# Decoding kinetic features of hand motor preparation from single-trial EEG using convolutional neural networks

Gatti, Ramiro; Atum, Yanina; Schiaffino, Luciano; Jochumsen, Mads; Biurrun Manresa, José

*Published in:*
European Journal of Neuroscience

*DOI (link to publication from Publisher):*
10.1111/ejn.14936

*Publication date:*
2021

*Document Version*
Accepted author manuscript, peer reviewed version

MR. RAMIRO GATTI (Orcid ID : 0000-0001-7367-0791)

DR. JOSÉ A. BIURRUN MANRESA (Orcid ID : 0000-0003-4060-9665)

Article type : Research Report

**European Journal of Neuroscience – Research article**

# Decoding kinetic features of hand motor preparation from single-trial EEG using convolutional neural networks

**Ramiro Gatti [1,2], Yanina Atum [2], Luciano Schiaffino [2], Mads Jochumsen [3] and José Biurrun Manresa [1,2,4]**

[1] Institute for Research and Development in Bioengineering and Bioinformatics (IBB), CONICET-UNER, Oro Verde, Argentina

[2] Laboratory for Rehabilitation Engineering and Neuromuscular and Sensory Research (LIRINS), Faculty of Engineering, National University of Entre Ríos, Oro Verde, Argentina

[3] Center for Sensory-Motor Interaction (SMI ®), Aalborg University, Aalborg, Denmark

[4] Center for Neuroplasticity and Pain (CNAP), Aalborg University, Aalborg, Denmark

**Corresponding author**

- **Name:** José Biurrun Manresa

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/EJN.14936

- **Institutional affiliation:** Institute for Research and Development in Bioengineering and Bioinformatics (IBB), CONICET-UNER
- **Address:** Route 11 km. 10, Oro Verde (Entre Ríos), Argentina.
- **Fax number:** (+54) 343 4975077 / 5078
- **e-mail address:** jbiurrun@ingenieria.uner.edu.ar


**Running title:** Decoding motor preparation from EEG using ConvNets

Number of pages: 34. Number of figures: 8 main figures and 3 supplementary figures. Number of tables: 1.

The total number of words in: (i) the whole manuscript: 9224; (ii) the Abstract: 222


**Keywords:** movement prediction, multi-class classification, neural engineering, brain computer interface, deep learning

## Abstract

Building accurate movement decoding models from brain signals is crucial for many biomedical applications. Predicting specific movement features, such as speed and force, before movement execution may provide additional useful information at the expense of increasing the complexity of the decoding problem. Recent attempts to predict movement speed and force from the electroencephalogram (EEG) achieved classification accuracy at or slightly above chance levels, highlighting the need for more accurate prediction strategies. Thus, the aims of this study were to accurately predict hand movement speed and force from single-trial EEG signals and to decode neurophysiological information of motor preparation from the prediction strategies. To these ends, a decoding model based on convolutional neural networks (ConvNets) was implemented and compared against other state-of-the-art prediction strategies, such as support vector machines and decision trees. ConvNets outperformed the other prediction strategies, achieving an overall accuracy of 84% in the classification of two different levels of speed and force (4-class classification) from pre-movement single-trial EEG (100 ms and up to 1600 ms prior to movement execution). Furthermore, an analysis of the ConvNet architectures suggests that the network performs a complex spatiotemporal integration of EEG data to optimize classification accuracy. These results show that movement speed and force can be accurately predicted from single-trial EEG, and that the prediction strategies may provide useful neurophysiological information about motor preparation.

## Introduction

Decoding brain signals to predict movements is useful in many research areas, such as neuromechanics, robotics and neural engineering, among others (Nordin *et al.*, 2017). Furthermore, it is also relevant in neurological rehabilitation, since it has potential to facilitate the assessment of the central nervous system in patients, promote neural plasticity, improve motor dysfunction and allow the control of assistive devices through brain-computer interfaces (BCI) (Brunner *et al.*, 2015). In this regard, motor commands generated prior to or during movement execution can be extracted from specific oscillatory patterns in the electroencephalogram (EEG) (Wolpaw *et al.*, 2002; Machado *et al.*, 2010). Particularly, the component waves of movement-related cortical potentials immersed in the EEG, such as the readiness potential and contingent negative variation, carry information about anticipatory behaviour, which can be used to predict movements before they are actually performed, i.e., in a time window ranging from 100 ms and up to 2000 ms prior to motor execution (Brunia, 1999; Ibáñez *et al.*, 2015; Shakeel *et al.*, 2015).

The movement decoding process is generally focused on detecting a predetermined final state and often ignores other relevant features of the execution, resulting in simple, rough commands (Uktveris & Jusas, 2017). Research on fine movements of body structures such as fingers (Liao *et al.*, 2014), or complex movement control (Jochumsen *et al.*, 2016) is comparatively scarce. It is straightforward to hypothesize that better commands can be achieved if movement kinematics and kinetics are considered in the decoding process (Jerbi *et al.*, 2011). In this regard, recent attempts to predict speed and force from a hand grasping task in a single-trial, single electrode strategy resulted in a classification accuracy at or slightly above chance level (Morash *et al.*, 2008; Jochumsen *et al.*, 2015; Jochumsen *et al.*, 2017). These results contrast with recent studies showing that it is indeed possible to decode hand movement velocities (Bradberry *et al.*, 2010; Lv *et al.*, 2010) and 3D trajectories (Kim *et al.*, 2015) as well as to prediction speed and force of a specific movement from EEG (Jochumsen *et al.*, 2013; Jochumsen *et al.*, 2015), albeit with limited accuracy. Thus, we hypothesized that a better prediction strategy could result in a higher performance in this multi-class classification problem.

In relation to this, the control strategies generated by the nervous system for goal-directed motor behaviour are extremely complex. Thus, pattern recognition systems used to decode and predict movements require careful engineering and domain expertise to transform raw EEG signals

(usually by means of a feature extraction subsystem) into a suitable representation for the classification stage (LeCun *et al.*, 2015). In this regard, several techniques have been proposed for feature extraction, e.g., common spatial patterns, independent component analysis, and joint time-frequency analysis, and also for classification, e.g., nearest neighbour classifier, linear discriminant analysis, support vector machines (SVMs), and ensemble strategies, among others (Lotte *et al.*, 2018). An alternative is to use representation learning methods that automatically perform a feature extraction and classification through optimization algorithms. Deep learning is a paradigmatic example, with multiple levels of representation obtained by combining simple but non-linear modules that transform the input into increasingly more abstract levels (LeCun *et al.*, 2015). In line with this, a decoding model based on deep learning implemented through convolutional neural networks (ConvNets) recently showed promising results in classification performance using different EEG paradigms (Lawhern *et al.*, 2018).

Therefore, the aims of this study were to accurately predict hand movement speed and force from single-trial EEG signals and to decode neurophysiological information of motor preparation from the prediction strategies. To these ends, a group of healthy subjects executed an isometric right hand palmar grasp task using two predefined levels of force (20% and 60% of the maximum voluntary contraction, MVC) and speed (a 3-s slow grasp and a 0.5-s fast grasp). EEG data were minimally pre-processed, to minimize experimenter bias. A prediction strategy using ConvNets was implemented and compared against state-of-the-art prediction strategies, such as support vector machines (SVMs) and decision trees. Overall classification accuracy, precision, recall and Cohen's kappa (κ) values were quantified to evaluate the performance of the proposed prediction strategies. Furthermore, the resulting predictions strategies were analysed to decode useful neurophysiological information related to motor preparation.

## Materials and methods

### Dataset

A dataset consisting of EEG recordings from sixteen healthy subjects was employed (Jochumsen *et al.*, 2015). Written informed consent was obtained from all subjects prior to participation, and the Declaration of Helsinki was respected. The study was approved by the local ethical committee of Region NordJylland (approval no. N-20100067). A Neuroscan NuAmp Express amplifier was used to record the EEG (Compumedics Ltd., Victoria, Australia) from the electrodes shown in Fig.

1, in accordance to the 10/10 system. The corresponding EEG channels were referenced to the right earlobe and grounded at nasion. EEG was recorded during four isometric right palmar grasp tasks with different execution speeds and force levels (expressed as percentage of MVC), categorized as follows: *Slow20*, 3 s to reach 20% MVC; *Slow60*, 3 s to reach 60% MVC; *Fast2*, 0.5 s to reach 20% MVC and *Fast60*, 0.5 s to reach 60% MVC. Forty externally cued repetitions (trials) were performed for each task, in which the cue was delivered 3 s before movement onset (Fig. 2, top). During the experiment, the impedance of all electrodes was kept below 5 kΩ, continuously sampled at 500 Hz and stored for offline analysis. For additional details of the experimental procedure, please refer to (Jochumsen *et al.*, 2015).

**Pre-processing**

EEG was notch-filtered (50 Hz) using a zero-phase filter to reduce power line interference and the baseline (1-s interval before the cue) was subtracted from all trials. No further pre-processing or filtering was applied to the EEG signals, and noisy epochs were not removed, to minimize experimenter bias. Forty trials per task were executed, resulting in 160 trials per subject. Using previous studies as reference, trials were segmented into 500-ms epochs, from 600 ms to 100 ms before movement onset (Fig. 2, bottom), and these segments were used to compare classification performance between different strategies. To corroborate data quality and ensure that classification performance was driven by electrocortical activity and not by contamination from muscle artifacts, an in-depth time-frequency analysis of the signals was performed (Supplementary Fig. 1), that did not show any signs of contamination by noise or artifacts that could compromise classification performance.

**Prediction strategies**

*Convolutional Neural Network (ConvNet)*

The model was based on the EEGnet described in Lawhern et al. (2018). The ConvNet was built in TensorFlow 1.11 (Abadi *et al.*, 2016) using the Keras API (Chollet, 2015) and trained on a Dell Precision 7910 workstation with an NVIDIA Titan Xp GPU, using CUDA 9 and cuDNN 7.3. The model consisted of two blocks (Table 1). The input of the first layer was a pre-processed three-dimensional (3D) matrix for each trial, which was reshaped to apply four temporal filters ($F_1$) to each channel. Following the original net architecture, convolutional kernels of size $(1, 64)$ were applied in the temporal dimension. Kernel weights were initialized with a Glorot uniform

technique, without applying a bias vector. The spatial dimension size was kept constant through zero padding without stride. Then, a batch normalization was applied. Afterwards, the matrix was reshaped, and its dimensions were permuted to apply a depthwise convolution to every temporal slice by means of the wrapper time distributed layer (Chollet, 2017). Two spatial filters of size ($C_x$ , $C_y$) for each feature map (depth multiplier parameter $D$) were applied, the matrix was then reshaped, and dimensions were permuted again. Afterwards, a batch normalization followed by an Exponential Linear Unit (ELU) activation with $alpha = 1$, an average pooling of size $(1, 4)$ and drop-out with a rate of 0.25 were applied. In the second block, a 2D separable convolution of size $(1, 16)$ with eight filters ($F_2$) was applied. Henceforth, ELU activation, batch normalization, average pooling of size $(1, 8)$ and drop-out were applied using the same hyperparameters as in the first block. Finally, the data was flattened to a single dimension and the four resulting scores of the dense layer were transformed to probabilities by means of a softmax activation.

