Neural Processing Letters (2021) 53:2917–2932
https://doi.org/10.1007/s11063-021-10533-7

Deep Learning with ConvNet Predicts Imagery Tasks Through EEG

Gokhan Altan[1] · **Apdullah Yayık**[2] · **Yakup Kutlu**[1]

Accepted: 18 May 2021 / Published online: 25 May 2021

Abstract
Deep learning with convolutional neural networks (ConvNets) has dramatically improved the learning capabilities of computer vision applications just through considering raw data without any prior feature extraction. Nowadays, there is a rising curiosity in interpreting and analyzing electroencephalography (EEG) dynamics with ConvNets. Our study focused on ConvNets of different structures, the efficiency of multiple machine learning algorithms with optimization on ConvNets, constructing for predicting imagined left and right movements on a subject-independent basis through raw EEG data. We adapted novel lower-upper triangularization based extreme learning machines (LuELM) to the ConvNet architecture. Results showed that recently advanced methods in machine learning field, i.e. adaptive moments and batch normalization together with dropout strategy, improved ConvNets predicting ability, outperforming that of conventional fully-connected neural networks with widely-used spectral features. The proposed prediction model achieved improvements in classification performances with the rates of 90.33%, 91.00%, and 89.67% for accuracy, recall, and specificity, respectively.

Keywords ConvNets · Deep learning · Predicting imagined hand movements · EEG

1 Introduction

Machine learning methods together with electroencephalography (EEG) data empower researchers to interpret neurological activities, and are key components of the brain-computer interface (BCI) research field. For instance, such systems can enable locked-in patients to type phone numbers [44], to use wheel-chair [5] and to operate computer explorer [4]. In addition, such systems may be used in prediction onset of stroke [3]. Although these successful and

✉ Gokhan Altan
gokhan.altan@iste.edu.tr

Apdullah Yayık
apdullah.yayik@huawei.com

1 Department of Computer Engineering, Iskenderun Technical University, Hatay, Turkey

2 Huawei R&D Center, Istanbul, Turkey

promising studies, a general framework for extracting features and learning mechanism with regard to recent advances in machine learning field is still needed.

Deep learning with convolutional neural networks (ConvNets) is of prominent recent advances in machine learning, particularly computer vision. They are the most successfully biologically inspired neural networks since their principles and structures rely on nonscientific hierarchical learning [12]. Following achievements in computer vision, it continued in a straight way in sentiment analysis from text [8] and audio processing [16]. Nowadays, handcrafted-features have lost their usefulness with ConvNets capability to reveal prominent features from input data via end-to-end hierarchical representation. In addition to the high classification performances of ConvNets in image and sound analysis, It has also been a popular focus for the various time-series analysis in recent years. The significant characteristics of Deep Learning including using many hidden layers, transfer learning, and extracting deterministic features for low-, middle- and high-levels by transferring the feature activation maps layer-by-layer, feature learning, and more make it easy to achieve effective generalization capabilities on time-series. ConvNet has been focused on identifying different neurological disorders and cognitive tasks using EEG recordings. Zhang et al. used ConvNet on Hilbert-Huang transform-based frequency-energy-time distribution to identify the sleep apnea disorder. They smoothed the frequency domain plot using the autoencoder model and applied ConvNet to the EEG channels with sampling rates of 128 Hz and 250 Hz. They proposed an orthogonal ConvNet algorithm and classified the recordings with classification accuracy rates of 88.4% and 87.6% for 128 Hz and 250 Hz, respectively [51]. Mousavi et al. used batch normalization and ConvNet on EEG recordings with a sampling rate of 100 Hz to detect the sleep stage on the different number of sleep stages. They utilized overlapped shifting segmentation method to overcome the problem of unbalanced sleep stages. They applied the increasing size of convolution filters for their model and used two fully connected layers with MLP on the supervised stage of the model. They achieved the classification accuracy rates from 92.95% to 98.10% for identifying 2–6 sleep stages [31]. Raghu et al. applied pre-trained ConvNet architectures including VGGNet, GoogleNet, DenseNet, ResNet and more using transfer learning flexibility of Deep learning to detect seizure type on 16-channel EEG recordings. They experimented with support vector machines (SVM) and MLP at supervised learning as different models with various optimizations. They reported the highest classification accuracy rate of 88.30% using InceptionV3 architecture with SVM with radial basis kernel function and MLP with Adam optimization [35]. Acharya et al. also proposed a ConvNet architecture to identify seizure on EEG recordings with a sampling rate of 173.61 Hz. They compared the efficiency of MLP and ConvNet on EEG. They achieved classification performance rates of 88.67%, 95.00%, and 90.00% for accuracy, sensitivity, and specificity using ConvNet, respectively. They reported the superiority of the ConvNet over simple MLP by feature learning capabilities [1]. Sun et al. proposed a ConvNet with long short term memory model neural networks for EEG-based human identification. They analyzed the EEG dataset on motor imagery tasks for their proposal on 16-channel EEG recordings from 109 subjects with a sampling rate of 160 Hz. They fed the ConvNet features to the long short term memory model and sequentially two fully connected layers at the supervised learning stage. They separated the subjects with an averaged accuracy of 99.58% using directly EEG signals to the ConvNet model [43]. San-Segundo et al. applied ConvNet to detect epilepsy on various transformation plots. They extracted frequency distributions using Fourier, wavelet and six intrinsic mode functions using empirical mode decomposition. They extracted the plots obtained from signal transformations were fed into the ConvNet with two fully connected layers based on root-mean-square propagation (RmsPROP). They reported the increase of seizure detection accuracy rate by 99% to 99.5% using empirical mode decomposition filter modulations with ConvNet [36].

