dataset: [32, 68] reported better accuracy using SAE compared to SVN or MLPNN; subsequently, [64] reported that DBN consistently outperformed the SAE architecture. Hajinoroozi *et al* [69] reported that the standard DBN significantly outperformed standard CNN (although less accurately than hybrid models), whereas [41] found several variations of CNN's able to outperform when compared against a standard DBN.

4.1.2. Emotion recognition tasks. The application of deep learning networks to this task had the highest number of studies and included the repeated dataset described in the results section. References [50, 70] found that DBN's were more capable for this task type when compared against SAE's and MLPNN's respectively. In the shared dataset, the four highest performing studies [50, 65–67] achieved accuracies ranging between 87% and 89%. These studies employed a DBN [50], CNN's [66, 67], and an RNN [65]. As the accuracy differences between these four studies is small, it is likely more research is needed to better understand the most effective deep learning algorithm for this type of task. These studies taken together led to the recommendation of DBN's, CNN's, or RNN's as good candidate architectures.

4.1.3. Motor imagery tasks. There were no direct comparisons between DBN and CNN architectures among motor imagery tasks and further research is needed to identify the more effective architecture. The hybrid architecture does shed some insight into whether SAE is a valid option. In [43], a hybrid CNN/SAE architecture was compared to a standard CNN and a standard SAE. While the hybrid architecture outperformed the others, the standard CNN far exceeded the SAE. Wang *et al* [45] compared the performance of a CNN against a pure LSTM RNN network, finding the CNN architecture to

significantly outperform the RNN. For these reasons, the recommendation for motor imagery tasks is to use either a CNN or DBN network.

4.1.4. Seizure detection tasks. The majority of seizure detection studies used either CNN's or RNN's and, unlike the previous three groups, there were no instances of studies using DBN's. The sole standard MLPNN architecture study [71] varied the number of hidden layers, but experienced large drops in accuracy with more than two hidden layers, indicating further improvements must come outside MLPNN architectures. On a shared seizure detection dataset from the University of Bonn [33], two studies achieved near perfect classification performance with [72] reaching 99% accuracy with a CNN and [73] reaching 100% with an RNN. More research into the possibility of using DBN's is needed before disregarding that option for this task, but both CNN's and RNN's are good candidate architectures for this type of task.

4.1.5. ERP tasks. Of the ERP task studies, three different types of deep learning architectures were used: SAE's, DBN's, and CNN's. The two SAE studies, [54, 74], compared performances between SAE architectures and MLPNN's, both finding SAE's to perform best. Kulasingham *et al* [75] specifically compared performances between DBN's and SAE's and found that DBN's had a slight advantage in classification accuracy. The remaining five studies used different variations of CNN's, but did not offer any direct comparisons between other architecture types. For these reasons, both CNN's and DBN's have been selected as good candidate architectures for ERP tasks.

4.1.6. Sleep tasks. Sleep scoring tasks had the lowest number of studies within this review and showed no consensus for architecture type. Due to the low study count and lack of consensus, no architecture recommendations can be given and sleep scoring tasks are not included in figure 7.

4.2. Architecture design and input formulation

Figure 7 also depicts the specific input formulation and architecture parameter recommendations. DBN's saw no use of images as inputs, so the options for DBN's in the *Input Formulation* stage are restricted to calculated features or signal values. Based on [26], which found that channel-wise DBN's outperformed a combined channel approach, DBN inputs should then be fed into the deep learning architecture in a channel-wise approach in that the signal values from each channel are fed into individual architectures before outputting into a common classifier.

In the *Calculated Features* branch of the DBN path, we recommend use either differential entropy features or PSD features. Zheng and Lu [76] compared the classification accuracy

when using differential entropy features against PSD features, finding differential entropy features outperformed PSD features over all frequency sub-bands. PSD features, the most prevalent choice for DBN input formulation, was used in [50], which was the highest performing study within the common emotion recognition dataset. Regardless of the chosen feature type, this review's recommendation is to calculate features from manually selected critical channels. Zheng and Lu [76] found that classification accuracies were higher when calculating features from set of manually selected critical channels rather than from all channels. Manual selection of channels requires significant task-specific prior knowledge, which may not always be feasible.

Figure 7 also includes suggestions for the major characteristic section of the DBN path; three studies motivated the recommendation that, regardless of input formulation, inputs should be sent through three RBM's. References [29, 67, 75] all achieved higher classification accuracies when using three RBM's as opposed to one, two, or four RBM's. In the final stage in the diagram, we recommend a single dense layer as the end classifier, which is the choice in most DBN studies. For the final fully-connected layer in all paths, the softmax activation layer is recommended, while for fully-connected non-classifier layers, the sigmoid activation layer is recommended.

In the CNN branch of figure 7, the input recommendation splits into two branches: signal values or images. CNN studies had the greatest proportion of studies using signal values as inputs and the majority of those studies did not limit the number of channels, indicating that CNN's are more capable of handling the high dimensionality and size of EEG signal value datasets when compared to other DL algorithms. For example, [63] achieved higher accuracy with raw signal values from all channels than other studies that required extensive effort creating inputs using the same dataset (table 1). Spectrograms were the most prevalent choice when images were used as CNNs' input. Within the common emotion recognition dataset group, the second and third highest accuracy was achieved by [66, 67], respectively, both choosing to form images as inputs. Qiao *et al* [66] used spectrograms created from a set of manually selected critical channels to generate EEG images, whereas [67] instead created a 3D grid of electrode values. As both studies achieved accuracies within a single percentage, both image formulation strategies seem like viable options for future research.

The two CNN input paths do not share the same major characteristic recommendation. For signal values, four studies compared accuracies while varying the number of convolutional layers. Yanagimoto and Sugimoto [63], an emotion recognition study, and [78], a seizure detection study, found that five convolutional layers achieved the best accuracies. Antoniades *et al* [79] found that accuracy peaked with four convolutional layers and trended downwards as convolutional layers increased. Schirrmeister *et al* [80] compared a shallow two convolutional layer CNN versus a deep four convolutional

CNN and found that the deep CNN consistently outperformed the shallow CNN. While there were no studies specifically comparing the different numbers of classifier layers, the identified studies mostly used one or two fully connected layers. This review therefore recommends four to five convolutional layers feed into one to two fully connected classifier layers. For the image branch of the CNN path, there were not any studies that specifically compared different numbers of convolutional layers or end classifier layers We therefore recommend following [66, 67] with two convolutional layers feeding into one or two fully connected layers, while cautioning that more research is needed to optimize the strategy to use images as CNN inputs.

