Brain Computer Interface for Neuro-rehabilitation With Deep Learning Classification and Virtual Reality Feedback

Karácsony, Tamás; Hansen, John Paulin; Iversen, Helle Klingenberg; Puthusserypady, Sadasivan

Published in:
Proceedings of the 10th Augmented Human International Conference 2019

Link to article, DOI:
10.1145/3311823.3311864

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Brain Computer Interface for Neuro-rehabilitation With Deep Learning Classification and Virtual Reality Feedback

Tamás Karácsony
Technical University of Denmark
Lyngby, Denmark
karacsonytamas.mail@gmail.com

John Paulin Hansen
Technical University of Denmark
Lyngby, Denmark
jpha@dtu.dk

Helle Klingenberg Iversen
University of Copenhagen
Stroke Centre Rigshospitalet, Department of Neurology
Copenhagen, Denmark
helle.klingenberg.iversen@regionh.dk

Sadasivan Puthusserypady
Technical University of Denmark
Lyngby, Denmark
sapu@dtu.dk

ABSTRACT

Though Motor Imagery (MI) stroke rehabilitation effectively promotes neural reorganization, current therapeutic methods are immeasurable and their repetitiveness can be demotivating. In this work, a real-time electroencephalogram (EEG) based MI-BCI (Brain Computer Interface) system with a virtual reality (VR) game as a motivational feedback has been developed for stroke rehabilitation. If the subject successfully hits one of the targets, it explodes and thus providing feedback on a successfully imagined and virtually executed movement of hands or feet. Novel classification algorithms with deep learning (DL) and convolutional neural network (CNN) architecture with a unique trial onset detection technique was used. Our classifiers performed better than the previous architectures on datasets from PhysioNet offline database. It provided fine classification in the real-time game setting using a 0.5 second 16 channel input for the CNN architectures. Ten participants reported the training to be interesting, fun and immersive. "It is a bit weird, because it feels like it would be my hands", was one of the comments from a test person. The VR system induced a slight discomfort and a moderate effort for MI activations was reported. We conclude that MI-BCI-VR systems with classifiers based on DL for real-time game applications should be considered for motivating MI stroke rehabilitation.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *Virtual reality*; Usability testing.

KEYWORDS

Motor Imagery, Brain Computer Interface, Deep learning, CNN, Virtual Reality, Online EEG classification

AH2019, March 11–12, 2019, Reims, France

ACM ISBN 978-1-4503-6547-5/19/03…$15.00
https://doi.org/10.1145/3311823.3311864

ACM Reference Format:
Tamás Karácsony, John Paulin Hansen, Helle Klingenberg Iversen, and Sadasivan Puthusserypady. 2019. Brain Computer Interface for Neuro-rehabilitation With Deep Learning Classification and Virtual Reality Feedback. In *Augmented Human International Conference 2019 (AH2019), March 11–12, 2019, Reims, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3311823.3311864

1 INTRODUCTION

Motor Imagery (MI) Brain Computer Interfaces (BCI) have been explored in several studies for neuro-rehabilitation of stroke patients. The subject imagines to move a body part and the neural activity generated by this imagination is captured by electroencephalogram (EEG), while an appropriate feedback is provided to the subject. Neuromechanisms underlying the MI practice activates and improves the motor pathways in both healthy and in post-stroke subjects, especially if it is combined with conventional therapies [30, 35]. The fundamental phenomena is the lifelong neuroplasticity, which allows the brain to adapt to various circumstances [8, 12, 15]. Moreover, sensory feedback has been suggested to improve the induced neuroplasticity, by means of involving a greater part of the sensorimotor system. Feedback may involve a variety of sensory systems, for instance, visual, tactile, auditory and haptic [4, 9, 20].

The basic principle of motor rehabilitation therapy is progressive and skilled motor practice. It can become monotonic, which may influence the motivation and engagement of patients and thereby declining the effectiveness of the therapy [10]. To overcome these limitations, there have been approaches in combining BCI and virtual reality (VR) systems to create a more immersive and motivating environments.

Besides the benefit of MI practice, MI-BCI systems offers a quantitative, measurable treatment. Conventional MI practice cannot be verified by a physician, but with MI-BCI, brain activity can be monitored [31]. The feedback is beneficial because it induces higher activation of the cortex, which has clinically significant effect on neuro-rehabilitation, measured on the Fugl-Meyer Assessment scale [30]. Besides, it can boost the interest, motivation and engagement of the subjects [3, 31].

Real-time visual feedback in VR provides a more immersive 3D environment than just a 2D animation. A higher task engagement reduce awareness on the therapeutic aspects and makes motor

learning more intrinsic [42]. In VR, the subject is able to perceive the imagined motor action, which activates the mirror neurons that are also employed by mirror therapy [28]. Use of VR in stroke rehabilitation have been investigated in several studies, which concluded the effectiveness of this type of intervention [1, 6, 24, 41].

In this paper, we first explore the convolutional neural network (CNN)-based classifiers and evaluate them on a real-time dataset. Then we present a feasibility study of a real-time application that provides MI control of a computer game with VR feedback on each successfully imagined and virtually executed movement of hands or feet. The main contributions of this paper include: (1) superior classification accuracy (offline tests) compared to previous works, (2) demonstrates the feasibility, and (3) very good user acceptance of the system.

2 RELATED WORKS

2.1 Classical machine learning

Most of the current BCI-VR systems employ classical machine learning (ML) techniques on the collected EEG signals. This approach requires signal pre-processing, feature extraction and classification steps.

Signal pre-processing improves the signal-to-noise-ratio (SNR) of the EEG signals and prepares it for the feature extraction. The most widely used techniques include band-pass (BP) filtering and notch filtering to avoid the power-line interference if the pass-band includes the power line frequency. Other popular noise reduction steps include Common Average Reference (CAR) and Weighted Average Reference (WAR) [1, 26, 38]. For feature extraction, the common spatial pattern (CSP) filter and its variants (e.g. filter bank CSP (FBCSP)) are the widely used approach [1, 44]. Other works employ independent or principal component analysis (ICA,PCA) [16, 43]. In case of binary problems, a band power or power spectrum is also an adequate approach [21, 38]. The most preferred classification is the supervised machine learning such as the support vector machine (SVM) and linear discriminant analysis (LDA) [2, 44]. In case of binary problems, a simple specified threshold could be an efficient choice [21, 38, 43].