This paper concentrated on a challenging task of predicting imagined left and right movements through raw EEG data with ConvNet on a subject-independent basis with considering 109 number of subjects. In the literature, studies on EEG motor movement/imagery (EEG-MMI) database aim to predict imagined movements through the use of SVM or multi-layer perceptron (MLP) achieved success on either only a subject-dependent basis or a subject-independent basis but for limited subjects. Mostly, it is claimed that these specialized tasks could uniquely be predicted just for each subject. Additional researches were performed in distinguishing executed and imaginary motor movements [41,48] that differ from our study in that we concern with predicting imagined motor left and right fist movements. Mohammed et al. proposed SVM learning model for predicting motor-imagery activities based on wavelet spectral analysis. They have reached an accuracy of 84% on a subject independent basis for only 20 subjects [2]. Schirrmeister et al. analyzed EEG to obtain task decoding. They compared the efficiency and robustness of the filter bank common spatial patterns(FBCSP) algorithm and ConvNet. They analyzed various large- and small-scaled EEG datasets with hybrid and ConvNet architectures. They achieved motor-imagery task classification accuracy rates of 71.2%, 72.2%, and 67.7% for FBCSP, ConvNet, and Shallow ConvNet, respectively. They reported the applicability of their proposal for the visualization of EEG bands for channels [38]. Cecotti and Gräser also used ConvNet on EEG to detect the P300 waves for event-related potentials. They analyzed two subjects from the P300 speller dataset in BCI Competition III. They evaluated the generalization performances of multiple machine learning algorithms on ConvNet. They identified the P300 waves in EEG with classification performance rates of 70.37–78.19%, 67.40–69.2%, and 31.7–40.9% for accuracy, recall, and precision, respectively. They reported the advantages of ConvNet with MLP against ConvNet with SVM (Linear and Gaussian kernels) [6].

Besides, studies not-using EEGMMI database reached promising results with considering artifact removal at preprocessing, energy, and power features [13], proposing Joint Approximate Diagonalization method for handling non-stationary characteristics of EEG that aids in predicting imagined movements [30], integrating magnetoencephalographic signals with EEG and converting EEG time-series into 2D mesh-like hierarchy together with convolutional recurrent neural network [50].

EEG is a non-stationary and nonlinear time-series signal which has recent advancements for neurological disabilities and more. It is commonly recorded various numbers of channels that make it analyzed and understood. Whereas increasing the number of EEG channels gives rise to challenging analysis, various studies are constantly developing novel algorithms to overcome this issue. Wu et al. proposed a Bayesian framework for easing the multi-channel EEG analysis and avoiding overfitting the machine learning models by exploiting the spatial patterns [46]. EEG data are physically dissimilar to typical 2-D or 3-D images input of ConvNets, they consist of time-series from several electrodes on the scalp surface, can be conceptualized as 2-D, the voltage varies over time and space, where space refers to electrodes. In the neuroscience field, EEG data are assumed to be originated from several dipolar current sources in the brain and they are linear combinations of them. From this perspective, spatial relations should be preserved and are of key components in EEG data to reveal data of high signal-to-noise-ratio from that of low signal-to-noise-ratio. Therefore, the adaptation of ConvNets inputs for EEG data should be handled. In addition, design-choices and learning strategies should be compared. Unlike ConvNet with many machine learning algorithms for the supervised learning stage of the models, advanced techniques were also proposed for motor imagery classification. Li et al. studied on modeling a hybrid algorithm to detect event-related potential on EEG by spatio-temporal patterns. They used restricted Boltzmann machines based temporal features on multi-channel EEG. They reported an average AUC score of 0.889 for 11 subjects [26]. Qi et al. proposed a regularized spatio-temporal filtering on EEG. In the first step, they enhanced spatial and high-order temporal filters. They applied the filters using eigenvalue decomposition. In the second step, they integrated the Fisher linear discriminant analysis as classifier and feature dimensionality reduction step on single-trial EEG recordings. They specified the advantages of optimization on filters and the robustness of their algorithm on various multi-channel EEG datasets [34].

The aim of the study is to compare the competence and efficiency of multiple machine learning algorithms and optimization techniques at the supervised learning stage of Deep Learning using ConvNet features for the prediction of motor-imagery tasks through multi-channel EEG. The paper addresses two classification problems using high generalization capacity and fast classification kernels in addition to conventional machine learning algorithms on a ConvNet structure with 3 convolutional layers, batch normalization, and max-pooling layers. The main contributions are highlighted as follows:

1. The proposal and analysis of ConvNet for extracting low- and high-level features from EEG signals and transferring them into the next layers for imagery task classification
2. ConvNet on EEG signals was evaluated in multiple machine learning algorithms with optimization. We achieved significant improvements in classification performances for predicting imagery tasks
3. Novel lower-upper triangularization based Extreme Learning machines (ELM) kernel, LuELM, which had high generalization capability and accelerated learning speed by the advantage of using no iterations, was adapted to the supervised learning stage of ConvNet.
4. The prediction score of motor-imagery tasks through EEG was improved by 88.90% to 90.33%.

In our study, a design-choice that preserves spatial information of multi-channel EEG data includes dropout layer [42] and batch-normalization [19,27] and with different back-propagation methods i.e. RmsPROP [17], Adam [22] and stochastic gradient descent with momentum were evaluated with the same hyper-parameters values i.e. learning rates, regularization constants. To see the impacts of the ConvNet model on EEG data results of classical spectral features together with traditional fully-connected multilayer perceptron were compared. Results showed that recently advanced methods in the machine learning field, i.e. Adam, batch normalization together with dropout strategy, improved predicting ability, outperformed that of conventional fully-connected neural networks with spectral features estimated with Welch periodogram.

2 Materials and Methods

First, information about EEG recordings and preprocessing were provided. This is followed by describing Welch and Morlet wavelet methods of spectral analysis. Next, we explained ConvNet constructed for this study in detail, particularly the design-choice for EEG data. Afterward, six training strategies were described.

