

Title: Unraveling Somatotopic Organization in the Human Brain using Machine Learning and Adaptive Supervoxel-based Parcellations

Running Title: Somatotopy: Adaptive Supervoxel-based Brain Parcellation

Author Names: Kyle See^{1^}, David J. Arpin^{2^}, David E. Vaillancourt^{1,2}, Ruogu Fang^{1,3,4,*},
Stephen A. Coombes^{1,2,*}

Author Affiliations:

¹J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL.

²Laboratory for Rehabilitation Neuroscience, Department of Applied Physiology and Kinesiology, University of Florida, Gainesville, FL.

³Center for Cognitive Aging and Memory, McKnight Brain Institute, University of Florida, Gainesville, USA

⁴Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL.

[^] co-first authors

***Co-Corresponding Authors:**

Stephen A. Coombes, Ph.D.
University of Florida
Laboratory for Rehabilitation Neuroscience
Department of Applied Physiology and Kinesiology
PO Box 118206
Email: scoombes@ufl.edu
Phone: 352-294-1768

Ruogu Fang, Ph.D.
J. Crayton Pruitt Family Department of Biomedical Engineering
Smart Medical Informatics Learning and Evaluation Lab
College of Engineering
PO Box 116131
Email: ruogu.fang@bme.ufl.edu
Phone: (352) 294-1375

Financial support: Grant support for See was provided by the National Center for Advancing Translational Sciences of the National Institutes of Health under the University of Florida and State University Clinical and Translational Science Awards TL1TR001428 and UL1TR001427. R01 NS052318 provided support in data acquisition. Fang was supported by National Science Foundation under Grant No. 1908299. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Conflict of interest: There are no conflicts of interest to report. S.A. Coombes and D.E. Vaillancourt are co-founders and managers of Neuroimaging Solutions, Gainesville, FL.

ABSTRACT

In addition to the well-established somatotopy in the pre- and post-central gyrus, there is now strong evidence that somatotopic organization is evident across other regions in the sensorimotor network. This raises several experimental questions: To what extent is activity in the sensorimotor network effector-dependent and effector-independent? How important is the sensorimotor cortex when predicting the motor effector? Is there redundancy in the distributed somatotopically organized network such that removing one region has little impact on classification accuracy? To answer these questions, we developed a novel experimental approach. fMRI data were collected while human subjects performed a precisely controlled force generation task separately with their hand, foot, and mouth. We used a simple linear iterative clustering (SLIC) algorithm to segment whole-brain beta coefficient maps to build an adaptive brain parcellation and then classified effectors using extreme gradient boosting (XGBoost) based on parcellations at various spatial resolutions. This allowed us to understand how data-driven adaptive brain parcellation granularity altered classification accuracy. Results revealed effector-dependent activity in regions of the post-central gyrus, precentral gyrus, and paracentral lobule. SMA, regions of the inferior and superior parietal lobule, and cerebellum each contained effector-dependent and effector-independent representations. Machine learning analyses showed that increasing the spatial resolution of the data-driven model increased classification accuracy, which reached 94% with over 1,755 supervoxels. Our SLIC-based supervoxel parcellation outperformed classification analyses using established brain templates and random simulations. Occlusion experiments further demonstrated redundancy across the sensorimotor network when classifying effectors. Our observations extend our understanding of effector-dependent and

effector-independent organization within the human brain and provide new insight into the functional neuroanatomy required to predict the motor effector used in a motor control task.

1. INTRODUCTION

Somatotopy is the point-to-point mapping of an area of the body to a specific point in the central nervous system (Saladin, 2012). Pioneering studies by Penfield (Penfield and Boldrey, 1937; Penfield and Rasmussen, 1950) used electrical stimulation on the surface of the cortex to map somatotopic organization in the human brain. Within both the primary motor cortex (M1) and primary somatosensory cortex (S1), lower extremities are located dorsally and close to the midline, the arms and hands are represented more laterally, and the face representation is most lateral and ventral. Somatotopic organization is not limited to the pre- and post-central gyri (M1/S1), with both fMRI and EEG studies revealing somatotopy in the supplementary motor area (SMA) (Chainay et al., 2004; Zeharia et al., 2012), cingulate motor area (Mayer et al., 2001), premotor cortex, superior and inferior parietal lobules (Cunningham et al., 2013), insula, basal ganglia (Zeharia et al., 2015), and cerebellum (Batson et al., 2015; Boillat et al., 2020; Grodd et al., 2001a; Nitschke et al., 1996; Schlerf et al., 2010; van der Zwaag et al., 2013). This raises several experimental questions: To what extent is activity in the sensorimotor network effector-dependent and effector-independent? How important is the sensorimotor cortex when predicting the motor effector? Is there redundancy in the distributed somatotopically organized network such that removing one region has little impact on classification accuracy? Current knowledge is unfortunately derived from experimental paradigms which often do not measure or control motor performance across effectors. This can be problematic given that the BOLD signal can be influenced by differing characteristics of force production (Spraker et al., 2012, 2007), and different effectors inherently produce different amounts of force even when flexing/extending at the same frequency (Lotze et al., 2000; Zeharia et al., 2015).

Previous studies have used brain signals to classify different effectors with varying levels of success. Although studies vary in which effectors are assessed, what movements are made, and how the movements are measured (Cunningham et al., 2013; Lotze et al., 2000; Zeharia et al., 2015; Zhao et al., 2019), it is clear that classification performance is higher at the individual level as compared to the group level (Sleight et al., 2009). Classifying effectors at the individual level is helpful in the context of brain-computer interfaces when there is an adaptation phase between the machine and the brain (Bhattacharyya et al., 2010; Morash et al., 2008; Shiman et al., 2017), but this approach does not always generalize well to new individuals. Other fMRI studies have implemented multivoxel pattern analyses (MVPA) to classify movements based on predefined regions of interest (Filimon et al., 2015; Gallivan et al., 2013, 2011; Shay et al., 2019; Zeharia et al., 2015). While effective in determining how well the BOLD signal can differentiate effectors within a specific brain region, it is challenging to implement this approach across the whole brain using a data-driven clustering approach.