2.2 Deep learning

In recent years, the developments in the field of deep learning (DL) has drawn the attention to the biomedical field as well. Bashivan et. al [5] proposed a robust transfer learning approach, with obtaining a sequence of topology-preserving multi-spectral images from EEG signals, and applying deep recurrent CNN (RCNN) architecture inspired by state-of-the-art video classification method. In this research, the spatial and spectral invariant representations were extracted with CNN and temporal patterns with a long short-term memory (LSTM) network combined with 1D convolution. The CNN network was adopted from the VGG network proposed by Simoyan et. al. [37] for Imagenet classification. Furthermore, the performance of the proposed RCNN was found superior to the commonly used classification methods, like the SVM, Random Forest, sparse Logistic Regression, and Deep Belief Networks (DBN).

A CNN and stacked auto-encoder (SAE) based approach proposed by Tabar et. al. [39], used the advantage of 2D generated pictures also. The three channel EEG signal after a short time Fourier transform (STFT) was converted to a 2D image on μ and β frequency bands and classified with CNN, SAE and the combination of the two architectures. The proposed CNN-SAE framework was tested on MI left/right hand tasks on 'BCI Competition IV dataset 2b' and 'BCI Competition II dataset III' and found to perform better and more robust to state-of-the-art methods, like FBCSP or Twin SVM.

Kumar et. al. [19] proposed a computationally efficient DNN classifier on CSP features. This research compares the DNN classifier to the commonly used classifiers, like CSP, common spatio-spectral pattern (CSSP), FBCSP, and discriminative FBCSP (DFBCSP). Even though the DNN architecture used a low number of variables, it outperformed the CSP, CSSP and FBCSP methods in terms of average error, but not the DFBCSP, which performed better at the expense of increased computational load.

Zhang et. al. [45] applied a 7 layer CNN for a 2-class MI task and investigated the influence of the activation functions in the CNN. Three activation functions were tested, the ReLU, exponential linear unit (ELU) and scaled exponential linear unit (SELU). Both in terms of accuracy and speed of convergence, SELU performed far superior to the other two activation functions and CSP+SVM methods.

An almost end-to-end approach designed by Shen et. al. [36] pre-processed the EEG signal only with BP filtering and propose two deep classification structures. RCNN and deep forests were trained and tested on the 3-class BCI Competition III Dataset V and a 2-class dataset acquired from five post-stroke patients. In order to ensure fast interaction with BCI systems, 8 overlapping 1 [s] frames were used as inputs, to get the final output label for each 0.5 [s]. The suggested classification methods both outperformed the SVM, Naive Bayes and MLP methods. On the 3-class dataset, the RCNN performed the best and on the dataset from the post-stroke patients, the deep forest method proved to be the favorable choice, because of the low number of samples.

Using the raw EEG signals, Schirrmeister et. al. [34] explored several end-to-end CNN architectures, their training strategies and optimization methods along with the visualization of the CNN architecture in order to achieve a better understanding of the system. They reported that the deep CNNs performed better than FBCSP and shallow CNNs, especially employing the recent advancements in DL field, as regularization techniques like dropout (DO) and batch normalization.

The multimodal DNN classification of EEG and fNIRS combination was explored by Chiarelli et. al. [7] with a 2-class MI (left and right hand) task. A fully connected (FC) DNN was designed and comparison of the DNN, SVM, and LDA classifiers were made with an input of standalone and multimodal signals. The best classification results were obtained with the EEG-fNIRS input and DNN classifier, even though it was employing FC layers and not CNN or RNN framework, which could improve the accuracy even more.

Further research include the works by Tang et. al. [40], who proposed a 5-layer CNN with more precise classification accuracy, than power+SVM, CSP+SVM, and AR+SVM methods. Similarly the research presented by Lu et. al. [25], a frequential deep belief network (FDBN) composing the stacked restricted Boltzmann machines, was compared to the state-of-the-art solutions and exceeded them in performance. In contrast, the recurrent spatio-temporal neural network (RSTNN) proposed by Ko et. al. [18], have not reached better

performance, than FBCSP with mutual information-based best individual (MIBIF) method, but it was suggested that the performance could be improved with transfer learning or domain adaptation.

In essence, the state-of-the-art MI-BCI systems successfully utilize the recent advancements in the DL field, applying different approaches and surpassing conventional ML and signal processing techniques. Adequate pre-processing of the signal can be as simple as a BP filter, but could be a generation of a 2D image sequence to make use of the developments of DL architectures on different fields also. Recent improvements in the DL field, like different activation functions, regularization techniques, optimization methods and architectures allows faster training of the DNNs, which is able to learn more complex patterns without overfitting the training set. Even though the main drawback of this method is the longer training time, it allows the network for faster classification on online testing and classification with lower computational requirements and better accuracy. Current literature on MI-BCI systems applying DL is very limited, thus further exploration of this particular field is required, like optimal pre-processing techniques and DNN architectures.

3 METHODS

3.1 Datasets

The implemented systems were trained and evaluated on the PhysioNet EEG motor movement/imagery data set, consisting of over 1500 one- and two-minute labeled EEG recordings, obtained from 109 volunteers [13, 33]. EEGs were recorded with 160 [Hz] sampling frequency from 64 electrodes as per the international 10-20 system (excluding electrodes Nz, F9, F10, FT9, FT10, A1, A2, TP9, TP10, P9, and P10) and for the 16 electrode setting, channels FC3, FCz, FC5, C1-C6, Cz, CP3, CPZ, CP4, P3, Pz and P4 were used [36]. These recordings include both motor execution and MI data, with classes of: opening and closing of left fist, right fist, booth feet, both fists and rest stage with open or closed eyes and between trials.

