Computers in Biology and Medicine 172 (2024) 108188

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Convolutional neural networks reveal properties of reach-to-grasp encoding in posterior parietal cortex

Davide Borra [a,*], Matteo Filippini [c], Mauro Ursino [a,b], Patrizia Fattori [b,c], Elisa Magosso [a,b]

[a] Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Cesena Campus, Cesena, 47522, Italy
[b] Alma Mater Research Institute for Human-Centered Artificial Intelligence, University of Bologna, Bologna, 40126, Italy
[c] Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, 40126, Italy

ARTICLE INFO

Keywords:
Explainable artificial intelligence
Convolutional neural networks
Layer-wise relevance propagation
Single-neuron recordings
Macaque
V6A area
Reach-to-grasping

ABSTRACT

Deep neural networks (DNNs) are widely adopted to decode motor states from both non-invasively and invasively recorded neural signals, e.g., for realizing brain-computer interfaces. However, the neurophysiological interpretation of how DNNs make the decision based on the input neural activity is limitedly addressed, especially when applied to invasively recorded data. This reduces decoder reliability and transparency, and prevents the exploitation of decoders to better comprehend motor neural encoding. Here, we adopted an explainable artificial intelligence approach – based on a convolutional neural network and an explanation technique – to reveal spatial and temporal neural properties of reach-to-grasping from single-neuron recordings of the posterior parietal area V6A. The network was able to accurately decode 5 different grip types, and the explanation technique automatically identified the cells and temporal samples that most influenced the network prediction. Grip encoding in V6A neurons already started at movement preparation, peaking during movement execution. A difference was found within V6A: dorsal V6A neurons progressively encoded more for increasingly advanced grips, while ventral V6A neurons for increasingly rudimentary grips, with both subareas following a linear trend between the amount of grip encoding and the level of grip skills. By revealing the elements of the neural activity most relevant for each grip with no a priori assumptions, our approach supports and advances current knowledge about reach-to-grasp encoding in V6A, and it may represent a general tool able to investigate neural correlates of motor or cognitive tasks (e.g., attention and memory tasks) from single-neuron recordings.

1. Introduction

Motor neural decoding consists in translating neural activity into motor behavior or into motor-related external variables (e.g., the shape of a specific object to reach and grasp). Neural decoding is performed by exploiting machine learning approaches and represents a fundamental processing stage for developing Brain-Computer Interfaces (BCIs), utilizing the decoder decision to provide feedback to the user, e.g., for guiding an external device [1]. Crucially, machine learning approaches have gained interest in neuroscience as tools not only able to translate neural activity [2] but also to shed light on the neural features underlying the decoded motor states [3]. Indeed, the knowledge learned by the decoder could be exploited to identify which elements of the input neural activity, in a specific domain of interest (e.g., space or time) are more relevant for driving the decoder's decision, thus obtaining a view in the corresponding domains, of how much information the neural activity contains about the decoded states. This could help to boost our comprehension of how movement properties are encoded in the brain.

The posterior parietal cortex (PPC) hosts areas involved in the sensorimotor processing required to plan actions [4–7], and is known to be crucially implicated in the control/planning of upper limb motor acts, in both non-human and human primates. Several studies have applied decoding techniques to neural signals collected from PPC to infer reaching goals and trajectories, as well as grasping properties, both in non-human primates [8–11] and in human patients [12,13]. Among PPC areas, area V6A is a key node located in the dorsomedial stream, serving as connection between the extrastriate visual area V6 and the superior parietal lobule. V6A exhibits a gradient in terms of cytoarchitecture and functionality. Visual features are predominantly observed in the ventral portion (V6Av), characterized by a cytoarchitecture resembling the occipital cortex and more connected with the occipital extrastriate visual areas (including V6) [14–16]. Sensorimotor features prevail in the

[*] Corresponding author.
E-mail address: davide.borra2@unibo.it (D. Borra).

https://doi.org/10.1016/j.compbiomed.2024.108188
Received 20 October 2023; Received in revised form 26 January 2024; Accepted 18 February 2024
Available online 19 February 2024

dorsal portion (V6Ad), characterized by a cytoarchitecture more similar to the parietal cortex and connected to the parietal and frontal cortex [14,15,17]. V6A was found to encode reaching goals and directions [18–22] in addition to grasp information [23–26], and it was proved that neural signals recorded from this area in non-human primates can be used to reliably decode reaching and grasping properties [27–32].

Deep neural networks (DNNs) have rapidly emerged in recent years as powerful tools for decoding neural signals, both from non-invasive recordings, e.g., electroencephalogram (EEG) acquired in humans [33–38], and from invasive recordings, e.g., activity of single cells in non-human primates [2,27–29,39]. By exploiting raw/lightly pre-processed neural signals, these decoders automatically learn the features that maximize the discriminability among the classes and proved to outperform traditional machine learning approaches, including linear discriminant analysis, support vector machine (SVM), XGBoost, and Naïve Bayes classifiers (see Refs. [2,36,40,41] for benchmarks and reviews). This holds also when decoders are applied to PPC recordings of non-human primates, as we obtained in recent studies (see Refs. [27,29] and in particular [28] for a benchmark study). DNNs are composed by the sequence of many layers of artificial neurons, and learn complex non-linear functions mapping the input multi-variate neural activity to the desired motor output by composing many simple non-linear functions, each learned within each network layer. Different DNNs exist depending on the connections between artificial neurons, such as convolutional neural networks (CNNs), fully-connected neural networks (FCNNs), and recurrent neural networks (RNNs). Crucially, in previous benchmark studies [28,36], CNNs resulted the most accurate DNN approach for decoding both non-invasive and invasive recordings. Because of their automatic feature learning on raw/minimally pre-processed neural signals, DNNs potentially represent good candidates to perform data-driven analyses of the elements of the neural activity that most encode a specific motor state. However, the main limitation of DNNs is that their automatically learned features are difficult to be directly interpreted in neurophysiological terms. This limit has two main negative implications in the field of neural motor decoding. First, the validation of DNNs for prospective BCI applications is lowered. Indeed, the validation would be only limited to decoding performance and not to the knowledge exploited by the decoder to produce a specific decision; thus, it would remain unknown what elements of the input neural activity the DNN focuses on to take the decision. Second, the scarce interpretability prevents to exploit the automatic feature learning of DNNs to identify and analyze the most relevant elements of the neural activity related to movement properties. To overcome this limitation, solutions enabling a neurophysiological interpretation of the DNN decision were proposed. Among these solutions, the most common one consists in coupling the neural network with an 'explanation technique', such as saliency maps [42] or layer-wise relevance propagation (LRP) [43,44], designing an approach of explainable artificial intelligence. Explanation techniques are devoted to explaining the network decision towards one specific decoded condition (e.g., one specific reached and grasped object). By doing so, they identify which elements of the input neural activity (in an interpretable domain e.g., spatial, temporal, frequency domain) mostly drive the network decision towards one specific decoded brain state. The crucial point in using such an approach in neuroscience is that it enables identifying aspects of the input neural signals that are most important for the underlying neural processes (e.g., motor planning and control). In this way, this approach could contribute to the validation and also to future advancements of motor/cognitive theories that functionally relate neural activity to movement properties.

So far, explanation techniques have been applied to non-invasive recordings [34,35,38,45–49], empowering the analysis of the neural activity with respect to traditional analyses (e.g., event-related potential analyses in case of EEG) and supporting motor/cognitive neuroscience with the characterization of novel useful DNN-based biomarkers. Conversely, when dealing with signals from single neurons, DNNs have been used only as 'black boxes', and the potentials of applying explanation techniques for interpreting DNN decision still remain unexploited and unexplored.

The aim of this study is to propose a computational framework, based on a CNN and on an explanation technique, to investigate the encoding of grip properties in area V6A of macaque monkey. In particular, a CNN is used to decode five different grip types from the activity of V6A cells recorded in 2 macaques during a reach-to-grasping task; the five grip types differ as to the degree of grip skills required, ranging from a highly rudimentary grip to a highly precise grip. The explanation technique is used to identify the spatial and temporal samples driving most the decision of the CNN when discriminating the different grips. Specifically, the adopted technique quantifies the impact (or relevance) of each cell (at different locations in space, e.g., located more dorsally or ventrally) and of each time step during the motor task in producing the decoding decision. This serves to explore possible differential contribution of different cells (e.g., dorsal vs. ventral) in encoding different grip types, as well as to evidence how grip encoding evolves along time. The proposed framework is based on successful methodologies previously adopted for non-invasive recordings (EEG), here transposed to single-neuron recordings. Specifically, the CNN is inspired by the design proposed by Lawhern et al. [33], and LRP [43,44] is used to explain network decision. We expected that the proposed framework, by highlighting the spatial and temporal samples in the input data that are most discriminative for decoding grip types, could not only support the current knowledge about reach-to-grasp information encoded in PPC, but also extend it by providing a more refined view of the temporal and spatial organization of reach-to-grasp encoding.

2. Materials and methods

2.1. Data description and pre-processing

The data used in this study were obtained in a previous study [32]. The study was performed in accordance with the guidelines of EU Directives (86/609/EEC; 2010/63/EU) and Italian national law (D.L. 116-92, D.L. 26–2014) for the care and use of animals for scientific purposes. Experimental protocols have been approved by the Ethical Committee of the University of Bologna, by the Animal Welfare Body of the University of Bologna, and by the Italian Ministry of Health.

Single-neuron activity was recorded extracellularly from the posterior parietal area V6A of two male Macaca fascicularis monkeys (monkey 1 and 2), see Fig. 1a. Animals were trained to perform reach-to-grasping movements toward an object with the arm contralateral to the recorded hemisphere. Specifically, the activity from 93 cells and from 75 cells was recorded in monkey 1 and monkey 2, respectively. Depending on their location, V6A cells were assigned to the ventral (V6Av) or dorsal (V6Ad) sector, by identifying these sectors as in Luppino et al. [15]. Specifically, 59/53 cells (monkey 1/monkey 2) fell in V6Av, while 34/22 cells fell in V6Ad (see right panel of Fig. 1a).

During recordings, monkeys sat on a primate chair with their head restrained in front of a rotating panel hosting one object to reach and grasp. Five objects with different shapes were used, presented to the monkey one at a time and evoking grip types with different hand configurations (see Fig. 1b): a ball (c_0: whole-hand prehension), handle (c_1: finger prehension), ring (c_2: hook grip), plate (c_3: primitive precision grip), stick-in-groove (c_4: advanced precision grip). The five grip postures differed in the level of coordination required: from more rudimentary grips involving the whole hand or all fingers (c_0, c_1) to more precise grips involving the index finger only (c_2), or fingers-thumb opposition (c_3) or index finger-thumb opposition (c_4) [25,26]. The presentation order of the objects was randomized. Animals performed 10 trials per object, thus resulting in 50 trials for each monkey and neuron.

Each trial was divided into different phases ('epochs'), which are described in the following and are illustrated in Fig. 1c. The trial begun when the animal pressed a 'home button' near to its chest in complete