Clustering methods that have been used in the neuroimaging literature include K-means (Chyzhyk and Graña, 2014), mean-shift (Ai and Xiong, 2016), and spectral clustering (Venkataraman et al., 2009). These data-driven methods can be effective in clustering; however, their flexibility and efficiency decrease as the dataset size increases. For example, K-means is an excellent general-purpose clustering algorithm, but it prefers spherical-shaped clusters and can fail to capture clusters with complicated geometric shapes. It also has randomized seed locations which makes reproducibility more challenging. Mean-shift works well with spherical data and automatically selects the optimal number of clusters. However, it is not highly scalable. Spectral clustering is a graph-based method that clusters regardless of shape or distribution, but it works best with fewer numbers of clusters and is costly to compute. In this study, we used a modified

simple linear iterative clustering (SLIC) algorithm (Achanta et al., 2010), to address these limitations by providing flexibility in the compactness of clusters (i.e., parcellations), improved computational efficiency, and grid-based initialization. These factors provide a controllable, generalizable, and repeatable method for generating data-driven parcellations.

We used fMRI to assess BOLD signal changes across the whole brain while force production was precisely controlled across three different effectors: the foot, hand, and mouth. Real-time visual feedback was provided to subjects while they completed a target matching task by executing a sustained isometric contraction with each effector. We first used conventional univariate analyses to identify effector-dependent and effector-independent brain activation during task performance. Next, we implemented a novel machine learning approach which allowed us to control the brain parcellation resolution by tuning the maximum number of features fed into the classifier, to understand how data-driven, adaptive brain parcellation granularity altered classification accuracy and key region identification. The findings provide new insight into the functional neuroanatomy required to predict the motor effector used in a motor control task.

2. METHODS

2.1 Subjects

This study included 20 right-handed adults (age = 20.50 ± 0.92 years; female = 10). All participants had normal or corrected to normal vision, self-reported no disability or neurological conditions that would prevent them from completing the required force tasks with their foot, hand, or mouth. All participants provided written informed consent before testing, and all procedures were approved by the Institutional Review Board at the University of Florida.

2.2 Force task and acquisition

Three force production tasks were performed by each participant inside an MRI scanner. Participants were supine for all conditions, and generated force against a custom fiber-optic force transducer with a resolution of 0.025 N (Neuroimaging Solutions, Gainesville, FL) using the right foot (see Figure 1A), the right hand (Figure 1B), and bilaterally by the mouth (Figure 1C). Force signals were transmitted through a fiber-optic cable to an SM130 Optical Sensing Interrogator (Micron Optics, Atlanta, Georgia). The interrogator digitized analog force data at 62.5 Hz. We used a customized program written in LabVIEW (National Instruments, Austin, TX) to collect the force data. Online visual feedback of the force output was visible to the participants at a 60 Hz refresh rate. Force data were low-pass filtered using a Butterworth, 20 Hz 4th-order dual-pass filter.

Before entering the scanner, each participant's maximum voluntary contraction (MVC) was recorded for each effector (foot, hand, mouth) during a practice session. During a series of three consecutive MVC trials, participants held a maximum force contraction (MFC) for 5 seconds. A 60 second rest period was provided between the trials. Average force during each sustained MFC was calculated and the mean of three trials was used as the participant's MVC value. fMRI

data were collected while each participant produced force to a target level set at 10% of their MVC for each of the three effectors separately. Effector order was counterbalanced across participants.

A representation of the task paradigm is shown in Figure 1D. Each task involved both a rest and force condition. During the force condition, two horizontal bars were displayed which represented the cue to produce force (green bar) as well as the target force (white bar). Subjects produced force for 4s when the bar turned green. The force produced caused the green bar to move vertically, providing visual feedback to the participant. The white target bar remained stationary and represented the participant's 10% MVC target. When the bar turned red, subjects rested for 2s. This pattern of 4s force and 2s rest repeated 5 times over the 30s force block. In the rest block, subjects rested the entire 30s with stationary bars on the screen. Each scan started and ended with a rest block and contained a total of 4 force blocks.

2.3 MRI acquisition

MRI images were collected using a 32-channel head coil inside a Phillips 3T magnetic resonance scanner (Achieva, Best, the Netherlands). T1-weighted images (resolution: 1 mm isotropic, repetition time: TR = 6.8 ms, echo time: TE = 3.3 ms, flip angle = 8°, field of view = 240 mm³, and acquisition matrix = 240 x 240) were collected in 170 contiguous axial slices. Functional data were collected in 46 contiguous axial slices using a single-shot gradient echo-planar imaging pulse sequence (resolution: 3 mm isotropic, TR = 2500 ms, TE = 30 ms, flip angle = 80°, field of view = 240 mm³, and acquisition matrix = 80 x 80).

2.4 Force Data Analysis

Force data were analyzed using customized algorithms in LabVIEW. Mean force generated as a percentage of MVC, variability of force generated, and target force error were calculated for each trial and then averaged separately for each effector (hand, foot, mouth). These measures were calculated from the data extracted from the middle 2s of each 4s force pulse. Variability of force generated was quantified by the standard deviation of the force data. Target force error was quantified using root mean square error (RMSE). The onset of a trial was identified by the time point at which the force rose above 10% of the peak force, and the offset of a trial was identified as the time point at which force fell below 10% of the peak force.

Shapiro-Wilk's test of normality was used to determine whether the data were normally distributed (Shapiro and Wilk, 1965), which is one of the key assumptions when implementing ANOVA. Those data that failed the test were subsequently logarithmically transformed for all statistical testing. Differences in mean force as a percent of MVC, variability of force generated, and target force errors between each effector were quantified using 1-way ANOVAs with Bonferroni post hoc. Statistical analyses were performed in SPSS version 26 (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp) at a 0.05 alpha level.