The recording sessions were organized into either MI or execution sessions, within one session either left and right fist classes or both fist and both feet classes were performed. The trials in the experimental paradigm are defined as: the subject sits relaxed in front of a screen, where visual cues are displayed to instruct the subject what to perform. The visual cues are alternating between instruction to take rest and instructing to imagine or executing one of the above described tasks. This visual cue is presented for 4[s] for each task.

3.2 Data subsets

Data subsets were created to investigate the influence of different number of classes, length of input data and position of samples, thus to find the optimal configuration for the online experiment. All configuration of data subsets were evaluated with 64 and 16 channel recordings. The 16 channel recording was necessary because of the limited number of channels of the used signal acquisition equipment (section 3.5.1).

3.2.1 Number of classes. Three data subsets were constructed in regard to number of classes from the available database, to investigate the performance of the classifier for different complexities of tasks. These class definitions correspond to the description by Hauke et. al. [11], in order to have a good base for comparison of the DNNs, as follows:

- **2-class:** This dataset was constructed form MI left and right fist opening and closing trials. Some subjects' recording was performed with different sampling frequency, therefore these have not been used. Furthermore, as the recordings feature some variability in the number of single trials, a subset of 105 subjects and 42 trials/subject were selected (21 for each side; 7 from each three recordings), although most subjects performed more than 42 trials. The discarded subjects performed less than 42 trials [11].
- **3-class:** This subset is an expansion of the 2-class subset, by random sections of trials from the available baseline recordings with open eyes to obtain a total of 63 trials/subject with 21 trials/class. This third class represents the resting state, where the subject is not performing any MI task [11].
- **4-class:** The fourth class corresponds to both feet MI task. Although these tasks were performed in sessions together with booth fist movements, the later ones were not used as they were expected to share several features with the single fist trials. Therefore the 4-class dataset contains 84 trials/subject with 21 trials/class [11].

3.2.2 Length and position of samples. Several starting positions and length of trial classification have been explored, in order to find the optimal solution for the online application of the system. The two most representative are presented in the following. The position of samples is defined by the labeled start of the trial as "trial onset" and labeled end "trial offset".

As initial studies presented [11] with trial onset and offset included the classifier manages to improve performance, a 6[s] (960 sample) trial length was included in the studies, which is padded with 1[s] rest labeled stage before the trial onset and after the offset. Therefore, including the full transition between rest stage and MI task in the beginning and vice-versa at the end. In order to optimize for online experiments, trials with 0.5[s] (80 sample) starting from the trial onset were also investigated.

3.3 Signal pre-processing

In BCI systems, EEG signals are heavily pre-processed, due to the very low SNR due to the presence of several artifacts (section 2.1). These procedures are computationally heavy and accordingly increases the delay of the system. Therefore, the designed BCI system employs minimal explicit pre-processing, with the aim of reducing this delay and computational load. Additional pre-processing, feature extraction and classification are carried out by the DNN architectures as explained in section 3.4. These pre-processing steps are explained in detail in the following sections.

3.3.1 Filtering. The first step was a 6th order Butterworth BP (0.5-75[Hz]) filter. The filter is applied with a forward and a backward pass, therefore having a zero phase distortion. Furthermore, a 50[Hz] notch filter was applied the same way for power-line interference cancellation.

3.3.2 Running standardization. Following the filtering, a channel-wise running standardization was performed on each recording,

with a decay factor of $\frac{639}{640}$, proportional to 4[s], with 160[Hz] sampling frequency, to compute exponential moving means and variances. It was used to standardize the continuous data, as described in the following equations [34]:

$$x'_t = \frac{x_t - \mu_t}{\sigma_t}, \quad (1)$$

$$\mu_t = \frac{1}{640}x_t + \frac{639}{640}\mu_{t-1}, \quad (2)$$

$$\sigma_t^2 = \frac{1}{640}(x_t - \mu_t)^2 + \frac{639}{640}\sigma_{t-1}^2, \quad (3)$$

where x'_t and x_t were the standardized and the original signal for one electrode at time t. For initialization of the recursive form, the first 640 mean (μ_t) and variance (σ_t^2) values were set to the mean and variance of the first 640 samples. This strategy for the normalization allows the online scheme as it uses only the previous values[34].

3.4 Implemented DNN architectures

In current EEG classification, there is no well established approach and network with hyperparameters to apply. Therefore the DNN architectures were developed from scratch inspired by works on the field, like [11, 34]. The NN architecture and hyperparameter design utilized the most recent advancements on the DL field. In the following, the base architecture is presented, which have been slightly modified in some aspects for the different data subsets. Hyperparameters for training included, but not limited to DO rate, learning rate, batch size. Some design specifications of the architectures were determined by random search and the best performing ones are reported. Architectural random search was performed in many dimensions, some of which were the number and dimensions of convolutional, recurrent and dense layers, application of different regularization layers, kernel initializer, number, size and stride of layers, where applicable.

3.4.1 CNN architectures. The blueprint of the implemented CNN architectures with the scale of the smallest 0.5[s] 16 channel input with 2-class classification is provided in Fig. 1.

The first convolutional layer was responsible for temporal filtering with 100 filters along the time axis with a kernel size of 25 samples. This layer has same padding in order to preserve the dimension of the architecture, therefore creating a more flexible network. These temporally filtered feature maps were fed into a second convolutional layer for spatial filtering across the EEG channels for every time-step, also providing 100 feature maps. The spatial filter also designed to reduce dimensions across channels, producing 1D feature maps. The spatially and temporally processed 1D feature maps were further propagated through two convolutional layers, for further filtering, feature abstraction and extraction, both of them delivering 50 feature maps. After each of the last two convolutional layers, dimension reduction was performed with max pooling layers, because it was important if some features were present in the feature maps to generalize well, and provide a more robust architecture. Furthermore to ensure generalization and resistance for variations in the signal, batch normalization (BN) was applied before every convolutional layer.