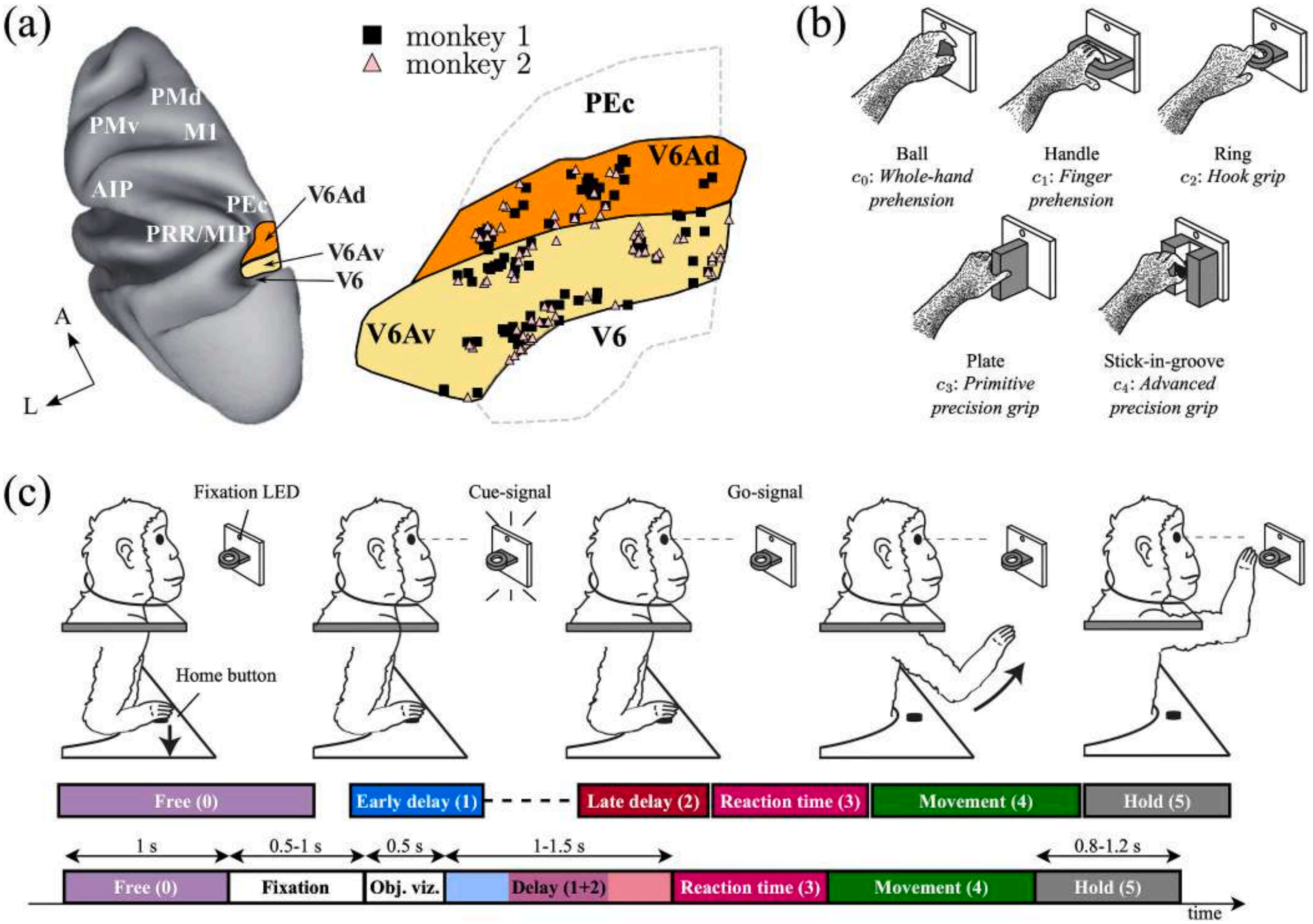


Fig. 1. Recording area and experimental paradigm. *Panel a* – 3D dorsal view of the left hemisphere of macaque brain (left, the shown directions are A = anterior, L = lateral) with the recording area highlighted (V6Av in yellow and V6Ad in orange), and 2D cortex projection of the medial parieto-occipital region with recording sites (right), as done in Refs. [14,50]. In the figure, besides area V6A, other areas are also marked: the ventral (PMv) and dorsal (PMd) premotor cortices, parietal reach region (PRR)/medial intraparietal area (MIP), anterior intraparietal area (AIP), primary motor cortex (M1), and PEc. *Panel b* – Objects and evoked grip types (reported in italic). These represent the motor conditions (i.e., classes $c_k, 0 \leq k \leq 4$) that were discriminated by the decoder. *Panel c* – Trial structure. On top, the epochs defining each trial are illustrated. On the bottom, epoch timings are displayed.

darkness. The monkey waited 1 s (*free epoch*, epoch 0) until the fixation LED mounted on top of the panel turned on (green). After a fixation period of 0.5–1.0 s (*fixation epoch*, random interval), in which the monkey had to maintain the fixation on the LED without performing any movement, the LEDs surrounding the object turned on, illuminating it (from that time on, the object remained illuminated until the end of the trial). Starting from the illumination of the object, an interval of 0.5 s (*object visualization epoch*) and a subsequent interval of 1–1.5 s (*delay epoch*, random interval) elapsed, during which the animal kept maintaining the fixation while attending the object, waiting for the go-signal. The delay epoch was divided into the *early delay epoch* (epoch 1, 1s-interval after the start of the delay epoch) and *late delay epoch* (epoch 2, 1s-interval before the end of the delay epoch). Finally, the fixation LED changed its color (from green to red), providing the go-signal to the animal. The monkey started the reach-to-grasping movement after a short reaction time; thus, the *reaction time epoch* and *movement epoch* can be identified (epochs 3 and 4, respectively). Once performed the movement, the animal kept holding (*hold epoch*, epoch 5) the object until the fixation LED switched off (0.8–1.2 s, random interval); this cued the monkey to release the object and to press the home button again, starting a new trial.

During recordings of each trial and each neuron, action potentials (spikes) were isolated and sampled at 100 kHz. These were initially binned within a window of 5 ms and were then re-binned to cope for inter-trial and inter-neuron variability in epoch duration, by using a window length such that epochs had the same number of time samples (i.e., bins) across trials and neurons. Firing rates were computed from the re-binned activity, and the activity during the free, early delay, late delay, reaction time, movement, and hold epochs was considered in this study. Firing rates obtained during the t-th trial of one monkey are denoted in this study by $X_t \in \mathbb{R}^{N \times T}$, $0 \leq t \leq N_t - 1$, where N is the number of recorded neurons, T is the number of time samples in the trial, t is the trial index, $N_t = 50$ is the total number of trials.

2.2. Framework for decoding single-neuron activity via neural networks and for explaining network decision

This section describes the methodologies adopted to decode neural activity via neural networks and to analyze the neural activity most relevant for decoding the grip types. At a high level, the presented methodologies respond to the following two needs. First, we were interested in finding the relationship that maps small portions of neural activity (i.e., chunks) of the observed neuron population (i.e., V6A neurons in this case) to the object that the monkey reached and grasped (among 5 possible objects), thus, solving a 5-class classification problem (*neural decoding*). Then, this relationship, once found, was exploited to derive useful insights about the encoding of grip properties during the motor task from the recorded neuron population (*explanation of network* *decision*). Fig. 2 schematizes the operations performed by the described framework.

Neural decoding was performed on sliding windows of firing rates over the trial course, hereafter referred as 'chunks' of neural activity. In this approach, overlapped chunks $X_{t,i} \in \mathbb{R}^{N\times T_z}$ were extracted with window size of $T_z = 60 \equiv 300\ ms$ and a stride of T_s time samples, (see the 'sliding window decoding' process in Fig. 2). This last parameter was coarser during training ($T_s\ = 10 \equiv 50\ ms$), to speed up network learning and was finer during inference ($T_s\ = 1 \equiv 5\ ms$), to provide inference with the highest time resolution for all possible chunks. The chunk of neural activity ($X_{t,i}$) represents the input of the decoder: it is a 2D feature map corresponding to the activities of all cells within the considered sliding window, with cells along the rows and time samples along the columns. The label $y_{t,i}$ associated to each chunk sampled from a trial was the label associated to that trial ($y_{t,i} \equiv y_t$), i.e., one of the $N_c = 5$ possible grip types. This represents the desired label that the network should reproduce as output. We have:

$$\begin{cases} X_{t,i} \in \mathbb{R}^{N\times T_z} \\ y_{t,i} \equiv y_t \in C = \{c_k\}, 0 \leq k \leq N_c - 1 \end{cases}, 0 \leq t \leq N_t - 1, 0 \leq i \leq M - 1, \quad (1)$$

where i is the chunk index, k is the class index, and $M = (T - T_z)/\ T_s + 1$ is the total number of chunks extracted per trial. Therefore, the dataset recorded from each monkey can be represented by the set:

$$D = \left\{ \left(X_{0,0}, y_{0,0}\right), \ldots, \left(X_{t,i}, y_{t,i}\right), \ldots, \left(X_{N_t-1,M-1}, y_{N_t-1,M-1}\right) \right\}, 0 \leq t \leq N_t - 1, 0 \leq i \leq M - 1 \quad (2)$$

The stage of neural decoding consists in training the decoder (here realized by a CNN) on a labelled training set (corresponding to a partition of D) to discriminate the grip types from each small chunk of neural activity (training stage), realizing the function $f(X_{t,i};\vartheta) : \mathbb{R}^{N\times T_z} \rightarrow C$ parametrized in the parameters ϑ (trainable parameters) to fit the training set. Once the decoder is trained, it encapsulates in ϑ the knowledge needed to optimally discriminate between the contrasted motor states and it is then tested (inference stage) on a held-out test set (corresponding to a partition of D, different from the training set). In our analyses, while decoders were trained using epochs from early delay epoch to hold epoch (the interval in which the animal performed the task), during inference these were tested also on the free epoch, representing a control interval in which the animal was not engaged in the task. The following Section 2.3 is devoted to the description of the CNN adopted for neural decoding.

It is worth remarking that the use of a sliding window decoding approach has several benefits [27,28]:

i. Fast decoding. The decoder is forced to produce a prediction based on a few hundred of milliseconds of neural activity (here 300 ms, chunk-level decoding) instead of producing a prediction once the entire trial (lasting a few seconds) is processed (i.e., trial-level decoding).
ii. Data augmentation. Sliding window decoding is equivalent to performing data augmentation via slicing and is commonly adopted when dealing with neural time series [51].
iii. Analysis of temporal dynamics of neural encoding. The discriminability of the contrasted conditions (e.g., grip types), as quantified by a performance metric (e.g., accuracy), can be analyzed as a function of time over the trial course.

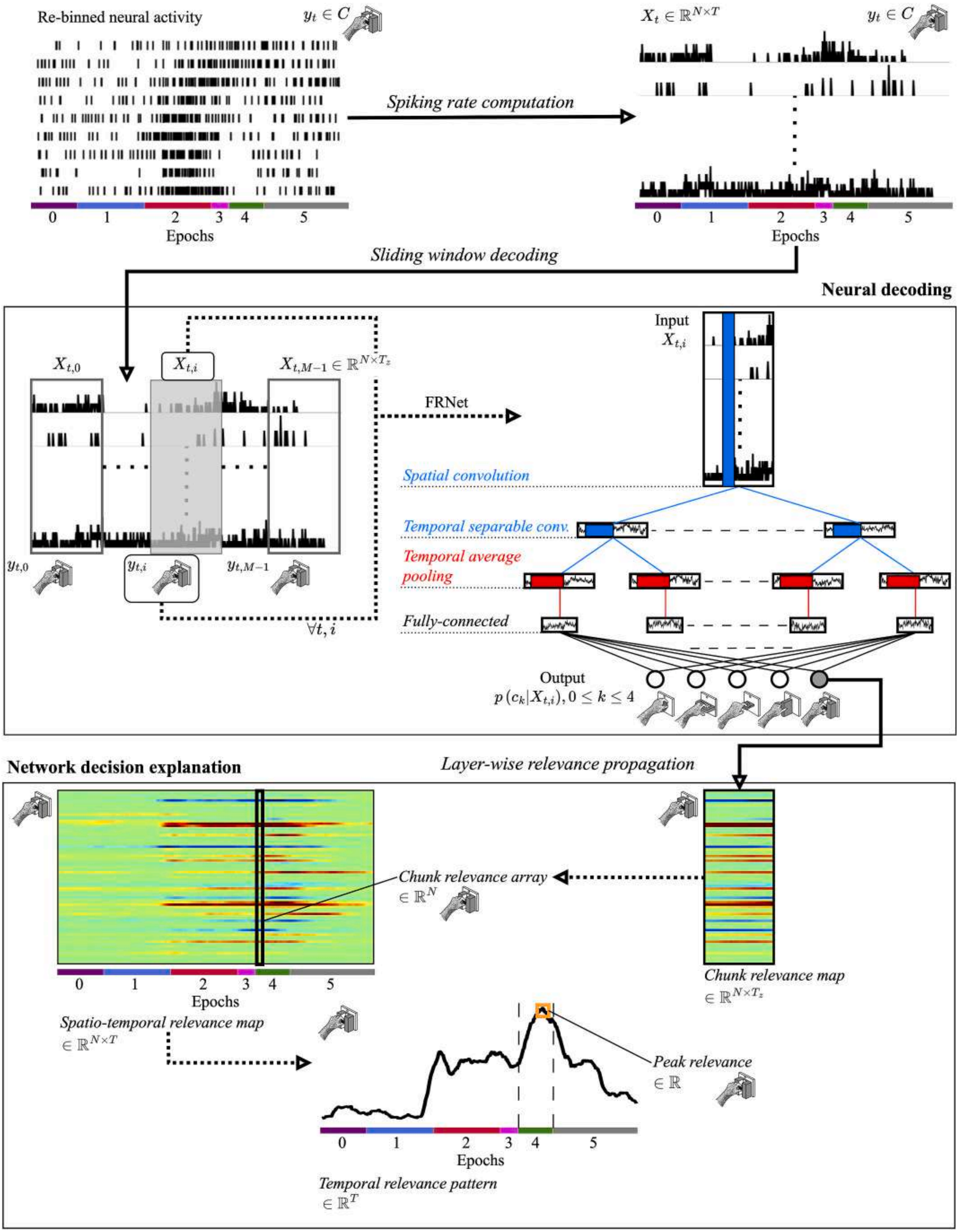
The stage of explanation of network decision was based on the use of an explanation technique (LRP) to identify the most relevant cells and temporal samples within the input $X_{t,i}$ driving the decision of the trained network towards one of the five grip types. By doing so, a relevance representation associated to each tested input chunk $X_{t,i}$ was obtained $g(X_{t,i}) : \mathbb{R}^{N\times T_z} \rightarrow \mathbb{R}^{N\times T_z}$. Then, these relevance maps, appropriately processed, enabled to analyze the temporal and spatial properties of grip encoding across area V6A. Section 2.4 describes the network decision explanation.