2.5 Imaging Data Analysis

Imaging data were analyzed with AFNI software (Analysis of Functional NeuroImages; National Institutes of Health, Bethesda, MD), SPM8 (Statistical Parametric Mapping) software, the SUIT (A spatially unbiased atlas template of the cerebellum and brainstem) (Diedrichsen, 2006; Diedrichsen et al., 2011, 2009; Diedrichsen and Zotow, 2015) toolbox of SPM8, and custom

UNIX shell scripts. AFNI software was used to compute whole-brain statistical maps. Cerebellum-specific fine-tuning of the statistical maps was performed with SPM8 and SUI.

2.6 Subject-level Analysis

Each effector (hand, foot, mouth) was analyzed separately. Three functional volumes were collected prior to the beginning of the experiment to allow for magnetization to stabilize, and these volumes were discarded before analysis. The remaining volumes were slice-acquisition-dependent slice-time corrected. The anatomical image was skull stripped using 3dSkullStrip in AFNI. Functional volumes were registered to a base volume via rigid body rotations and aligned with the anatomical image in a single transformation, therefore avoiding repeated image resampling. For the whole-brain analysis, functional volumes were warped into MNI space and smoothed with a 4 mm full-width-half-maximum Gaussian kernel for increasing the signal-to-noise ratio. The blood-oxygen-level-dependent (BOLD) signal in each voxel at every time point was scaled by the mean of its respective time series to normalize the data. The BOLD signal during the 30s force periods was modeled using a boxcar regressor convolved with the hemodynamic response function. Head motion parameters calculated during the registration step were included in the general linear model as regressors. Head motion between adjacent volumes that was greater than 0.6 mm resulted in the exclusion of both volumes from the regression analysis. For the cerebellum analysis, warping to MNI space and smoothing occurred using the SUI template after the BOLD signal was normalized and head motion was regressed out of the signal. Head motion above the threshold limit resulted in more than 35% of the functional volumes being excluded from analysis in three individuals. These three subjects were excluded, resulting in 17 participants being included in the final imaging and behavioral analyses. Across the remaining 17 participants, after correcting for

head motion, an average of 97.5%, 99.8%, and 89.7% of the volumes remained for the foot, hand, and mouth tasks, respectively. The resulting regression-coefficient (beta-value) maps for each individual were then passed to the group level and were analyzed using two approaches: group-level voxelwise analyses, and machine learning classification analysis.

2.7 Group Level Voxelwise Analyses

2.7.1 Creating an Activation Mask and Determining Cluster Extent Threshold

We first created an activation mask to constrain the number of voxels and therefore the number of statistical analyses performed at the group level. To do this we conducted t-tests for each effector (foot, hand, mouth) compared to rest for all voxels in the cerebrum, and then separately for all voxels in the cerebellum. The output from each t-test was then binarized using a liberal threshold ($p < 0.05$ and cluster size ≥ 2 voxels) to include as many voxels as possible in the mask. The binarized maps for foot, hand, and mouth for the cerebrum were then summed, and any voxel with a value greater than 1 was included in the activation mask for the cerebrum. The binarized maps for foot, hand, and mouth for the cerebellum were also summed, and any voxel with a value greater than 1 was included in the activation mask for the cerebellum. At this stage, we make no statistical inferences on whether brain activation is effector-dependent or effector-independent.

Next, we identified the cluster-extent threshold using 3dClustSim to control for Family-wise error rate (FWER) in the statistical maps that are outlined below in sections 2.7.2 and 2.7.3. The number of voxels required for a cluster to reach significance when $p = 0.005$ and $\alpha = 0.05$ was 24.

2.7.2 Identifying Effector-Independent Brain Regions

Task (foot, hand, mouth) vs. rest t-tests were computed using activation-masked beta-coefficient maps for the cerebrum, and significantly active voxels were identified using FWER-corrected parameters outlined above. Statistical maps for each t-test were then binarized such that significantly active voxels were assigned a value of 1 and all other voxels were assigned a value of zero. To identify voxels that were commonly active across the t-tests we computed a conjunction analysis by summing the binarized maps for each effector. Voxels that had a value of 3 were retained and all other voxels were assigned a value of zero. Clusters of voxels with a value of 3 were labeled as effector-independent regions because they were significantly active across all t-tests. Size, coordinates, and anatomical label were recorded for each effector-independent cluster. The same analysis was performed for the SUI normalized cerebellum data.

2.7.3 Identifying Effector-Dependent Brain Regions

To determine differences in brain activity between effectors, activation-masked beta-coefficient maps for each effector were directly compared using a 1-way ANOVA with three levels (foot, hand, mouth). The resulting F-statistic map was thresholded to retain voxels with P-values <0.005 and clusters that were >24 voxels in volume to keep the FWER <0.05 . The F-statistic map identified regions of the brain that were significantly different between at least two of the three effectors. Size, coordinates, and anatomical label were recorded for each effector-dependent cluster. Follow-up t-tests (foot vs. hand, foot vs. mouth, hand vs. mouth) were then used to identify which of the effectors differed within the clusters identified by the ANOVA, as well as the direction of the difference. A separate ANOVA and follow-up t-tests were run for the SUI normalized cerebellum data using the same procedures.

2.8 Identifying Effector-Dependent Regions using Machine Learning Classification Analysis

A machine learning classification analysis was used to determine whether brain activity from multiple regions could together predict which effector was being used and to determine which brain regions have the largest contribution to that prediction. Activation-masked beta coefficient maps for each effector (i.e., task vs rest) for each subject were used as inputs to the pipeline. Subject-level whole-brain maps were first averaged across effectors for each subject (Figure 2a) and then across all subjects (Figure 2b) to create a single whole-brain map. Each voxel in the map represented an average beta coefficient value, which represents the mean percent signal change in the BOLD signal across all subjects, all effectors, and all tasks. Supervoxels were then identified based on a data-driven algorithm which we outline next (Figure 2c). Supervoxels reduce computational complexity, and by iteratively increasing the number of allowed supervoxels we were able to determine the relationship between spatial resolution and performance of the classifier.