When assessing trends of activation functions of convolutional layers, two studies performed an analysis on the performance differences when using different activation functions. In [45], performances were compared between a CNN with ReLU, ELU, and SELU and found that SELU performed best. Abbas and Khan [44] did a similar analysis and assessed performance differences of a CNN when using ReLU, sigmoid, and tanh activation functions, finding that ReLU performed best. Due to the high number of studies employing ReLU (70% of convolutional architecture studies employed a ReLU activation), it is this review's recommendation that convolutional layer construction begin by using ReLU activation before investigating the performance changes due to different activation functions.

The RNN branch of figure 7 also splits into two branches for the input formulation section: signal values and images. While no study directly compared performances between different input formulations, this recommendation is based on the lack of consensus for feature selection, high accuracy reports from image studies, and the high performing raw signal value study within the shared dataset, [65]. Unlike the image branch for CNN's, the image recommendation for RNN's instead points towards the use of 2D or 3D grids as opposed to spectrograms based on two well performing seizure detection studies [47, 49]. These input recommendations are given with caution as more research is needed to know the most effective EEG input formulation for RNN architectures.

Regardless of the input formulation, the RNN branch merges into one for the major characteristic section of figure 7. The majority of RNN studies used two LSTM layers with two studies specifically varying this parameter. In [81], the authors first compared performances between LSTM or GRU recurrent layers and found that LSTM layers performed better. Next, the authors varied the number of recurrent layers between one and eight, concluding that two layers led to the highest classification. In [82], the authors also compared performances while varying the number of LSTM layers and found that there was significant improvement when using two layers versus one layer, whereas accuracy essentially remained even with the introduction of a third LSTM layer. Based on these two studies, the recommendation outlined

in figure 7 is that future research uses two layers of LSTM recurrent units for pure RNN architectures. Finally, the RNN branch concludes with the recommendation for end classifier layers. In the case of RNN's, no study specifically varied the number of classifier layers, with all RNN studies choosing either one or two fully-connected layers. Due to this lack of research, this review recommends future research use either one or two fully-connected layers for classification.

4.3. Hybrid architectures

There were ten studies that applied hybrid architectures, which are architectures that use a combination of two or more standard deep learning algorithms. Among these studies, several performed accuracy comparisons between the proposed hybrid architectures versus architectures based on the component standard deep learning algorithms. Nine of the ten hybrid studies combined convolutional layers with another type of layer, with eight of those nine employing one or more RNN layers. While there were no task-specific trends for hybrid designs, results from these hybrid studies suggest that LSTM RNN modules outperform both standard and GRU RNN modules and that there are consistent improvements with the addition of non-convolutional layers to standard CNN's.

LSTM is a popular extension of the standard RNN module [83]. In [62], the researchers designed two hybrid architectures: a two convolutional layer CNN feeding into a dense layer that fed into either an LSTM RNN module or a standard RNN module. The LSTM CNN/RNN hybrid architecture consistently outperformed the standard CNN/RNN architecture. Bresch *et al* [81] also compared performance differences between different types of recurrent units. In this study, the classification performance when using LSTM units was compared against performance when employing a GRU unit, finding LSTM units to be the more effective recurrent unit. The hybrid architectures varied significantly as to the number of convolutional or recurrent layers as well as to the order of these different architecture types. Further structured research is needed in order to better understand how varying different aspects of hybrid architectures affects performance.

Generally, the hybrid architectures that used a CNN as part of the design benefited from the addition of non-convolutional layers. Tabar and Halici [43] compared the accuracies between a CNN/SAE architecture against a CNN and an SAE and found that the hybrid architecture significantly outperformed against the rest. Jiao *et al* [41] designed channel-wise architectures, a standard DBN, fused CNN systems (where two architectures feed into a common fully connected layer), and a fused CNN/DBN hybrid, which outperformed all other architectures. Hajinoroozi *et al* [30, 69] analyzed the same collections of mental load task datasets, comparing channel-wise CNN's with an RBM in place of the first

convolutional layer against a channel-wise CNN, a standard CNN, an MLPNN, and several types of standard DBN's. The channel-wise hybrid significantly outperformed against the rest for cross-subject classification. This is significant because it indicates that this particular architecture type is effective for transfer learning of EEG analysis. Hybrid CNN architecture designs show promise for EEG classification, while further research is needed to adequately assess the effectiveness of each design.

5. Conclusions

Deep learning classification has been successfully applied to many EEG tasks, including motor imagery, seizure detection, mental workload, sleep stage scoring, event related potential, and emotion recognition tasks. The design of these deep network studies varied significantly over input formulization and network design. Several public datasets were analyzed in multiple studies, which allowed us to directly compare classification performances based on their design. Generally, CNN's, RNN's, and DBN's outperformed other types of deep networks, such as SAE's and MLPNN's. Additionally, CNN's performed best when using signal values or (spectrogram) images as inputs, whereas DBN's performed best when using signal values or calculated features as inputs. We further discussed deep network recommendations for each specific type of task. This recommendation diagram has been provided in the hope that it will guide the deployment of deep learning to EEG datasets in future research. Hybrid designs incorporating convolutional layers with recurrent layers or restricted Boltzmann machines showed promise in classification accuracy and transfer learning when compared against standard designs. We recommend more in-depth research of these combinations, particularly the number and arrangement of different layers including RBM's, recurrent layers, convolutional layers, and fully connected layers. Outside of network design, we also encourage further research to compare how deep networks interpret raw versus denoised EEG, as this has not yet been specifically assessed.

6. Declarations

6.1. Competing interests

The authors declare that there are no conflicts of interest regarding the publication of this paper.