After the convolutional layers, a flattening layer was applied, which reduced the dimension to one, preparing for computations with FC dense layers. Advancing forward on the architecture, FC layers were organized in blocks of three layers. Such blocks consisted of a BN layer, followed by a DO fed into a FC layer. There were 6 blocks arranged in a shape of an "upside-down pyramid", starting from 1024 neurons in the first FC layer to 32 in the sixth block, with halving down the number of neurons every consequent FC layer. This part of the NN can be implied as the feature ranking and classification of the high level features extracted by the previous layers. The final one also a BN-DO-FC block, but the number of neurons were representing the classes with a softmax activation to provide an output of probability distribution as classification output of the required categories.

3.4.2 Training strategy, hyperparameters and further considerations. The aforementioned architecture was inspired by other works, but the final layout of the NN was determined by empirical results. Random search was performed with the objective of maximizing the validation accuracy on the PhysioNet dataset. The results of this search and further considerations and architectural choices are described below:

- **5-fold-cross validation** was performed across subjects, thus training a **global classifier**.
- Increasing further, the **number of convolutional layers with combination of max pooling layers**, than described before, was not found to be more effective, but decreasing them would result in loss of effectiveness. The following parameters for these layers were included in random search:
  - **convolutional kernel number** for the first two convolutional layers in the range of 20-100 and for the third and fourth between 5 and 80,
  - **convolutional kernel size** for all convolutional layers in the range of 1-40, and **stride** was set between 1 and the kernel size, meaning it was able to not overlap,
  - **max pooling kernel size** was explored in the range from 1 to 10, with **stride** set from 1 to kernel size, with the same strategy as in the convolutional kernel exploration, in consequence the max pooling layers were eliminated after the first and second convolutional layers,
  - even though **dropout** was tested before the convolutional layers also, it proved to be a too drastic regularization technique in this case and prevented learning, therefore it has been eliminated.
- The **BN-DO-FC blocks** are designed with considerations respect to Li et. al. [22]. The number and dimensions of them were designated for establishing necessary learning abilities, while restrain the chance of overfitting. In random search, the following hyperparameters were explored:
  - the **number of neurons** in the first FC layer was between 128 and 2048, consequently the number of neurons in the following layers were also modified accordingly,
  - **dropout rate** in range of 0.0 and 0.5. In the final architecture, 0.15 was applied, which was lower than 0.2 as Li et. al. [22] suggested. This architecture however applied L2 regularization also.
- **L2 regularization** was applied with $\lambda = 0.01$ on every FC and convolutional layer additionally to refrain overfitting

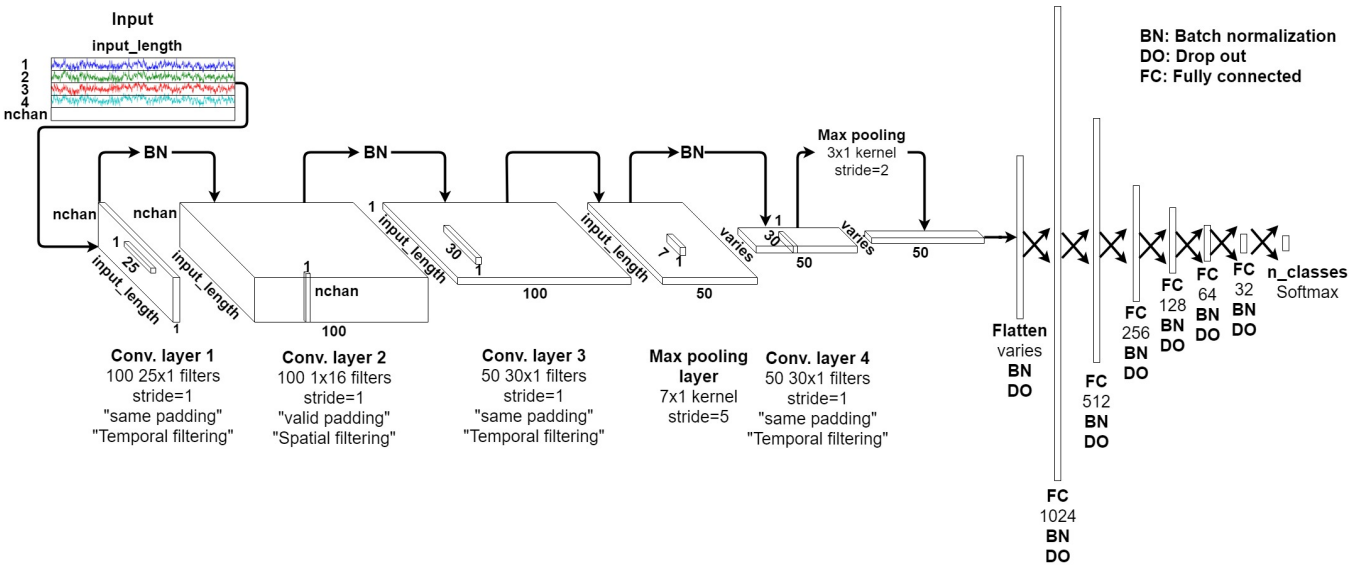

Figure 1: Overview of the designed CNN architecture, dimensions approximately scaled for the architecture with 0.5 sec 16 channel input and 2-class output

and exploding gradients, as the architecture is considerably deep.

- Every FC and convolutional layer, except the last softmax layer applies:
  - **ReLU activation**,
  - **He uniform variance scaling initializer** to speed up learning process and also to prevent vanishing and exploding gradients.
- **Categorical cross-entropy** was used as cost function.
- **AMSGrad** variant of Adam optimizer was used with the suggested default variables as described in Reddi et. al. [32].
- One training set was performed with **learning rate** of $1e-3$ and one applying **learning rate decay** on validation loss plateau with a factor of 0.5, after every 20 epoch without improvement. The minimum value of learning rate for this setting was set to $1e-6$.
- **Batch size** of 16 samples proved to be the most efficient choice, where the architecture was able to learn and generalize well at the same time.
- The architectures were trained for 500 **epochs**, although it has to be noted in most of the cases a lower number of epoch could be sufficient, but there were slight improvements in the later steps also.
- **Early stopping** was used as final regularization, therefore architectures with the best validation accuracies were reported.