The learning process consisted of a fixed number of learning steps using mini batches of 16 randomly selected trials and the Adam optimization. The initial number of learning steps was set to 500, and validation accuracy and loss curves as a function of the number of learning steps were obtained in order to derive the smallest number of learning steps required to achieve an acceptable classification accuracy. The loss obtained from the validation set was used as metric, and the model was updated if the loss decreased compared to the last saved model. To prevent model overfitting, only the model with the lowest validation loss was kept. In this regard, the relationship between training set size and performance was also analysed to verify that the training set size was appropriate in relation to the dataset size (Goodfellow *et al.*, 2016).

*Alternative ConvNet architectures*

Additionally, we tested the performance of three different architectures to evaluate the effect of the number of parameters on classification performance: 4 temporal filters with a depth multiplier of 1 (ConvNet-4,1), 2 temporal filters with a depth multiplier of 2 (ConvNet-2,2) and 2 temporal filters with a depth multiplier of 1 (ConvNet-2,1), against the original configuration (ConvNet-4,2). These architectures presented different numbers of parameters: ConvNet-4,2 had 916 parameters in total, from which 876 were trainable, whereas ConvNet-4,1 had 580 parameters (556 trainable), ConvNet-2,2 had 444 parameters (424 trainable), and finally, ConvNet-2,1 had 288 parameters (276 trainable).

*Feature decoding*

One of the advantages of ConvNets is that feature extraction and classification is intrinsically optimized, so ConvNets do not require an additional feature extraction stage before classification, so raw spatiotemporal EEG data can be directly classified without additional pre-processing. Thus, the input data for the ConvNets were three-dimensional matrices of $5 \times 4 \times T$ elements, where $5 \times 4$ represents the number of channels and $T$ could be 250 or 500 time samples, depending on the window sizes being tested (see below).

The original ConvNet architecture presented by Lawhern (2018) was devised considering temporal and spectral characteristics of the EEG signals recorded for that study, such as sampling rate, window size and frequency resolution of the filters resulting from the convolutional kernels. For this reason, a dataset-specific alternative architecture was also tested here, whose parameters were derived from extrapolating the original criteria to match the characteristics of the dataset used in this study. The resulting architecture had the first convolutional kernels of size $(1, 250)$, an average pooling of size $(1, 4)$ in the second block, while the rest of the parameters remained unchanged.

Furthermore, we also explored the effect of the EEG window onset and length, so we also tested four alternatives: 500-ms epochs, from 1600 to 1100 ms before movement onset; 500-ms epochs, from 1100 to 600 ms before movement onset; 1000-ms epochs, from 1100 to 100 ms before movement onset; and finally, 1500-ms epochs, from 1600 to 100 ms before movement onset. Finally, we performed a spectral analysis to test the discriminative information content of each EEG frequency band: delta (0-4 Hz), theta, (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (30-150 Hz). To do this, EEG was band-pass filtered using a zero-phase 2$^{nd}$ order Butterworth filter according to the corresponding frequency bands prior to classification.

*State-of-the-art prediction strategies*

Three state-of-the-art prediction strategies were selected, to compare their performance against ConvNets: support vector machines (SVMs), and decision trees (DT) using the bootstrap aggregating (DT-BA) and random forest (DT-RF) ensemble algorithms. SVMs were implemented in this study for reproducibility purposes, as they would allow a direct comparison with prior studies using the same data (Jochumsen *et al.*, 2013; Jochumsen *et al.*, 2015). With regards to the SVM parameters, a radial basis function was used as kernel. Based on a heuristic search, the cost

hyperparameter of the SVM was set to 0.001 and the gamma hyperparameter of the kernel was set to 0.0002. Furthermore, a one-against-one strategy was used to implement the multi-class SVM prediction strategy. The open source library tool for classification and regression problems (LIBSVM) was used to build the SVMs (Chang & Lin, 2011). With regards to DT, Bayesian optimization was used to determine the parameters of BA and RF that minimized the classification error (Cho *et al.*, 2020). The CART (classification and regression tree) model was used for both algorithms, with a maximum split of 2221 and minimum leaf size of 2. After optimization, 359 trees were used for both algorithms; for RF, the number of randomly extracted features was set to 3. All DTs were implemented using Matlab® Statistics and Machine Learning Toolbox (R2019b).

*Feature extraction*

Unlike ConvNets, the other prediction strategies in this study do require a separate feature extraction stage before classification (Chaovalitwongse *et al.*, 2011; Geethanjali *et al.*, 2012; Jochumsen *et al.*, 2015; Rajpurohit *et al.*, 2015). For this purpose, ten features were calculated for each 500-ms epoch, based on *a priori* neurophysiological knowledge about EEG signals: 1) Basal amplitude value, using the Hilbert transform to estimate the area envelope, 2) Kurtosis, 3) Curve length, as the sum of consecutive distances between amplitudes, 4) Noise level, as 3 times the standard deviation of the amplitudes, 5) Number of positive peaks, 6) Average nonlinear energy, using the Teager energy operator, 7) Number of zero crossings, 8) Maximum negativity peak, 9) Root-mean-square amplitude, and 10) Average power in the interval from 0 to 5 Hz, using Welch power spectral density estimator with a Hamming window and a 50% overlap. Therefore, each feature vector for the SVMs and the DTs comprised $5 \times 4 \times 10$ elements, where $5 \times 4$ is the number of channels and $10$ is the number of extracted features.

## Data analysis

Data were divided in 128 trials (80%) for training and validation and 32 trials (20%) for testing. The training and validation set was further split into 102 trials (80%) for training and 18 trials (20%) for validation. To estimate an unbiased generalization performance, a 5-fold nested cross-validation procedure was carried out (Cawley & Talbot, 2010). In this nested scheme, there is an inner-loop cross-validation nested in an outer-loop cross-validation, both using 5-fold partitions. The inner loop is responsible for model selection/hyperparameter tuning (using the validation sets), while the outer loop is for error estimation (using test sets). Individual prediction strategies

were trained for each subject, and the same data partitioning for training, validation, and testing was used for all prediction strategies. Additionally, ConvNets were trained with the same dataset partitioning, but with randomly scrambled labels, to determine the chance classification accuracy level. Furthermore, ConvNets were also trained with increasing number of examples to test the effect of training set size on classification accuracy. The overall classification accuracy and Cohen's κ (a metric that compares observed accuracy with expected accuracy due to random chance), were quantified for each subject. In relation to κ, it is suggested that scores below 0.40 are poor, between 0.41-0.75 are fair to good, and above 0.75 are excellent (Fleiss *et al.*, 2003). Additionally, per-class performance was assessed using precision and recall. For all indexes, the point estimate was calculated as the mean value of the 5-fold cross-validation procedure results computed from the test sets. To ensure reproducibility (Roy *et al.*, 2019), the source code for the ConvNet is accessible at https://github.com/ragatti/STSnet and the dataset is available upon request to the corresponding author.

## Statistics

The Shapiro-Wilk test was performed to assess the assumption of normality, which in general held for all indexes. Performance indexes are reported as mean ± standard deviation unless stated otherwise. A paired t-test was used to assess differences in overall classification accuracy and Cohen's κ between the original ConvNet architecture and the dataset-specific alternative. A one-way repeated-measures analysis of variance (RMANOVA) was used to assess differences in accuracy and κ, with *Classifier*, *Architecture*, *Time window*, and *Frequency band* as factors. Furthermore, per-class differences were also analysed using precision and recall as outcome measures in a three-way RMANOVA; either *Classifier* or *Architecture,* together with *Speed* and *Force,* were selected as factors. Main effects and two-way interactions (when appropriate) were analysed, and the Greenhouse-Geisser correction was applied to account for deviations in sphericity. Furthermore, the Tukey test was used for post-hoc comparisons. In line with current recommendations for statistical analysis (Wasserstein *et al.*, 2019), no fixed threshold for statistical significance was set, and the results are otherwise analysed in terms of the effect sizes and their experimental relevance.

## Results

### ConvNet validation

The evolution of the validation accuracy and validation loss as a function of the number of learning steps is shown in Fig. 3 (top left). It can be observed that the accuracy reaches a stable value after 100 steps while the loss stabilizes after approximately 300 steps, indicating that more training steps would not improve the results and that the ConvNet is not overfitting the data. As a reference, training the ConvNets using 500 steps took approximately 3 min per subject and classifying each new trial took approximately 7 ms. Furthermore, Fig. 3 (top right) shows the chance level accuracy obtained after training the model with randomly scrambled labels. In this case, the prediction accuracy is close to the theoretical chance level of 25% for a 4-class classification problem and it does not improve with the number of training steps. Furthermore, Fig. 3 (bottom center) shows the relationship between test accuracy and training set size, from which it can be deduced that the ConvNet strategy can be trained with as few as 80 examples and still achieve an acceptable classification accuracy (above 80%). In practical terms, this means that it would be viable to perform short recording sessions (in the range of minutes) in order to obtain enough data to train the ConvNets, therefore making it possible to use them in experimental or clinical sessions involving a BCI.

**ConvNet architecture analysis**

We observed that accuracy and κ depended on the selection of *Architecture*, as shown in Fig. 4 (One-way RMANOVA; $\varepsilon = 0.64, F_{1.92,28.8} = 28.17, p < 0.001$ and $\varepsilon = 0.62, F_{1.86,\,27.9} = 38.76, p < 0.001$, respectively). The post hoc analysis revealed that there were no differences in performance between ConvNet-4,2, ConvNet-4,1 and ConvNet-2,2 (Tukey, $p > 0.971$), and that these three strategies outperformed ConvNet-2,1 (Tukey, $p < 0.001$). Furthermore, per-class classification indexes (Supplementary Fig. 2), showed the same behaviour. Essentially, *Architecture* was the only significant factor for precision (Three-way RMANOVA; $\varepsilon = 0.67$, $F_{2.1,\,30.2} = 29.08, p < 0.001$) and recall (Three-way RMANOVA; $\varepsilon = 0.64, F_{1.9,\,28.8} = 28.18,\ p < 0.001$). No other main effect or interactions significantly affected per-class indexes. Post hoc analysis showed that ConvNet-2,1 performed worse than all other strategies (Tukey, $p < 0.001$ for both indexes), with no significant differences among them (Tukey, $p > 0.971$).