2.1 Database

We evaluated predicting imagery left and right movements on publicly available EEGMMI dataset [37] in Physionet [11]. Dataset consists of 160 Hz sampled EEG recordings through 64 electrodes from 109 subjects in the course of 4 motor/imaginary tasks. Each subject performed 14 experimental runs: two one-minute baseline runs (one with eyes open, one with eyes closed), and three two-minute runs of each of the 4 following tasks.

In this study EEG recordings in the course of one of the tasks were considered. The procedure in the selected task is as follows: A target appears on either the left or right side of the screen, the subject imagines opening and closing the corresponding fist until the target disappears. Then the subject relaxes. This trial is repeated 3 times, each repetition has 15 number of right and left labeled segments. Therefore, for each subject there exist 45 number of labeled segments.

2.2 Preprocessing

Preprocessing was performed at a minimum level to enable ConvNet to capture the dynamics and characteristics of EEG recordings itself without bias. EEG recordings were filtered above 30 Hz using a designed high-pass filter with an ordinary 3^{rd} order Butterworth filter.

2.3 Multi-Layer Perceptron

In this study, the network contained two fully-connected hidden layers comprising 100 and 75 nodes, respectively. The training set was segmented in estimation and validation subsets (85 and 15% of the training set respectively). The tangent hyperbolic activation function was used for the hidden layers and the output layer. The sequential (in other words, batch size is one) learning strategy was performed for computing gradients. Gradients were computed with the steepest descent algorithm and a learning rate of 0.01 was set and kept constant throughout the training process. The training of the network was stopped either at the 100^{th} epoch or whenever the updates of the weights failed to reduce the loss (mean sum squared error) of the validation subset for 15 consecutive times. The status of the neural network was then reverted to the last most successful epoch.

2.4 Welch Method

Welch method includes dividing time series data into overlapped segments, estimating periodograms of windowed each segment using fast Fourier transform and averaging [45]. Dividing trials into overlapped segments provides a more accurate estimation from non-stationary time series. However, using the same repetitive information cause problems in spectral analysis. To eliminate such repetitive information due to overlapping segments, non-rectangular windowing methods are used. In this way, the amplitude of the data is attenuated at the initial and last parts of segments therefore their unnecessary (repetitive) information is decreased. Of several windowing methods, Hann tapering is mostly preferred because it makes the initial and last parts of segments fully equal to zero [7]. Also, averaging enables estimating periodograms that have relatively lower variance than the entire time series.

Each trial, which had a duration of 0.4 seconds (656 number of data) was split into Hann windowed segments of 0.15 ms length that overlaps 50% with the previous segment −except for the first one and periodograms were estimated with a resolution of 1.67. The estimated periodograms of alpha bands (8−12 Hz) with the Welch method were used as features to train a multi-layer perceptron.

2.5 Deep Convolutional Neural Network

Deep learning with ConvNets [10,25] is of a specialized type of neural networks that particularly processes grid-like shaped data. They have a strong ability to learn non-linearly separated features by means of discrete convolutions and non-linear activation. In addition, employing deep (multiple) layers allows them to represent high-level features as a combination of low-level features. For affine transformation, they simply use widely-known discrete convolution operation in at least one of their layer rather than general matrix multiplication. Discrete convolution with weight-sharing enables convolutional layers to be efficient in the representation of scale large scale of data (images, audio, etc) and equivariance to translation (that means shifting of input can easily be captured by naturally shifting discrete convolution). Following, element-wise non-linear activation functions i.e. ReLU, LeakyReLU are applied to improve the separability of data. The pooling layer is typically applied following the convolution layer that compresses (in a way of down-sampling) output groups of discrete convolutions in-line. Changing the level of striding in the convolutional layer also provides such compression. Pooling operations are generally performed with a function of L2 norm, maximum, mean or weighted mean. Such pooling operations make outputs gain almost invariant to tiny translations of the network input.

In order to predict imagery tasks through EEG signals, we designed a deep ConvNet architecture in Fig. 1 inspired by the successful study in [38]. It consists of three convolution max-pooling layers, with the first layer was dedicated to preserving spatial characteristics of EEG, followed by two traditional convolution layers, two fully-connected layers and a dropout layer (probability was set to 0.5). Batch normalization [19,27] (1) and rectified linear unit (ReLU) (2) activation were applied following each discrete convolution operation at convolution layers.

$$ReLU(x) = max(0, x) \tag{1}$$

$$H^{'} = \frac{H - \mu}{\sigma} \tag{2}$$

where H is the activation output of any layer to normalize, is a vector including the means of each neuron and is a vector including the standard deviation of each neuron.

Gradients were computed at every 100 batches, and weights were updated according to them with a learning rate of 0.001 that decreases at a level of 0.1 in every 10 epochs. Updating weights were separately realized with using three different approaches; stochastic gradient descent with momentum (SGDM) (momentum value was 0.9) optimization and adaptive moments (Adam) (gradient decay factor, squared gradient decay factor, and epsilon constant were 0.9, 0.99 and 10^{-8}, respectively) and RmsPROP (squared gradient decay factor and epsilon constant were 0.99 and 10^{-8}, respectively) adaptive learning optimization. The training of the network was stopped either at the 100^{th} epoch or whenever the updates of the weights failed to reduce the loss (cross−entropy) of the validation subset for 15 consecutive times. The status of the ConvNet was then reverted to the last most successful epoch. (Codes for downloading data form remote servers and guides for implementing this study in detail are available at https://github.com/apdullahyayik/).