3.5 Experimental setup

The online system was evaluated by 10 subjects (8 males and 2 females, average age 25.3±3.4 years) with the 0.5[s] length 16 channel setup for 2,3 and 4 classes.

3.5.1 Equipment. EEG data collection was performed with a 16 active channel g.tec system. g.LADYbird active-wet sintered Ag/AgCl electrodes were used, where a g.GAMMAcap provided the layout for the electrodes (section 3.1). Furthermore, g.GAMMAbox supplied the power and driver/interface for the electrodes. g.USBamp bio-signal amplifier was used with a sampling frequency of 256[Hz].

3.5.2 Online classification. Implemented global DNN architectures were designed to be robust to variations and generalize well, especially between subjects as described in section 3.4. It should be noted that the base of the training, the PhysioNet dataset, was recorded with one experimental setup, therefore the classifier were not necessarily prepared for artifacts specific to other experimental setups. Global classifiers were chosen with 0.5[s] input to optimize online classification response time, thus promoting subject engagement. Moreover, classifiers with the highest validation accuracy on the 16 channel layout were utilized for the 2,3 and 4-class trials. In order to enhance user experience, three different classifier decision boundaries were tested. It was essential to provide ideal feedback for the users, which relies on the proportion of the true and false positives and negatives. These were empirically set up with the selected architectures as follows:

- the **2-class** trials decision boundary was set 0.6, where 0.5 was random chance to produce a **very easy** setting,
- **3-class** trials were **medium** level, with decision limit of 0.4, where 0.33 was random chance,
- and finally **4-class** trials used **very hard** setting with decision limit of 0.3, where 0.25 was random chance.

3.5.3 VR environment. Visual feedback was provided for subjects by FOVE™ VR headset. The virtual environment developed in Unity 2018.1 was a game, where participants had to catch falling fruits and kick footballs, called "Bichael May: Fruits and Footballs" (Fig. 2). The

Figure 2: VR game: hands are floating in a natural position allowing the brain fill in the gap between the torso and hands. The goal is to catch the falling fruits or kick the appearing balls. Continuous feedback is provided on activation levels, by bars filling up accordingly. Successful hits (i.e. grasping or foot kicking) trigger explosion of the object and are counted on the scoreboard.

surroundings were designed to provide a comfortable first person setting. The controlled character was combined of a full humanoid model and detailed models of left and right hands, floating in a natural position. As the position of arms were not controlled or measured in any way the best strategy was to let the brain fill in the gap.

Instructions, like specified trial start or end, break time, and session end were showed to the subject in the middle of their view for a few seconds. After the trial have started, fruits and footballs started to randomly spawn every 4 seconds on one of the three locations. Two of them were located above the hands (dropping fruit) and one was located between the player and the scoreboard (incoming footballs). After the objects were spawned, the subject was not able to grasp or hit them for the first 2 seconds, offering them time to react. Subsequently a 1.5 [s] window of opportunity was provided to interact with the targets. Then a 0.5[s] brake was left until the next fruit or football appeared.

Participants received several feedback from the game on their performance. Three progress bars were filling up proportional to the activation levels/class; one for each hand and one for the feet. When the activation threshold was reached for a class, the corresponding animation was played. Hands were opening and closing and the feet was kicking. If the player successfully hits one of the targets, this object exploded, providing feedback on a successfully imagined and virtually executed movement. Score and accuracy percentages were continuously updated and presented on a board in front of the player. These real-time feedback systems were essential for subject engagement and motivation and may introduce a competitive setting, where the participants strive to brake records.

3.5.4 Experimental protocol. Participants played the VR game for 5 minutes to familiarize themselves with the environment and the goals of the system. Then subjects were seated comfortably, with their hands resting on the table and equipped with the BCI-VR system. They were asked to limit their head and eye movements to the minimal, in order to keep the contact of the electrodes as good as possible. After the preparations, 3 trials were performed, namely 2, 3 and 4-classes, each for 5 minutes, with 30 seconds break between them. After the recording session was finished, the BCI-VR system was removed and subjects were asked to fill out a post experimental questionnaire to evaluate the system.

4 RESULTS

4.1 Offline classification on PhisioNet dataset

Architectures were validated with 5-fold cross-validation (CV) accuracies and are reported in form of mean±std% (max%) and referred as accuracies.

Best performing 64 channel 6[s] architectures achieved 5-fold CV accuracies of 85.94±2.71% (90.14%) for 2-classes, 88.50±1.27% (89.86%) for 3-classes and 76.37±2.15% (77.71%) for 4-classes. The 16 channel 0.5[s] variants performed on the offline tests with 72.81±1.81% (75.40%), 78.62±1.46% (80.81%) and 60.37±0.97% (61.63%) respectively. These results are presented in Table 1 along with previous works on the dataset and the counterparts of the architectures. Our architectures performed better as it can be seen by the following improvements on 5-fold CV classification accuracies. The global whole trial classifications (6[s] input) with 2-classes both improved compared to the best previous work on the same offline dataset, in case of 64 channel setup with 5.84 % and with the reduced 16 channel configuration compared to the 14 configuration with 3 %. Even with the trial onset detection (0.5[s] input) the 64 channel setup managed to improve classification results with 0.35% compared to the best 6[s] classifier in prior works.

4.2 Real-time system

4.2.1 Performance. Performance of the real-time system was measured with goal achievement accuracy (goal accuracy), defined by the achieved score in the game. This accuracy was not the same as the classification accuracy, as measuring the proportions of false positives is problematic in real-time systems.

4.2.2 2-class real-time performance. In 2-class trials, all subjects achieved 100% goal accuracy, due to the low activation threshold. The 0.6 decision boundary caused a high amount of false positives. This was perceived by subjects as slightly random. In spite of high false positive rates, the classifier managed to categorize distinct classes. It was identified by continuous grasp for the corresponding MI activity, as the activation levels were sustained correctly. The high number of false positives could be decreased, by increasing the decision boundary to approximately 0.8.