6.2. Funding

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[19]	CNN 1 conv 1 FC 2 OUT	ReLU (conv) Softmax (FC)	f—SWD CC (256)	0.3–30 Hz	MR	BI 18 subj. 14 vids. Per	94.0%
[44]	CNN 1 conv 1 FC 4 OUT	ReLU (conv) Softmax (FC)	C2I—FFM All (64)	8–30 Hz	ALI	MI 9 1.44 h	61.0%
[43]	CNN 1 conv 6 FC 2 OUT	N/A	C2I—FFM All (3)	6–30 Hz	MR	MI 9 subj. 0.77 h	75.1%
[84]	CNN 16 conv 1 FC 2 OUT	ReLU (conv) Softmax (FC)	C2I—color scale All (6)	2–9 Hz	N/A	ERP—P300 66 subj. 25.7 h	81.0%
[23]	CNN 2 conv 2 FC 2 OUT	ReLU (conv) N/A	C2I—Spect All (8)	.5–7.5 Hz	ALI	ND 13 subj. 0.65 h	69.0%
[85]	CNN 2 conv 1 FC 2 out	Tanh (conv) N/A	SV All (28)	8–30 Hz	N/A	MI 2 subj. 1.03 h	86.0%
[86]	CNN 2 conv 1 FC 2 OUT	L-ReLU (conv) Softmax (FC)	SV All (64)	N/A	ALI	MI 109 subj. 10.9 h	80.0%
[87]	CNN 2 conv 1 FC 2 OUT	N/A	AA SV All (12)	N/A	N/A	SD 25 subj. 0.68 h	87.5%
[67]	CNN 2 conv 1 FC 2 OUT	L-ReLU (conv) Softmax (FC)	C2I— 3D Grid All (32)	1–50 Hz	ALI	ER 22 subj. 14.7 h	88.0%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[16]	CNN 2 conv 1 FC 3 OUT	Sigmoid (conv) Softmax (FC)	f—Stats All (19)	0.1–30 Hz	MR	AD 142 subj. 11.8 h	82.0%
[88]	CNN 2 conv 1 FC 5 OUT	Split Tanh (conv) N/A	CVT SV (1)	.5–30 Hz	N/A	SS 25 subj. 175 h	87.0%
[45]	CNN 2 conv 2 FC 2 OUT	SELU (conv) Softmax (FC)	C2I—FFM All (11)	8–30 Hz	N/A	MI 14 subj. 1 h	93.0%
[66]	CNN 2 conv 2 FC 2 OUT	ReLU (conv) Softmax (FC)	C2I—FFM All (32)	4–45 Hz	MR	ER 22 subj. 14.7 h	87.3%
[21]	CNN 2 conv 2 FC 4 OUT	L-ReLU (conv) Softmax (FC)	f—PSD All (62)	.5–45 Hz	MR	Gait 27 subj. 184 gait cycles	78.0%
[89]	CNN 2 conv 3 FC 2 OUT	ReLU (conv) tanh (FC) sigmoid (FC)	N SV All (64)	.1–20 Hz	AR	ERP—P300 — 27.7 h	75.0%
[90]	CNN 3 conv 1 FC 4 OUT	N/A	f—CSP All (22)	4–40 Hz	N/A	MI 9 1.44 h	74.0%
[72]	CNN 3 conv per 2 FC per 2–3 OUT	ReLU (conv) Softmax (FC)	N SV (1)	N/A	N/A	SD 5 subj. 4097 samples	99.0%
[91]	CNN 3 conv 1 FC 1 OUT	ReLU (conv) ___	C2I 3D Grid All (40)	0–50 Hz	N/A	ER 32 subj. 21.3 h	99.0%
[92]	CNN 3 conv 1 FC 12 OUT	ELU (conv) Softmax (FC)	SV All (8)	9–30 Hz	N/A	ERP—RSVP 10 subj. 2 h	80.0%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[93]	CNN 3 conv 1 FC 2 OUT	N/A	SV (1)	0.5–40 Hz	MR	MW 120 subj. 20 h	80.0%
[94]	CNN 3 conv 1 FC 2 OUT/ 2 OUT/ 4 OUT	ELU (conv) Softmax (FC)	SV All (64/56/22)	P300(1) 1–40 Hz P300(2) 1–40 Hz MI 4–40 Hz	N/A	P300(1) 15 0.5 h P300(2) 26 3 h MI 9 1.44 h	P300(1) 93% P300(2) 83% MI 70%
[80]	CNN 4 Conv 1 FC 4 OUT	N/A	SV All MS (3 of 62)	N/A	N/A	MI mult. sets	92.0%
[95]	CNN 4 conv 1 FC 4 OUT	N/A	N SV All (19)	N/A	ALI	MI 1 subj. 560 trials	100.0%
[46]	CNN 4 conv 2 FC 3 OUT	L-ReLU (conv) Softmax (FC)	C2I—3D Grid All (22)	N/A	N/A	SD 13 patients 336 h	90.0%
[41]	CNN 4 conv 2 FC 4 OUT	ReLU (conv) Softmax (FC)	C2I—Spect MS 21 (of 64)	4–30 Hz	MR	MW 15 subj. 1 h	90.0%
[79]	CNN 4 conv 2 FC 4 OUT	Tanh (conv) Softmax (FC)	SV All (26)	0.3–70 Hz	MR	SD 18 subj. 6 h total	89.0%
[96]	CNN 4 conv 3 FC 2 OUT	ReLU (conv) Softmax (FC)	SV All (64)	0.5–50 Hz	ALI	ERP—RSVP 18 subj. 18 h	73.0%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[97]	CNN 4 conv 4 OUT	ReLU (conv) Softmax (FC)	f—DE CW All (8)	4–40 Hz	N/A	MI 9 subj. 0.02 h	70.6%
[98]	CNN 5 conv 1 FC 2 OUT	ReLU (conv) Softmax (FC)	N SV All (16)	1–40 Hz	MR	MW 10 subj. 30 h	93.0%
[63]	CNN 5 conv 1 FC 2 OUT	PReLU (conv) Softmax (FC)	SV All (16)	N/A	ALI	ER 22 subj. 14.7 h	81.2%
[24]	CNN 5 conv 1 FC 2 OUT	ELU (conv) Softmax (FC)	SV All (21)	0–100 Hz	ALI	ND — 8522 ∗ 20 Min+	82.0%
[78]	CNN 5 conv 2 FC 3 OUT	ReLU (conv) Softmax (FC)	N SV (1)	N/A	MR	SD 10 subj. 3.3 h	88.7%
[20]	CNN 5 conv 3 FC 2 OUT	L-ReLU (conv) Softmax (FC)	N SV All (4)	N/A	MR	Depression 30 subj. 5 h	96.0%
[40]	CNN 5 conv 3 FC 5 OUT	ReLU (conv) Softmax (FC)	C2I—Spect All (2)	N/A	N/A	SS 19 subj. 304 h	86.0%
[99]	CNN 6 conv 1 FC 2 OUT	ReLU (conv) Softmax (FC)	SV All (8)	N/A	ALI	SD 18 subj. 835 h	AUC 97.1%
[100]	CNN 6 conv 1 FC 2 OUT	ReLU (conv) Softmax (FC)	SV All (24)	0.5–25 Hz	N/A	Gender 1308 subj. 29 h	63.0%
[101]	CNN 7 conv 3 FC 2 OUT	ReLU (conv) Sigmoid (FC) Softmax (FC)	N SV All (8)	1–15 Hz	N/A	SS 26 subj. 492 h	92.0%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[102]	DBN 1 RBM 1 FC 2 OUT	Softmax (FC)	N SV MS (3 of 13)	0.5–100 Hz	AR	MI 12 subj. 4.8 h	78.0%
[103]	DBN 1 RBM per ch. 8 FC Per ch. 2 OUT	N/A	SV CW (3)	8–30 Hz	MR	MI 4 subj. 0.47 h	83.0%
[76]	DBN 2 RBM 1 FC 3 OUT	N/A	f—Entropy MS (12 of 64)	0.3–50 Hz	MR	ER 15 subj. 2 h	86.1%
[104]	DBN 2 RBM 1 FC 7 OUT	Softmax (FC)	f—MAD SS crit channels	0–40 Hz	N/A	MW 6 subj. 10 h	93.0%
[26]	DBN 2 RBM 2 FC 3 OUT	N/A	f—PSD CW (62)	0.5–100 Hz	N/A	TE 12 subj. 12 h	77.0%
[31]	DBN 2 RBM 2 FC 3 OUT	N/A	f—PSD MS (8 of 32)	0.5–60 Hz	AR—DWT	MW 15 subj. 60 h	97.0%
[77]	DBN 3 RBM 1 FC 2 OUT	Softmax (FC)	f—WD All (3)	N/A	N/A	MI 3 subj. 140 trials	93.6%
[105]	DBN 3 RBM 1 FC 2 OUT	Softmax (FC)	f—WD All (3)	0.5–100 Hz	ALI	MI 9 subj. 1.4 h	83.0%
[70]	DBN 3 RBM 1 FC 3 OUT	N/A	f—stat All (64)	2–44 Hz	MR	ER 15 subj. 15 h	94.9%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[50]	DBN 3 RBM 1 FC 3 OUT	Softmax	f—PSD Ch. Avg (32)	4–45 Hz	N/A	ER 32 subj. 21.3 h	89.0%
[106]	DBN 3 RBM 1 FC 4 OUT	N/A	SV All (14)	N/A	MR	ER 21 subj. ——	68.4%
[107]	DBN 3 RBM N/A 4 OUT	N/A	f—SVD All (22)	5–30 Hz	MR	MI 9 subj. ——	80.4%
[18]	DBN 3 RBM SVM 2 OUT	N/A	SV CW (16)	0.5–30 Hz	N/A	AD 30 subj. 10 h	92.0%
[75]	DBN 3 RBM's 1 FC 2 OUT	N/A	AA SV MS (2 of 4)	1–30 Hz	N/A	ERP—P300 14 subj. 10 exp. per subj.	87.0%
[108]	DBN 3 RBM's 1 FC 3 OUT	N/A	f—PSD All (64)	8–35 Hz	AR	MI 9 subj. 15 h	71.0%
[109]	DBN 4 RBMs 1 FC 2 OUT	Softmax (FC)	f—WD All (16)	.3–30Hz	ALI	ERP—P300 10 subj. 5 h	81.0%
[110]	DBN Adapt. RBM 1 FC 3 OUT	N/A	f—Entropy Channel pairs (2 of 11)	N/A	AR—IC	MW 8 subj. total 22 h	77.4%
[69]	Hybrid CNN-DBN 1 RBM 3 conv 1 FC 2 OUT	N/A	SV CW All (30)	N/A	N/A	MW 37 subj. 9.7 h	82.8%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[61]	Hybrid CNN-RNN 1 conv 1 RNN layer 2 FC 2 OUT	Softmax (FC)	f—WD CW (32)	N/A	N/A	ER 32 subj. 21.3 h	73.0%
[62]	Hybrid CNN-RNN 2 conv 1 FC 1 RNN layer 1 FC 4 OUT	ReLU (conv) Softmax (FC)	C2I—color scale All (32)	N/A	N/A	ER 32 subj. 21.3 h	75.3%
[25]	Hybrid CNN-RNN 3 conv 3 GRU layers 1 FC 2 OUT	N/A softmax	SV All (32)	N/A	N/A	ND — 550 h	82.0%
[81]	Hybrid CNN-RNN 3 conv 3 LSTM layers 5 OUT	ReLU (conv) N/A	SV (1)	.3–40 Hz	N/A	SS 29 subj. 1073 h	80.0%
[111]	Hybrid CNN-RNN 3 paths, 2 conv per path, 1 LSTM 1 FC 2 OUT	ReLU (conv) sigmoid (FC)	f—PSD All (128)	3–55 Hz	MR	MW 8 subj. 6 h	87.0%
[112]	Hybrid CNN-RNN 4 conv 2 RNN layers 1 FC 5 OUT	ReLU (conv) Softmax (FC)	SV CW MS (2 of 20)	0.3–100 Hz	MR	SS 62 subj. 837.6 h	91.3%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[113]	Hybrid CNN-RNN 7 conv 1 LSTM layer 2 FC 4 OUT	ReLU (conv) Softmax (FC)	C2I—Spect All (64)	4–30 Hz	MR	MW (RSVP) 13 subj. 3.5 h	91.0%
[39]	Hybrid CNN-RNN 9 conv 1 FC 1 LSTM layer 1 FC 4 OUT	ReLU (conv) Softmax (FC)	C2I—Spect All (64)	4–30 Hz	ALI	MW 25 subj. 8.1 h	93.0%
[114]	Hybrid MLPNN-RNN 2 hid. 1 RNN layer 1 FC 5 OUT	ReLU (FC) Softmax (FC)	f—PSD MS (2 of 20)	0.06–30 Hz	MR	SS 62 subj. 838 h	83.6%
[115]	MLPNN 2 hid. 2 OUT	ReLU	f—STFT All (9)	1–64 Hz	MR	ER 16 subj. 0.35 h	64.0%
[71]	MLPNN 2 hid. 2 OUT	Sigmoid	f—PSD All (18 / 23)	N/A	ALI	SD 23 subj. 2.65 h	94.2%
[55]	MLPNN 2 hid. 2 OUT	Softmax	SV All (118 / 58)	9–13 Hz	N/A	MI 10 subj. —	75.0%
[116]	MLPNN 3 hid. 2 OUT	N/A	f—CSP All (22)	8–30 Hz	N/A	MI 9 1.44 h	85.0%
[117]	MLPNN 3 hid. 2 OUT	N/A	f—Stats MS (2 of 6)	0.5–40 Hz	ALI	SS 83 subj. 664 h	90.0%
[17]	MLPNN 4 hid. 2 OUT	Sigmoid	f—PSD All (32)	4–30 Hz	ALI	AD 20 patients 0.33 h	59.0%
[118]	MLPNN 4 hid. 2 OUT	N/A	f—PSD All (8)	.1–100 Hz	N/A	MI 7 subj. 1.15 h	85.0%