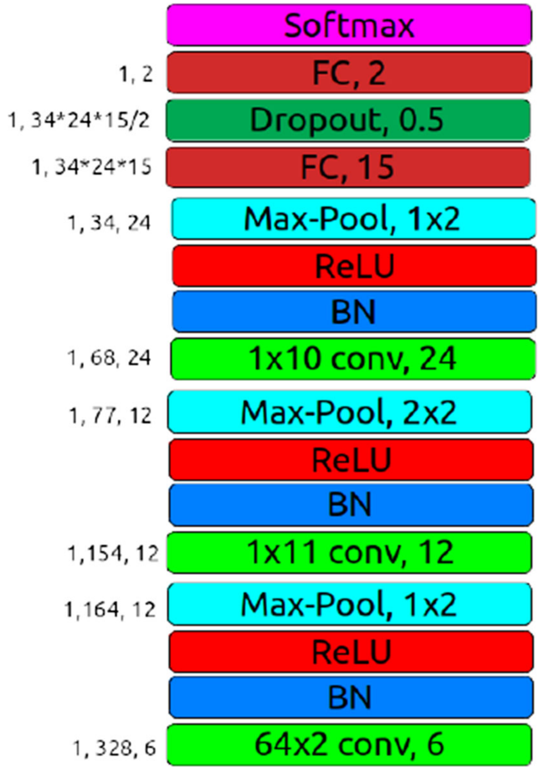

Fig. 1 Deep ConvNet Architecture. Following EEG input, 3 number of convolutional layers one of which is 2-D and allows to preserve spatial relations. Output sizes (width, height and depth) corresponding to layers are given on the left side. Each pooling is performed without overlapping. BN and FC refer to batch normalization and fully-connected, respectively

2.6 Extreme Learning Machines

ELM is a single layer feed-forward network (SLFN) that uses simple matrix inversion solutions to obtain the output weights. It utilizes random assignment between input and hidden layer. The initialized neuron weights on a single hidden layer are used to calculate the optimal output weights between the hidden and output layer by single-step matrix inversions without optimization, learning rate, backpropagation, and iteration [18]. Therefore, the training time of the ELM can be shortened conspicuously. The conventional ELM is based on Moore-Penrose inversion with singular value decomposition.

$$\beta = H^T \left(\frac{1}{\lambda} + HH^T \right)^{-1} T \tag{3}$$

where β, H, and T represent for output weight matrix, randomly assigned hidden layer matrix, and target matrix, respectively. Due to the efficient generalization capability with short training time, ELM is preferred by the researchers with more effective kernels.

2.6.1 Lower Upper Triangularization ELM —LuELM

LuELM kernel is a novel ELM classifier that is based on a lower-upper triangularization matrix inversion solution [24]. It is calculated by simple forward and backward substitutions of $H = LU$.

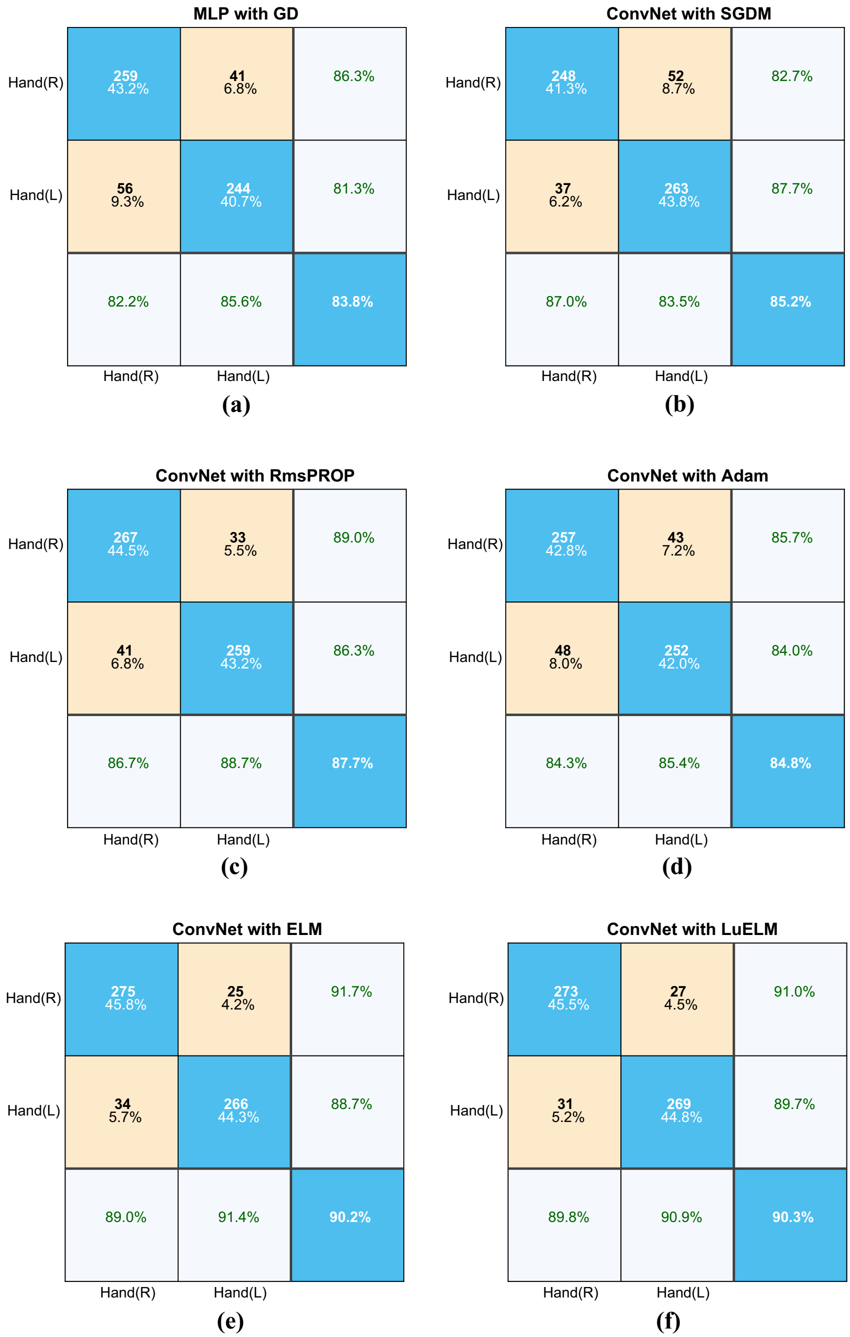

Fig. 2 Confusion matrices for (**a**) Multilayer perceptron (MLP) with gradient descent (GD), and for deep ConvNets with (**b**) stochastic gradient descent with momentum (SGDM), **c** RmsPROP, **d** adaptive momentum (Adam), **e** ELM, and **f** LuELM. Diagonal values correspond to accurately predicted numbers of trial for each class. Bottom rows correspond to sensitivity and right-most columns correspond to precision values. Lower-right values are overall accuracies