**Performance of the classification strategies**

Overall performance in terms of accuracy and κ was affected by the choice of *Classifier*, as shown in Fig. 5 (One-way RMANOVA; $\varepsilon = 0.53, F_{1.60,23.9} = 27.35, p < 0.001$ and $\varepsilon = 0.54, F_{1.57,23.5} = 29.63, p < 0.001$, respectively). Specifically, ConvNets showed higher accuracy and κ values compared to all other strategies (Tukey; $p < 0.001$ for both indexes), which in time did not show any relevant differences among them (Tukey, $p > 0.985$). With regards to per-class classification indexes (Supplementary Fig. 3), precision and recall were likewise only affected by the choice of *Classifier* (Three-way RMANOVA; $\varepsilon = 0.53,\ F_{1.6,\ 24.1} = 28.77, p < 0.001$ and $\varepsilon = 0.55, F_{1.6,\ 24.7} = 26.58, p < 0.001$, respectively). No other main effect or interactions significantly affected per-class indexes. Post hoc analysis showed that ConvNets outperformed all other strategies (Tukey, $p < 0.001$ for both indexes), which did not show any relevant differences among them either (Tukey, $p > 0.894$).

**Feature decoding**

No differences in accuracy and κ were observed between the original ($84.0 \pm 7.0\%$ and $0.79 \pm 0.09$, respectively) and the dataset-specific alternative architecture ($83.8 \pm 5.6\%$; $t_{15} = 0.200, p = 0.844$ and $0.78 \pm 0.07$; $t_{15} = 1.206, p = 0.246$). Considering that the alternative architecture required over 500 additional parameters compared to the original, we decided to use the latter for all further comparisons. Furthermore, Fig 6 (left) shows the performance of the selected architecture, in which the performance of the ConvNet using EEG epochs with different *Time windows* were compared against each other (One-way RMANOVA; accuracy: $\varepsilon = 0.77, F_{3.1,46.5} = 6.575, p < 0.001$ and κ: $\varepsilon = 0.74, F_{2.95,44.3} = 7.553,\ p < 0.001$). Post hoc analysis did not show significant differences due to window onsets and lengths, except for the window from 1.6 to 1.1 s before movement onset, which consistently showed the worst performance. Finally, Fig. 6 (right) shows that accuracy and κ depended on the discriminative information in each *Frequency band* (One-way RMANOVA; $\varepsilon = 0.58, F_{2.32,34.8} = 162.6, p < 0.001$ and $\varepsilon = 0.54, F_{2.17,32.6} = 161.1, p < 0.001$, respectively). Delta (accuracy: $84.0 \pm 7.0\%$; κ: $0.79 \pm 0.08$), gamma (accuracy: $63.7 \pm 8.9\%$; κ: $0.52 \pm 0.12$) and beta bands (accuracy: $48.0 \pm 14.3\%$; κ: $0.31 \pm 0.19$) contained the most discriminative information regarding speed and force (in that order), whereas classification performance using theta (accuracy: $26.7 \pm 5.4\%$; κ: $0\ [0 - 0.08]$ IQR[1]) and alpha bands (accuracy: $28.5 \pm 7.9\%$; κ:

[1] IQR: interquartile range

$0.04\ [0 - 0.13]$ IQR) was not better than chance level. The post hoc analysis revealed that classification performance was different between all frequency bands (Tukey, $p < 0.01$), except for theta and alpha (Tukey, $p > 0.196$).

## Discussion

### Neurophysiological aspects of motor preparation

Building efficient movement decoding models from brain signals is crucial for many biomedical applications, particularly in the BCI field that require precision in online control of assistive devices. Moreover, decoding specific movement features, such as speed, force and/or direction, provides additional degrees of freedom, resulting in more accurate and natural motor commands at the expense of increasing the complexity of the decoding problem (Bradberry *et al.*, 2010; Lv *et al.*, 2010; Agashe & Contreras-Vidal, 2011; Kim *et al.*, 2015). Early attempts to decode movement from brain signals recorded non-invasively during movement execution or imagination were focused on classifying between limb movements (Pfurtscheller *et al.*, 1998, 2006; Yorn-Tov & Inbar, 2001; Olivas-Padilla & Chacon-Murguia, 2018). Classification accuracy for these studies was close to 80% for 2 classes (Pfurtscheller *et al.*, 1998; Yorn-Tov & Inbar, 2001), and close to 75% for 4 classes (Olivas-Padilla & Chacon-Murguia, 2018). Other studies have tried to decode movement of specific body parts from surface EEG, such as wrist (Gu, Dremstrup, *et al.*, 2009), or individual finger movements (Liao *et al.*, 2014), obtaining similar results.

On the other hand, prediction of movement, i.e., decoding movement not during, but before its execution, is a much more difficult task. Considering the brain as a predictive neural system, expectation can be seen as a representation of prediction that serve to sensory or motor areas as preparatory processing prior to an event, particularly in short time scales (Bubic, 2010). It is well known that information about motor preparation is encoded in the movement-related cortical potentials, around 1.5 s prior to movement onset (Shakeel *et al.*, 2015). The timing of the prediction is a relevant feature to study, since it has been shown that a sensory stimulus delivered synchronously with the peak negativity of the movement-related cortical potential maximizes neural plasticity (Mrachacz-Kersting *et al.*, 2012). In this regard, we observed that discriminative information for classification was mostly located in the interval between 1.1 s and movement onset, in line with previous results (Ibáñez *et al.*, 2015).

Furthermore, kinetic information encoded in movement planning could be particularly useful; for example, by decoding these movement parameters it would be possible to introduce task variability in the rehabilitation training, which has been shown to maximize the motor learning (Krakauer, 2006). It has been already shown that pre-movement EEG contains valuable information about motion. Indeed, detection of voluntary movement from non-invasive, single trial EEG using a matched filter approach demonstrated relatively good performance in a 2-class classification scheme (sensitivity ≈ 82.5% for healthy subjects) (Niazi *et al.*, 2011). However, classification rates for multi-class classification problems are still relatively low in healthy subjects. As an example, recent studies directed towards the extraction of additional information from surface EEG regarding movement intention beyond simple detection, such as the prediction of the body part that is about to perform the movement (Morash *et al.*, 2008), or the classification between different types of movement used in daily life, such as palmar, lateral and pinch grasps (Jochumsen, Niazi, *et al.*, 2015), resulted in classification accuracies at or slightly above chance levels for the 4-class classification attempts.

In particular, previous work with the same dataset used in this study obtained mean accuracy values of approximately 32-40% for the 4-class classification (Jochumsen *et al.*, 2015), which is on par or slightly above chance level for that type of problem (Müller-Putz *et al.*, 2008). These results might be partially explained by the fact that the aim of the study was to obtain a fast prediction scheme using few electrodes and a simple classifier that did not require extensive calibration. As such, only one channel was used as input, and the signals were band filtered using low cut-off frequencies values. However, it was recently suggested that information from the entire EEG spectrum is needed to discriminate between task-related parameters from single-trial movement intention (Jochumsen *et al.*, 2017).

Based on this idea, in this study it was possible to significantly improve the movement prediction accuracy using twenty available surface EEG channels without additional pre-processing, such as artifact removal or epoch selection. Accuracy levels reached values close to 85% in healthy subjects, representing an improvement of almost 45% compared to previous results using a single EEG channel. Furthermore, previous studies have shown that classification of speed tasks achieved higher accuracy in the prediction of ankle dorsiflexion movements (Jochumsen *et al.*, 2013). However, it was not the case for hand grasping tasks, since no significant effects of speed or force and no interactions were observed, in line with previous results (Jochumsen, Khan Niazi,

*et al.*, 2015). In light of this, it could be hypothesized that the decoding of complex movement from surface EEG requires more information (in terms of number of channels or features) in order to achieve a classification accuracy comparable with that obtained for simpler movements, such ankle or wrist flexion/extension (binary classification problems) (Shibasaki & Hallett, 2006a; Gu, do Nascimento, *et al.*, 2009; Gu, Dremstrup, *et al.*, 2009; Jochumsen *et al.*, 2013).

In this regard, we altered the original ConvNet architecture to better match the time-frequency characteristics of our database, without a significant improvement in performance. We also observed that neither shifting nor increasing the length of the time window improved the ConvNet performance (Fig. 6, left). Filtering the EEG in pre-defined physiological frequency bands did have an impact on classification performance, with delta, gamma and beta frequency bands providing the best classification performance results, in that order (Fig. 6, right). Indeed, preparatory activity in electrocortical motor signals is well documented in the delta band, particularly related to the readiness potential and the movement-related cortical potentials (Slobounov & Ray, 1998; Ray *et al.*, 2000; Shibasaki & Hallett, 2006b). More recently, evidence of pre-motor activation in the gamma (Gunduz *et al.*, 2016; Schirrmeister *et al.*, 2017) and beta (Tzagarakis *et al.*, 2015) ranges has also been reported. Whereas these results provide a direct estimate of the predictive power of delta, gamma and beta oscillations for motor preparation, it remains a challenge to decode the specific neural processes contributing to the information content stored in each frequency band.

It follows that if sufficient class-discriminative information can be obtained from the delta band alone, and the different time windows onsets and lengths tested did not influence classification performance, then the improvements in classification performance compared to previous attempts might be encoded in the additional spatial information provided by all channels. Fig. 7 shows the resulting network architecture for a randomly selected subject. We attempted to find common patterns across subjects in the filter layers, but we did not detect any patterns that could be generalized across subjects. However, we did identify a common pattern across subjects in the 2D separable convolution layer. Indeed, Fig. 8 shows that the kernels in this layer have weights that are either predominantly positive (in blue) or negative (in red) along the temporal direction (represented as columns in Fig. 8). We interpreted this as a sign that the network performs a complex integration between time, frequency, and spatial features of the EEG signal in the first two layers, that are enhanced in the 2D separable convolution layer. Furthermore, the high channel

density over a relatively small scalp area in this particular dataset might be crucial in order to reach a high classification accuracy, which has potential implications in the selection of electrode distributions in experimental or clinical setups, when the number of available channels is limited.