2.3. *Neural decoding via FiringRateNet (FRNet)*

Here we used a CNN to realize the function $f(X_{t,i};\vartheta)$ (i.e., to perform neural decoding), since in our previous study [28] this type of decoder resulted the best compared to both traditional decoders (e.g., XGBoost, Support Vector Machine, Naïve Bayes) and other DNNs (e.g., fully-connected and recurrent neural networks) on a variety of single-neuron motor datasets, including the reach-to-grasping dataset used in this study.

The CNN structure adopted here, named FiringRateNet (FRNet), was inspired from the successful design proposed by Lawhern et al. [33] for decoding EEG signals. Here, spatial and temporal convolutions were performed in two separate convolutional layers and not in a mixed way within one mixed spatio-temporal convolutional layer, as in the CNN used in our previous benchmark study [28]. Furthermore, here convolutions were performed using a specialized layer (separable convolutional layer [52]) that reduces the number of parameters and limits the network size, an expedient useful for contrasting overfit of small datasets. A schematic representation of the adopted CNN is reported in Fig. 2 ('neural decoding' part). We adopted this design as it represents one of the most successful CNNs for EEG decoding, providing high decoding performance (e.g., it also won an international neural decoding challenge, see Ref. [36]), and also because it is characterized by a peculiar compact design, introducing a limited number of parameters to fit (here 2485, on average across monkey-specific decoders, see later for further details about network training). Furthermore, the network proposed by Lawhern et al. [33] was applied on neural time series other than EEG, e.g., magnetoencephalography [53], thus on time series with different nature (reflecting magnetic fields vs. electric potential). Therefore, while in our previous study [28] a CNN topology as general as possible was considered, with the aim of comparing different DNN families, here we used a topology inspired from a previous CNN known to decode other neural time series with high accuracy and in an efficient way (low number of parameter to fit) [33]. See also Section 4 for further comments on this point.

In the following, we provide a summarized description of the network; more details about the network structure and its hyperparameters (that is, the parameters that define the functional form of the decoder, e.g., number of convolutional filters) can be found in Section 2.3.1. The code for defining FRNet is available at https://github.com/ddavidebb/macaque-single-neuron-decoding.git. First, the network performed spatial convolution on the input chunk of neural activity, $X_{t,i}$, learning 16 spatial filters, each describing how to combine the information across all cells. Then, the network performed temporal separable convolution with 16 temporal filters learning temporal patterns within approximately 100 ms (as in Ref. [28]). All the previous units were activated via Rectified Linear Units (ReLUs). Lastly, feature maps were downsampled 10-times along the time-axis by applying an average moving window (average pooling), and were provided as input to the output layer (fully-connected layer) with $N_c = 5$ output units (one per class). Output units were activated via a softmax activation function, producing as output the probability that the input chunk $X_{t,i}$ contained neural activity linked to a specific grip type, i.e., $p(c_k|X_{t,i}), 0 \leq k \leq N_c -$



(caption on next page)

Fig. 2. Proposed framework to decode reach-to-grasping from single-neurons using a CNN and to investigate the most relevant spatial and temporal samples for different grip types, by applying an explanation technique to the CNN. Once computed the firing rates from the neural activity recorded in each trial, we applied an explainable artificial intelligence framework, composed by two main stages. In a first stage (*neural decoding*), 300 ms-length sliding windows (i.e., chunks) of neural activity were decoded within each trial by using a CNN (named FRNet). The CNN accepted as input a 2D matrix ($X_{t,i} \in \mathbb{R}^{N\times T_s}$) containing the activity of all cells within the considered i-th chunk (reporting cells by rows and time samples by columns) of the t-th trial, and provided as output the probability that the input chunk belonged to each grip type ($p(c_k|X_{t,i}), 0 \le k \le 4$). For brevity, only the main layers, i.e., convolutional, fully-connected, and pooling layers, are displayed. Boxes contain layer outputs, and the internal colored rectangles represents convolutional (blue) and pooling (red) filters. Connections of convolutional layers, pooling layers, and fully-connected layers are colored in blue, red, and black, respectively. See Section 2.3 and 2.3.1 for further details. In a second stage (*network decision explanation*), the relevance of each input cell and time samples (belonging to the input chunk) for decoding the associated grip type was derived (obtaining a chunk-level relevance map with the same shape of the input chunk, i.e., $\in \mathbb{R}^{N\times T_s}$). This representation was termed chunk relevance map. Then, by aggregating this information across the chunks composing the trial, we also obtained the relevance of the cells in all time points composing the entire trial (obtaining a trial-level relevance map, with the same shape of the trial, i.e., $\in \mathbb{R}^{N\times T}$). This representation was termed spatio-temporal relevance map. By averaging the spatio-temporal relevance map across cells, a temporal relevance pattern was derived ($\in \mathbb{R}^{T}$), resuming the relevance in the temporal domain only. Similarly, ventral/dorsal and grip-sensitive/not grip-sensitive temporal patterns were derived (not shown in the figure for brevity), by averaging the spatio-temporal relevance map across a selection of cells (e.g., only V6Av cells), and not across the totality of cells. Lastly, the peak relevance value (maximum in the figure) within the movement epoch (epoch 4) was extracted from the temporal relevance pattern. All previous relevance representations were derived specifically for each decoded grip type (e.g., advanced precision grip in the figure). See Section 2.4 and 2.4.1 for further details.

Table 1
Data description.

Property	Set	Value (monkey 1/monkey 2)
No. of decoded classes (N_c)		5
No. of recorded cells (N)		93/75
Epochs	train. and valid.	[1,2,3,4,5]
	test	[0,1,2,3,4,5]
No. of time steps (T)	train. and valid.	812/816
	test	1013/1017
No. of examples	train.	3040/3040
	valid.	380/380
	test	4770/4790

1. To increase the network generalization, batch normalization [54] was performed (immediately after each convolutional layer), and dropout [55] was applied (immediately before temporal convolution and before output layer).

Networks were trained by using the cross-entropy as loss function and Adam as optimizer (500 training epochs, learning rate of 1e-3 and mini-batch size of 64). Network training and evaluation was performed separately for each monkey (realizing monkey-specific networks). This is commonly adopted in previous studies [27–32], in order to deal with the differences across monkeys in the recording setups (e.g., different positions and number of resulting recording sites due to different microelectrode penetrations, see Fig. 1a), and with the cross-animal variability in the neural activity, as these aspects hinder cross-animal model evaluations. The high inter-participant variability is also known to affect neural decoding in human, requiring the adoption of participant-specific models for achieving high decoding performance (e.g., for BCI applications) [45,56,57]. Using monkey-specific networks might introduce a polarization of results towards each considered animal. However, this did not represent a drawback in our study, but rather an intentional choice. Indeed, in line with past studies analyzing neuronal firing rates separately for each monkey (via traditional statistical tools, i.e., not deep learning-based approaches) [14,23–26], here we were specifically interested into designing a framework for performing still an analysis at the single monkey level (i.e., monkey-specific), but using a deep learning approach. In particular, the focus here is the exploitation of the monkey-specific features learned by the neural networks, to derive useful insights about neural correlates and encoding of reach-to-grasping in a data-driven way. A subject-specific approach, being tuned on each subject, usually provides higher classification performance than other approaches (e.g., cross-subject) [45,56,57], with consequent higher reliability of the results deriving from the analysis of the features learned by the networks.

The dataset of each animal was split according to a 10-fold cross-validation scheme to perform network training (on the training set) and performance evaluation (on the test set). For each monkey and each cross-validation fold, a different neural network was trained, evaluated in terms of performance, and analyzed for highlighting the most relevant elements of the input neural activity. Specifically, for each cross-validation fold, the monkey-specific dataset (consisting of 50 trials) was split into 45 trials for generating the training set and 5 trials for generating the test set. A 10% of training trials (i.e., 5 trials) was held back from the training set for designing the validation set, that was exploited to define the number of training epochs. Indeed, for each cross-validation fold the model was trained until it maximized the performance on the validation set (early stopping). Considering the sliding window approach, the examples forming the training set, validation set, and test set were the chunks extracted from the corresponding trials. In particular, based on the cross-validation procedure, and on the sliding window size and stride, for each cross-validation fold, 380 examples (i.e., chunks) derived from 5 trials (one per grip type) were used in the validation set, 3040 examples derived from 40 trials (8 trials per grip type) were used in the training set, and 4780 examples derived from the remaining 5 trials (one per grip type) were used in the test set, on average across monkeys. Even though test examples were extracted from less trials than training examples, test examples were more abundant than training examples also due to the use of $T_s = 1$ while testing networks (i.e., providing predictions at each time step), as explained in Section 2.2. It is worth remarking that we did not perform trial-level decoding (not feasible using deep neural networks in this case, due to the small dataset size) but rather we performed chunk-level decoding (sliding window decoding). Considering the details presented here and in Section 2.2, this corresponds to perform data augmentation via slicing, augmenting training data 76-times (from trial-level decoding to chunk-level decoding). Dataset properties are resumed in Table 1.

For each cross-validation fold, each monkey-specific network was tested on the corresponding held-out test set. Specifically, for each trial, predictions were provided chunk by chunk by the network; thus, we computed the accuracy over time within the trial course. Furthermore, we also computed the confusion matrix within the trial epoch with highest performance, that is, the movement epoch (epoch 4); this was computed by considering the predictions of chunks falling into the movement epoch. Therefore, as 10 networks (one per cross-validation fold) were trained for each monkey, during model evaluation 10 patterns of accuracy over time and 10 confusion matrices for each monkey were computed (each relative to the held-out test set of each fold). Note that, the performance results presented in this study (see Section 3) always refer to the held-out test set (across the 10 cross-validation folds).

Finally, to provide a comparison of the CNN adopted here with respect to the state-of-the-art, we compared the performance of FRNet with the mixed spatio-temporal CNN that was proposed in our previous benchmark study [28] (see also Section 2.3.2). This mixed spatio-temporal CNN is a CNN composed by only 1 convolutional hidden layer that learns features both in the spatial and temporal domains ('mixed' spatio-temporal feature learning), and introduces 61285 parameters to fit (on average, across monkeys). We selected this decoder as state-of-the-art decoder since in our previous study [28], it significantly outperformed others machine learning and deep learning approaches (specifically, XGBoost, SVM, Naïve Bayes, RNNs, FCNNs), on motor datasets involving both reaching and reach-to-grasping, and also on the same dataset adopted in this study.

It is crucial to remark that readers can find in our past study [28] an extensive benchmark on motor decoding, including reach-to-grasping with the same dataset adopted here. Therefore, in this study we compare FRNet only with the top-performing state-of-the-art algorithm as resulting from Ref. [28], while we focus mainly on the explainable approach to identify the temporal and spatial samples most guiding the decision. The code for designing the state-of-the-art CNN used as reference decoder is available online at https://github.com/ddavidebb/macaque-single-neuron-decoding.git.

Neural networks were developed in Python (version 3.8.5) with PyTorch (version 1.9.0) [58] and network decisions were explained with Captum (version 0.5.0) [59], using a workstation equipped with an AMD Threadripper 1900X, NVIDIA TITAN V and 48 GB of RAM.

2.3.1. Details about FRNet

The details of FRNet are presented in the following and are summarized in Table 2. In FRNet, the input layer simply replicates the input neural activity in a single feature map; thus, the output shape of this layer is (1,N,T_z)=(1,93/75,60). Then, the first convolutional layer performs 2D convolution in the spatial domain using $K_0 = 16$ filters with size $F_0 = (93/75, 1)$ depending on the monkey (i.e., depending on the number of recorded cells N), unitary stride and no padding. Neuron activations were then normalized via batch normalization [54], and passed through a ReLU non-linearity. Lastly, neuron activations were dropped out during training using a dropout probability of $p = 0.5$ [55]. The second convolutional layer performs 2D separable convolution [52] in the temporal domain. This convolution is composed by a first depthwise temporal convolution and a second pointwise convolution. Depthwise temporal convolution learns a set of $D_1 = 1$ temporal filters for each spatially filtered version of the input ($K_1 = K_0 \bullet D_1 = 16$ in total) with size $F_1 = (1, 21)$, unitary stride, and zero-padding such that the layer output shape matches its input shape, i.e., $P_1 = (0, 10)$. Then, pointwise convolution learns how to optimally recombine the 16 feature maps provided by depthwise convolution. Neuron activations were normalized via batch normalization, passed through a ReLU non-linearity, and downsampled 10-times in time by using an average pooling layer with pooling size $F_p = (1, 10)$ and pooling stride $S_p = (1, 10)$. Then, neuron activations were dropped out during training using a dropout probability of $= 0.5$, flattened into a 1D array and provided as input to the fully-connected layer with $N_c = 5$ neurons, providing the class scores as output. Finally, class scores were converted into the conditional probabilities by using the softmax activation function. Network hyper-parameters (e.g., number of convolutional filters, convolutional filter size, etc.) were set via empirical evaluations.