Hidden layer output matrix $\mathbf{H}$ can be decomposed as $\mathbf{H} = LU$ where L is a lower triangular matrix and U is an upper triangular matrix using (LU) triangularization.

$\mathbf{H}w = t$ is the base solution for LuELM. The overall steps for this solution presented as follows:

- Decompose $\mathbf{H}$ such that $\mathbf{H} = LU$. Hence $LUw = t$
- Let $Uw = y$, so that $Ly = t$. Solve this system using forward substitution.

$$\begin{aligned} y_1 &= t_1/L_{1,1} \\ y_2 &= (t_2 - (L_{2,1}y_1))/L_{2,2} \\ y_3 &= (t_3 - (L_{3,1}y_1) + (L_{3,2}y_2))/L_{3,3} \\ &\vdots \\ y_i &= (t_i - \sum_{j=1}^{i-1} L_{ij}y_j)/L_{ii} \end{aligned} \tag{4}$$

- Solve the triangular system $Uw = y$ using backward substitution.

$$\begin{aligned} w_e &= y_e/U_{e,e} \\ w_{e-1} &= (y_{e-1} - (U_{e-1,e}w_e))/U_{e-1,e-1} \\ w_{e-2} &= (y_{e-2} - (U_{e-2,e-1}w_{e-1}) + (U_{e-1,e}w_e))/U_{e-2,e-2} \\ &\vdots \\ w_i &= (y_i - \sum_{j=i+1}^{n} U_{ij}w_j)/U_{ii} \end{aligned} \tag{5}$$

Due to training Deep ConvNets needs considerable time and a big dataset, advantages of ELM commonly transferred to the supervised learning stage of the ConvNets and ELM achieved high classification performances [20,23,32,47,49]. Therefore, we suggest integrating novel LuELM kernel learning capabilities for ConvNets on motor imagery task prediction through EEG.

3 Results and Conclusions

The automatic prediction of imagery tasks using EEG comprises the same ConvNet features and applying several iterations on learning procedures for constituting an optimized classifier model. Analyzing multi-channel EEG recordings for each subject enhances the capability of assessing brain activities in detail.

The training set was segmented in estimation and validation subsets for all classifiers (85 and 15% of the training set respectively).

The ConvNet feature vector was tested on multiple machine learning algorithms and optimization techniques including MLP with GD, Deep ConvNet with SGDM, RmsPROP, Adam, ELM, and LuELM. Tested classifiers except ELM and LuELM kernel need iterations and optimization in many parameters. Therefore, the proposed Deep ConvNet and MLP models were tested within a limited variety of layer size and neuron numbers. Furthermore, the classification parameters at iterated variety for the best motor-imagery task prediction rates were reported in the text.

The independent statistical test characteristics enable evaluating system performance for many criteria. Moreover, the analysis of the subject-based population proves the reliability in real life and clinical use. Hereby, we calculated accuracy, precision, specificity, recall, negative predictive value (NPV), and F1 score from the predictions of the proposed classifier models using BDPV package in R.

The MLP classifier was built from two hidden layers for binary classification (Left-Right). The number of neurons for each hidden layer was experimentally iterated at 50$\sim$250 neurons increased by 10. The highest prediction performance for MLP with GD was achieved using 90 at the 1st layer and 230 neurons at the 2nd layer. The ELM and LuELM classifiers were experimentally built at 50$\sim$500 neurons increased by 10. The highest prediction performance models had 410 neurons and 370 neurons for ELM and LuELM, respectively. The highest achievements for the test characteristics depending on the classifiers are presented in Table 1, separately.

Previous works in the literature predicted imagined hand movements on a subject-independent basis with considering only 20 number of subjects [2]. In the proposed model, the design and generalization capacity of the tasks were enhanced. We proposed a deep ConvNet approach for this challenging task through raw EEG data on a subject-independent basis considering 109 number of subjects. Hand-crafted spectral features of Welch method with MLP and ConvNet features with multiple classifiers were compared. Confusion matrices and performance measures are detailed in Fig. 2 and Table 1.

In the case of using the ConvNet features, the proposed models were observed to predict the motor-imagery tasks through EEG with overall accuracy rates of 83.83$\sim$90.33%, 82.67$\sim$91.67%, 81.33$\sim$89.67%, 82.22$\sim$89.80%, 83.49$\sim$91.41%, and 0.8423$\sim$0.9040 for accuracy, recall, specificity, precision, NPV, and F1 score, respectively.

The fact that MLP with spectral features failed to predict motor-imagery tasks at a fixed MLP model. Therefore, we experimented with the ConvNet classifier models to improve the classification performances at a variety of classification parameters. Using a variety in neuron sizes had enabled reaching optimum models for the EEG issue. Nevertheless, the MLP was less successful than deep ConvNet models. The Deep ConvNet models achieved high enough performances that the hierarchical feature representation and training strategies in deep ConvNets are suitable for modeling imagined motor movements on a subject dependent basis. The reason of why their performance varies considerably for MLP and Deep ConvNet is the feature learning stage advantage of ConvNet. ConvNet provides analyzing low-, middle and high-level features from the raw EEG plot. Although spectral feature extraction is a method that proved its efficiency on EEG, various-level features from Deep ConvNet are more responsible for the prediction of motor imagery tasks at an iterated variety of proposed models. In addition RmsPROP optimization technique was provided to reach an accuracy rate of 87.67% that is higher than Adam and SGDM.