2.8.1 Supervoxel Parcellations

The supervoxels were generated using a 3-dimensional simple linear iterative clustering (3D-SLIC) algorithm. The algorithm segments data points with consistent uniformity and is computationally efficient (Achanta et al., 2010). In the current study, we implement the 3D-SLIC algorithm to segment clusters of voxels in three-dimensional space (x, y, z). The algorithm takes a single input K , which is the desired number of equally sized supervoxels. The approximate size of each supervoxel is N/K , where N is the number of voxels in the 3-dimensional map of beta coefficients calculated from the fMRI volume. A supervoxel center is found at every grid step $S =$

$\sqrt{\frac{N}{K}}$ for roughly equally sized supervoxels. Clustering begins with an initial seeding of K regularly spaced cluster centers. Each cluster center is then moved to the lowest gradient position within a $3 \times 3 \times 3$ neighborhood. The voxel map gradient is computed with Equation 1.

$$G(x, y, z) = \|I(x + 1, y, z) - I(x - 1, y, z)\|^2 + \|I(x, y + 1, z) - I(x, y - 1, z)\|^2 + \|I(x, y, z + 1) - I(x, y, z - 1)\|^2 \quad (1)$$

$I(x, y, z)$ is the beta coefficient that corresponds to the voxel position of (x, y, z) . $\|\cdot\|^2$ is the L_2 norm which measures the magnitude of signal difference between two voxels. The gradient accounts for intensity while avoiding the placement of centers on edges or noisy voxels. Each voxel in the map is then associated with the nearest cluster center based on a $2S \times 2S \times 2S$ neighborhood around each cluster center. The distance, D , between two voxels i and j is measured using both signal intensities and spatial distances (Equation 2). The compactness variable, m , defines the shape of the supervoxels. Higher values of m ensure uniform shapes, while lower values allow supervoxels to adhere to boundaries resulting in irregular shapes. We set m to a constant value of 0.001 which is considered low enough to allow irregular shapes that would be expected given brain anatomy. Using SLIC offers more flexibility and adaptiveness to the large variance in the anatomical brain regions. This versatility allows us to freely conform to activations with little restriction.

$$\begin{aligned} d_{intensity} &= \sqrt{(l_i - l_j)^2} \\ d_{spatial} &= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \\ D &= \sqrt{\left(\frac{d_{intensity}}{m}\right)^2 + \left(\frac{d_{spatial}}{s}\right)^2} \quad (2) \end{aligned}$$

Cluster centers are recomputed using the mean positions of respectively matched voxels. Residual errors are taken using the L_1 distance between the old and the new centers. The process of assigning voxels and recomputing centers was repeated until the error was minimized. Connectivity is enforced by relabeling disjointed segments to the largest neighboring cluster. The voxels are then masked using the activation mask that is detailed in section 2.7.1 above (Figure 2d). This may remove supervoxels entirely or remove a portion of the voxels contributing to a supervoxel. This approach guaranteed that the same number of supervoxels was used for each subject in the classification analysis, while also ensuring that we did not bias supervoxel generation using between effector contrasts. For each supervoxel, an average beta coefficient was then calculated for each subject across all effectors (Figure 2e). Each supervoxel, therefore, represented a single feature, reflecting activity in an area of the brain. Hence, the number of supervoxels indicated how many features were fed to the classification analysis, where an increase in feature number reflects an increase in spatial resolution. The goal of the classification analysis was to determine which supervoxels (i.e., features or brain regions) contributed most to predicting which effector was being used, while also determining the overall performance accuracy of the combination of supervoxels. We repeated the supervoxel identification process 37 times, changing the maximum number of *allowed* supervoxels between 1-10,000. Specifically, we sampled 9 models between 1 and 9 with a step size of 1, 9 models between 10 and 90 with a step size of 10, 9 models between 100 and 900 with a step size of 100, and finally 10 models between 1,000 and 10,000 with a step size of 1,000. A total of 37 models of different spatial resolution were generated. Note here the maximum number of allowed supervoxels is input to the SLIC algorithm as a parameter, however, SLIC does not always generate the maximum number during the optimization problem. Supervoxels are merged during the optimization process when their centers overlap. This

results in a lower number of supervoxels than was specified. The number of generated supervoxels describes the output of the SLIC algorithm. The 10,000 *allowed* supervoxels *generated* 1,755 supervoxels which is the largest number of features used among the models, as shown in Table 5. Machine learning was run separately for each iteration that allowed a different number of supervoxels.

For comparison, random uniformly generated supervoxels and pre-existing brain atlases were used to compare against the proposed supervoxel-based parcellation. Each voxel in the random maps was randomly assigned a supervoxel number. Since the maximum number of actual supervoxels from the SLIC algorithm is 1,755, as a comparison, we generated 4 different random supervoxel models with 1, 10, 100, and 1000 random supervoxels. We compare these random supervoxels to those generated by the SLIC algorithm. We used the automated anatomical labeling (AAL) (n=117 subregions), Brainnetome (n=247 subregions), and Harvard-Oxford cortical atlases (n=68 subregions) (Desikan et al., 2006; Fan et al., 2016; Tzourio-Mazoyer et al., 2002). These were resampled into 3 mm voxel sizes and a $61 \times 73 \times 61$ matrix to match our activation mask.

The supervoxel features were fed into a machine learning algorithm implemented in Python 3.7 to classify effectors (Figure 2f). We use a modified implementation of gradient boosting decision trees (GBDT), known as extreme gradient boosting (XGBoost), which utilizes an ensemble of weak learners or shallow decision trees (Friedman, 2001; Chen, 2016).

2.8.2. XGBoost Machine Learning Classification

XGBoost (Extreme Gradient Boosting) is an ensemble of gradient boosted decision trees that we implemented to predict effectors using beta coefficient maps as features. A decision tree uses a tree-like model of sequential and hierarchical decisions. Decision nodes test for a condition

and pass the instance through a branch to one of its children nodes based on the condition met by the instance. The tree stops growing when the nodes cannot be further split, or the stopping criteria are met. The final outcome or leaf determines the class label. Every instance begins at the decision stump or base of the tree. This instance follows a series of decision nodes until it reaches a leaf predicting a class label.