Table 2
Details of FRNet. Each layer is provided with its name, main hyper-parameters, number of parameters to fit and output shape. Where not specified, stride (S) and padding (P) were set to $(1, 1)$ and $(0, 0)$, respectively. Values are reported for both monkey 1 and monkey 2, separated by a forward dash symbol. The total number of parameters to fit was 2629/2341 (monkey 1/monkey 2).

Layer name	Hyper-parameters	No. of tr. parameters	Output shape
Input		0	(1,93/75,60)
Conv2D	$K_0 = 16$, $F_0 = (93/75, 1)$	1488/1200	(16, 1, 60)
BatchNorm2D		32	(16, 1, 60)
ReLU		0	(16, 1, 60)
Dropout	$p = 0.5$	0	(16, 1, 60)
Separable-Conv2D	$D_1 = 1$, $K_1 = K_0 \bullet D_1 = 16$, $F_1 = (1,21)$, $P_1 = (0,10)$	592	(16, 1, 60)
BatchNorm2D		32	(16, 1, 60)
ReLU		0	(16, 1, 60)
AvgPool2D	$F_p = (1,10)$, $S_p = (1,10)$	0	(16, 1, 6)
Dropout	$p = 0.5$	0	(16, 1, 6)
Flatten		0	(96)
Fully-Connected	$N_c = 5$	485	(5)
Softmax		0	(5)
		2629/2341	

2.3.2. Details about the state-of-the-art decoder

The reference state-of-the-art decoder adopted in this study consisted of a shallow CNN proposed in Borra et al. [28], and its details are summarized in the following (see Ref. [28] for further details), while its main parameters are reported in Table 3. In this network, after the input layer, a mixed spatio-temporal convolutional layer was used, learning $K_0 = 32$ filters with size $F_0 = (93/75, 21)$ and applying zero-padding such that the layer output shape matches its input shape, i.e., $P_0 = (0, 10)$. This corresponded to learning filters within approximately 100 ms of temporal window for all cells in a mixed way in the spatio-temporal domain (i.e., without disentangling spatial and temporal contributions). Neurons were activated via an Exponential Linear Unit (ELU) activation function [60], and neuron activations were pooled using an average pooling layer, halving the dimension of feature maps in the temporal domain, i.e., with pooling size $F_p = (1, 2)$ and pooling stride $S_p = (1, 2)$. Dropout was applied with a dropout probability $p = 0.5$ [55]. Lastly, as for FRNet (see Section 2.3.1), neuron activations were flattened into a 1D array and provided as input to the fully-connected layer with $N_c = 5$ neurons, activated via softmax activation function.

2.4. Network decision explanation via layer-wise relevance propagation

Layer-wise relevance propagation (LRP) [43,44] was used as explanation technique to realize the function $g(X_{t,i})$ (i.e., to explain network decision). Successful applications of LRP to neural time series are reported in the literature, e.g., for explaining DNNs applied to EEG [34,48,49]. LRP is a backward propagation technique that propagates one network class score of interest (one grip type, e.g., whole-hand prehension) – consisting of the activation of the output layer immediately before the softmax function – back to the input layer (replicating the input chunk of neural activity), by exploiting propagation rules applied locally at each layer of the network. Theoretical details of LRP are

Table 3
Details of the mixed spatio-temporal CNN proposed in Borra et al. [28]. Each layer is provided with its name, main hyper-parameters, number of parameters to fit and output shape. Where not specified, stride (S) and padding (P) were set to $(1, 1)$ and $(0, 0)$, respectively. The total number of parameters to fit was 67333/55237 (monkey 1/monkey 2).

Layer name	Hyper-parameters	No. of tr. parameters	Output shape
Input		0	(1,93/75,60)
Conv2D	$K_0 = 32$, $F_0 = (93/75, 21)$, $P_0 = (0,10)$	62528/50432	(32,1,60)
ELU		0	(32,1,60)
AvgPool2D	$F_p = (1,2)$, $S_p = (1,2)$	0	(32,1,30)
Dropout	$p = 0.5$	0	(32,1,30)
Flatten		0	(960)
Fully-Connected	$N_c = 5$	4805	(5)
Softmax		0	(5)
		67333/55237	

described in Section 2.4.1. By doing so, given an input chunk $X_{t,i}$, LRP produces a representation containing one *relevance* value per spatio-temporal sample of that input chunk, thus, overall producing a relevance 2D map with the same dimension as the input. This representation – termed in this study as *chunk relevance map* ($\in \mathbb{R}^{N\times T_z}$) – quantifies how much each spatio-temporal sample of the input chunk $X_{t,i}$ contributed to the prediction of a grip type under analysis. Relevance values are not bounded in their values and can be both positive and negative, associated to positive and negative evidence for the model prediction. That is, the relevance quantifies to what extent each spatio-temporal sample contributed positively to the classification result (increasing the output class score, thus improving classification) or contributed negatively to the classification result (decreasing the output class score, thus worsening classification) [43]. Due to their unbounded nature, to understand whether relevance values are (significantly) relevant or not, they should be compared with the null relevance value (relevance equal to 0, see points iii., iv., and vi. in Section 2.5 for the statistical analyses conducted to this aim).

Here we applied LRP on each trained decoder (i.e., for each monkey and cross-validation fold) using the test examples (i.e., test chunks) as input and the score of the associated grip as target class score (e.g., the class score of the whole-hand prehension in case of test chunks extracted from whole-hand prehension trials). Then, for each monkey and each cross-validation fold, by considering a specific grip type, the following processing on the chunk relevance map was performed (see also the schematization reported in Fig. 2 in the 'explaining network decision' part).

i. From chunk relevance map (chunk-level) to spatio-temporal relevance map (trial-level). Each chunk relevance map (i.e., chunk-level representation) associated to the grip type under analysis was averaged across time samples within the chunk, obtaining a *chunk relevance array* ($\in \mathbb{R}^N$), summarizing the relevance for each cell within the sliding window. By concatenating together, chunk-by-chunk, the chunk relevance arrays belonging to the same trial, a *spatio-temporal relevance map* ($\in \mathbb{R}^{N\times T}$) was obtained (i.e., trial-level representation), highlighting the importance of each cell and time sample for discriminating the grip type across the entire trial duration.

ii. From spatio-temporal relevance map to temporal relevance patterns. The spatio-temporal relevance map obtained in the previous processing point was averaged across all cells, thus obtaining *a temporal relevance pattern* ($\in \mathbb{R}^T$). In addition, we also averaged the spatio-temporal relevance map separately across cells belonging to the dorsal and ventral sectors of V6A (see Fig. 1a) or separately across cells that resulted modulated (grip-sensitive) or not (not grip-sensitive) during the movement epoch (epoch 4), obtaining *dorsal and ventral temporal relevance patterns* ($\in \mathbb{R}^\mathrm{T}$), and *grip-sensitive and not grip-sensitive temporal relevance patterns* ($\in \mathbb{R}^\mathrm{T}$). Here, a cell was classified as grip-sensitive if its average discharge frequency during movement epoch (epoch 4), for the movement condition evoking the highest discharge, was significantly different compared to the free epoch (epoch 0), similarly to the analyses performed in previous studies [24,61]. Wilcoxon signed-rank tests (one for each neuron, separately for each monkey) were performed, and the Benjamini-Hochberg procedure was applied for correcting p-values for multiple tests [62]. From this procedure, 65.6% and 74.7% of cells were classified as grip-sensitive, respectively for monkey 1 and monkey 2.

iii. From temporal relevance to peak relevance values. By separately considering the dorsal/ventral and grip-sensitive/not grip-sensitive temporal relevance patterns obtained from the previous processing point, we derived the peak relevance value (i.e., the maximum or minimum relevance value) inside the movement epoch (epoch 4), obtaining the *peak relevance* ($\in \mathbb{R}$). This scalar value summarized the relevance, separately of V6Av and V6Ad cells, or of grip-sensitive and not grip-sensitive cells, for the grip type under analysis during the interval of movement execution (including the phase of transport of the hand, change of wrist orientation, and finger pre-shaping into the appropriate grip). We selected the movement epoch (epoch 4) to extract the peak relevance, as this interval was the one associated with highest performance (see Section 3), and as it was also considered in a previous study [14] when investigating the functional spatial segregation of V6A during reach-to-grasping and in other studies [24,61] when classifying grip-sensitive cells based on their nature.

At the end of these processing steps, overall (across cross-validation folds), each trial of each monkey-specific dataset (50 trials in total) was associated to its own spatio-temporal relevance map, temporal relevance patterns and peak relevance values, derived from the 10 trained monkey-specific networks.

In order to complete and enrich the validation, we also validated the framework from a computational perspective, by analyzing how the model accuracy changes when using only a subset of cells among the most relevant cells or among the least relevant cells during model training and testing, rather than using all cells. To this aim, we considered the peak relevance values across all cells, and we sorted cells from the least relevant to the most relevant. Then, we trained and tested the model using a subset of cells (with 5 or 15 or 25 cells) composed of the least or most relevant cells. Furthermore, we included also a third experimental condition (control condition), in which the model was trained and tested using a randomly selected subset of cells (i.e., without sorting cells by relevance). For this analysis, we considered the average decoding accuracy in the interval of movement execution (i.e., epoch 4), as this interval resulted the one associated with highest performance (see Section 3).

2.4.1. Details about layer-wise relevance propagation

Let us denote with o_k the class score of the k-th class ($0 \le k \le 4$). Layer-wise relevance propagation [43] propagates the prediction of the network, represented by the class score o_k (e.g., the predicted score associated by the network to the advanced precision grip, o_4), backward in the network. To do so, propagation rules must be defined for each layer of the network. Let m and n be the indices of two neurons of two consecutive layers ($l-1$ and l) of the network and let $R_n^{(l)}$ be the relevance for the neuron n of the layer l in predicting o_k. The backward propagation of the relevance at a given layer back to a preceding layer of the network is achieved by applying the rule:

$$R_m^{(l-1)} = \sum_n \frac{z_{mn}}{\sum_m z_{mn}} R_n^{(l)}, \tag{3}$$

where z_{mn} weights how much the neuron m contributed to make the neuron n relevant, and the denominator $\sum_m z_{mn}$ forces the conservation of the relevance during the propagation. Indeed, the conservation is ensured locally by $\sum_m R_m^{(l-1)} = \sum_n R_n^{(l)}$, and thus, globally throughout the network, as $\sum_i R_i^{(0)} = \ldots = \sum_n R_n^{(l)} = \ldots = o_k$.

The propagation rule applied in this study is the LRP-ε rule [44], as previously done in Ref. [34] when using LRP with EEG data:

$$R_m^{(l-1)} = \sum_n \frac{a_m w_{mn}}{\varepsilon + \sum_m a_m w_{mn}} R_n^{(l)}, \tag{4}$$

where the term a_m denotes the activation of the neuron m, w_{mn} denotes the weight of the connection from unit m to unit n, and ε is a small positive term that ensures that $R_m^{(l-1)}$ is bounded for small or null values of neuron activations in the denominator $\sum_m a_m w_{mn}$. Compared to other ERP rules, LRP-ε rule reduces noise (high fidelity in the obtained representations) by absorbing weak or contradictory contributions of neurons to the relevance (e.g., with respect to LRP-0 rule) [44]. Furthermore, LRP-ε treats also negative contributions of neurons and not only positive contributions (e.g., with respect to LRP-γ rule) [44].

2.5. Statistical analyses

For each monkey-specific decoder, the following statistical analyses were conducted on the performance and relevance measures.

i. Comparison of the decoding accuracy over time scored by FRNet against the chance level (0.2). A permutation *t*-test with tmax correction (5000 iterations) [63] was performed.
ii. Comparisons of each entry of the confusion matrix scored by the state-of-the-art decoder against the same entry of the confusion matrix scored by FRNet. Pairwise Wilcoxon signed-rank tests were performed and the Benjamini-Hochberg correction [62] was applied to correct for multiple tests (25 tests).
iii. Comparisons of the temporal relevance patterns (averaged across grip types) against the null relevance (0). A permutation *t*-test with tmax correction (5000 iterations) [63] was performed.
iv. Comparisons of the peak relevance (averaged across grip types) of grip-sensitive cells and of not grip-sensitive cells against the null relevance (0). Pairwise Wilcoxon signed-rank tests were performed and the Benjamini-Hochberg correction [62] was applied to correct for multiple tests (2 tests).
v. Comparison of the peak relevance (averaged across grip types) of grip-sensitive cells against not grip-sensitive cells. A pairwise Wilcoxon signed-rank test was performed.
vi. Comparison of the peak relevance of V6Av cells and of V6Ad cells against the null relevance (0), separately for each grip type. Pairwise Wilcoxon signed-rank tests were performed and the Benjamini-Hochberg correction [62] was applied to correct for multiple tests (10 tests).
vii. Comparisons of peak relevance of V6Av cells against V6Ad cells, separately for each grip type. Pairwise Wilcoxon signed-rank tests were performed and the Benjamini-Hochberg correction [62] was applied to correct for multiple tests (5 tests).
viii. Correlation analysis between peak relevance and the degree of grip skills required in the reach-to-grasp movement (increasing from $k = 0$ to $k = 4$, related to the c_k grip type), separately for each sector (V6Av and V6Ad). The Pearson correlation coefficient was computed over the values of peak relevance obtained for each grip type and each cross-validation fold (10 folds and 5 grip types, resulting in 50 data points in total).
ix. Comparison of the decoding accuracy between the different types of subsets used when training and testing the model only with a small subset of cells. Accuracies were compared between subsets formed by the most relevant cells vs. the least relevant cells vs. randomly selected cells. For each number of cells defining the subset (3 total subset sizes, i.e., 5, 15, 25 cells), pairwise Wilcoxon signed-rank tests were performed considering all possible combinations across the subset types, i.e., random vs. most relevant, random vs. least relevant, most relevant vs. least relevant. The Benjamini-Hochberg correction [62] was applied to correct for multiple tests (9 tests).