**Methodological aspects of movement prediction using ConvNets**

Deep learning methods recently gained popularity in EEG analysis by improving classification performance over more traditional approaches, such as linear discriminant analysis, k-nearest neighbours or SVMs (Lotte *et al.*, 2018). ConvNets are a type of feed-forward deep learning networks that are useful when data have a known topological structure (LeCun *et al.*, 2015; Goodfellow *et al.*, 2016). The ConvNet implemented in this study is based on a recently proposed architecture that demonstrated good performance employing a small number of parameters in the classification of EEG signals recorded using different paradigms (Lawhern *et al.*, 2018).

In the ConvNet, the first convolutional layer works as a frequential filter, in which the outcome consists of four different band-pass filters that minimize the error at the output. In accordance with the input structures used in image processing, the EEG input to a ConvNet is usually reshaped into a 2D distribution, by arranging channels along the rows and time samples in the columns (Tang *et al.*, 2016; Schirrmeister *et al.*, 2017) or by transforming the input into a new space (Uktveris & Jusas, 2017), e.g., to a time-frequency domain through Fourier transform and averaging along the channels (Soare, 2016; Lu *et al.*, 2017). Taking this into consideration, only minimal and automatic pre-processing (baseline correction and notch filtering) was performed in this study prior to the classification stage, and no epochs were removed. Results showed that the ConvNet reached the same performance without pre-processing compared to the case in which we specifically selected the best time window and frequency bands according to pre-existing physiological knowledge, implying that the intrinsic optimization built in the ConvNet already performs the best possible feature extraction strategy (including appropriate spatial and temporal filtering) to extract discriminative information, and that this information largely aligns with pre-existing neurophysiological knowledge about EEG signals. Furthermore, our results showed that this strategy did not require extremely large datasets for training and the training time was negligible compared to the average setup time for a BCI, which makes it viable for use in rehabilitation.

We also compared the results obtained using ConvNets with state-of-the-art prediction strategies. The first comparison is with a strategy based on SVMs, to allow for a comparison with previously published results (Jochumsen *et al.*, 2013; Jochumsen *et al.*, 2015). The prediction results of the ConvNet were better than the SVM for all tasks and all performance indexes in healthy subjects by an average of 12 percentage points (Fig. 5). This is even more relevant considering that the SVMs implemented in this study (using twenty available channels) already improved the classification accuracy by approximately 32 percentage points compared to the previous study with the same dataset (using only a single channel, C3, plus an eight-channel Laplacian filter) (Jochumsen *et al.*, 2015), further supporting the notion that the additional spatial information plays a crucial role in classification performance. Furthermore, the performance of classification strategies based on decision trees was also significantly lower compared to ConvNets. Indeed, accuracy and κ values obtained with DTs were not different from those obtained with SVMs, using the same features.

In relation to this, a systematic investigation regarding movement prediction performed with combinations of spatial filtering (principal component analysis, independent component analysis, common spatial patterns analysis, and surface Laplacian derivation), temporal filtering (power spectral density estimation and discrete wavelet transform), pattern classification (linear and quadratic Mahalanobis distance classifier, Bayesian classifier, multi-layer perceptron neural network, probabilistic neural network, and SVM), and multivariate feature selection strategy using a genetic algorithm, achieved a maximum accuracy of 75% for binary classification (Bai *et al.*, 2007). Taken together, these results might imply that the differences in performance between ConvNets and other strategies are probably due to the feature selection and generation methods.

In this regard, most of the current methods for feature extraction are defined by a human investigator based on *a priori* knowledge of the neurophysiology of the brain, for example in terms of time-frequency characteristics of the signals or how the sources of electrical activity are spatially distributed in the cortex. In fact, even standard EEG pre-processing (e.g. band-pass filtering or channel selection based on predefined brain activation patterns) could be inadvertently discarding relevant information for classification. These processes can be time-consuming, prone to experimenter bias and may result in the loss of relevant or interesting information. In contrast, ConvNets require minimal pre-processing and are not limited by feature selection or generation constrains. We changed the number of input features (by changing the window length), and tested several different architectures by changing the number and parameters of the temporal and spatial

filters (Fig. 4), and the ConvNet still maintained excellent performance, even with a reduced number of parameters ($83.9 \pm 5.9\%$ accuracy using 444 total parameters in ConvNet-2,2).

**Limitations and future work**

Several constraints need to be considered: attempts to use a single ConvNet to predict movements from all subjects resulted in low performance indexes during pilot tests (average accuracy of $27.4 \pm 4.4\%$). This is not an issue in most real-life applications where the decoding is used to control a device for a single subject (and thus an individual ConvNet is trained for each subject), but nevertheless highlights the difficulty in describing a general behaviour of the EEG signal in terms of decoding force and speed. The same issue can be observed when attempting to understand and visualize of the specific features that allow a good classification, since it is not straightforward to extract and interpret physiological information from the network, and these features vary between subjects, as for example in the depthwise layer (Fig. 7). Furthermore, even if high accuracy was achieved offline, it is crucial to perform real-time tests with adequate feedback. Future work will be directed towards testing the strategy with a real application, for which an accurate detection of the movement onset is necessary, and an idle state should be considered (Lew *et al.*, 2012). Finally, once the definitive scheme has been defined, efficient hardware implementations should be tested in chips or field-programmable gate arrays (LeCun *et al.*, 2015).

## Conclusion

The results from this study suggest that hand movement speed and force can be accurately predicted from pre-movement EEG using ConvNets. Furthermore, even with minimal pre-processing, the neurophysiological information decoded from the prediction of motor intention aligns with our current understating of motor neuroscience. Nevertheless, additional considerations are still required to transfer these protocols from laboratory to clinic.

## Acknowledgment

The workstation used in this study was provided by the Center for Neuroplasticity and Pain (CNAP), which is supported by the Danish National Research Foundation (DNRF121). The Titan Xp GPU used for this research was donated by the NVIDIA Corporation.

## Competing interests

The authors do not have any potential sources of conflict of interest to disclose.

## Authors' contributions

RG, YA, MJ and JBM conceived the study; MJ collected the data; RG, YA, LS and JBM performed the data analysis; and all authors wrote the manuscript and critically reviewed it.

## Abbreviations

BA: bootstrap aggregating (bagging).

BCI: brain computer interface.

CART: classification and regression tree.

ConvNet: convolutional neural network.

DT: decision tree.

EEG: electroencephalography.

ELU: exponential linear unit.

MVC: maximum voluntary contraction.

RF: random forest.

RMANOVA: repeated measures analysis of variance

SVM: support vector machine.

## References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., & Zheng, X. (2016) TensorFlow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, 265–284.

Agashe, H.A. & Contreras-Vidal, J.L. (2011) Reconstructing hand kinematics during reach to grasp movements from electroencephalographic signals. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 5444–5447.

Bai, O., Lin, P., Vorbach, S., Li, J., Furlani, S., & Hallett, M. (2007) Exploration of computational methods for classification of movement intention during human voluntary movement from single trial EEG. *Clin. Neurophysiol.*, **118**, 2637–2655.

Bradberry, T.J., Gentili, R.J., & Contreras-Vidal, J.L. (2010) Reconstructing Three-Dimensional Hand Movements from Noninvasive Electroencephalographic Signals. *J. Neurosci.*, **30**, 3432–3437.

Brunia, C.H.M. (1999) Neural aspects of anticipatory behavior. *Acta Psychol.*, **101**, 213–242.

Brunner, C., Birbaumer, N., Blankertz, B., Guger, C., Kübler, A., Mattia, D., Millán, J. del R., Miralles, F., Nijholt, A., Opisso, E., Ramsey, N., Salomon, P., & Müller-Putz, G.R. (2015) BNCI Horizon 2020: towards a roadmap for the BCI community. *Brain-Computer Interfaces*, **2**, 1–10.

Bubic (2010) Prediction, cognition and the brain. *Front. Hum. Neurosci.*, **4**, 1–15.

Cawley, G.C. & Talbot, N.L.C. (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.*, **11**, 2079–2107.

Chang, C.-C. & Lin, C.-J. (2011) LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.*, **2**, 27:1--27:27.

Chaovalitwongse, W., Jeong, Y., Jeong, M.K., Danish, S., & Wong, S. (2011) Pattern Recognition Approaches for Identifying Subcortical Targets during Deep Brain Stimulation Surgery. *IEEE Intell. Syst.*, **26**, 54–63.

Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., & Rhee, W. (2020) Basic Enhancement Strategies When Using Bayesian Optimization for Hyperparameter Tuning of Deep Neural Networks. *IEEE Access*, **8**, 52588–52608.

Chollet, F. (2017) Xception: Deep learning with depthwise separable convolutions. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017),* pp. 1800–1807.

Chollet, F. et al. (2015) Keras [WWW Document]. URL https://keras.io/

Fleiss, J.L., Levin, B., & Paik, M.C. (2003) *Statistical Methods for Rates and Proportions*, Statistical Methods for Rates and Proportions, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Geethanjali, P., Mohan, Y.K., & Sen, J. (2012) Time domain Feature extraction and classification of EEG data for Brain Computer Interface. In *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 1136–1139.

Goodfellow, I., Bengio, Y., & Courville, A. (2016) *Deep Learning*. MIT Press.

Gunduz, A., Brunner, P., Sharma, M., Leuthardt, E.C., Ritaccio, A.L., Pesaran, B., & Schalk, G. (2016) Differential roles of high gamma and local motor potentials for movement preparation and execution. *Brain-Computer Interfaces*, **3**, 88–102.

Gu, Y., do Nascimento, O.F., Lucas, M.F., & Farina, D. (2009) Identification of task parameters from movement-related cortical potentials. *Med. Biol. Eng. Comput.*, **47**, 1257–1264.

Gu, Y., Dremstrup, K., & Farina, D. (2009) Single-trial discrimination of type and speed of wrist movements from EEG recordings. *Clin. Neurophysiol.*, **120**, 1596–1600.

Ibáñez, J., Serrano, J.I., del Castillo, M.D., Minguez, J., & Pons, J.L. (2015) Predictive classification of self-paced upper-limb analytical movements with EEG. *Med. Biol. Eng. Comput.*, **53**, 1201–1210.

Jerbi, K., Vidal, J.R., Mattout, J., Maby, E., Lecaignard, F., Ossandon, T., Hamamé, C.M., Dalal, S.S., Bouet, R., Lachaux, J.-P., Leahy, R.M., Baillet, S., Garnero, L., Delpuech, C., & Bertrand, O. (2011) Inferring hand movement kinematics from MEG, EEG and intracranial EEG: From brain-machine interfaces to motor rehabilitation. *IRBM*, **32**, 8–18.