4.2.3 3-class real-time performance. The online mean goal accuracies are reported in Table 2. The highest achieved goal accuracy was an impressive 87%. Subjects enjoyed this protocol most and reported experiencing a sense of high control over the system. The reported goal accuracies also included the difficulties in the game, like timing the movement.

Table 1: Overview of works performing L/R classification tasks on the Physionet [13] EEG dataset with the number of EEG channels used, the training mode (one global classifier or one classifier per subject), the reached mean accuracy, and the methods applied for feature extraction and classification, (Extended from Dose et. al. [11])

Author	#ch	training	max. acc.	Methods	
Handiru & Prasad [14]	16	global	63.62%	FB-CSP SVM	
Tolic & Jovic [27]	3	subject	68.21%	Wavelet transform Feed-forward DNN	
Loboda et al.[23]	9	global	71.55%	Phase information	
Park et al. [29]	58	global	72.37%	SUT-CCSP SVM	
Kim et al. [17]	14	subject	80.05%	SUT-CCSP Random forest	
Dose et.al. [11]	64	global	80.10%	CNN	
		subject	86.13%		
	14	global	76.66%	CNN	
		subject	82.66%		
This work			6 [sec]	0.5 [sec]	
	64	global 2cl	85.94%	80.45%	CNN
		global 3cl	88.50%	84.08%	
		global 4cl	76.37%	72.28%	
	16	global 2cl	79.66%	72.81%	CNN
		global 3cl	84.13%	78.62%	
		global 4cl	65.96%	60.37%	

Table 2: 3-class online experiments, mean goal accuracies

Left hand	Right hand	Overall
44.9±17.8% (77%)	75.6±11.4% (95.0%)	60.2±14.6% (87.0%)

Table 3: 4-class online experiments, mean goal accuracies

Left hand	Right hand	Feet	Overall
38.6±22.0% (77.0%)	19.3±9.5% (30.0%)	100.0±0.0% (100.0%)	53.6±9.6% (70.0%)

4.2.4 4-class real-time performance. This online experiment had consistent high feet activation levels, accordingly this class achieved a 100% accuracy (Table 3). It can be led back to the combination of serious artifacts on the signal, by the physical contact of the VR system with the electrodes. Regardless of the above described issue, the classifier still managed to react for MI controls, but it was remarkably more challenging to control the system. With this in mind, the reported performance was tolerable (Table 3). It should be noted that there was a slight improvement for the 4 subject subset in this run also.

4.2.5 User Experience. In the last stage of the experiment, the participants had been provided with a post experimental questionnaire to measure the user acceptance and experience. Results of this survey were summarized in the following.

Participants evaluated five categories on a Likert-scale from 0 to 10, where 0 was "not at all" and 10 was "extremely". They found controlling the VR game with their mind on average very much fun, extremely interesting, somewhat strenuous, very much immersive and experienced a little discomfort (Table 4). They found the game to be fun or interesting but also somewhat strenuous, because of the demand for constant concentration, which was tiring after a while. The immersiveness of the game was likely to increase the deep engagement of MI practise. The slight discomfort was mostly caused by the conductive gel and the VR game environment causing a slight dizziness (i.e. cybersickness) . The participants enjoyed the 3-class trials very much; one of the feedback was "It is a bit weird, because it feels like it would be my hands", which perfectly summarize our intentions with the system.

Table 4: Mean rating of user experience measured on a scale of 0 to 10 (0 means *not at all* and 10 means *extremely*)

Fun	Interesting	Strenous	Immersive	Discomfort
8.40±1.65 (10)	9.50±0.71 (10)	5.90±2.47 (10)	8.20±1.55 (10)	3.70±2.31 (7)

5 CONCLUSION

In this work, a real-time MI-BCI was developed with VR feedback. This system performed well on 3-class experiments and received very good feedback from the users. The designed CNN classifiers' offline performance surpassed the previous works on the same dataset. One of the most important aspect of the current work compared to the previous works was that the DL was applied in a real-time experiment. Only Schirrmeister et. al. [34], and Dose et. al.[11] suggested, that their network might be used for real-time applications. Our most promising results is most likely that we get a 78.6 % hitrate with a 0.5 [s] 16 channel 3-class configuration, that may even become more efficient if we use a subject setting instead of a global. A contemporary approach was suggested with detection of trial onsets, thus optimizing the classification for real-time performance. These classifiers had a slightly decreased performance compared to the whole trial counterparts, but as they can produce a better response time it is implied to employ them on game-like applications.

The conducted experiment confirms a clinical potential of such a system by addressing the problem of low motivation. Therefore, we suggest using MI-BCI-VR systems with classifiers based on DL for real-time applications, like a motivated MI stroke rehabilitation.

REFERENCES

[1] David Achanccaray, Kevin Acuna, Erick Carranza, and Javier Andreu-Perez. 2017. A virtual reality and brain computer interface system for upper limb rehabilitation of post stroke patients. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE. https://doi.org/10.1109/fuzz-ieee.2017.8015726

[2] Roxana Aldea and Monica Fira. 2014. Classifications of motor imagery tasks in brain computer interface using linear discriminant analysis. *International Journal of Advanced Research in Artificial Intelligence* 3, 7 (2014), 5–9.