4 Discussion

Most of the studies focused on analyzing spectral domain features using conventional machine learning algorithms. However, the achievements are incompetent to be used as a predictor application for motor-imagery tasks through EEG and have no just-noticeable performance for real-time applications with signal processing stages. Alomari et al. proposed an EEG-based mouse controller application. They analyzed EEG recordings from 100 subjects at a range of 0.5-50Hz using Coiflet wavelets of Discrete Wavelet Transform (DWT) features

Table 1 The best three achievements (%) for each machine learning algorithms with ConvNet for imagery task prediction on EEG

Classifier	Model	Accuracy	Recall	Specificity	Precision	NPV	F1 Score
MLP with GD	110–140 neurons	80.17	78.67	81.67	81.10	79.29	0.7986
	60–180 neurons	82.17	84.67	79.67	80.63	83.86	0.8260
	90–230 neurons	83.83	86.33	81.33	82.22	85.61	0.8423
ConvNet with SGDM	140–90 neurons	83.50	82.67	84.33	84.07	82.95	0.8336
	110–210 neurons	85.17	83.33	87.00	86.51	83.92	0.8489
	70–110 neurons	85.17	82.67	87.67	87.02	83.49	0.8479
ConvNet with RmsPROP	110–110 neurons	84.67	83.67	85.67	85.37	83.99	0.8451
	120–130 neurons	86.83	85.67	88.00	87.71	85.99	0.8668
	90–220 neurons	87.67	89.00	86.33	86.69	88.70	0.8783
ConvNET with Adam	210–100 neurons	83.67	83.00	84.33	84.12	83.22	0.8356
	170–120 neurons	83.83	84.67	83.00	83.28	84.41	0.8397
	70–230 neurons	84.83	85.67	84.00	84.26	85.42	0.8496
ConvNET with ELM	330 neurons	88.00	85.67	90.33	89.86	86.31	0.8771
	300 neurons	88.83	87.33	90.33	90.03	87.70	0.8866
	410 neurons	90.17	91.67	88.67	89.00	91.41	0.9031
ConvNET with LuELM	100 neurons	88.50	89.67	87.33	87.62	89.42	0.8863
	410 neurons	89.33	91.33	87.33	87.82	90.97	0.8954
	370 neurons	90.33	91.00	89.67	89.80	90.88	0.9040

on SVM. They predicted the motor-imagery tasks with an accuracy rate of 86.79% [14]. Furthermore, They reached a classification accuracy rate of 88.90% using power, mean, and energy features from independent component analysis (ICA) on MLP [15]. Similarly, Major and Conrad analyzed the EEG-based motor tasks using ICA features on MLP. They applied the 8^{th} order Butterworth filter at 8-30 Hz. They utilized MLP with scaled conjugate gradient backpropagation and reported an accuracy of 72.81% [29]. Sita and Nair applied a band-pass filter at the range of 42-50 Hz and task-based segmentation as preprocessing of their model. They fed ICA features to linear and quadratic discriminant analysis (LDA and QDA) algorithms. They reported a motor-imagery prediction accuracy rate of 75.84% on the QDA classifier [40]. Filho et al used the functional connectivity matrix algorithm and power spectral density (Welch's transform) as the feature extractor and fed the features into the LDA classifier. They achieved a classification performance rate of 87.24% [9]. Kim et al. applied the multivariate empirical mode decomposition and extracted intrinsic mod function modulations. They fed the modulation features into random forests classifier and reached a motor task prediction rate of 81.15% [21]. Deep learning has the advantages of minimizing preprocessing and passing feature extraction approaches on time-series for the classification. The closest paper is organized as encoding spatial and temporal information from EEG using recurrent neural networks algorithm. Ma et al. used a sliding window method to augment data for analysis. They fed the classifier using long short term memory supports. They reported an average accuracy rate of 68.20% for motor-imagery tasks [28]. To the best of our knowledge, the proposed deep ConvNet model has a higher generalization performance than the literature.

As seen in Table 1, especially, ConvNet with LuELM has superiority on the prediction of motor-imagery tasks on EEG against other methods considering classification performance metrics including accuracy, specificity, precision, and F1 score. ConvNet with ELM has higher achievements in recall and NPV. The results show that ConvNet with both ELM and LuELM has the advantages of random feature mapping and least square fitting. The proposed models reached the highest achievements with simple learning procedures, no iterative adaptation, and no backpropagation.

Whereas adapting LuELM into the ConvNet architecture indicates the main significance and novelty of this study, the proposed ConvNet uses transfer learning advancements on EEG without the necessity of feature extraction and signal processing stages. The prediction score of motor-imagery tasks through EEG was improved to 90.33%, 91.00%, 89.67%, 89.80%, 90.88% and 0.9040 for accuracy, recall, specificity, precision, NPV, and F1 score. Although the deep learning algorithms need a big number of data, the proposed ConvNet with LuELM is convenient for small-scale datasets.

The weakest aspect of this study is the variety in ConvNet architectures. It is possible to reach better prediction performances using large number of hidden layers and neurons at each layer. This study shows that ConvNets allow accurate imagery hand movement predicting, that recent techniques; Adam optimization, batch normalization together with dropout strategy boost performance with raw EEG data, outperforming conventional fully-connected MLP with hand-crafted spectral features.

Thus, ConvNets can provide robust learning from EEG data with the only use of minimum preprocessing. This study also shows that ConvNets can offer promising achievements in the neuroscience research field.