Jochumsen, M., Khan Niazi, I., Taylor, D., Farina, D., & Dremstrup, K. (2015) Detecting and classifying movement-related cortical potentials associated with hand movements in healthy subjects and stroke patients from single-electrode, single-trial EEG. *J. Neural Eng.*, **12**, 056013.

Jochumsen, M., Niazi, I.K., Dremstrup, K., & Kamavuako, E.N. (2016) Detecting and classifying three different hand movement types through electroencephalography recordings for neurorehabilitation. *Med. Biol. Eng. Comput*, **54**, 1491–1501.

Jochumsen, M., Niazi, I.K., Mrachacz-Kersting, N., Farina, D., & Dremstrup, K. (2013) Detection and classification of movement-related cortical potentials associated with task force and speed. *J. Neural Eng.*, **10**, 056015.

Jochumsen, M., Rovsing, C., Rovsing, H., Niazi, I.K., Dremstrup, K., & Kamavuako, E.N. (2017) Classification of Hand Grasp Kinetics and Types Using Movement-Related Cortical Potentials and EEG Rhythms. *Comput. Intell. Neurosci.*, **2017**, 7470864.

Kim, J.-H., Biessmann, F., & Lee, S.-W. (2015) Decoding Three-Dimensional Trajectory of Executed and Imagined Arm Movements From Electroencephalogram Signals. *IEEE Trans. Neural Syst. Rehab. Eng.*, **23**, 867–876.

Krakauer, J.W. (2006) Motor learning: Its relevance to stroke recovery and neurorehabilitation. *Curr. Opin. Neurol.*, **19**(1), 84–90.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., & Lance, B.J. (2018) EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.*, **15**, 056013.

LeCun, Y., Bengio, Y., & Hinton, G. (2015) Deep learning. *Nature*, **521**, 436–444.

Lew, E., Chavarriaga, R., Silvoni, S., & Millán, J. del R. (2012) Detection of self-paced reaching movement intention from EEG signals. *Front. Neuroeng.*, **5**, 13.

Liao, K., Xiao, R., Gonzalez, J., & Ding, L. (2014) Decoding individual finger movements from one hand using human EEG signals. *PLoS ONE*, **9**(1), e85192.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018) A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J. Neural Eng.*, **15**, 031005.

Lu, Y., Jiang, H., & Liu, W. (2017) Classification of EEG Signal by STFT-CNN Framework: Identification of Right-/left-hand Motor Imagination in BCI Systems. In *Proceedings of The 7th International Conference on Computer Engineering and Networks — PoS(CENet2017)*. Sissa Medialab, Trieste, Italy, Corpus ID 65431442.

Lv, J., Li, Y., & Gu, Z. (2010) Decoding hand movement velocity from electroencephalogram signals during a drawing task. *Biomed. Eng. Online*, **9**, 64.

Machado, S., Araújo, F., Paes, F., Velasques, B., Cunha, M., Budde, H., Basile, L.F., Anghinah, R., Arias-Carrión, O., Cagy, M., Piedade, R., de Graaf, T.A., Sack, A.T., & Ribeiro, P. (2010) EEG-based Brain-Computer Interfaces: An Overview of Basic Concepts and Clinical Applications in Neurorehabilitation. *Rev. Neurosci.*, **21**, 451–468.

Morash, V., Bai, O., Furlani, S., Lin, P., & Hallett, M. (2008) Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries. *Clin. Neurophysiol.*, **119**, 2570–2578.

Mrachacz-Kersting, N., Kristensen, S.R., Niazi, I.K., & Farina, D. (2012) Precise temporal association between cortical potentials evoked by motor imagination and afference induces cortical plasticity. *J. Physiol.*, **590**, 1669–1682.

Müller-Putz, G.R., Scherer, R., Brunner, C., Leeb, R., & Pfurtscheller, G. (2008) Better than random? A closer look on BCI results. *Int. J. Bioelectromagn.*, **10**, 52–55.

Niazi, I.K., Jiang, N., Tiberghien, O., Nielsen, J.F., Dremstrup, K., & Farina, D. (2011) Detection of movement intention from single-trial movement-related cortical potentials. *J. Neural Eng.*, **8**, 066009.

Nordin, A.D., Rymer, W.Z., Biewener, A.A., Schwartz, A.B., Chen, D., & Horak, F.B. (2017) Biomechanics and neural control of movement, 20 years later: what have we learned and what has changed? *J. Neuroeng. Rehab.*, **14**, 91.

Olivas-Padilla, B.E. & Chacon-Murguia, M.I. (2018) Classification of multiple motor imagery using deep convolutional neural networks and spatial filters. *Appl. Soft Comput. J.*, **75**, 461–472.

Pfurtscheller, G., Brunner, C., Schlögl, A., & Lopes da Silva, F.H. (2006) Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *Neuroimage*, **31**, 153–159.

Pfurtscheller, G., Neuper, C., Schlogl, A., & Lugger, K. (1998) Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans. Rehab. Eng.*, **6**, 316–325.

Rajpurohit, V., Danish, S.F., Hargreaves, E.L., & Wong, S. (2015) Optimizing computational feature sets for subthalamic nucleus localization in DBS surgery with feature selection. *Clin. Neurophysiol.*, **126**, 975–982.

Ray, W.J., Slobounov, S., Mordkoff, J.T., Johnston, J., & Simon, R.F. (2000) Rate of force development and the lateralized readiness potential. *Psychophysiology*, **37**, 757–765.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., & Faubert, J. (2019) Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.*, **16**, 051001.

Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017) Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.*, **38**, 5391–5420.

Shakeel, A., Navid, M.S., Anwar, M.N., Mazhar, S., Jochumsen, M., & Niazi, I.K. (2015) A Review of Techniques for Detection of Movement Intention Using Movement-Related Cortical Potentials. *Comput. Math. Methods Med.*, **2015**, 1–13.

Shibasaki, H. & Hallett, M. (2006a) What is the Bereitschaftspotential? *Clin. Neurophysiol.*, **117**, 2341–2356.

Slobounov, S.M. & Ray, W.J. (1998) Movement-related potentials with reference to isometric force output in discrete and repetitive tasks. *Exp. Brain Res.*, **123**, 461–473.

Soare, C. (2016) Brain Computer Interface using Machine Learning. In Iftene, A. & Vanderdonckt, J. (eds), *13th International Conference on Human Computer Interaction, RoCHI*. Matrix Rom, Iasi, Romania, pp. 65–68.

Tang, Z., Li, C., & Sun, S. (2016) Single-trial EEG classification of motor imagery using deep convolutional neural networks. *Optik – Int. J. Light Electron Opt.*, **130**, 11–18.

Tzagarakis, C., West, S., Pellizzer, G. (2015) Brain oscillatory activity during motor preparation: Effect of directional uncertainty on beta, but not alpha, frequency band. *Front. Neurosci.* **9**, 1–13.

Uktveris, T. & Jusas, V. (2017) Convolutional Neural Networks for Four-Class Motor Imagery Data Classification. In: Ivanović M., Bădică C., Dix J., Jovanović Z., Malgeri M., Savić M. (eds) *Intelligent Distributed Computing XI. IDC 2017. Studies in Computational Intelligence*, vol 737. Springer, Cham pp. 185–197.

Wasserstein, R.L., Schirm, A.L., & Lazar, N.A. (2019) Moving to a World Beyond " p < 0.05." *Am. Stat.*, **73**, 1–19.

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., & Vaughan, T.M. (2002) Brain–computer interfaces for communication and control. *Clin. Neurophysiol.*, **113**, 767–791.

Yorn-Tov, E. & Inbar, G.F. (2001) Selection of relevant features for classification of movements from single movement-related potentials using a genetic algorithm. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 1364–1366.

## Figure captions

**Figure 1:** Experimental design. EEG signals were recorded from twenty surface electrodes centred around the area above the motor cortex contralateral to the hand performing the grasping task. EEG signal were arranged in three-dimensional matrices, where $x$ and $y$ correspond to the spatial location of each channel and $z$ corresponds to either features ($F_1, F_2$ , ..., $F_N$) extracted from the EEG epochs, as input to the support vector machines (SVM) or decision trees (DT), or raw EEG time samples, as input to the convolutional neural network (ConvNet). In this case, features are intrinsically optimized in the temporal and spatial layers of the ConvNet, and can be later decoded into neurophysiological information.

**Figure 2:** Representative examples of 3-s (top) and 500-ms (bottom) averages of 40 EEG trials recorded during four isometric right palmar grasp tasks, categorized as follows: *Slow20*, 3 s to reach 20% maximum voluntary contraction (MVC); *Slow60,* 3 s to reach 60% MVC; *Fast20*, 0.5 s to reach 20% MVC and *Fast60*, 0.5 s to reach 60% MVC. Vertical lines represent cue (solid green, $t = -3\ s$) and movement onset (dashed green, $t = 0\ s$) times. The solid trace and shading represent mean and 95% confidence intervals for each class, respectively, derived using 5000 bootstrap iterations. Note that classification was performed using pre-movement, single trial EEG (the figure depicts average EEG signals for clarity).

**Figure 3:** Convolutional neural network (ConvNet) validation tests. Top left: Evolution of the validation accuracy (blue) and validation loss (orange) as a function of the number of learning steps. Top right: validation accuracy from all subjects as a function of the number of learning steps with randomly scrambled labels. Dark lines represent the mean validation accuracy / loss across all subjects ($n = 16$), and light lines represents the mean validation accuracy/loss for single subjects, derived from the inner-loop 5-fold cross-validation procedure. Note that validation accuracy does not improve significantly after 100 learning steps, and that the average classification accuracy obtained with randomly scrambled level matches the expected change level accuracy for 4-class classification (approx. 25%). Bottom center: classification accuracy as a function of the number of examples, using the test set. The dark line represents the mean test accuracy for all subjects ($n = 16$), and each light line represents the mean test accuracy for a single subject, derived from the outer-loop 5-fold cross-validation procedure. Note that test accuracy does not improve significantly with more than 80 training examples, indicating that the training set size was appropriate. a.u.: arbitrary units.

**Figure 4:** Overall classification performance (left: accuracy, right: Cohen's κ) of different convolutional neural networks (ConvNets), using the test set (the numerical designation denotes the number of temporal filters and depth multiplier, respectively). Boxes represent the median and the 25$^{th}$ and 75$^{th}$ percentiles, whiskers represent 5$^{th}$ and 95$^{th}$ percentiles, diamonds represent values outside of the 5$^{th}$ – 95$^{th}$ percentile range and the individual dots represent the average accuracy / κ for each individual subject, calculated from the outer-loop 5-fold cross-validation procedure ($n = 16$). Note that the ConvNets required at least four filters in the first layer to maintain a high classification performance.