We can obtain the performance of decision trees with an objective function $\mathcal{L}^m$. This objective function compares the true label and predicted label from the decision trees or base learners. F is a set of f base learners up to M total learners (i.e. the decision trees) (Equation 3).

$$F = \{f_1, f_2, f_3, f_4, \dots, f_M\} \quad (3)$$

The prediction scores from individual trees are summed and averaged to get a final prediction label $\hat{y}_i$ (Equation 4), where x_i is the i^{th} training instance.

$$\hat{y}_n = \frac{1}{m} \sum_{m=1}^M f_m(x_n) \quad (4)$$

The objective function is calculated with a loss function l comparing the true label y_i and predicted label $\hat{y}_i$ across N instances. A regularization term $\Omega(f_m)$ is added to control model complexity and reduce overfitting of the m^{th} tree (Equation 5).

$$\mathcal{L}^m = \sum_{n=1}^N l(y_n, \hat{y}_n) + \Omega(f_m) \quad (5)$$

Equation 6 is the logistic regression loss l between the true and predicted label.

$$l = -(y_n * \log(\hat{y}_n) - (1 - y_n) * (\log(1 - \hat{y}_n))) \quad (6)$$

The regularization term $\Omega(f_m)$ tunes the bias and variance trade-off. T is the total number of leaves and ω_t^2 are the weights from each leaf (Equation 7). Larger weights and larger numbers of leaves

are penalized. Terms γ and λ are tunable in the algorithm. The larger γ and λ are the more conservative the algorithm would be.

$$\Omega(f_m) = \gamma T_t + \frac{1}{2} \lambda \sum_{t=1}^T w_{mt}^2 \quad (7)$$

Boosting is an ensemble learning method of sequentially combining weak learners (i.e. shallow decision trees) to form a strong learner. Each sequential tree aims to reduce the error of the previous tree. The f^{th} tree in boosting is chosen from a forest of decision trees. A forest is generated with a fixed number of decision trees. XGBoost determines the maximum number of trees as each new tree is sequentially tested. The f^{th} tree is selected as the best performing tree in the forest. Equation 8 is the updated prediction label that takes the result from the previous tree. t is the iteration in boosting.

$$\hat{y}_n^{(t)} = \sum_{m=1}^M f_m(x_n) = \hat{y}_n^{(t-1)} + f_t(x_n) \quad (8)$$

Equation 9 is the updated objective function for boosted decision trees. This equation substitutes Equation 8 $\hat{y}_i$ into Equation 5. Now the current model considers the results from the previous model.

$$\mathcal{L}^m = \sum_{n=1}^N l(y_n, \hat{y}_n^{(t-1)} + f_m(x_n)) + \Omega(f_m) \quad (9)$$

The gradient in XGBoost describes the method of minimizing the loss function with gradient descent. The direction of gradient descent is approximated with a 2nd order Taylor series expansion (Equation 10).

$$f(x) \approx f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 \quad (10)$$

Here, $a = \hat{y}_n^{(t-1)}$, $h = f_m(x_n)$, and $f(a) = l(y_n, \hat{y}_n^{(t-1)})$. These are substituted into the Taylor series approximation in Equation 11.

$$\mathcal{L}^m = \sum_{n=1}^N l(y_n, \hat{y}_n^{(t-1)}) + \left(\frac{\partial l(y_n, \hat{y}_n^{(t-1)})}{\partial \hat{y}_n^{(t-1)}} \right) f_m(x_n) + \left(\frac{\partial^2 l(y_n, \hat{y}_n^{(t-1)})}{\partial \hat{y}_n^{(t-1)^2}} \right) f_m(x_n)^2 + \Omega(f_m) \quad (11)$$

Equation 12 simplifies the expansion where $g_i = \left(\frac{\partial l(y_n, \hat{y}_n^{(t-1)})}{\partial \hat{y}_n^{(t-1)}} \right)$ and $h_i = \left(\frac{\partial^2 l(y_n, \hat{y}_n^{(t-1)})}{\partial \hat{y}_n^{(t-1)^2}} \right)$. The final objective function (Equation 12) describes the gradient boosting decision trees.

$$\mathcal{L}^m = \sum_{n=1}^N \left(l(y_n, \hat{y}_n^{(t-1)}) + g_n f_m(x_n) + \frac{1}{2} h_n f_m^2(x_n) \right) + \Omega(f_m) \quad (12)$$

Traditional gradient descent minimizes a cost function $f(x)$ by following the negative gradient of the function at each step. However, the exact minimum is not found after a single step. Convergence can be quite slow and large steps can overshoot the minimum. The 2nd order approximation gives more information in the direction of the minimum. A tree that is evaluated with a full cost function l , has a computation every new split. XGBoost is faster than other gradient boosted tree algorithms because $l(y_n, \hat{y}_n^{(t-1)})$, g_i , and h_i are constants in the approximation and are computed once and reused for all decision tree splits. $f_m(x_n)$ and $\Omega(f_m)$ remain to be calculated each split.

XGBoost enhances the gradient boosting implementation with the inclusion of additional features. These include but are not limited to hardware optimization, parallelized tree building, L1

and L2 regularization, and handling of missing data. XGBoost yields improved model performance and parallelized computational speed compared to other decision tree-based models. This is ideal in comparing multiple models with many features. The algorithm outputs weights for supervoxels (Figure 2g), providing information on which supervoxels (and therefore brain areas) are identified as the most important when classifying effectors (Figure 2h).