3. Results

The CNN adopted here is first validated in terms of decoding capabilities; then, the results of the analysis on the most relevant spatial and temporal samples driving the network decision are presented.

As concerning the network decoding capabilities, Fig. 3 reports the accuracy scored by FRNet while decoding the 5 different grip types. The decoder was able to accurately discriminate between grip types significantly above the chance level from the delay epoch (ramp up in accuracy between epochs 1 and 2), reaching the maximum of accuracy at the end of the movement epoch (epoch 4), and returned at the chance level at approximately half of the hold epoch (ramp down in accuracy at epoch 5). As expected, the decoder performed at the chance level within the time interval in which the animal was not engaged in the task (epoch 0, free epoch).

Fig. 4 displays the confusion matrix scored by the state-of-the-art decoder (panel a) and by FRNet (panel b) within epoch 4, representing the time interval with the highest decoding accuracy. Notably, no significant differences ($p > 0.05$) were observed between the confusion matrices scored by FRNet and by the state-of-the-art decoder across the two monkey-specific decoders. However, it is worth remarking that FRNet was slightly more accurate (but not statistically significant) than the state-of-the-art when decoding monkey 2 for some motor conditions, with accuracies (FRNet vs. state-of-the-art) of 0.8 vs. 0.73, 0.97 vs. 0.92, and 0.98 vs. 0.91, respectively in decoding whole-hand prehension,

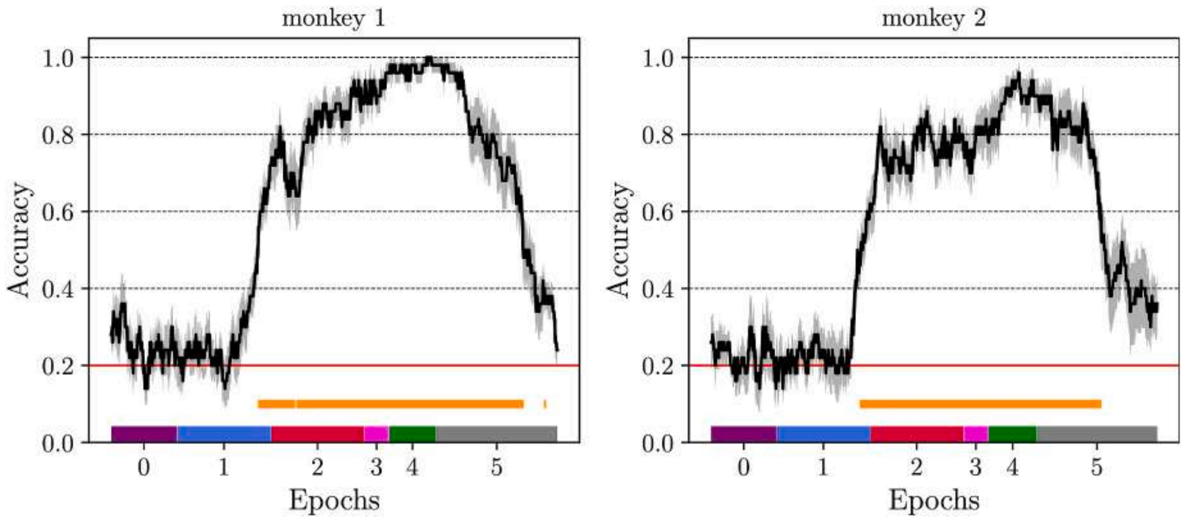


Fig. 3. Decoding accuracy over time scored by FRNet separately for monkey 1 and monkey 2. The thick black line represents the mean value of decoding accuracy (computed on the test set) across the ten cross-validation folds and the grey overlayed area represents the standard error of the mean. Orange stripes denote time samples at which the network classified the grip types significantly ($p < 0.05$) above the chance level (0.2, identified by the horizontal red line). The epochs outlining the time sequence of the task are color-coded as: purple: free; blue: early delay; red: late delay; magenta: reaction time; green: movement; grey: hold.

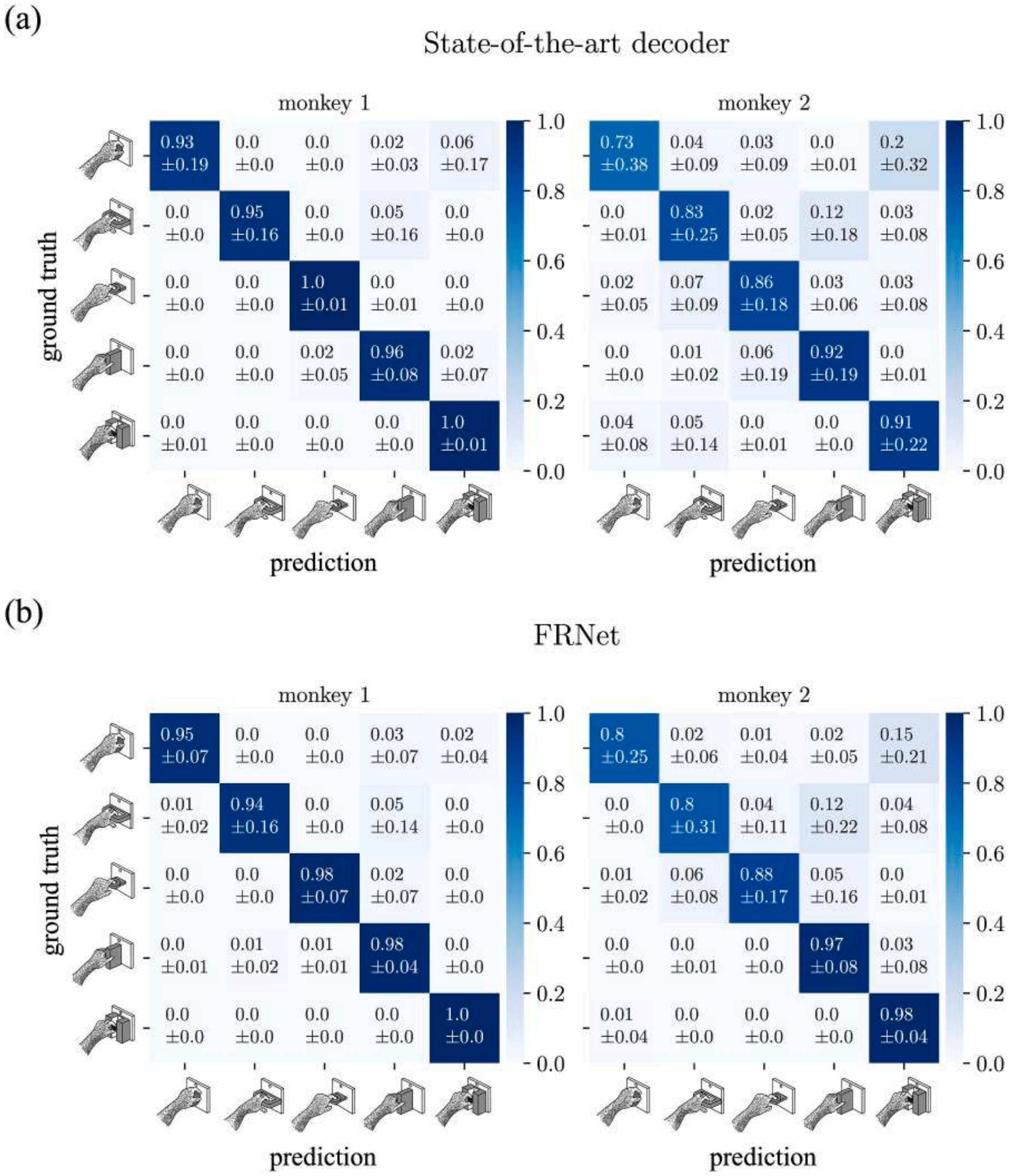


Fig. 4. Confusion matrices scored by the state-of-the-art decoder (panel a) and by FRNet (panel b) during movement execution (i.e., epoch 4), separately for monkey 1 and monkey 2. Matrices were computed on examples belonging to the test set. The state-of-the-art decoder used here was the mixed spatio-temporal CNN that scored the best performance across a wide set of machine learning and deep learning approaches in Borra et al. [28], also on the reach-to-grasping dataset adopted in this study. The (i, j)-th entry of the confusion matrix contains the ratio between the number of examples belonging to class i and predicted of class j, and the total number of examples belonging to class i. This ratio is displayed as mean ± standard deviation across the ten cross-validation folds.

primitive precision grip and advanced precision grip (on average across cross-validation folds). Thus, the adopted CNN was comparable to the state-of-the-art or even slightly more accurate (for some grip types in monkey 2) than the state-of-the-art, while being a CNN structure 25-times more parsimonious in terms of parameters to fit, i.e., 2485 (FRNet) vs. 61285 (state-of-the-art) parameters, on average across monkeys.

Remarkably, in this study we aimed not only at decoding reach-to-grasping with a DNN, but also at detecting the spatial and temporal samples most relevant for each grip type decoding by using an explainable artificial intelligence approach, and the related results are presented in the following.

Spatio-temporal relevance maps over the trial course, averaged across grip types, are displayed in Fig. 5a as heatmaps. The relevance for discriminating grips resulted strongly modulated across cells, with a group of cells in each monkey clearly appearing not relevant (i.e., having relevance value close to 0). This result was related to the inclusion of the entire set of recorded cells into the neural activity given as input to the CNN, as done in Refs. [27,28], without removing cells not modulated by movement. Indeed, by analyzing the peak relevance aggregated across cells that resulted grip-sensitive or not (reported in Fig. 5b), the former resulted significantly relevant, while the latter was