**Figure 5:** Overall classification performance (left: accuracy, right: Cohen's κ) of the prediction strategies, using the test set. Boxes represent the median and the 25$^{th}$ and 75$^{th}$ percentiles, whiskers represent 5$^{th}$ and 95$^{th}$ percentiles, diamonds represent values outside of the 5$^{th}$ – 95$^{th}$ percentile range and the individual dots represent the average accuracy / κ for each individual subject, calculated from the outer-loop 5-fold cross-validation procedure ( $n = 16$). Note that the state-of-the-art prediction strategies implemented here significantly improved previously reported results using this dataset; nevertheless, convolutional neural networks significantly outperformed them all. ConvNet: convolutional neural network. SVM:

support vector machine. DT: decision tree. BA: bagging algorithm. RF: random forest algorithm.

**Figure 6:** Overall classification performance (top: accuracy, bottom: Cohen's κ) of the convolutional neural network (ConvNet) as a function of time window (left) and frequency band (right), using the test set. Time windows were selected from intervals *before* movement onset. Boxes represent the median and the 25$^{th}$ and 75$^{th}$ percentiles, whiskers represent 5$^{th}$ and 95$^{th}$ percentiles, diamonds represent values outside of the 5$^{th}$ – 95$^{th}$ percentile range and the individual dots represent the average accuracy / κ for each individual subject, calculated from the outer-loop 5-fold cross-validation procedure ($n = 16$). Note that most discriminative information is encompassed in an interval from 1.1 s to 0.1 s before movement onset, predominantly on the delta and gamma frequency bands.

**Figure 7:** Example of resulting convolutional neural network (ConvNet) architecture for a randomly selected subject. Left: frequency response of the four temporal filters of the first layer. Center: spatial kernels for the depthwise layer. Right: kernels for the separable convolutional layer. Note that temporal filters' frequency response appears to enhance low frequencies (predominantly in the delta range), and that it is difficult to decode an interpretable physiological pattern from the spatial filters.

**Figure 8:** Kernels from 2D separable convolution layer for all volunteers ($n = 16$). Each convolutional neural network (ConvNet) has 8 spatial kernels (columns) with 16 temporal values each (rows) in this layer. Note that there is a clear pattern in the weights, that are either predominantly positive (in blue) or negative (in red) along the temporal direction (represented as columns), enhancing the differences between complex spatiotemporal patterns produce in previous layers.

**Supplementary Figure 1:** Representative examples of single trial EEG data (odd columns) along with the corresponding scalograms (even columns) for two randomly selected subjects. Each row represents a different task, categorized as follows: *Slow20*, 3 s to reach 20% maximum voluntary contraction (MVC); *Slow60,* 3 s to reach 60% MVC; *Fast20*, 0.5 s to reach 20% MVC and *Fast60*, 0.5 s to reach 60% MVC. For single trial EEG data, the $x$ axis represents time, the $y$ axis represent the number of epoch, and the colorbar depicts EEG amplitude in $\mu V$. For the scalograms, the $x$ axis represents time, the $y$ axis represent

frequency, and the colorbar depicts scalogram magnitude, computed using generalized analytic Morse wavelets with gamma factor γ = 3. Vertical black lines denote movement onset ($t = 0\ s$), and the shaded grey zones represent the cone of influence.

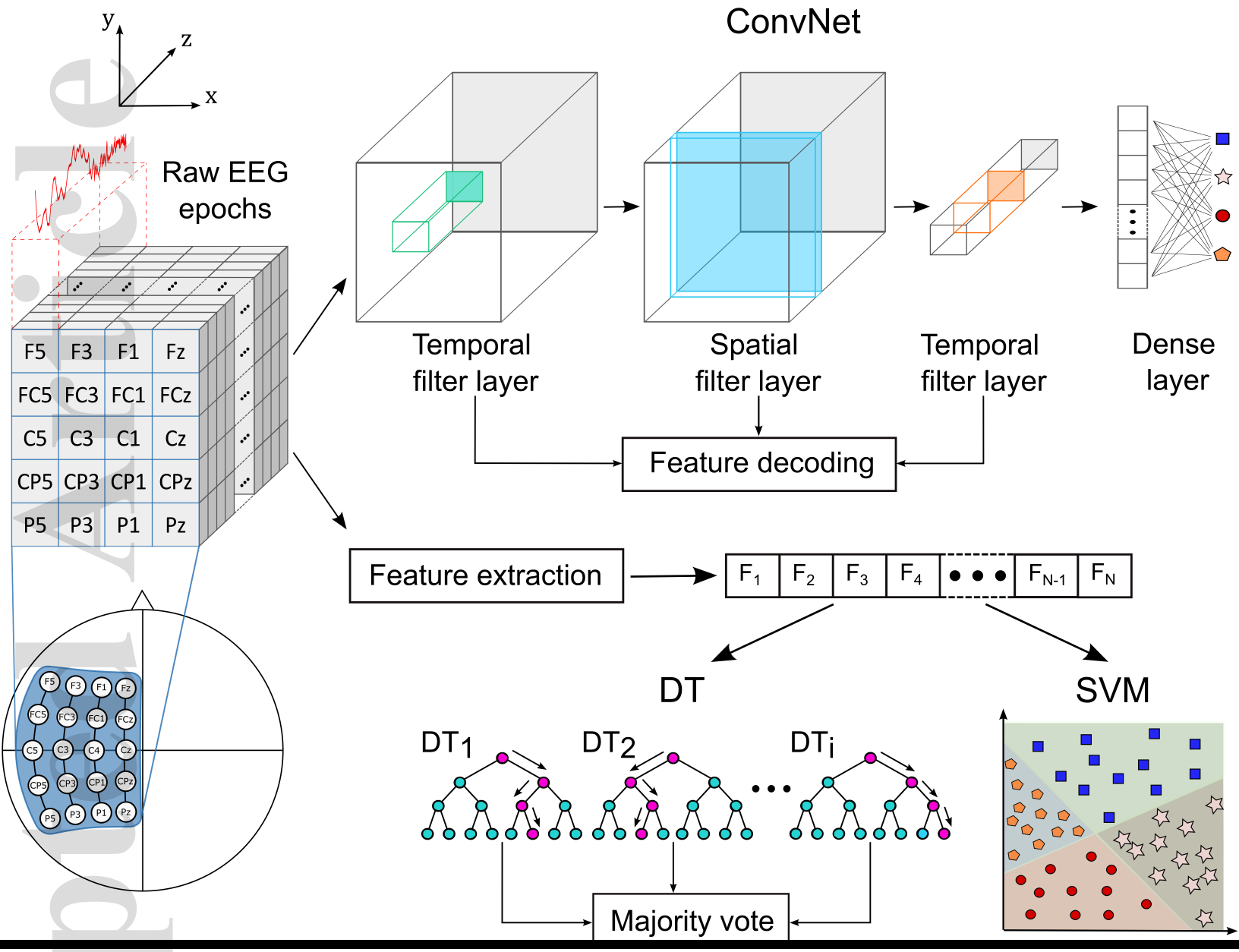
**Supplementary Figure 2:** Per-class classification performance (top: precision, bottom: recall) of different convolutional neural networks (ConvNets) architectures, using the test set (the numerical designation denotes the number of temporal filters and depth multiplier, respectively). Boxes represent the median and the 25$^{th}$ and 75$^{th}$ percentiles, whiskers represent 5$^{th}$ and 95$^{th}$ percentiles, diamonds represent values outside of the 5$^{th}$ – 95$^{th}$ percentile range and the individual dots represent the average precision / recall for each individual subject, calculated from the outer-loop 5-fold cross-validation procedure ($n = 16$). Note that per-class classification performance was not different across ConvNet architectures.

**Supplementary Figure 3:** Per-class classification performance (top: precision, bottom: recall) of the prediction strategies, using the test set. Boxes represent the median and the 25$^{th}$ and 75$^{th}$ percentiles, whiskers represent 5$^{th}$ and 95$^{th}$ percentiles, diamonds represent values outside of the 5$^{th}$ – 95$^{th}$ percentile range and the individual dots represent the average precision / recall for each individual subject, calculated from the outer-loop 5-fold cross-validation procedure ($n = 16$). Note that per-class classification performance was not imbalanced across strategies. ConvNet: convolutional neural network. SVM: support vector machine. DT: decision tree. BA: bagging algorithm. RF: random forest algorithm.