(*Continued*)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[64]	MLPNN 4 hid. 4 OUT	ReLU (FC) Softmax (FC)	f—PSD All (5)	4–45 Hz	MR	ER 32 subj. 21.3 h	82.5%
[119]	RNN—GRU 1 GRU layer 1 FC 4/2 OUT	N/A	f—CSP All (22)	8–30 Hz	ALI	MI (1) 9 1.44 h MI (2) 9 subj. 4 h	77% 82%
[49]	RNN—LSTM 1 layer 2 FC 2/3/5 OUT	Softmax (FC)	C2I—3D Grid (1)	0.53–40 Hz	ALI	SD 5 subj. 3.2 h	100.0%
[73]	RNN—LSTM 2 layer 2 FC 2 OUT	N/A	f—PSD/Stats All (18)	N/A	ALI	SD 23 subj. 983 h	100.0%
[120]	RNN—LSTM 2 layers 1 FC 2 OUT	N/A	SV (1)	N/A	N/A	SD 5 subj. 3.2 h	91.0%
[121]	RNN—LSTM 2 layers 1 FC 2 OUT	Sigmoid (FC)	f—PSD All (19)	3–55 Hz	N/A	MW 6 subj. 10 h	93.0%
[122]	RNN—LSTM 2 layers 2 FC 5 OUT	N/A Softmax	SV All (64)	N/A	ALI	MI 12 subj. 98 trials per	68.0%
[65]	RNN—LSTM 2 LSTM layers 1 FC 2 OUT	Sigmoid (FC)	SV All (-)	N/A	MR	ER 22 subj. 14.7 h	87.0%
[82]	RNN—LSTM 3 layers 2 OUT	N/A	f—Stats All (22)	8–35Hz	N/A	MI 9 1.44 h	79.6%
[47]	RNN 2 layer 1 FC 2 OUT	N/A	C2I—2D Grid channel pairs (2 of 23)	3–30 Hz	AR	SD 5 subj. 20 h	100.0%