Our main finding was that Deep ConvNet with both ELM and LuELM classifiers is a powerful deep architecture for raw EEG, whenever the proposed model only has a minimum preprocessing stage. Despite deep learning algorithms often need large datasets, the ConvNet with ELM kernel has advantages of ELM classifier that has high generalization performance for even 109 subjects. Herein, extracting convolution-based low- and high-level features

Table 2 Related works on motor-imagery task prediction on EEGMMI dataset

Related works	Preprocessing	Methods	Classifier	Accuracy	Precision	Recall	F1 Score
Ma et al. [28]	Sliding window method	–	RNN	68.20	69.71	73.25	0.7144
Alomari et al. [14]	BPF (0.5–50 Hz) AAR	DWT	SVM	86.79	87.88	87.86	0.8787
Sita and Nair [40]	Task-based segmentation BPF (42–50Hz)	ICA	QDA	75.84	76.31	77.02	0.7666
Shenoy et al. [39]	–	FBCSP	SVM	83.08	83.01	84.25	0.8363
Filho et al. [9]	FIR μ(7–13 Hz) β(13–30 Hz)	FCM PSD	LDA	87.24	88.74	88.74	0.8874
Pinheiro et al. [33]	BPF (0.5–42 Hz)	SFFT	SVM	84.88	85.13	85.69	0.8541
Kim et al. [21]	–	MEMD	RF	81.15	81.28	80.87	0.8107
Alomari et al. [15]	BPF (0.5–90Hz) AAR	ICA	MLP	88.90	–	–	–
Major and Conrad [29]	BPF (8–30 Hz)	ICA	MLP	72.81	70.77	62.29	0.6626
This study	HPF (30 Hz)	–	MLP with GD	83.83	82.22	86.33	0.8423
			ConvNET with SGDM	85.17	87.02	82.67	0.8479
			ConvNET with RmsPROP	87.67	86.69	89.00	0.8783
			ConvNET with Adam	84.83	84.26	85.67	0.8496
			ConvNET with ELM	90.17	89.00	91.67	0.9031
			ConvNET with LuELM	90.33	89.80	91.00	0.9040

DWT: Discrete Wavelet Transform, ICA: Independent Component Analysis, PSD: Power Spectral Density (Welch's transform), FCM: Functional connectivity matrix, MEMD: Multivariate empirical mode decomposition, SFFT: Sparse Fast Fourier Transform, QDA: Quadratic discriminant Analysis, RF: Random Forests, AAR: Automatic Artifact Removal, BPS: Band-pass filter, HPF: High-Pass Filter by using ConvNet supported obtaining characteristic information for motor-imagery task prediction. Feeding the Deep ConvNet models through EEG as an input extends the training capability for even complex tasks on ELM kernels. Table 2

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H (2018) Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comp Biol Med. https://doi.org/10.1016/j.compbiomed.2017.09.017
2. Alomari MH, Baniyounes AM, Awada EA (2014) EEG-based classification of imagined fists movements using machine learning and wavelet transform analysis. Int J Adv Electron Electr Eng 23:83–87
3. Altan G, Kutlu Y, Allahverdi N (2016) Deep Belief Networks Based Brain Activity Classification Using EEG from Slow Cortical Potentials in Stroke. Int J Appl Mathemat Electron Comput 4:205–205
4. Bai L, Yu T, Li Y (2015) A brain computer interface-based explorer. J Neurosci Methods 244:2–7
5. Carlson T, Millan JDR (2013) Brain-controlled wheelchairs: a robotic architecture. IEEE Robotics Automat Magazine 20(1):65–73
6. Cecotti H, Gräser A (2011) Convolutional neural networks for P300 detection with application to brain-computer interfaces. IEEE Transact Pattern Anal Machine Intell. https://doi.org/10.1109/TPAMI.2010.125
7. Cohen MX (2014) Analyzing neural time series data: theory and practice. MIT press, Cambridge
8. Dos Santos, C., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 69–78 (2014)
9. Filho CA, Attux R, Castellano G (2017) EEG sensorimotor rhythms' variation and functional connectivity measures during motor imagery: Linear relations and classification approaches. PeerJ. https://doi.org/10.7717/peerj.3983
10. Fukushima K, Miyake S. (1982) Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets, pp. 267–285. Springer
11. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220
12. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press, Newyork
13. Gregory, K., George, K., George, H.: An EEG pre-processing technique for the fast recognition of motor imagery movements. In: 2016 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 90–94. IEEE (2016)
14. Alamori H, M., AbuBaker, A., Turani, A., M., A. (2014) EEG Mouse: A Machine Learning-Based Brain Computer Interface. Int J Adv Comput Sci Appl. doi.org/10.14569/ijacsa.2014.050428
15. H., M., Samaha, A., AlKamha, K. (2013) Automated Classification of L/R Hand Movement EEG Signals using Advanced Feature Extraction and Machine Learning. International Journal of Advanced Computer Science and Applications
16. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition. Ieee Sig Process Magazine 29(6):82–97. https://doi.org/10.1109/MSP.2012.2205597
17. Hinton G, Srivastava N, Swersky K (2012) Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. Cited on 14
18. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: Theory and applications. Neurocomputing. https://doi.org/10.1016/j.neucom.2005.12.126
19. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
20. Kim J, Kim J, Jang GJ, Lee M (2017) Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection. Neural Netw. https://doi.org/10.1016/j.neunet.2016.12.002