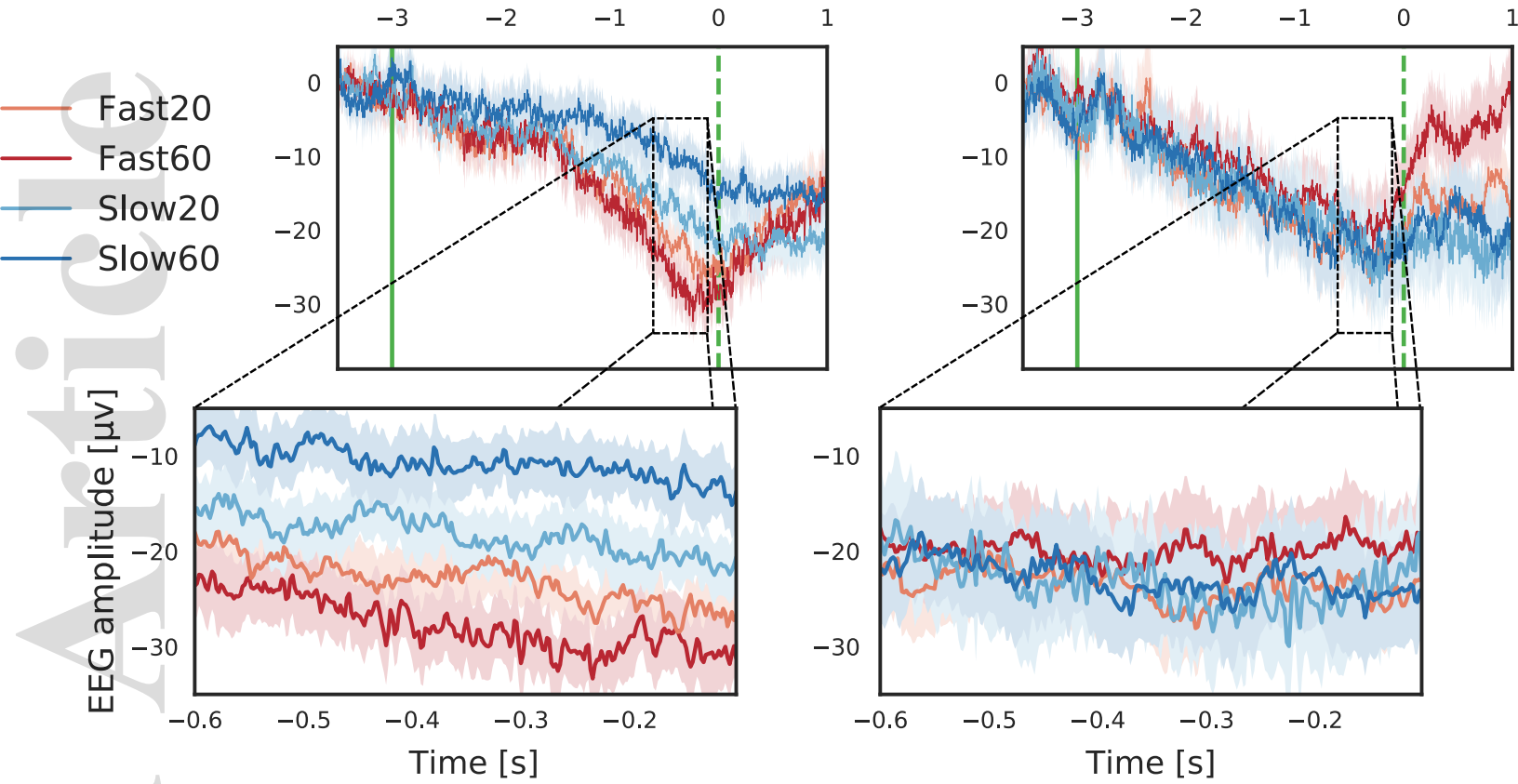
Table 1: ConvNet architecture

| Block | Layer | Output size | Param. # |
|---|---|---|---|
| **1** | Input | $(C_y \times C_x \times T)$ | 0 |
| | Reshape | $(1 \times C_y * C_x \times T)$ | 0 |
| | Conv2D $(1,64) \times F_1$ | $(4 \times C_y * C_x \times T)$ | 256 |
| | Batch normalization | $(4 \times C_y * C_x \times T)$ | 16 |
| | Reshape | $(4 \times C_y \times C_x \times T)$ | 0 |
| | Permute | $(T \times 4 \times C_y \times C_x)$ | 0 |
| | TD $(C_y, C_x) \times D * F_1$ | $(T \times D * F_1 \times 1 \times 1)$ | 160 |
| | Permute | $(D * F_1 \times 1 \times 1 \times T)$ | 0 |
| | Reshape | $(D * F_1 \times 1 \times T)$ | 0 |
| | Batch normalization | $(D * F_1 \times 1 \times T)$ | 32 |
| | Activation (ELU) | $(D * F_1 \times 1 \times T)$ | 0 |
| | AveragePooling2D $(1,4)$ | $(D * F_1 \times 1 \times T/4)$ | 0 |
| | Dropout (.25) | $(D * F_1 \times 1 \times T/4)$ | 0 |
| **2** | SeparableConv2D $(1,16) \times F_2$ | $(F_2 \times 1 \times T/4)$ | 192 |
| | Batch normalization | $(F_2 \times 1 \times T/4)$ | 32 |
| | Activation (ELU) | $(F_2 \times 1 \times T/4)$ | 0 |
| | AveragePooling2D $(1,8)$ | $(F_2 \times 1 \times T/32)$ | 0 |
| | Dropout (.25) | $(F_2 \times 1 \times T/32)$ | 0 |
| | Flatten | $(F_2 * T/32)$ | 0 |
| | Dense | $(N)$ | 228 |
| | Activation (Softmax) | $(N)$ | 0 |
| | **Total** | | **916** |

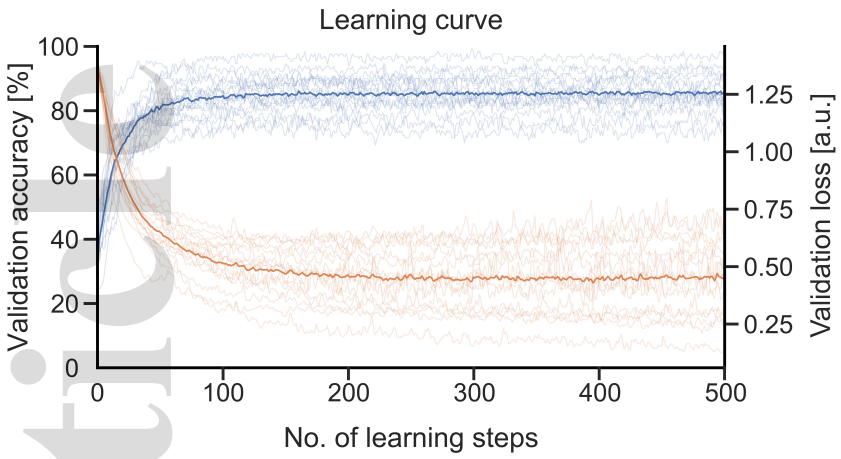
$C_x$ = channels (mediolateral direction), $C_y$ = channels (anteroposterior direction), $T$ = time samples, $F_1$ = number of temporal filters, $TD$ = TimeDistributed (DepthwiseConv2D), $D$ = depth multiplier (number of spatial filters), $F_2$ = number of pointwise filters, $N$ = number of classes.