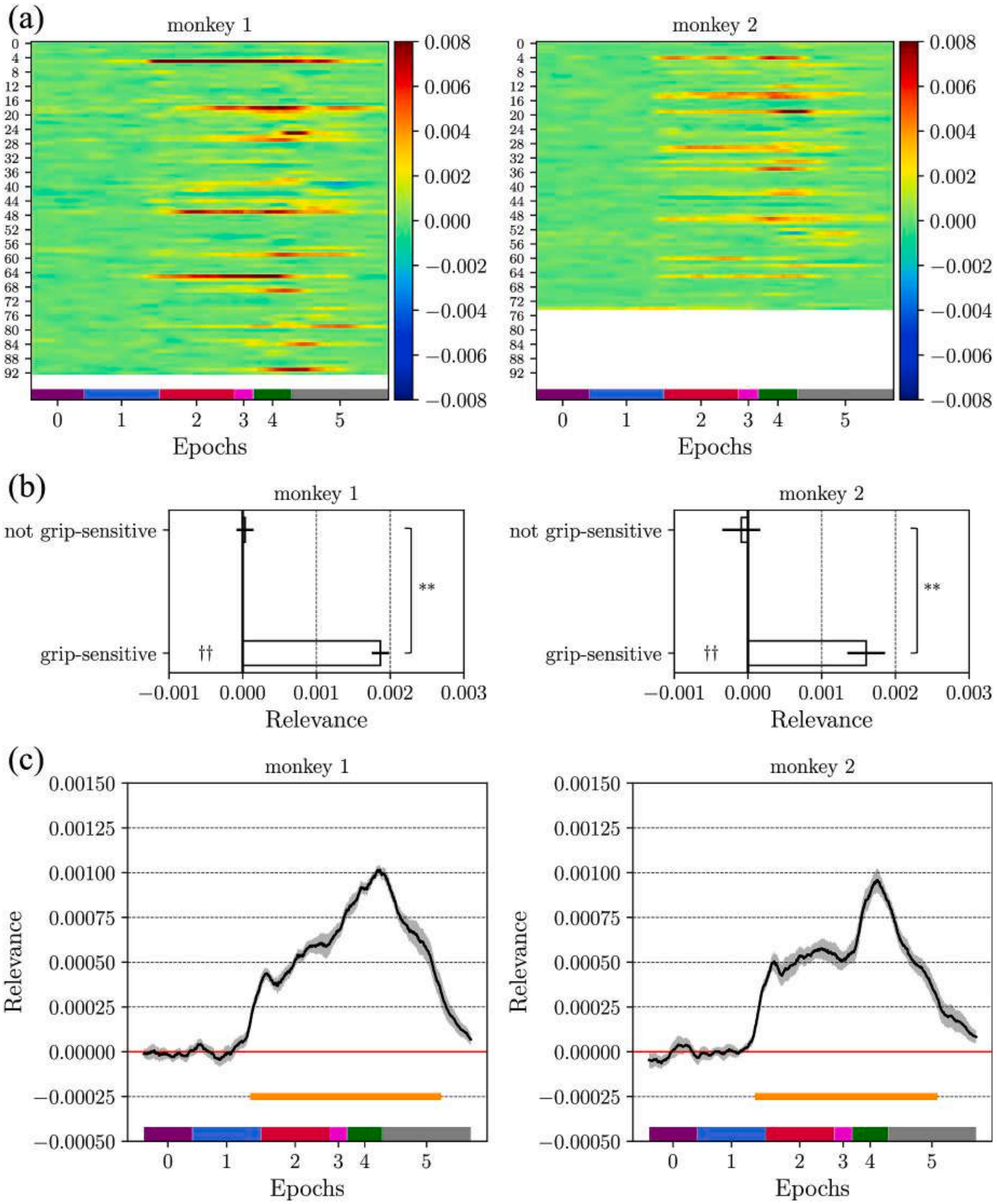


Fig. 5. Spatio-temporal relevance maps (panel a), peak relevance of grip-sensitive and not grip-sensitive cells (panel b), and temporal relevance patterns (panel c), separately for monkey 1 and monkey 2. Panel a. For each monkey, spatio-temporal relevance maps were averaged across grip types and across cross-validation folds. This resulted into one average spatio-temporal relevance map for each monkey, and it was displayed as an heatmap with cells reported along rows and temporal samples across columns. Panel b. The peak relevance of grip sensitive cells and not grip sensitive cells was averaged across grip types. Grip-sensitive cells were 65.6% and 75.7% of the total population, respectively for monkey 1 and monkey 2. Bar heights denote the mean value and error bars represent the standard deviation across cross-validation folds. Results from the performed pairwise comparisons are reported. Distributions with relevance values significantly different compared to the null value (0) are marked on the left of each barplot († $p < 0.05$, †† $p < 0.01$, ††† $p < 0.001$). Relevance significantly different between grip-sensitive and not grip-sensitive cells is marked on the right of each barplot (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Panel c. For each monkey, temporal relevance patterns were averaged across grip types and the mean value (thick line) and standard error of the mean (overlayed area) across cross-validation folds are displayed. Orange stripes denote time samples at which the network attributed a relevance value significantly ($p < 0.05$) different from the null relevance value (0, identified by the horizontal red line). For both panels a and c, the epochs outlining the time sequence of the task are color-coded as: purple: free; blue: early delay; red: late delay; magenta: reaction time; green: movement; grey: hold.

not; furthermore, grip-sensitive cells were significantly more relevant than not grip-sensitive cells. Thus, despite the network was fed with neural signals including 25–35% of cells not modulated by reach-to-grasping (as 65.6% and 74.7% of cells were grip-sensitive), it was able to exploit at most the useful task-related information contained in data (high and significant relevance for grip-sensitive cells), filtering out task-unrelated information (null relevance for not grip-sensitive cells).

The temporal relevance patterns are reported in Fig. 5c, obtained by averaging the spatio-temporal relevance maps across all cells. In accordance with the results in Fig. 3 (temporal dynamics of decoding accuracy), the time steps within the free epoch (epoch 0) resulted not relevant, while time steps from the delay epoch (between epochs 1 and 2) up to approximately half of the hold epoch (epoch 5) resulted significantly relevant, with the movement epoch (epoch 4) showing maximum relevance. However, it is worth noticing that the time pattern of relevance exhibits a sharper peak in epoch 4 (Fig. 5c), compared to time pattern of decoding accuracy (Fig. 3).

In Fig. 6, the peak relevance aggregated across cells falling in V6Av and V6Ad is reported separately for each grip type. In most of the cases, the relevance resulted significantly different from the null value (0), except for the relevance of V6Av cells for hook grip in monkey 1 and primitive precision grip in monkey 2. More interestingly, the relevance attributed by the network to the different grip types was modulated depending on the cytoarchitectonic sector of V6A, i.e., depending on the spatial location within V6A of the decoded cells. Indeed, cells belonging to V6Av were more relevant than the ones belonging to V6Ad for rudimental grips (whole-hand prehension and finger prehension, consistently across monkeys); conversely, V6Ad cells resulted more relevant than V6Av cells for skilled grips (hook grip significantly for one monkey, and primitive and advanced precision grips consistently across monkeys). Furthermore, the relevance for V6Av cells was significantly and negatively correlated to the degree of grip skills, while the relevance for V6Ad cells resulted significantly and positively correlated to the degree of grip skills (with strong and very strong correlations, for monkey 1 and monkey 2 respectively).

Fig. 7 reports the results of the analysis conducted on the decoding accuracy within the movement interval (epoch 4) when using a small subset of cells instead of the whole neuron populations, that is, the most relevant cells, the least relevant cells, and randomly selected cells (control condition). As expected, the most relevant cells conveyed most of the grip-discriminative information for the network. Indeed, from Fig. 7 the highest decoding accuracies were achieved when using only these cells (for almost all the subset sizes tested), when compared to the least relevant cells and to randomly selected cells. On the other hand, the lowest decoding accuracies were scored when using the least relevant cells (for almost all the subset sizes tested), confirming that these cells conveyed less information for discriminating among different grip types. Furthermore, as the number of cells used for decoding increases, the accuracy appeared to increase faster for the randomly selected cells and the most relevant cells, while increased slower for the least relevant cells, with accuracies (on average) upper bounded at values < 0.5.

4. Discussion

This study proposes an explainable artificial intelligence framework based on a CNN aimed at decoding reach-to-grasping from V6A single-neuron recordings with high decoding capabilities, and also at deriving and analyzing spatial and temporal properties of reach-to-grasp encoding in area V6A, based on the features learned by the CNN trained to discriminate among different grip types. To the best knowledge of the authors, this is the first time that an explanation technique is applied to provide a neurophysiological interpretation of the DNN decision when using single-neuron recordings as input. As such, this study contributes: i) to not only further validate the performance of DNNs when decoding single-cell activity (as in past studies [2,27–29,39]), but also to uncover and inspect DNN decision, to corroborate the use of appropriate evidence for prediction. This can be of value to trust more network decisions for future practical applications (e.g., online decoding in BCIs), without using networks only as ‘black boxes’; ii) to propose an approach that can automatically disclose the input neural activity mostly implicated in encoding the predicted motor states at the level of single cells, by leveraging the knowledge learned by a neural network free to explore the entire information contained in the neural activity.

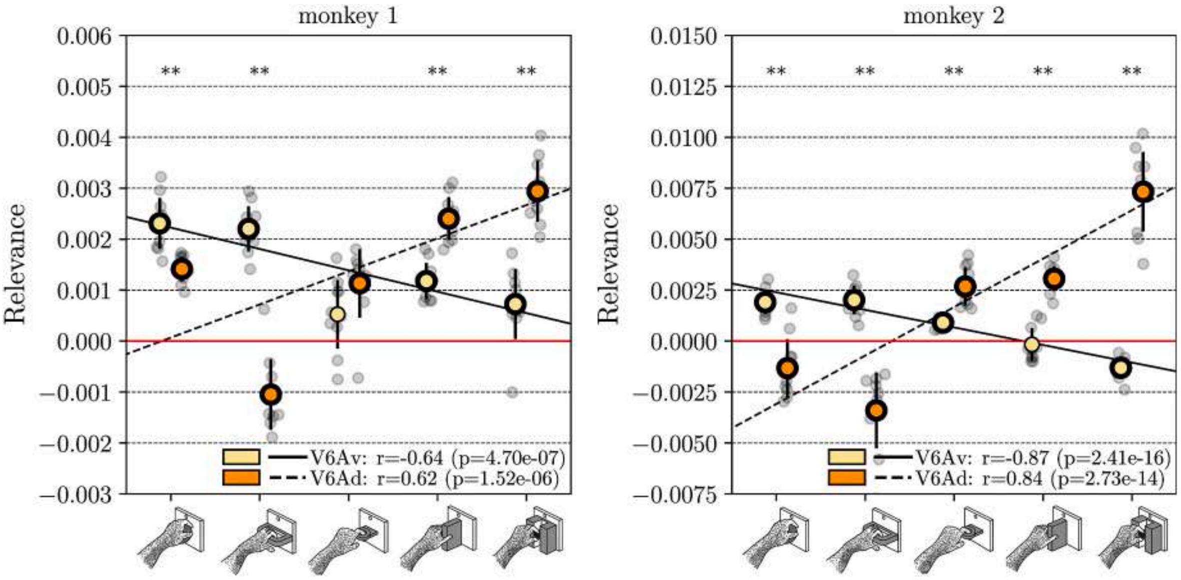


Fig. 6. Peak relevance of V6Av cells and V6Ad cells, separately for monkey 1 and monkey 2 and for each grip type. For each distribution displayed, smaller black dots denote the single relevance observations scored in each cross-validation fold, while the bigger colored dot (yellow: V6Av, orange: V6Ad) denotes the mean value across folds. Error bars represent the standard deviation across folds. Results from the performed pairwise comparisons are reported. Distributions with relevance values significantly different compared to the null value (0, identified by the horizontal red line) are marked by bigger dots with a thicker edge. In addition, grip types with relevance values significantly different between V6Av and V6Ad are marked on top of each panel (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Furthermore, results from the correlation analysis are reported too. Specifically, regression lines are displayed in the figure by means of black continuous (V6Av) and dashed (V6Ad) lines, while the Pearson correlation coefficient (r) and the p-value obtained from the correlation analysis are reported in the legend.

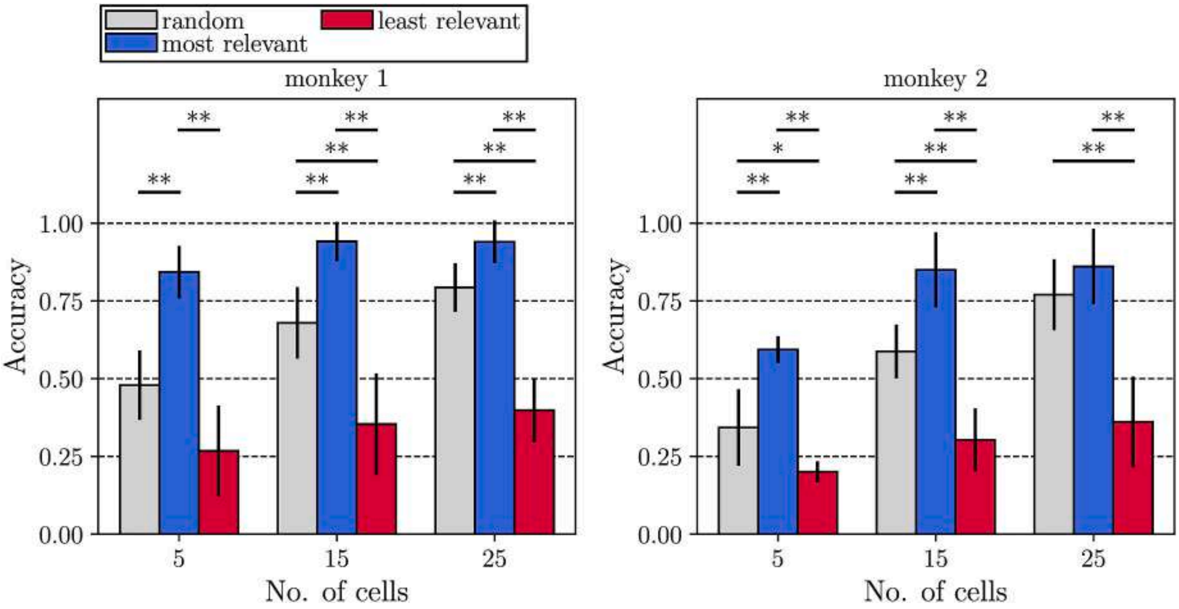


Fig. 7. Decoding accuracy when using a subset of cells, separately for monkey 1 and monkey 2. The accuracy was computed when training and testing models using a subset of 5, 15, 25 cells, and was averaged in the interval with the highest accuracy as obtained from Fig. 3 (movement epoch, epoch 4). Three types of subsets were considered: randomly selected cells ('random' condition), the most relevant cells ('most relevant' condition), the least relevant cells ('least relevant' condition). Bar heights denote the mean value and error bars represent the standard deviation across cross-validation folds. Results from the performed pairwise comparisons are reported (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

From the analysis of the decoding capabilities of the adopted CNN (FRNet), the first important result is that the network was able to accurately decode (well above the chance level) the different grip types. This indicates that the network learned some features from the data able to discriminate between the contrasted motor states, and thus, that the network is suited for the application of a post-hoc explanation technique for gaining a neurophysiological understanding about the elements of the input neural activity the CNN relies on for taking its decision. The second notable result is that the adopted network, inspired by a successful network design for decoding non-invasive recordings (e.g., EEG) [33], performed on par with the state-of-the-art decoder of single-neuron motor activity, as represented by the mixed spatio-temporal CNN proposed in our previous benchmark study [28]. Remarkably, the CNN adopted here represents a better compromise between performance and model size, as it introduces a much less number of parameters to fit compared with the state-of-the-art [28]. Thus, the present study further supports the use of CNNs as high-performing decoders for invasively recorded neural time series and highlights that high accuracies are not constrained to a specific CNN configuration, but rather is common to different designs, even highly parsimonious and even transposed from different recording modalities. This also provides the general indication that successful methodologies adopted for decoding neural signals with different nature and invasiveness (e.g., scalp electric potentials or magnetic fields), may be effectively used for decoding single-neuron recordings as well. Therefore, future studies may benefit also from existing decoders proposed for other recording modalities, reducing the need to design from scratch new neural networks depending on the specific application.