[3] Kai Keng Ang, Cuntai Guan, K S G Chua, Beng Ti Ang, C Kuah, Chuanchu Wang, Kok Soon Phua, Zheng Yang Chin, and Haihong Zhang. 2010. Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE. https://doi.org/10.1109/iembs.2010.5626782

[4] Derek B. Archer, Nyeonju Kang, Gaurav Misra, Shannon Marble, Carolynn Patten, and Stephen A. Coombes. 2018. Visual feedback alters force control and functional activity in the visuomotor network after stroke. *NeuroImage: Clinical* 17 (2018), 505–517. https://doi.org/10.1016/j.nicl.2017.11.012

[5] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. 2015. Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks. *CoRR* (2015). arXiv:cs.LG/1511.06448v3

[6] M. S. Cameirao, S. B. i. Badia, E. Duarte, A. Frisoli, and P. F. M. J. Verschure. 2012. The Combined Impact of Virtual Reality Neurorehabilitation and Its Interfaces on Upper Extremity Functional Recovery in Patients With Chronic Stroke. *Stroke* 43, 10 (aug 2012), 2720–2728. https://doi.org/10.1161/strokeaha.112.653196

[7] Antonio Maria Chiarelli, Pierpaolo Croce, Arcangelo Merla, and Filippo Zappasodi. 2018. Deep learning for hybrid EEG-fNIRS brain–computer interface: application to motor imagery classification. *Journal of Neural Engineering* 15, 3 (apr 2018), 036028. https://doi.org/10.1088/1741-2552/aaaf82

[8] Michael A. Dimyan and Leonardo G. Cohen. 2011. Neuroplasticity in the context of motor rehabilitation after stroke. *Nature Reviews Neurology* 7, 2 (jan 2011), 76–85. https://doi.org/10.1038/nrneurol.2010.200

[9] Bruce H. Dobkin. 2007. Brain-computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation. *The Journal of Physiology* 579, 3 (mar 2007), 637–642. https://doi.org/10.1113/jphysiol.2006.123067

[10] Bruce H. Dobkin and Andrew Dorsch. 2013. New Evidence for Therapies in Stroke Rehabilitation. *Current Atherosclerosis Reports* 15, 6 (apr 2013). https://doi.org/10.1007/s11883-013-0331-y

[11] Hauke Dose, Jakob Skadkær Møller, Helle K. Iversen, and Sadasivan Puthusserypady. 2018. An End-to-end Deep Learning Approach to MI-EEG Signal Classification for BCIs. *Expert Systems with Applications* 114 (2018), 532–542. https://doi.org/10.1016/j.eswa.2018.08.031

[12] Bogdan Draganski, Christian Gaser, Volker Busch, Gerhard Schuierer, Ulrich Bogdahn, and Arne May. 2004. Changes in grey matter induced by training. *Nature* 427, 6972 (jan 2004), 311–312. https://doi.org/10.1038/427311a

[13] A L Goldberger, L A Amaral, L Glass, J M Hausdorff, P C Ivanov, R G Mark, J E Mietus, G B Moody, C K Peng, and H E Stanley. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101 (June 2000), E215–E220. Issue 23.

[14] Vikram Shenoy Handiru and Vinod A. Prasad. 2016. Optimized Bi-Objective EEG Channel Selection and Cross-Subject Generalization With Brain–Computer Interfaces. *IEEE Transactions on Human-Machine Systems* 46, 6 (dec 2016), 777–786. https://doi.org/10.1109/thms.2016.2573827

[15] Steven G. Kernie and Jack M. Parent. 2010. Forebrain neurogenesis after focal Ischemic and traumatic brain injury. *Neurobiology of Disease* 37, 2 (feb 2010), 267–274. https://doi.org/10.1016/j.nbd.2009.11.002

[16] Jasmin Kevric and Abdulhamit Subasi. 2017. Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomedical Signal Processing and Control* 31 (jan 2017), 398 – 406. https://doi.org/10.1016/j.bspc.2016.09.007

[17] Youngjoo Kim, Jiwoo Ryu, Ko Keun Kim, Clive C. Took, Danilo P. Mandic, and Cheolsoo Park. 2016. Motor Imagery Classification Using Mu and Beta Rhythms of EEG with Strong Uncorrelating Transform Based Complex Common Spatial Patterns. *Computational Intelligence and Neuroscience* 2016 (2016), 1–13. https://doi.org/10.1155/2016/1489692

[18] Wonjun Ko, Jeeseok Yoon, Eunsong Kang, Eunji Jun, Jun-Sik Choi, and Heung-Il Suk. 2018. Deep recurrent spatio-temporal neural network for motor imagery based BCI. In *2018 6th International Conference on Brain-Computer Interface (BCI)*. IEEE. https://doi.org/10.1109/iww-bci.2018.8311535

[19] Shiu Kumar, Alok Sharma, Kabir Mamun, and Tatsuhiko Tsunoda. 2016. A Deep Learning Approach for Motor Imagery EEG Signal Classification. In *2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*. IEEE. https://doi.org/10.1109/apwc-on-cse.2016.017

[20] Peter Langhorne, Fiona Coupar, and Alex Pollock. 2009. Motor recovery after stroke: a systematic review. *The Lancet Neurology* 8, 8 (aug 2009), 741–754. https://doi.org/10.1016/s1474-4422(09)70150-4

[21] Robert Leeb, Doron Friedman, Gernot R. MÃijller-Putz, Reinhold Scherer, Mel Slater, and Gert Pfurtscheller. 2007. Self-Paced (Asynchronous) BCI Control of a Wheelchair in Virtual Environments: A Case Study with a Tetraplegic. *Computational Intelligence and Neuroscience* 2007 (2007), 1–8. https://doi.org/10.1155/2007/79642

[22] Xiang Li, Shuo Chen, Xiaolin Hu, and Jian Yang. 2018. Understanding the Disharmony between Dropout and Batch Normalization by Variance Shift. *arXiv:1801.05134v1* (2018). arXiv:cs.LG/http://arxiv.org/abs/1801.05134v1

[23] Ana Loboda, Alexandra Margineanu, Gabriela Rotariu, and Anca Mihaela. 2014. Discrimination of EEG-Based Motor Imagery Tasks by Means of a Simple Phase Information Method. *International Journal of Advanced Research in Artificial Intelligence* 3, 10 (2014). https://doi.org/10.14569/ijarai.2014.031002

[24] Keith R. Lohse, Courtney G. E. Hilderman, Katharine L. Cheung, Sandy Tatla, and H. F. Machiel Van der Loos. 2014. Virtual Reality Therapy for Adults Post-Stroke: A Systematic Review and Meta-Analysis Exploring Virtual Environments and Commercial Games in Therapy. *PLoS ONE* 9, 3 (mar 2014), e93318. https://doi.org/10.1371/journal.pone.0093318