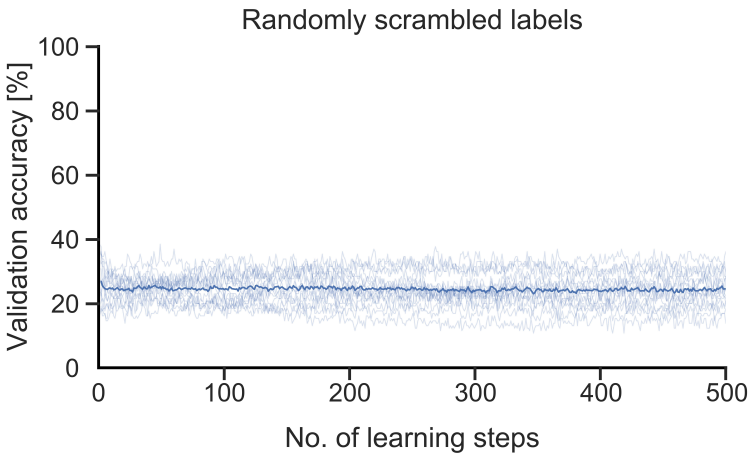


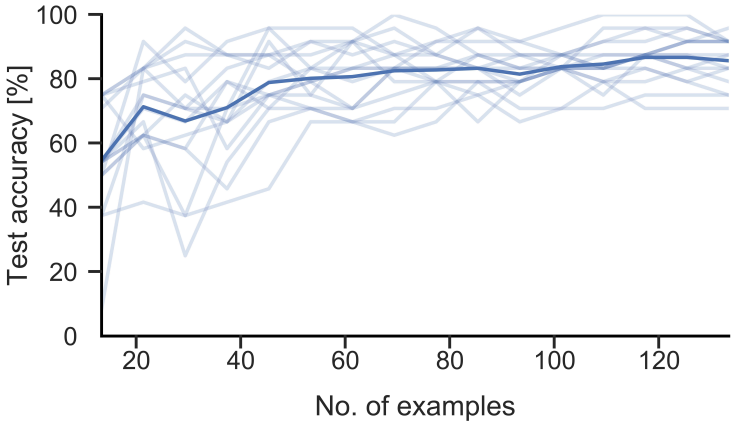

ejn_14936_f1.eps



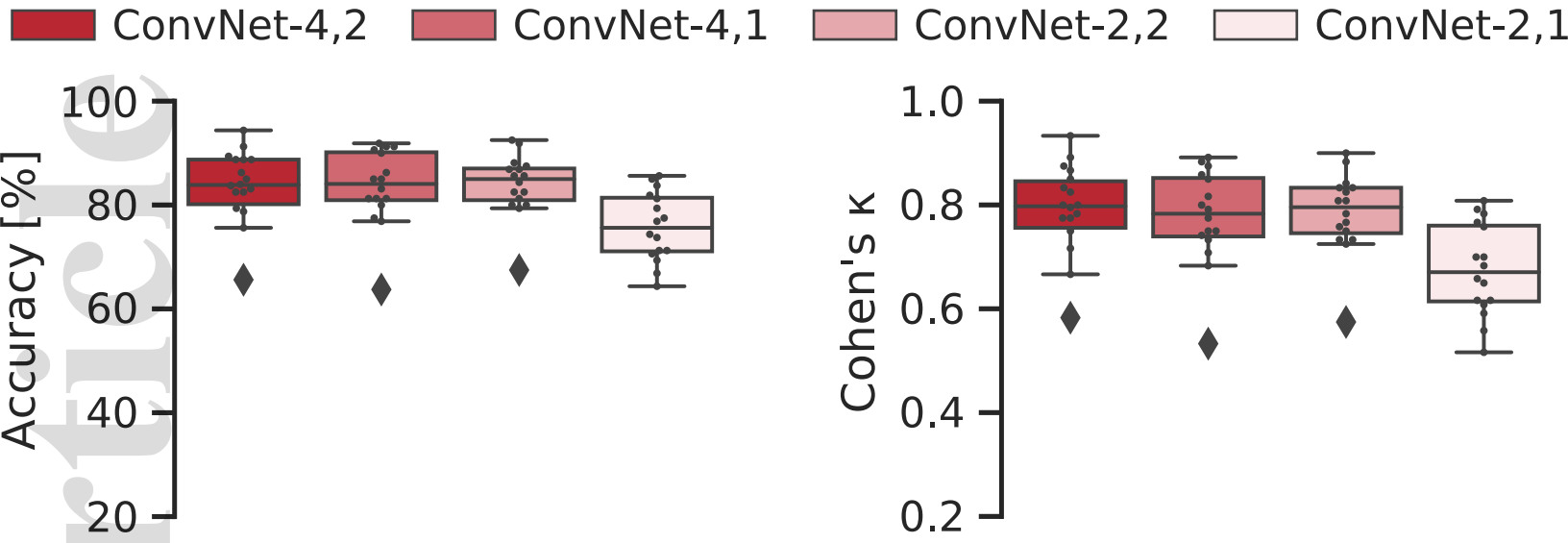

ejn_14936_f2.eps



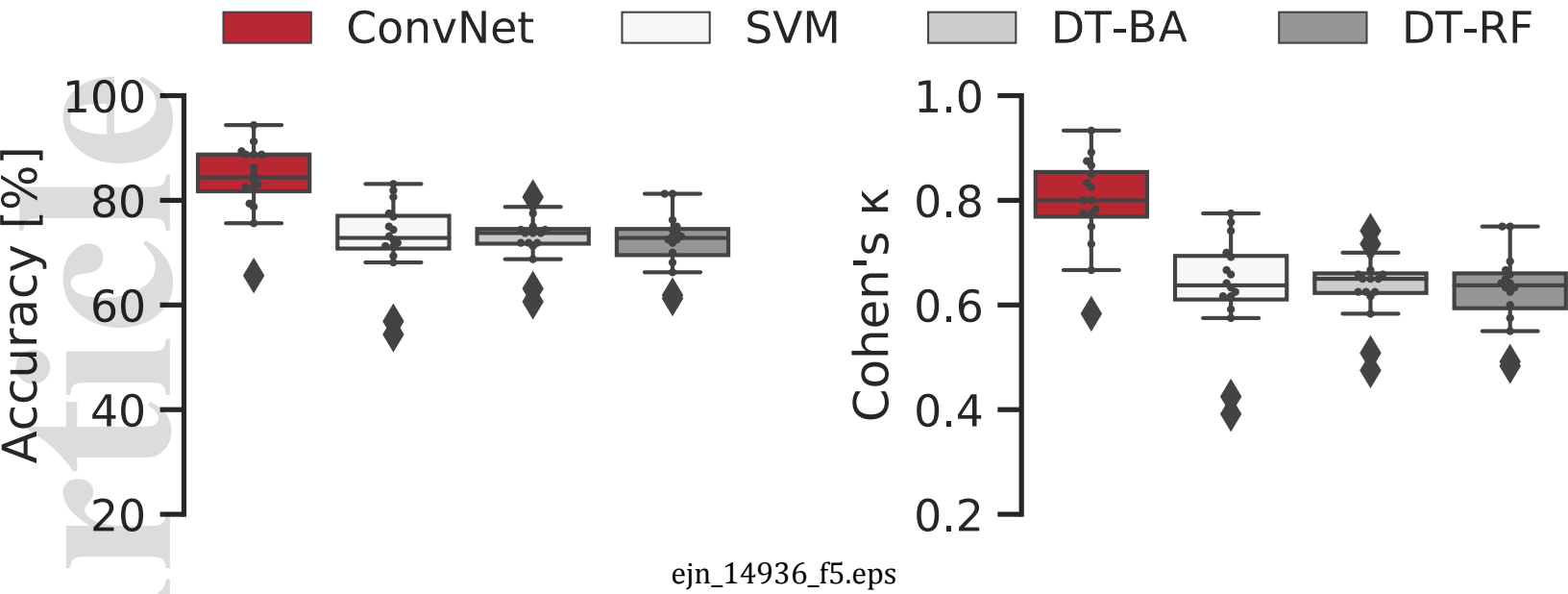



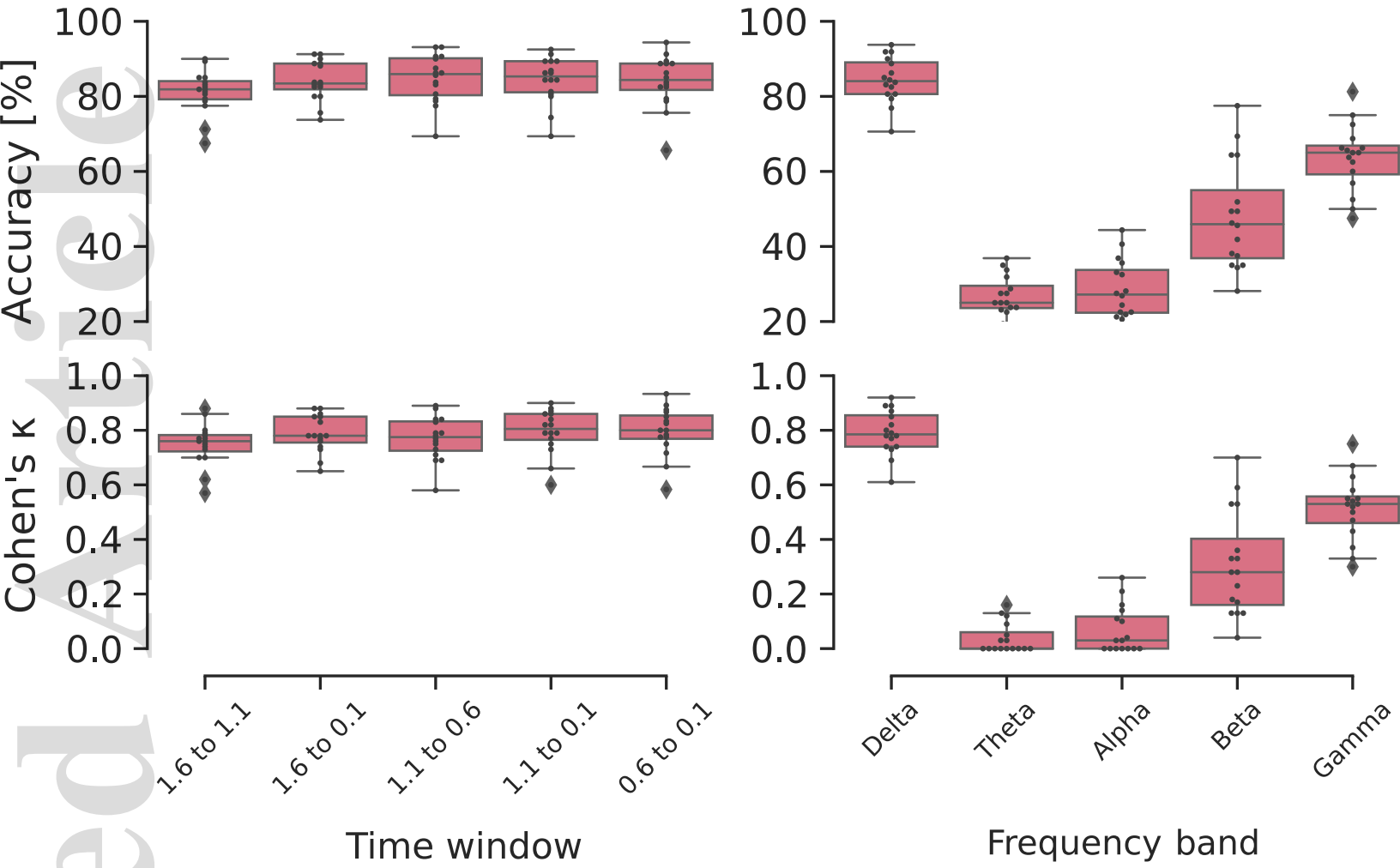

ejn_14936_f3.eps



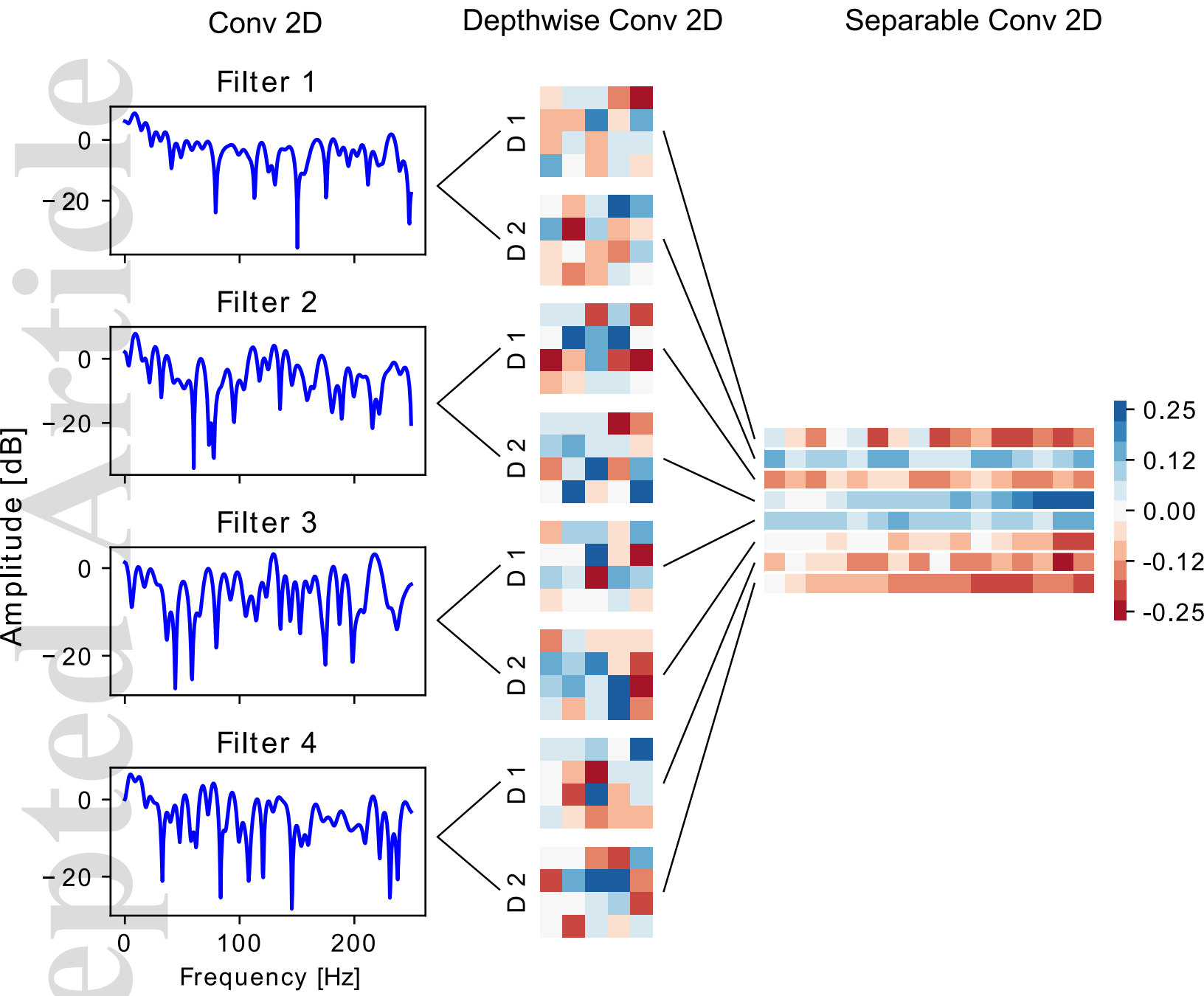

ejn_14936_f4.eps




ejn_14936_f5.eps




ejn_14936_f6.eps




ejn_14936_f7.eps