The main novel point of this study is the analysis of the relevance the network attributes to the input neural activity in the spatial and temporal domains for decoding grip types. Regarding the most relevant neural activity in the temporal domain, the relevance temporal dynamics (with a sudden increase at the end of epoch 1, then a plateau and a clear peak in epoch 4, Fig. 5c) reflected the discharge dynamics of grasp-related V6A cells, as observed by Filippini et al. [32] who recorded and previously analyzed the dataset utilized here. Indeed, grasp-related V6A cells were found to discriminate among the different grips as soon as the motor plan can be formulated following the illumination of the object to reach and grasp, when an intermediate visuomotor transformation stage occurs, converting the visual information into motor commands [25]. Then, the discrimination power of V6A population increases while the monkey is preparing the reach-to-grasping action and peaks when the action is performed [26,32]. A similar pattern, although with the peak in epoch 4 less pronounced, emerged also by looking at the time course of network accuracy (Fig. 3); the latter is obtained thanks to the sliding window decoding approach that enables the analysis of the temporal dynamics of the chosen performance measure (the accuracy in this study), reflecting the V6A grip discrimination power. The peak of the temporal relevance patterns in epoch 4 (Fig. 5c) appears related to the involvement of a larger number of cells encoding information in that epoch, as it results from the spatio-temporal relevance maps in Fig. 5a. Indeed, while a set of cells presented high relevance already during movement preparation, i.e., from the end of epoch 1 up to epochs 4 and 5, another set of cells were highly relevant only later during movement and object holding, i.e., in epochs 4 and 5. Thus, due to this different temporal dynamics across cells (Fig. 5a), some cells added up their contribution to grip discrimination only during movement (epoch 4), resulting in the sharp peak of importance observed in Fig. 5c. It is worth remarking that the network, even from a small set of cells with high grip-discriminative power during movement preparation (epochs 1 and 2), was able to decode grip types with high accuracies (approximately at 0.8, see Fig. 3); then, later during the task, when also the second set of cells contributed to the discrimination, the network reached peak accuracies at values > 0.9. Notably, the most relevant cells were grip-sensitive (Fig. 5b), confirming that the network was able to automatically focus on the set of grip-sensitive neurons, while attributing less importance to the not grip-sensitive set. In conclusion, relevance representations based on LRP, by enabling the characterization of the temporal dynamics of grip-discriminative power specifically for each single cell, provided a richer description of temporal dynamics of reach-to-grasp encoding when compared to simpler performance measures tracked over time (e.g., the accuracy, as reported in Fig. 3). Overall, the previous results in the temporal domain are interesting as they show that the network is able to catch and capitalize on the temporal discharge pattern of grip-sensitive V6A cells during the reach-to-grasping task, although the network is fed by the entire set of cells, comprising 25–35% of not grip-sensitive cells that convey mainly noise.

Besides the results in the temporal domain, probably the most intriguing result of this study is provided by the analysis of the relevance in the spatial domain, considering the two cytoarchitectonic sectors of V6A. Indeed, from our findings, V6Ad cells resulted more relevant in more skilled grips (hook, primitive and advanced precision grips, c_2-c_4) and V6Av cells in more rudimentary grips (whole-hand and finger prehensions, c_0, c_1). These results are in agreement with previous studies in the literature: in particular, grip-sensitive neurons preferring a rudimentary grip (specifically, whole-hand prehension) were found to be more concentrated in V6Av, while the ones preferring a skilled grip (specifically, advanced precision grip) were found more concentrated in V6Ad [14]. The difference in grip properties encoding between the two V6A sectors, observed in our results and also in past literature [14], can be related to two main neurophysiological factors. First, V6A is characterized by a significant higher concentration of visual neurons responding to complex visual stimuli (e.g., light/dark gratings and corners with different orientation, direction and speed of movement) in its dorsal sector, and of visual neurons responding to elementary stimuli (e.g., light/dark borders, spots and bars) in its ventral sector [14,64]. These visual properties of V6A cells are likely reflected in the motor domain by a functional segregation of the preferred grip types, with the grip-sensitive neurons more activated by complex objects (requiring precision grip skills) being located in V6Ad, and the ones more activated by simple objects (requiring rudimentary grip skills) being located in V6Av [14]. Second, the anatomical connectivity of the two sectors is strongly different, mirroring their architectural organization. Specifically, V6Av is strongly interconnected with occipital visual areas, e.g., with afferent connections from V6, and it is considered more a visual area as part of a dorsomedial cortical network serving a fast motion analysis essential for the visual guidance of reach-to-grasping [16]. Conversely, V6Ad, strongly interconnected with areas of the superior parietal lobule, e.g., with the medial intraparietal area (MIP) and the anterior intraparietal area (AIP), and with the frontal cortex, is considered more a parietal area, as part of a parietofrontal network involved in the motor control of reaching and grasping [17]. Therefore, more skilled grips likely require more activation of the dorsal sector of V6A than more rudimentary grips, as obtained in our results (see Fig. 6). This can be a consequence of the higher degree of coordination required for prehension, that is, to control reach-to-grasp movements involving the selective use of only one finger (hook grip) or the opposition between the index finger/all fingers with the thumb (primitive and advanced precision grips). Notably, the results found here on the relative importance of the two V6A sectors in encoding grip types, do not only support the current knowledge about the functional segregation of this PPC area in reach-to-grasping, but they add an important advancement. Indeed, past analyses focused on investigating segregation of cells preferring only the most rudimentary grip (whole-hand prehension) or the most skilled grip (advanced precision grip), while no analysis was performed on intermediate skilled grips (e.g., finger prehension, hook grip, and primitive precision grip in this study). To this regard, our analysis provides a more refined view of the spatial properties of grip encoding across V6A area, showing that similar relevance differences between the two sectors, although to a less extent, hold also for intermediate skilled grips, with V6Av encoding more than V6Ad finger-prehension (other than whole-hand prehension) and V6Ad encoding more than V6Av hook grip and primitive precision grip (other than advanced precision grip). Even further, our analysis suggests the existence of a positive/negative linear trend between the amount of grip encoding in V6Ad/V6Av and the level of skills required by the specific grip. This result is in line with the hypothesis that V6Av encodes more simple features and V6Ad encodes more complex features; indeed, as the level of complexity of the reach-to-grasping movement gradually increases from the simplest whole-hand prehension to the most complex advanced precision grip, our results indicate a different amount of involvement of these two subareas.

Moreover, by analyzing the impact of the most relevant or of the least relevant cells on the decoding accuracy, we also validated LRP-derived measures from a computational point of view. Indeed, accuracy distributions and statistical analyses reported in Fig. 7 confirmed that the most relevant cells, as identified by LRP, were the ones with the highest grip-discriminative power (as quantified by the decoding accuracy) within the recorded neuron population, vice versa for the least relevant cells. Crucially, such analysis could also prospectively lead to a methodological improvement to invasive BCIs. Indeed, the most relevant cells (identified by our framework) were able to provide high accuracies (significantly above the chance level) while using only few recorded cells (e.g., only 15 cells to achieve >80% decoding accuracies, on average), and also achieved higher accuracies compared to randomly selected cells, i.e., with a non-informed selection of the cells used for decoding. Thus, our framework could also guide researchers to properly optimize the BCI recording setup in the future, e.g., by reducing the overall number of implanted electrodes, while at the same time maintain high decoding accuracies.

5. Conclusion

In conclusion, the adopted explainable artificial intelligence framework reveals, inside the single-neuron activities, meaningful neurophysiological aspects related to grip-type prediction, in a similar way as obtained in prior studies applying comparable methodologies to non-invasive recordings (EEG) and to other tasks [38,45,47]. Thus, this study encourages the use of techniques devoted at analyzing the knowledge learned by DNNs, not only to design decoders in a more robust, transparent, and reliable way but also for boosting our comprehension into the neural correlates underlying the investigated task. The obtained results support the current knowledge on PPC neural encoding of reach-to-grasping, and also expand it by providing more refined representations. This suggests that theories about the neural processes related to movement control may be not only validated but also advanced by explaining neural networks applied to neural signals invasively recorded from single cells. As such, the proposed framework would be extremely useful to validate motor theories/models and, in contexts in which there are no predetermined hypotheses or known neural correlates, to pave the way towards new theories and concepts on the neural control of movement. Notably, the presented framework was applied here for investigating encoding properties of reach-to-grasping; however, it could be easily transposed in the future to other motor (e.g., reaching) or cognitive (e.g., attention and memory) tasks involving single-neuron recordings.

CRediT authorship contribution statement

Davide Borra: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Matteo Filippini:** Data curation, Writing – review & editing. **Mauro Ursino:** Investigation, Writing – review & editing, Funding acquisition. **Patrizia Fattori:** Funding acquisition, Resources, Supervision, Writing – review & editing. **Elisa Magosso:** Conceptualization, Funding acquisition, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

None Declared

Acknowledgments

The authors declare no conflict of interest.

This study was supported by MAIA project. MAIA project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 951910. This article reflects only the author's view, and the Agency is not responsible for any use that may be made of the information it contains.

This work is also supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) — A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022).

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the TITAN V used for this research. The provider was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (2002) 767–791, https://doi.org/10.1016/S1388-2457(02)00057-3.

[2] J.I. Glaser, A.S. Benjamin, R.H. Chowdhury, M.G. Perich, L.E. Miller, K.P. Kording, Machine learning for neural decoding, eNeuro 7 (2020), https://doi.org/10.1523/ENEURO.0506-19.2020. ENEURO.0506-19.2020.

[3] J.I. Glaser, A.S. Benjamin, R. Farhoodi, K.P. Kording, The roles of supervised machine learning in systems neuroscience, Prog. Neurobiol. 175 (2019) 126–137, https://doi.org/10.1016/j.pneurobio.2019.01.008.

[4] R.A. Andersen, S. Kellis, C. Klaes, T. Aflalo, Toward more versatile and intuitive cortical brain–machine interfaces, Curr. Biol. 24 (2014) R885–R897, https://doi.org/10.1016/j.cub.2014.07.068.

[5] H. Cui, Forward prediction in the posterior parietal cortex and dynamic brain-machine interface, Front. Integr. Neurosci. 10 (2016), https://doi.org/10.3389/fnint.2016.00035.

[6] E.P. Gardner, Neural pathways for cognitive command and control of hand movements, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) 4048–4050, https://doi.org/10.1073/pnas.1702746114.

[7] E. Santandrea, R. Breveglieri, A. Bosco, C. Galletti, P. Fattori, Preparatory activity for purposeful arm movements in the dorsomedial parietal area V6A: beyond the online guidance of movement, Sci. Rep. 8 (2018) 6926, https://doi.org/10.1038/s41598-018-25117-0.

[8] S. Musallam, B.D. Corneil, B. Greger, H. Scherberger, R.A. Andersen, Cognitive control signals for neural prosthetics, Science 305 (2004) 258–262, https://doi.org/10.1126/science.1097938.

[9] G.H. Mulliken, S. Musallam, R.A. Andersen, Decoding trajectories from posterior parietal cortex ensembles, J. Neurosci. 28 (2008) 12913–12926, https://doi.org/10.1523/JNEUROSCI.1463-08.2008.

[10] S. Schaffelhofer, A. Agudelo-Toro, H. Scherberger, Decoding a wide range of hand configurations from macaque motor, premotor, and parietal cortices, J. Neurosci. 35 (2015) 1068–1081, https://doi.org/10.1523/JNEUROSCI.3594-14.2015.

[11] M. Hauschild, G.H. Mulliken, I. Fineman, G.E. Loeb, R.A. Andersen, Cognitive signals for brain-machine interfaces in posterior parietal cortex include continuous 3D trajectory commands, Proc. Natl. Acad. Sci. USA 109 (2012) 17075–17080, https://doi.org/10.1073/pnas.1215092109.

[12] T. Aflalo, S. Kellis, C. Klaes, B. Lee, Y. Shi, K. Pejsa, K. Shanfield, S. Hayes-Jackson, M. Aisen, C. Heck, C. Liu, R.A. Andersen, Decoding motor imagery from the posterior parietal cortex of a tetraplegic human, Science 348 (2015) 906–910, https://doi.org/10.1126/science.aaa5417.