(Continued)

Study	Deep learning strategy	Activation function	Input formulation	Frequency range	Artifact strategy	Task information	Accuracy
[123]	SAE 20 ind. AE's Avg. of AE's 5 OUT	Sigmoid (AE) N/A	f—WD All (4)	0.5–100 Hz	MR	SS 20 subj. 800 h	89.0%
[48]	SAE 3 hid. 1 FC 2 OUT	N/A	f—PCA, CSA, PSD All (32)	4–100 Hz	N/A	ER 32 subj. 21.3 h	53.4%
[53]	SAE 3 hid. 1 FC 2 OUT	Sigmoid (AE) Softmax (FC)	SV (1)	N/A	MR	SD 10 subj. 3.3 h	96.0%
[54]	SAE 3 hid. 3 OUT	ReLU (AE) Softmax (FC)	f—WD MS (3 of -)	N/A	N/A	ERP—P300 239 subj. 39.8 h	74.0%
[74]	SAE 4 hid. 2 OUT	N/A	f—Stats All (19)	N/A	MR	ERP—P300 15 subj. —	69.0%
[68]	SAE 5 hid. 1 FC 2 OUT	N/A	f—PSD (1)	1.5–40 Hz	N/A	MW 8 subj. 12 h	74.3%
[32]	SAE 6 hid. 1 FC 2 OUT	N/A	f—PSD MS (11 of -)	1–40 Hz	AR IC	MW 7 subj. 1 h	95.5%

ORCID iDs

Alexander Craik ⦿ https://orcid.org/0000-0003-3871-1839
Yongtian He ⦿ https://orcid.org/0000-0002-3700-4406
Jose L Contreras-Vidal ⦿ https://orcid.org/0000-0002-6499-1208

References

[1] He Y, Eguren D, Azorín J M, Grossman R G, Luu T P and Contreras-Vidal J L 2018 Brain–machine interfaces for controlling lower-limb powered robotic systems *J. Neural Eng.* **15** 021004

[2] Motamedi-Fakhr S, Moshrefi-Torbati M, Hill M, Hill C M and White P R 2014 Signal processing techniques applied to human sleep EEG signals—a review *Biomed. Signal Process. Control* **10** 21–33

[3] Chen G 2014 Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features *Expert Syst. Appl.* **41** 2391–4

[4] Kandel T M, Schwartz E R and Jessell J H (ed) 2000 *Principles of Neural Science* (New York: McGraw-Hill)

[5] Velu P D and de Sa V R 2013 Single-trial classification of gait and point movement preparation from human EEG *Front. Neurosci.* **7** 840

[6] Subasi A, Gursoy M I and Ismail Gursoy M 2010 EEG signal classification using PCA, ICA, LDA and support vector machines *Expert Syst. Appl.* **37** 8659–66

[7] Lee K, Liu D, Perroud L, Chavarriaga R and del Millán J R 2017 A brain-controlled exoskeleton with cascaded event-related desynchronization classifiers *Rob. Auton. Syst.* **90** 15–23

[8] Bengio Y, Simard P and Frasconi P 1994 Learning long-term dependencies with gradient descent is difficult *IEEE Trans. Neural Netw.* **5** 157–66

[9] Lecun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44

[10] Guerra E, de Lara J, Malizia A and Díaz P 2009 Supporting user-oriented analysis for multi-view domain-specific visual languages *Inf. Softw. Technol.* **51** 769–84

[11] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R and Li F F 2014 Large-scale video classification with convolutional neural networks *Proc. IEEE Computer Society Conf. Computer Vision Pattern Recognition* pp 1725–32

[12] Graves A, Mohamed A and Hinton G 2013 Speech recognition with deep recurrent neural networks *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* pp 6645–9

[13] Sutskever I, Martens J and Hinton G E 2011 Generating text with recurrent neural networks *Proc. of the 28th Int. Conf. on Machine Learning* pp 1017–24