[25] Na Lu, Tengfei Li, Xiaodong Ren, and Hongyu Miao. 2017. A Deep Learning Scheme for Motor Imagery Classification based on Restricted Boltzmann Machines. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 6 (jun 2017), 566–576. https://doi.org/10.1109/tnsre.2016.2601240

[26] Kip A Ludwig, Rachel M Miriani, Nicholas B Langhals, Michael D Joseph, David J Anderson, and Daryl R Kipke. 2009. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of neurophysiology* 101 (March 2009), 1679–1689. Issue 3. https://doi.org/10.1152/jn.90989.2008

[27] F. Jović M. Tolić. 2013. Classification of wavelet transformed EEG signals with neural network for imagined mental and motor tasks. *Kinesiology* 45, 1 (2013), 130–138. https://hrcak.srce.hr/104591

[28] Th. Mulder. 2007. Motor imagery and action observation: cognitive tools for rehabilitation. *Journal of Neural Transmission* 114, 10 (jun 2007), 1265–1278. https://doi.org/10.1007/s00702-007-0763-z

[29] Cheolsoo Park, Clive Cheong Took, and Danilo P. Mandic. 2014. Augmented Complex Common Spatial Patterns for Classification of Noncircular EEG From Motor Imagery Tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22, 1 (jan 2014), 1–10. https://doi.org/10.1109/tnsre.2013.2294903

[30] Floriana Pichiorri, Giovanni Morone, Manuela Petti, Jlenia Toppi, Iolanda Pisotta, Marco Molinari, Stefano Paolucci, Maurizio Inghilleri, Laura Astolfi, Febo Cincotti, and Donatella Mattia. 2015. Brain-computer interface boosts motor imagery practice during stroke recovery. *Annals of Neurology* 77, 5 (mar 2015), 851–865. https://doi.org/10.1002/ana.24390

[31] Floriana Pichiorri, Natalie Mrachacz-Kersting, Marco Molinari, Sonja Kleih, Andrea KÃijbler, and Donatella Mattia. 2016. Brain-computer interface based motor and cognitive rehabilitation after stroke – state of the art, opportunity, and barriers: summary of the BCI Meeting 2016 in Asilomar. *Brain-Computer Interfaces* 4, 1-2 (oct 2016), 53–59. https://doi.org/10.1080/2326263x.2016.1246328

[32] Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. 2018. On the Convergence of Adam and Beyond. In *International Conference on Learning Representations*. https://openreview.net/forum?id=ryQu7f-RZ

[33] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw. 2004. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions on Biomedical Engineering* 51, 6 (jun 2004), 1034–1043. https://doi.org/10.1109/tbme.2004.827072

[34] Robin Tibor Schirrmeister, Lukas Gemein, Katharina Eggensperger, Frank Hutter, and Tonio Ball. 2017. Deep learning with convolutional neural networks for decoding and visualization of EEG pathology. *Human Brain Mapping* 38, 11 (aug 2017), 5391–5420. https://doi.org/10.1002/hbm.23730 arXiv:cs.LG/1708.08012v1

[35] N. Sharma, V. M. Pomeroy, and J.-C. Baron. 2006. Motor Imagery: A Backdoor to the Motor System After Stroke? *Stroke* 37, 7 (jun 2006), 1941–1952. https://doi.org/10.1161/01.str.0000226902.43357.fc

[36] Yurun Shen, Hongtao Lu, and Jie Jia. 2017. Classification of Motor Imagery EEG Signals with Deep Learning Models. In *Lecture Notes in Computer Science*. Springer International Publishing, 181–190. https://doi.org/10.1007/978-3-319-67777-4_16

[37] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR* (2014). arXiv:cs.CV/1409.1556v6

[38] Ryan Spicer, Julia Anglin, David M. Krum, and Sook-Lei Liew. 2017. REINVENT: A low-cost, virtual reality brain-computer interface for severe stroke upper limb motor recovery. In *2017 IEEE Virtual Reality (VR)*. IEEE. https://doi.org/10.1109/vr.2017.7892338

[39] Yousef Rezaei Tabar and Ugur Halici. 2016. A novel deep learning approach for classification of EEG motor imagery signals. *Journal of Neural Engineering* 14, 1 (nov 2016), 016003. https://doi.org/10.1088/1741-2560/14/1/016003

[40] Zhichuan Tang, Chao Li, and Shouqian Sun. 2017. Single-trial EEG classification of motor imagery using deep convolutional neural networks. *Optik - International Journal for Light and Electron Optics* 130 (feb 2017), 11–18. https://doi.org/10.1016/j.ijleo.2016.10.117

[41] Wei-Peng Teo, Makii Muthalib, Sami Yamin, Ashlee M. Hendy, Kelly Bramstedt, Eleftheria Kotsopoulos, Stephane Perrey, and Hasan Ayaz. 2016. Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? – A Narrative Review of the Literature. *Frontiers in Human Neuroscience* 10 (jun 2016). https://doi.org/10.3389/fnhum.2016.00284

[42] L.E.H. van Dokkum, T. Ward, and I. Laffont. 2015. Brain computer interfaces for neurorehabilitation – its current status as a rehabilitation strategy post-stroke. *Annals of Physical and Rehabilitation Medicine* 58, 1 (feb 2015), 3–8. https://doi.org/10.1016/j.rehab.2014.09.016

[43] Athanasios Vourvopoulos, André Ferreira, and Sergi Bermudez i Badia. 2016. NeuRow: An Immersive VR Environment for Motor-Imagery Training with the Use of Brain-Computer Interfaces and Vibrotactile Feedback. In *Proceedings of the 3rd International Conference on Physiological Computing Systems*. SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0005939400430053

[44] Yijun Wang, Shangkai Gao, and Xiaornog Gao. 2005. Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* 5 (2005), 5392–5395. https://doi.org/10.1109/IEMBS.2005.1615701

[45] Jin Zhang, Chungang Yan, and Xiaoliang Gong. 2017. Deep convolutional neural network for decoding motor imagery based brain computer interface. In *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. IEEE. https://doi.org/10.1109/icspcc.2017.8242581