[13] C. Klaes, S. Kellis, T. Aflalo, B. Lee, K. Pejsa, K. Shanfield, S. Hayes-Jackson, M. Aisen, C. Heck, C. Liu, R.A. Andersen, Hand shape representations in the human posterior parietal cortex, J. Neurosci. 35 (2015) 15466–15476, https://doi.org/10.1523/JNEUROSCI.2747-15.2015.

[14] M. Gamberini, C. Galletti, A. Bosco, R. Breveglieri, P. Fattori, Is the medial posterior parietal area V6A a single functional area? J. Neurosci. 31 (2011) 5145–5157, https://doi.org/10.1523/JNEUROSCI.5489-10.2011.

[15] G. Luppino, S.B. Hamed, M. Gamberini, M. Matelli, C. Galletti, Occipital (V6) and parietal (V6A) areas in the anterior wall of the parieto-occipital sulcus of the macaque: a cytoarchitectonic study, Eur. J. Neurosci. 21 (2005) 3056–3076, https://doi.org/10.1111/j.1460-9568.2005.04149.x.

[16] L. Passarelli, M.G.P. Rosa, M. Gamberini, S. Bakola, K.J. Burman, P. Fattori, C. Galletti, Cortical connections of area V6Av in the macaque: a visual-input node to the eye/hand coordination system, J. Neurosci. 31 (2011) 1790–1801, https://doi.org/10.1523/JNEUROSCI.4784-10.2011.

[17] M. Gamberini, L. Passarelli, P. Fattori, M. Zucchelli, S. Bakola, G. Luppino, C. Galletti, Cortical connections of the visuomotor parietooccipital area V6Ad of the macaque monkey, J. Comp. Neurol. 513 (2009) 622–642, https://doi.org/10.1002/cne.21980.

[18] A. Bosco, R. Breveglieri, M. Filippini, C. Galletti, P. Fattori, Reduced neural representation of arm/hand actions in the medial posterior parietal cortex, Sci. Rep. 9 (2019) 936, https://doi.org/10.1038/s41598-018-37302-2.

[19] A. Bosco, R. Breveglieri, K. Hadjidimitrakis, C. Galletti, P. Fattori, Reference frames for reaching when decoupling eye and target position in depth and direction, Sci. Rep. 6 (2016) 21646, https://doi.org/10.1038/srep21646.

[20] A. Bosco, R. Breveglieri, E. Chinellato, C. Galletti, P. Fattori, Reaching activity in the medial posterior parietal cortex of monkeys is modulated by visual feedback, J. Neurosci. 30 (2010) 14773–14785, https://doi.org/10.1523/JNEUROSCI.2313-10.2010.

[21] R. Breveglieri, C. Galletti, G. Dal Bò, K. Hadjidimitrakis, P. Fattori, Multiple aspects of neural activity during reaching preparation in the medial posterior parietal area V6A, J. Cognit. Neurosci. 26 (2014) 878–895, https://doi.org/10.1162/jocn_a_00510.

[22] K. Hadjidimitrakis, F. Bertozzi, R. Breveglieri, A. Bosco, C. Galletti, P. Fattori, Common neural substrate for processing depth and direction signals for reaching in the monkey medial posterior parietal cortex, Cerebr. Cortex 24 (2014) 1645–1657, https://doi.org/10.1093/cercor/bht021.

[23] R. Breveglieri, M. De Vitis, A. Bosco, C. Galletti, P. Fattori, Interplay between grip and vision in the monkey medial parietal lobe, Cerebr. Cortex 28 (2018) 2028–2042, https://doi.org/10.1093/cercor/bhx109.

[24] R. Breveglieri, A. Bosco, C. Galletti, L. Passarelli, P. Fattori, Neural activity in the medial parietal area V6A while grasping with or without visual feedback, Sci. Rep. 6 (2016) 28893, https://doi.org/10.1038/srep28893.

[25] P. Fattori, R. Breveglieri, V. Raos, A. Bosco, C. Galletti, Vision for action in the macaque medial posterior parietal cortex, J. Neurosci. 32 (2012) 3221–3234, https://doi.org/10.1523/JNEUROSCI.5358-11.2012.

[26] P. Fattori, V. Raos, R. Breveglieri, A. Bosco, N. Marzocchi, C. Galletti, The dorsomedial pathway is not just for reaching: grasping neurons in the medial parieto-occipital cortex of the macaque monkey, J. Neurosci. 30 (2010) 342–349, https://doi.org/10.1523/JNEUROSCI.3800-09.2010.

[27] D. Borra, M. Filippini, M. Ursino, P. Fattori, E. Magosso, A bayesian-optimized convolutional neural network to decode reach-to-grasp from macaque dorsomedial visual stream, in: G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. Pardalos, G. Di Fatta, G. Giuffrida, R. Umeton (Eds.), Machine Learning, Optimization, and Data Science, Springer Nature Switzerland, Cham, 2023, pp. 473–487, https://doi.org/10.1007/978-3-031-25891-6_36.

[28] D. Borra, M. Filippini, M. Ursino, P. Fattori, E. Magosso, Motor decoding from the posterior parietal cortex using deep neural networks, J. Neural. Eng. 20 (2023) 036016, https://doi.org/10.1088/1741-2552/acd1b6.

[29] M. Filippini, D. Borra, M. Ursino, E. Magosso, P. Fattori, Decoding sensorimotor information from superior parietal lobule of macaque via Convolutional Neural Networks, Neural Network. 151 (2022) 276–294, https://doi.org/10.1016/j.neunet.2022.03.044.

[30] M. Filippini, A.P. Morris, R. Breveglieri, K. Hadjidimitrakis, P. Fattori, Decoding of standard and non-standard visuomotor associations from parietal cortex, J. Neural. Eng. 17 (2020) 046027, https://doi.org/10.1088/1741-2552/aba87e.

[31] M. Filippini, R. Breveglieri, K. Hadjidimitrakis, A. Bosco, P. Fattori, Prediction of reach goals in depth and direction from the parietal cortex, Cell Rep. 23 (2018) 725–732, https://doi.org/10.1016/j.celrep.2018.03.090.

[32] M. Filippini, R. Breveglieri, M.A. Akhras, A. Bosco, E. Chinellato, P. Fattori, Decoding information for grasping from the macaque dorsomedial visual stream, J. Neurosci. 37 (2017) 4311–4322, https://doi.org/10.1523/JNEUROSCI.3077-16.2017.

[33] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces, J. Neural. Eng. 15 (2018) 056013, https://doi.org/10.1088/1741-2552/aace8c.

[34] D. Borra, F. Bossi, D. Rivolta, E. Magosso, Deep learning applied to EEG source-data reveals both ventral and dorsal visual stream involvement in holistic processing of social stimuli, Sci. Rep. 13 (2023) 7365, https://doi.org/10.1038/s41598-023-34487-z.

[35] D. Borra, S. Fantozzi, E. Magosso, A lightweight multi-scale convolutional neural network for P300 decoding: analysis of training strategies and uncovering of network decision, Front. Hum. Neurosci. 15 (2021) 655840, https://doi.org/10.3389/fnhum.2021.655840.

[36] M. Simões, D. Borra, E. Santamaría-Vázquez, Gbt-Upm, M. Bittencourt-Villalpando, D. Krzemiński, A. Miladinović, Neural_Engineering_Group, T. Schmid, H. Zhao, C. Amaral, B. Direito, J. Henriques, P. Carvalho, M. Castelo-Branco, BCIAUT-P300: a multi-session and multi-subject benchmark dataset on autism for P300-based brain-computer-interfaces, Front. Neurosci. 14 (2020) 568104, https://doi.org/10.3389/fnins.2020.568104.

[37] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (2017) 5391–5420, https://doi.org/10.1002/hbm.23730.

[38] A. Vahid, M. Mückschel, S. Stober, A.-K. Stock, C. Beste, Applying deep learning to single-trial EEG data provides evidence for complementary theories on action control, Commun. Biol. 3 (2020) 112, https://doi.org/10.1038/s42003-020-0846-z.

[39] J.A. Livezey, J.I. Glaser, Deep learning approaches for neural decoding across architectures and recording modalities, Briefings Bioinf. 22 (2021) 1577–1591, https://doi.org/10.1093/bib/bbaa355.

[40] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, J. Neural. Eng. 16 (2019) 031001, https://doi.org/10.1088/1741-2552/ab0ab5.

[41] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T.H. Falk, J. Faubert, Deep learning-based electroencephalography analysis: a systematic review, J. Neural. Eng. 16 (2019) 051001, https://doi.org/10.1088/1741-2552/ab260c.

[42] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, 2014 arXiv:1312.6034 [Cs].

[43] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, PLoS One 10 (2015) e0130140, https://doi.org/10.1371/journal.pone.0130140.

[44] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, K.-R. Müller, Layer-wise relevance propagation: an overview, in: W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, K.-R. Müller (Eds.), Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Springer International Publishing, Cham, 2019, pp. 193–209, https://doi.org/10.1007/978-3-030-28954-6_10.

[45] D. Borra, E. Magosso, M. Castelo-Branco, M. Simoes, A Bayesian-optimized design for an interpretable convolutional neural network to decode and analyze the P300 response in autism, J. Neural. Eng. 19 (2022) 046010, https://doi.org/10.1088/1741-2552/ac7908.

[46] D. Borra, S. Fantozzi, E. Magosso, Interpretable and lightweight convolutional neural network for EEG decoding: application to movement execution and imagination, Neural Network. 129 (2020) 55–74, https://doi.org/10.1016/j.neunet.2020.05.032.

[47] D. Borra, E. Magosso, Deep learning-based EEG analysis: investigating P3 ERP components, J. Integr. Neurosci. 20 (2021) 791–811, https://doi.org/10.31083/j.jin2004083.

[48] A.I. Korda, E. Ventouras, P. Asvestas, M. Toumaian, G.K. Matsopoulos, N. Smyrnis, Convolutional neural network propagation on electroencephalographic scalograms for detection of schizophrenia, Clin. Neurophysiol. 139 (2022) 90–105, https://doi.org/10.1016/j.clinph.2022.04.010.

[49] I. Sturm, S. Lapuschkin, W. Samek, K.-R. Müller, Interpretable deep neural networks for single-trial EEG classification, J. Neurosci. Methods 274 (2016) 141–145, https://doi.org/10.1016/j.jneumeth.2016.10.008.

[50] M. Gamberini, G. Dal Bò, R. Breveglieri, S. Briganti, L. Passarelli, P. Fattori, C. Galletti, Sensory properties of the caudal aspect of the macaque's superior parietal lobule, Brain Struct. Funct. (2017), https://doi.org/10.1007/s00429-017-1593-x.

[51] E. Lashgari, D. Liang, U. Maoz, Data augmentation for deep-learning-based electroencephalography, J. Neurosci. Methods 346 (2020) 108885, https://doi.org/10.1016/j.jneumeth.2020.108885.

[52] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1800–1807.

[53] R. Shi, Y. Zhao, Z. Cao, C. Liu, Y. Kang, J. Zhang, Categorizing objects from MEG signals using EEGNet, Cogn. Neurodyn. 16 (2022) 365–377, https://doi.org/10.1007/s11571-021-09717-7.

[54] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: F. Bach, D. Blei (Eds.), Proceedings of the 32nd International Conference on Machine Learning, PMLR, Lille, France, 2015, pp. 448–456.

[55] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.

[56] D. Borra, V. Mondini, E. Magosso, G.R. Müller-Putz, Decoding movement kinematics from EEG using an interpretable convolutional neural network, Comput. Biol. Med. 165 (2023) 107323, https://doi.org/10.1016/j.compbiomed.2023.107323.

[57] A. Farahat, C. Reichert, C. Sweeney-Reed, H. Hinrichs, Convolutional neural networks for decoding of covert attention focus and saliency maps for EEG feature visualization, J. Neural. Eng. 16 (2019) 066010. http://iopscience.iop.org/10.1088/1741-2552/ab3bb4.

[58] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic Differentiation in PyTorch, NIPS-W, 2017.

[59] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan, O. Reblitz-Richardson, Captum: A Unified and Generic Model Interpretability Library for PyTorch, 2020. http://arxiv.org/abs/2009.07896. (Accessed 18 July 2022).

[60] D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and Accurate Deep Network Learning by Exponential Linear Units (Elus), arXiv Preprint, 2015.

[61] A. Murata, V. Gallese, G. Luppino, M. Kaseda, H. Sakata, Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP, J. Neurophysiol. 83 (2000) 2580–2601, https://doi.org/10.1152/jn.2000.83.5.2580.

[62] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Stat. Soc. B 57 (1995) 289–300.

[63] T.E. Nichols, A.P. Holmes, Nonparametric permutation tests for functional neuroimaging: a primer with examples, Hum. Brain Mapp. 15 (2002) 1–25, https://doi.org/10.1002/hbm.1058.

[64] P. Fattori, R. Breveglieri, A. Bosco, M. Gamberini, C. Galletti, Vision for prehension in the medial parietal cortex, Cerebr. Cortex 27 (2017) 1149–1163, https://doi.org/10.1093/cercor/bhv302.