[14] Greenspan H, van Ginneken B and Summers R M 2016 Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique *IEEE Trans. Med. Imaging* **35** 1153–9

[15] Moher D, Liberati A, Tetzlaff J and Altman D G 2010 Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement *Int. J. Surg.* **8** 336–41

[16] Morabito F C *et al* 2016 Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings *2016 IEEE 2nd Int. Forum Research Technologies Society Industry Leveraging a better Tomorrow* pp 1–6

[17] Kim D and Kim K 2018 Detection of early stage Alzheimer's disease using EEG relative power with deep neural network *40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)* pp 352–5

[18] Zhao Y 2014 Deep learning in the EEG diagnosis of Alzheimer's disease *2014 Asian Conf. Computer Visison* pp 1–15

[19] Baltatzis V, Bintsi K-M, Apostolidis G K and Hadjileontiadis L J 2017 Bullying incidences identification within an immersive environment using HD EEG-based analysis: a swarm decomposition and deep learning approach *Sci. Rep.* **7** 17292

[20] Acharya U R, Lih S, Hagiwara Y, Hong J, Adeli H and Subha D P 2018 Automated EEG-based screening of depression using deep convolutional neural network *Comput. Methods Progr. Biomed.* **161** 103–13

[21] Gait-pattern E, Goh S K, Abbass H A, Member S and Tan K C 2018 Spatio—spectral representation learning for classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **26** 1858–67

[22] Guo Y, Friston K, Faisal A, Hill S and Peng H 2015 *Brain Informatics and Health: 8th International Conference, BIH 2015 London, UK, August 30–September 2, 2015 Proceedings (Lecture Notes Computer Science (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics* vol 9250) pp 306–16

[23] Vrbancic G and Podgorelec V 2018 Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks *Elektron. Elektrotech.* **24** 3–7

[24] Van Leeuwen K G, Sun H, Tabaeizadeh M, Struck A F, Van Putten M J A M and Westover M B 2019 Detecting abnormal electroencephalograms using deep convolutional networks *Clin. Neurophysiol.* **130** 77–84

[25] Roy S, Kiral-kornek I and Harrer S 2018 Deep learning enabled automatic abnormal EEG identification *40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)* pp 2756–9

[26] Al-kaysi A M, Al-Ani A and Boonstra T W 2015 A multichannel deep belief network for the classification of EEG data *Neural Inf. Process.* **9492** 38–45

[27] Teplan M 2002 Fundamentals of EEG measurement *Meas. Sci.* **2** 1–11

[28] Barrett L F 1998 Discrete emotions or dimensions? The role of valence focus and arousal focus *Cogn. Emot.* **12** 579–99

[29] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain–computer communication **89** 1123–34

[30] Hajinoroozi M, Mao Z, Jung T P, Lin C T and Huang Y 2016 EEG-based prediction of driver's cognitive performance by deep convolutional neural network *Signal Process. Image Commun.* **47** 549–55

[31] Li F *et al* 2017 Deep models for engagement assessment with scarce label information *IEEE Trans. Hum.-Mach. Syst.* **47** 598–605

[32] Yin Z and Zhang J 2017 Cross-session classification of mental workload levels using EEG and an adaptive deep learning model *Biomed. Signal Process. Control* **33** 30–47

[33] Andrzejak R G, Lehnertz K, Mormann F, Rieke C, David P and Elger C E 2001 Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state *Phys. Rev.* E **64** 8

[34] Benson W E, Brown G C, Pp T and London S 1990 Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine *J. Ophthal.* **74** 255

[35] Nathan K and Contreras-Vidal J L 2016 Negligible motion artifacts in scalp electroencephalography (EEG) during treadmill walking *Front. Hum. Neurosci.* **9** 1–12

[36] Delorme A, Sejnowski T and Makeig S 2007 Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis *NeuroImage* **34** 1443–9

[37] Kline J E, Huang H J, Snyder K L and Ferris D P 2015 Isolating gait-related movement artifacts in electroencephalography during human walking *J. Neural Eng.* **12** 046022

[38] Kilicarslan A, Grossman R G and Contreras-Vidal J L 2016 A robust adaptive denoising framework for real-time artifact removal in scalp EEG measurements *J. Neural Eng.* **13** 026013

[39] Kuanar S, Athitsos V, Pradhan N, Mishra A and Rao K R Cognitive analysis of working memory load from EEG, by a deep recurrent neural network *2018 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 352–5

[40] Vilamala A, Madsen K H and Hansen L K 2017 Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring *2017 IEEE 27th Int. Workshop on Machine Learning for Signal Processing (MLSP)*

[41] Jiao Z, Gao X, Wang Y, Li J and Xu H 2018 Deep Convolutional Neural Networks for mental load classification based on EEG data *Pattern Recognit.* **76** 582–95

[42] Wagner J, Solis-escalante T, Grieshofer P, Neuper C, Müller-putz G and Scherer R 2012 NeuroImage Level of participation in robotic-assisted treadmill walking modulates midline sensorimotor EEG rhythms in able-bodied subjects *NeuroImage* **63** 1203–11

[43] Tabar Y R and Halici U 2017 A novel deep learning approach for classification of EEG motor imagery signals *J. Neural Eng.* **14** 016003

[44] Abbas W and Khan N A 2018 DeepMI : deep learning for multiclass motor imagery classification *40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)* pp 219–22

[45] Wang Z, Zhang Z, Gong X, Sun Y and Wang H 2018 Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition *Concurrency and Computation: Practice and Experience* **30** e4413

[46] Wei X, Zhou L, Chen Z, Zhang L and Zhou Y 2018 Automatic seizure detection using three- dimensional CNN based on multi-channel EEG **18** 111

[47] Vidyaratne L, Glandon A, Alam M and Iftekharuddin K M 2016 Deep recurrent neural network for seizure detection *2016 Int. Joint Conf. Neural Networks* pp 1202–7

[48] Jirayucharoensak S, Pan-Ngum S, Israsena P, Jirayucharoensak S, Pan-Ngum S and Israsena P 2014 EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation, EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation *Sci. World J.* **2014** e627892

[49] Hussein R, Palangi H, Ward R K and Wang Z J 2019 Deep neural network architecture for robust detection of epileptic seizures using EEG signals *Clin. Neurophysiol.* **130** 25–37

[50] Xu H and Plataniotis K N 2016 Affective states classification using EEG and semi-supervised deep learning approaches *2016 IEEE 18th Int. Workshop Multimedia Signal Processing* pp 1–6