We use a nested 5-fold cross-validation (CV) to estimate the unbiased accuracy of each iteration of supervoxel combinations. A traditional training (80% of the data) and testing (20% of the data) split can bias the results depending on the choice of the test set. Nested 5-fold CV provides a more unbiased performance measure. Nested 5-fold CV divides the data into five folds or groups of subjects with an equal number of subjects per fold. Each group is stratified to retain an equal ratio from each effector (hand, foot, and mouth). Nested CV uses an outer loop to estimate the error (Figure 3a), while the inner loop is used for model selection and to tune hyperparameters (Figure 3b). A single iteration of CV uses a single fold for testing and the rest for training. Each loop rotates the training and test folds until all folds have been used for testing once. Performance of the models and hyperparameters are recorded for each loop iteration. We used a broad parameter dictionary to account for the wide range of supervoxels (i.e., 1-10,000) (Table 1). A randomized search of parameters using different combinations was used to determine the ideal parameters in optimizing performance. The parameters with the best classification performance were used in the final model.

A final machine learning model is created for each supervoxel model to generate the supervoxel weights for interpretable analysis. Each final model is trained on the entire dataset. Results and parameters from CV are only used to estimate the performance and are not used to build the final model to avoid bias and overfitting. Each final model produces weights for every

supervoxel included in the model. These weights represent the importance of a supervoxel in determining an effector. The center of mass, volume, and weighting of each supervoxel are then recorded.

2.9 Classification Evaluation Metrics

In the classification analysis, we used a 3D-SLIC algorithm to automatically segment task-based fMRI using supervoxels. The supervoxels computationally group regions based on voxel similarities alone. We compared multiple models of generated supervoxels in classifying different effectors (hand, foot, and mouth). Increasing the number of supervoxels also increases the spatial resolution of brain parcellation. The task-based fMRI was utilized from 17 subjects, each with hand, foot, and mouth fMRI data yielding 51 total data sets. The classification used nested 5-fold CV for each supervoxel model to estimate the performance. An additional and separate model is generated from each supervoxel group using all 51 data sets to produce the weights. To compare the performances between models, we calculated the balanced accuracy, macro-precision, macro-recall, macro-F1 score, and confusion matrices for any given model.

Balanced accuracy is the mean of Sensitivity and Specificity, which is more appropriate for unbalanced data, i.e. data with an unbalanced number of samples for each label. Here sensitivity is the same as recall defined below, and specificity (a.k.a. true negative rate) is the ratio of negative samples correctly predicted among the total number of negative samples. This is because even though we have an equal number of labels for hand, foot, and mouth, stratified 5-fold CV still has slightly unbalanced data in each fold because there are 51 samples in total.

$$Accuracy_B = \frac{Sensitivity + Specificity}{2} \quad (13a)$$

Precision is the fraction of examples predicted to be a certain label that belongs to that label (Eq. 13b). For example, the machine learning model predicts 50 samples to be Foot and among them, only 40 are Foot samples. Then Precision equals $40/50=80\%$. A precision score is calculated for each label, then averaged to obtain a macro-precision score

$$Precision_{l \in \{Hand, Foot, Mouth\}} = \frac{\text{Correctly Predicted \# of Label } l}{\text{Predicted \# of Label } l} \quad (13b)$$

Recall (a.k.a. Sensitivity) measures the fraction of samples belonging to a certain label that are correctly predicted (Eq. 13c). For example, among 60 samples of label Hand, 30 samples are predicted as Hand. Then recall equals $30/60=50\%$. One recall score is calculated for each label. The recall score is averaged across all labels to give a macro-recall score.

$$Recall_{l \in \{Hand, Foot, Mouth\}} = \frac{\text{Correctly Predicted \# of Samples with Label } l}{\text{True \# of Samples with Label } l} \quad (13c)$$

The F1 score takes the harmonic mean (Eq. 13d) of precision and recall of each label, then these F1 scores are averaged to obtain a macro-F1 score across labels (Eq. 13d). F1 score conveys the balance between precision and recall.

$$F1 \text{ Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13d)$$

Lastly, confusion matrices are generated for each supervoxel model to show the distribution of predicted and true labels. A 3-by-3 confusion matrix is generated where each row and column correspond to the Hand, Foot, and Mouth labels. The rows contain the true labels whereas the columns contain the predicted labels, showing the number of different types of predictions. Labels that are correctly predicted (predicting a hand label to be a hand) lie upon the main diagonal. All other values in the confusion matrix represent misclassified samples. All previously mentioned metrics can be calculated from the confusion matrix.

2.10 Occlusion Tests

An occlusion test is used to show the change in the performance of a model by occluding a specific brain region (e.g. set values to zeros). The larger the contribution a brain region has in the classification, the larger the performance degradation in the occluded model. To evaluate the contribution of a certain brain region(s), the values of the brain region(s) are set to zero for all data (both training and test data). Then we use the same machine learning to build a model using data where the region has been lesioned and compute the classification performance on the test data. The change in the classification performance reflects the importance of the corresponding brain region(s). In this work, we use the AAL atlas to occlude specific brain regions (Tzourio-Mazoyer et al., 2002) and set the supervoxel number to 5000 when creating the adaptive supervoxel-based parcellation, as a good balance between the spatial resolution and computation time, leading to 1038 supervoxels in the actual parcellation model.

3. RESULTS

3.1 Force Data

Table 2 shows the mean $\pm$ SD of all behavioral variables assessed. No significant difference was found between effectors for the mean force generated as a percentage of MVC ($F(2,48) = 2.7$; $p = 0.079$), or for the target force error ($F(2,48) = 0.4$; $p = 0.661$). A significant effect of effectors was found for the variability of force generated ($F(2,48) = 43.1$; $p < 0.001$). A Bonferroni post hoc revealed that the variability of force generated by the hand (0.40 ± 0.09 ; $p < 0.001$) and foot (0.62 ± 0.44 ; $p < 0.001$) was greater than the variability of force generated by the mouth (0.15 ± 0.11). No significant difference was found between the variability of force generated by the hand and foot ($p = 0.069$).