21. Kim Y, Ryu J, Kim KK, Took CC, Mandic DP, Park C (2016) Motor Imagery Classification Using Mu and Beta Rhythms of EEG with Strong Uncorrelating Transform Based Complex Common Spatial Patterns. Computat Intell Neurosci. https://doi.org/10.1155/2016/1489692
22. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
23. Kolsch, A., Afzal, M.Z., Ebbecke, M., Liwicki, M.: Real-Time Document Image Classification Using Deep CNN and Extreme Learning Machines. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR (2018). https://doi.org/10.1109/ICDAR.2017.217
24. Kutlu Y, Yayık A, Yildirim E, Yildirim S (2019) LU triangularization extreme learning machine in EEG cognitive task classification. Neural Comput Appl. https://doi.org/10.1007/s00521-017-3142-1
25. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Others: Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
26. Li J, Yu ZL, Gu Z, Wu W, Li Y, Jin L (2018) A hybrid network for ERP detection and analysis based on restricted boltzmann machine. IEEE Transact Neural Syst Rehabilit Eng. https://doi.org/10.1109/TNSRE.2018.2803066
27. Liu M, Wu W, Gu Z, Yu Z, Qi FF, Li Y (2018) Deep learning based on Batch Normalization for P300 signal detection. Neurocomputing. https://doi.org/10.1016/j.neucom.2017.08.039
28. Ma, X., Qiu, S., Du, C., Xing, J., He, H.: Improving EEG-Based Motor Imagery Classification via Spatial and Temporal Recurrent Neural Networks. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS (2018). https://doi.org/10.1109/EMBC.2018.8512590
29. Major TC, Conrad JM (2017) The effects of pre-filtering and individualizing components for electroencephalography neural network classification. Conf Proc IEEE SOUTHEASTCON. https://doi.org/10.1109/SECON.2017.7925289
30. Meisheri, H., Ramrao, N., Mitra, S.: Multiclass common spatial pattern for EEG based brain computer interface with adaptive learning classifier. arXiv preprint arXiv:1802.09046 (2018)
31. Mousavi Z, Yousefi Rezaii T, Sheykhivand S, Farzamnia A, Razavi SN (2019) Deep convolutional neural network for classification of sleep stages from single-channel EEG signals. J Neurosci Methods. https://doi.org/10.1016/j.jneumeth.2019.108312
32. Pang S, Yang X (2016) Deep Convolutional Extreme Learning Machine and Its Application in Handwritten Digit Classification. Comput Intell Neurosci. https://doi.org/10.1155/2016/3049632
33. Pinheiro, O.R., Alves, L.R., Romero, M.F., De Souza, J.R.: Wheelchair simulator game for training people with severe disabilities. In: TISHW 2016 - 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing, Proceedings (2016). https://doi.org/10.1109/TISHW.2016.7847792
34. Qi F, Li Y, Wu W (2015) RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG. IEEE Transact Neural Netw Learning Syst. https://doi.org/10.1109/TNNLS.2015.2402694
35. Raghu S, Sriraam N, Temel Y, Rao SV, Kubben PL (2020) EEG based multi-class seizure type classification using convolutional neural network and transfer learning. Neural Netw. https://doi.org/10.1016/j.neunet.2020.01.017
36. San-Segundo R, Gil-Martín M, D'Haro-Enríquez LF, Pardo JM (2019) Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. Comput Biol Med. https://doi.org/10.1016/j.compbiomed.2019.04.031
37. Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR (2004) BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Transact Biomed Eng 51(6):1034–1043
38. Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W, Ball T (2017) Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapp. https://doi.org/10.1002/hbm.23730
39. Shenoy, H.V., Vinod, A.P., Guan, C.: Shrinkage estimator based regularization for EEG motor imagery classification. In: 2015 10th International Conference on Information, Communications and Signal Processing, ICICS 2015 (2016). https://doi.org/10.1109/ICICS.2015.7459836
40. Sita, J., Nair, G.J.: Feature extraction and classification of EEG signals for mapping motor area of the brain. In: 2013 International Conference on Control Communication and Computing, ICCC 2013 (2013). https://doi.org/10.1109/ICCC.2013.6731699
41. Sleight J, Pillai P, Mohan S (2009) Classification of executed and imagined motor movement EEG signals. University of Michigan, Ann Arbor Michigan, pp 1–10
42. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Machine Learn Res 15(1):1929–1958
43. Sun Y, Lo FP, Lo B (2019) EEG-based user identification system using 1D-convolutional long short-term memory neural networks. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2019.01.080

44. Wang YT, Wang Y, Jung TP (2011) A cell-phone-based brain-computer interface for communication in daily life. Journal of neural engineering 8(2):25,018
45. Welch P (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transact Audio Electroacoustics 15(2):70–73
46. Wu W, Chen Z, Gao X, Li Y, Brown EN, Gao S (2015) Probabilistic common spatial patterns for multichannel EEG analysis. IEEE Transact Pattern Anal Machine Intell. https://doi.org/10.1109/TPAMI.2014.2330598
47. Yu W, Zhuang F, He Q, Shi Z (2015) Learning deep representations via extreme learning machines. Neurocomputing. https://doi.org/10.1016/j.neucom.2014.03.077
48. Zakaria, M.H.F., Mansor, W., Lee, K.Y.: Time-frequency analysis of executed and imagined motor movement EEG signals for neuro-based home appliance system. In: TENCON 2017-2017 IEEE Region 10 Conference, pp. 1657–1660. IEEE (2017)
49. Zeng, Y., Xu, X., Fang, Y., Zhao, K.: Traffic sign recognition using extreme learning classifier with deep convolutional features. The 2015 international conference on intelligence science and big data engineering (IScIDE 2015), Suzhou, China (2015). https://doi.org/10.1007/978-3-319-23989-7_28
50. Zhang, D., Yao, L., Zhang, X., Wang, S., Chen, W., Boots, R.: EEG-based intention recognition from spatio-temporal representations via cascade and parallel convolutional recurrent neural networks. arXiv preprint arXiv:1708.06578 (2017)
51. Zhang J, Yao R, Ge W, Gao J (2020) Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG. Comp Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2019.105089