[51] Bojarski M *et al* 2016 End to end learning for self-driving cars (arXiv:1604.07316)

[52] Malafeev A *et al* 2018 Automatic human sleep stage scoring using deep neural networks *Front. Neurosci.* **12** 781

[53] Qin L, Ye S-Q, Huang X-M, Li S-Y, Zhang M-Z and Xue Y 2016 *Classification of Epileptic EEG Signals with Stacked Sparse Autoencoder Based on Deep Learning* vol 10363 (Dongguan: Guangdong Medical University)

[54] Vareka L 2017 *Application of Stacked Autoencoders to P300 Experimental Data (Int. Conf. on Artificial Intelligence and Soft Computing)* (Cham: Springer)

[55] Sturm I, Lapuschkin S, Samek W and Müller K R 2016 Interpretable deep neural networks for single-trial EEG classification *J. Neurosci. Methods* **274** 141–5

[56] Koelstra S, Christian Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A and Patras I 2012 DEAP: a database for emotion analysis; using physiological signals *IEEE Trans. on Affective Computing* **3** 18–31

[57] Zhang J, Yin Z and Wang R 2015 Recognition of mental workload levels under complex human-machine collaboration by using physiological features and adaptive support vector machines *IEEE Trans. Hum.-Mach. Syst.* **45** 200–14

[58] Tangermann M *et al* 2012 Review of the BCI competition IV *Front. Neurosci.* **6** 1–31

[59] Kemp B, Zwinderman A H, Tuk B, Kamphuisen H A C and Oberyé J J L 2000 Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG *IEEE Trans. Biomed. Eng.* **47** 1185–94

[60] O'Reilly C, Gosselin N, Carrier J and Nielsen T 2014 Montreal archive of sleep studies: An open-access resource for instrument benchmarking and exploratory research *J. Sleep Res.* **23** 628–35

[61] Li X, Song D, Zhang P, Yu G, Hou Y and Hu B 2016 Emotion recognition from multi-channel EEG data through convolutional recurrent neural network *2016 IEEE Int. Conf. Bioinformatics Biomedicine* pp 352–9

[62] Li Y, Huang J, Zhou H and Zhong N 2017 Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks *Appl. Sci.* **7** 1060

[63] Yanagimoto M and Sugimoto C 2016 Recognition of persisting emotional valence from EEG using convolutional neural networks *2016 IEEE 9th Int. Workshop Computational Intelligence Applications* pp 27–32

[64] Yuen T C, San W S and Rizon M 2009 Classification of human emotions from electroencephalogram (EEG) signal using deep neural network *Int. J. Integr. Eng.* **1** 71–9

[65] Alhagry S 2017 Emotion recognition based on EEG using LSTM recurrent neural network *Emotion* **8** 8–11

[66] Qiao R, Qing C, Zhang T, Xing X and Xu X 2017 A novel deep-learning based framework for multi-subject emotion recognition *ICCSS 2017—2017 Int. Conf. Information, Cybernetics and Computational Social Systems* pp 181–5

[67] Salama E S, El-khoribi R A, Shoman M E and Shalaby M A W 2018 EEG-based emotion recognition using 3D convolutional neural networks *Int. J. Adv. Comput. Sci. Appl.* **9** 329–37

[68] Yin Z and Zhang J 2016 Recognition of cognitive task load levels using single channel EEG and stacked denoising autoencoder *Chinese Control Conf. CCC* vol 2016 pp 3907–12

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

[70] Huang J, Xu X, Zhang T and Chen L 2017 Emotion classification using deep neural networks and emotional patches *2017 IEEE Int. Conf. Bioinformatics Biomedicine* pp 958–62

[71] Birjandtalab J, Heydarzadeh M and Nourani M 2017 Automated EEG-based epileptic seizure detection using deep neural networks *2017 IEEE Int. Conf. Healthcare Informatics* pp 552–5

[72] Ullah I, Hussain M, Qazi E and Aboalsamh H 2018 An automated system for epilepsy detection using EEG brain signals based on deep learning approach *Expert Syst. Appl.* **107** 61–71

[73] Tsiouris K M, Pezoulas V C, Zervakis M, Konitsiotis S, Koutsouris D D and Fotiadis D I 2018 A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals *Comput. Biol. Med.* **99** 24–37

[74] Vă L 2017 Stacked autoencoders for the P300 component detection *Front. Neurosci.* **11** 302

[75] Kulasingham J P, Vibujithan V and Mieee A C D S 2016 Deep belief networks and stacked autoencoders for the P300 guilty knowledge test *2016 IEEE EMBS Conf. Biomedical Engineering Sciences* pp 127–32

[76] Zheng W L and Lu B L 2015 Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks *IEEE Trans. Auton. Ment. Dev.* **7** 162–75

[77] Li M A, Zhang M and Sun Y-J 2016 A novel motor imagery EEG recognition method based on deep learning *Proc. of the 2016 Int. Forum on Management, Education and Information Technology Application* pp 728–33

[78] Acharya U R, Oh S L, Hagiwara Y, Tan J H and Adeli H 2018 Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals *Comput. Biol. Med.* 270–8

[79] Antoniades A *et al* 2017 Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 2285–94

[80] Schirrmeister R T *et al* 2017 Deep learning with convolutional neural networks for EEG decoding and visualization *Hum. Brain Mapp.* **38** 5391–420

[81] Bresch E, Großekathöfer U and Garcia-molina G 2018 Recurrent deep neural networks for real-time sleep stage classification from single channel EEG *Front. Comput. Neurosci.* **12** 85

[82] Wang P, Jiang A, Liu X, Shang J and Zhang L 2018 LSTM-Based EEG classification in motor imagery task *IEEE Trans. Neural Syst. Rehabil. Eng.* **26** 2086–95

[83] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80

[84] Pereira A *et al* 2018 Cross-subject EEG event-related potential classification for brain–computer interfaces using residual networks *Preprint* (HAL-id:hal-01878227)

[85] Tang Z, Li C and Sun S 2017 Single-trial EEG classification of motor imagery using deep convolutional neural networks *Optik* **130** 11–8

[86] Dose H, Møller J S, Iversen H K and Puthusserypady S 2018 An end-to-end deep learning approach to MI-EEG signal classification for BCIs *Expert Syst. Appl.* **114** 532–42

[87] Antoniades A, Spyrou L, Took C C and Sanei S 2016 Deep learning for epileptic intracranial EEG data *2016 IEEE 26th Int. Workshop Machine Learning Signal Processing* pp 1–6