3.2 Group Level Voxelwise Analyses

Figure 4A shows activation maps for each of the three effectors based on whole-brain t-test contrasts of each task versus rest ($p = 0.005$; $\alpha = 0.05$; cluster size = 24 voxels). Qualitatively, the well-established somatotopic organization of the foot, hand, and mouth areas of the sensorimotor cortex can be seen. For the foot, activation was localized medially in the paracentral lobule, consistent with the foot and leg area of the sensorimotor cortex. Hand activation was localized laterally in the left hemisphere, consistent with the right hand being used to perform the task. Finally, the mouth task resulted in more lateral and inferior activation in the left hemisphere when compared to the hand, as well as more widespread and bilateral activity. Figure 4B shows cerebellar activation maps for each of the three effectors ($p = 0.005$; $\alpha = 0.05$; cluster size = 24 voxels). For the foot, the cerebellum activity was localized to the right anterior lobules IV-V, as well as bilaterally in posterior lobule VI. The majority of activity during the hand task was

localized to the ipsilateral right anterior and posterior lobule VI. Finally, the mouth task resulted in bilateral activation in posterior lobule VI.

Figure 4C shows results of the separate 1-way ANOVAs used to determine differences in brain activity between effectors. Follow-up t-tests (rows 2, 3, and 4) show which effector(s) differed within each region identified in the ANOVA. For example, the Foot vs Hand follow-up t-test shows significantly greater activation for the foot in the medial sensorimotor cortex, indicated by the warm colors (yellow-red). This medial paracentral lobule region is consistent with prior work localizing peak BOLD responses during ankle movement to MNI coordinates of -7, -38, 74 (Alkadhi et al., 2002) and -7, -35, 73 (Cunningham et al., 2013) (after conversion of coordinates to MNI space), as well as -3, -42, 69 (Debaere et al., 2001). Significantly greater activation for the hand was identified laterally in the left sensorimotor cortex, indicated by the cool colors (blue), and is consistent with hand movement experiments localizing peak BOLD responses to MNI (-38, -24, 56) (Hardwick et al., 2013). Similarly, the Foot vs Mouth follow-up t-test shows significantly greater activation for the foot in the medial sensorimotor cortex, while significantly greater activation for the mouth effector was more lateral, aligning with localization of tongue movement in the left hemisphere (-54, -5, 30; after conversion to MNI space) (Alkadhi et al., 2002). The bottom row shows the Hand vs Mouth comparison revealed significantly greater bilateral activation for the mouth. For contrasts that include the mouth, the left-lateralized clusters extend into the dorsal and ventral premotor cortex, with greater activation for the mouth compared to the foot and hand.

In the cerebellum, Figure 4D shows significantly greater activation for the foot in right lobule IV-V, while the hand effector shows significantly greater activation in right lobule VI. Greater activation for the mouth as compared to the foot was evident in left lobule VI. Finally, the

cerebellar Hand vs Mouth comparison shows significantly greater activation for the hand effector in right lobule VI, while the mouth effector shows significantly greater activation in left lobule VI. Overall, our results showed that 11 whole-brain and 3 cerebellar regions significantly differed between the effectors ($p < 0.005$; $\alpha = 0.05$; cluster size = 24 voxels). These regions, along with the t-value for each of the three effector comparisons, are shown in Table 3.

Effector-independent areas of common activation are shown in Figure 5 and Table 4. We further refine these effector independent regions based on the extent to which they represent abstract motor representations. The activation in SMA and regions of IPL, SPL, and cerebellum (Table 4) are interpreted as likely representing abstract movement representations, whereas common activity in the inferior and middle occipital gyri more likely represent the general processing of online visual information across all tasks.

3.3 Classification Analysis

Table 5 shows the performance of a subset of different models, including balanced accuracy, macro-precision, macro-recall, and macro-F1 score. Across all of these measures, the numbers increase with an increase in the number of supervoxels. For instance, balanced accuracy increased from 0.42 to 0.94 from 1 supervoxel to 1,755 supervoxels. As we increased the number of allowed features, the model is better able to predict which effector was being used. This pattern of accuracy is also shown in Figure 6 across all supervoxel models and indicates that performance surpasses 80% at around 100 supervoxels. With 1,755 supervoxels, our model achieves 94% balanced accuracy. Figure 6 also shows that random supervoxels ($n < 100$) performed worse than random guessing and overall had accuracy under the average segmented supervoxel accuracies. Furthermore, AAL, Brainnetome, and Harvard-Oxford cortical atlases also yield lower accuracies

than the SLIC supervoxel accuracies. Table 5 and Figure 6B show macro-F1 scores (along with macro-precision and macro-recall), which account for the performance across effector labels. The macro-F1 scores are in line with the balanced accuracy scores and increase as the number of allowable features increased. Together these data show that the supervoxel models do well with differentiating effectors with increased spatial resolution.

The individual F1 score for each effector in each model shows the different trends in classification complexity (Figure 6B). Closer inspection shows the difficulty of classifying foot effector maps when lower numbers of supervoxels are used (i.e., lower spatial resolution). This suggests the foot requires greater spatial resolution to better classify. The performance improvement in classifying the mouth effector is not as steep as the foot effector, indicating that the mouth effector is more distinctive in the fMRI data even at relatively lower spatial resolution in the brain parcellation. The hand effector maps fall in between the other two effectors and follow a similar trend to the foot effector maps. Additionally, the scores of all three labels meet around 50 supervoxels suggesting that this is a minimum resolution to classify between the three effectors.

Confusion matrices show the prediction relationships between the true effector labels and predicted effector labels (Figure 7). The supervoxel model with only 1 supervoxel showed poor accuracy in a seemingly random distribution in the confusion matrix because only one supervoxel is used, which grouped the the entire activation mask as as one region. As the supervoxels increase from 1 to 10,000 the main diagonal scores become higher with less misclassification across labels. The misclassifications are fairly distributed as supervoxels increase except for the 10,000 supervoxel models. Three mouth labels were incorrectly classified in the 10,000 supervoxel confusion matrix. The distribution of the misclassifications could simply be a result of noise from the single model. Lastly, an overall confusion matrix was averaged from all confusion matrices to

determine if there were difficult labels across resolutions. The matrix suggests that all models performed equally across all three labels with nearly even scores across the main diagonal.