[88] Zhang J and Wu Y 2018 Complex-valued unsupervised convolutional neural networks for sleep stage classification *Comput. Methods Programs Biomed.* **164** 181–91

[89] Liu M, Wu W, Gu Z, Yu Z, Qi F and Li Y 2018 Deep learning based on Batch Normalization for P300 signal detection *Neurocomputing* **275** 288–97

[90] Sakhavi S, Yan S and Guan C 2018 Learning temporal information for brain–computer interface using convolutional neural networks *IEEE Trans. Neural Netw. Learn. Syst.* **29** 5619–29

[91] Moon S-E, Jang S and Lee J-S 2018 Convolutional neural network approach for eeg-based emotion recognition using brain connectivity and its spatial information *2018 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*

[92] Waytowich N R *et al* 2018 Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials *J. Neural Eng.* **15** 066031

[93] Ang K K and Guan C 2019 Inter-subject transfer learning with end-to-end deep convolutional neural network for EEG-based BCI *J. Neural Eng.* **16** 026007

[94] Lawhern V J, Solon A J, Waytowich N R, Gordon S M and May L G 2018 EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces *J. Neural Eng.* **15** 056013

[95] Liu J, Cheng Y and Zhang W 2015 Deep learning EEG response representation for brain computer interface *Chinese Control Conf. CCC* vol 2015 pp 3518–23

[96] Shamwell J, Lee H, Kwon H, Marathe A R, Lawhern V and Nothwang W 2016 Single-trial EEG RSVP classification using convolutional neural networks *Proc. SPIE* **9836** 983622

[97] Sakhavi S, Guan C and Yan S 2015 Parallel convolutional-linear neural network for motor imagery classification *2015 23rd European Signal Processing Conf. EUSIPCO 2015* pp 2736–40

[98] Zeng H, Yang C, Dai G, Qin F, Zhang J and Kong W 2018 Classification of driver mental states by deep learning *Cogn. Neurodyn.* **12** 597–606

[99] O'Shea A, Lightbody G, Boylan G and Temko A 2017 Neonatal seizure detection using convolutional neural networks *2017 IEEE 27th Int. Workshop on Machine Learning for Signal Processing (MLSP)*

[100] Van Putten M J A M, Olbrich S and Arns M 2018 Predicting sex from brain rhythms with deep learning *Sci. Rep.* **8** 1–7

[101] Ansari M L A and de Wel O 2018 Quiet sleep detection in preterm infants using deep convolutional neural networks *J. Neural Eng.* **15** 066006

[102] Kobler R J and Scherer R 2016 Restricted boltzmann machines in sensory motor rhythm brain–computer interfacing: a study on inter-subject transfer and co-adaptation *2016 IEEE Int. Conf. Syst. Man, Cybernetics SMC 2016—Conf. Proc.* pp 469–74

[103] An X, Kuang D, Guo X, Zhao Y and He L 2014 A deep learning method for classification of EEG data based on motor imagery *Int. Conf. on Intelligent Computing* (Cham: Springer) pp 203–10

[104] Zhang J and Li S 2017 A deep learning scheme for mental workload classification based on restricted Boltzmann machines *Cogn. Technol. Work* **19** 607–31

[105] Lu N, Li T, Ren X and Miao H 2017 A deep learning scheme for motor imagery classification based on restricted Boltzmann machines *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 566–76

[106] Gao Y, Lee H J and Mehmood R M 2015 Deep learning of EEG signals for emotion recognition *2015 IEEE Int. Conf. Multimedia Expo Workshops ICMEW 2015* vol 2 pp 1–5

[107] Yu Z and Song J 2017 Multi-class motor imagery classification by singular value decomposition and deep Boltzmann machine *2017 IEEE 3Rd Inf. Technology Mechatronics Engineering Conf.* pp 376–9

[108] Chu Y, Zhao X, Zou Y, Xu W and Han J 2018 A decoding scheme for incomplete motor imagery EEG with deep belief network *Front. Neurosci.* **12** 680

[109] Bablani A, Edla D R and Kuppili V 2018 Deceit identification test on EEG Data using deep belief network *2018 9th Int. Conf. Computing Communication Networking Technologies* pp 1–6

[110] Yin Z and Zhang J 2017 Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights *Neurocomputing* **260** 349–66

[111] Hefron R, Borghetti B, Schubert Kabban C, Christensen J and Estepp J 2018 Cross-participant EEG-based assessment of cognitive workload using multi-path convolutional recurrent neural networks *Sensors* **18** 1339

[112] Supratak A, Dong H, Wu C and Guo Y 2017 DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 1998–2008

[113] Bashivan P, Rish I, Yeasin M and Codella N 2015 Learning representations from eeg with deep recurrent-convolutional neural networks (arXiv:1511.06448)

[114] Dong H, Supratak A, Pan W, Wu C, Matthews P M and Guo Y 2018 Mixed neural network approach for temporal sleep stage classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **26** 324–33

[115] Teo J, Hou C L and Mountstephens J 2017 Deep learning for EEG-based preference classification *AIP Conf. Proc.* **1891** 020141

[116] She Q, Hu B, Luo Z, Nguyen T and Zhang Y 2019 A hierarchical semi-supervised extreme learning machine method for EEG recognition *Med. Biol. Eng. Comput.* **57** 147–57

[117] Shahin M, Ahmed B, Ben Hamida S T, Mulaffer F L, Glos M and Penzel T 2017 Deep learning and insomnia: assisting clinicians with their diagnosis *IEEE J. Biomed. Heal. Inform.* **21** 1546–53

[118] Yohanandan S A C *et al* 2018 A robust low-cost EEG motor imagery-based brain–computer interface *2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)* pp 5089–92

[119] Luo T, Zhou C and Chao F 2018 Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network (BMC) *Bioinformatics* **19** 344

[120] Ahmedt-aristizabal D, Fookes C, Nguyen K and Sridharan S 2018 Deep classification of epileptic signals *2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)* pp 332–5

[121] Hefron R G, Borghetti B J, Christensen J C and Schubert C M 2017 Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation *Pattern Recognit. Lett.* **94** 96–104

[122] Ma X, Qiu S, Du C, Xing J and He H 2018 Improving EEG-based motor imagery classification via spatial and temporal recurrent neural networks *2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)* pp 1903–6

[123] Tsinalis O, Matthews P M and Guo Y 2016 Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders *Ann. Biomed. Eng.* **44** 1587–97