We identified the top 10 supervoxels with the highest weights from the final machine learning models. These top 10 supervoxels were determined by their final weights after training. We only considered models that were trained by the entire dataset to limit bias from randomized subjects. Then, we localized each supervoxel based on the supervoxel's center of mass. This localization was performed for each of the models with 30 to 10,000 specified supervoxels (a total of 26 models), as the models including < 30 supervoxels produced less than 10 supervoxels. To determine how often a brain region was ranked in the top 10 across all models we counted the number of times a supervoxel's center of mass was identified within an anatomical region. Within a model, it was, therefore, possible that multiple supervoxels could be located within the same anatomical region. Figures 8A and 8B show a supervoxel overlap count on an anatomical template and then over the human motor area template (Mayka et al., 2006). Figure 8A shows that as with the voxelwise analysis the postcentral gyrus cluster extended anteriorly across the precentral gyrus into the primary motor cortex, and premotor cortex, whereas the paracentral lobule was largely centered in medial regions of the primary motor cortex (Figure 8B). Counts within the brain regions are listed in Supplemental Table 6, with a subset of these regions shown in Figure 8C and D. Figure 8C shows that supervoxels were most often localized to the postcentral gyrus ($n = 29$), paracentral lobule ($n = 21$), and cerebellum ($n = 16$). The majority of regions were localized to the left hemisphere consistent with movement of the right effector. We next used an occlusion paradigm to determine the redundancy in somatotopy. We occluded each region and then re-ran the model to see its effect on balanced accuracy. Although Figure 8D shows that removing the left post-central gyrus resulted in accuracy dropping below 80%, in general, the Figure shows that

accuracy remained in the 80-90% range, indicating that no one area was essential in predicting effector.

4. DISCUSSION

This study used fMRI to assess whole-brain and cerebellar specific somatotopic brain activity while force production was precisely controlled by the foot, hand, and mouth. We implemented univariate analyses and multivariate machine learning analyses to characterize, differentiate, and classify brain activation associated with each effector. Together our voxelwise and classification analyses identified effector-dependent activity in regions of the post-central gyrus, precentral gyrus, and paracentral lobule. Univariate voxelwise analyses identified the middle occipital gyrus and putamen as regions showing effector-independent activation. Finally, SMA, regions of the inferior and superior parietal lobule, and cerebellum each contained effector-dependent and effector-independent representations. Machine learning analyses showed that increasing the spatial resolution of the data-driven model increased classification accuracy, which reached 84-95% with supervoxel numbers of around 1000. Our 3D-SLIC-based supervoxel parcellation outperformed classification analyses using established brain templates and random simulations. The occlusion experiments further demonstrated redundancy across the sensorimotor network when classifying effectors. Our observations extend our understanding of effector-dependent and effector-independent organization within the human brain.

Within the cortex, the post-central gyrus, paracentral lobule, precentral gyrus, and inferior and superior parietal lobe were most commonly identified as showing effector-dependent activation. The left post-central gyrus was the region most often identified across classification models and the center of mass of the largest cluster of activation in the voxelwise analysis was centered in the post-central gyrus. The clusters of activation in the post-central gyrus extended anteriorly into the precentral gyrus, while also conforming to the medial to lateral hand to mouth organization in the sensorimotor cortex (Alkadhi et al., 2002; Hardwick et al., 2013; Meier et al.,

2008). The activation in the precentral gyrus extended through M1 into the premotor cortex, with activation differences localized across its dorsal and ventral border. Between effector differences in the premotor cortex were driven by greater activation during the mouth task as compared to both the foot and the hand. Our classification algorithm also exploited the BOLD signal in the medial wall of the pre- and post-central gyrus, with the paracentral lobule region being identified most often after the post-central gyrus. The paracentral lobule underlies the sensorimotor function of the lower limbs (Alkadhi et al., 2002; Cunningham et al., 2013), and was the second-largest cluster of activation identified in the univariate voxelwise analysis.

In humans, somatotopic organization is more discrete and segregated in S1 as compared to the more integrated and overlapping organization in M1 (Hlustík et al., 2001), and our findings are consistent with this. One explanation for the repeated identification of the post-central gyrus across our analyses is sensory feedback (Cunningham et al., 2013; Hlustík et al., 2001; Sanes and Schieber, 2001; Schellekens et al., 2018). Using a population receptive field approach to differentiate within-limb somatotopy of finger digits, Schellekens et al. (Schellekens et al., 2018) demonstrated that receptive fields in M1 are larger compared to S1, such that sensory information occurs at a more specific level in S1 as compared to motor activity in M1. The demonstration of larger receptive fields with less segregation in the precentral gyrus is consistent with primate studies showing extensive intrinsic horizontal connections in M1 (Huntley and Jones, 1991; Wu et al., 2000). Additional support for our findings comes from evidence showing that upper extremity nerve fibers are 75-95% sensory, with the ratio of sensory to motor fibers increasing when moving distally. As such, greater amounts of sensory feedback are required to perform more precise movements (Gesslbauer et al., 2017), as was the case in the current study given the real-time visual feedback presented to the subjects. Previous somatotopy studies often use visual

stimuli to cue individuals when to move which effector (Grodd et al., 2001b; Lotze et al., 2000; Zeharia et al., 2015), but do not always provide real-time visual feedback to the subject as in the current study. Here, for the first time, we controlled and normalized target force amplitude across effectors using precise visual feedback and extend the literature by identifying a preference for between limb somatotopic organization in the post-central gyrus and paracentral lobule as compared to the precentral gyrus.

Both effector independent and effector dependent activation was identified in the regions of SMA, IPL, and SPL. The effector-independent activation in SMA and bilateral middle-to-anterior intraparietal sulcus in the parietal lobe has been shown in individuals with congenitally absent upper limbs when pointing with their foot (Liu et al., 2020). IPL and SPL regions in the current study were within 5mm of each other in the Y and Z direction and near the intraparietal sulcus (Liu et al., 2020). The parietal lobe was first identified as a brain region with the capacity to store movement representations over 100 years ago (Liepmann, 1905). More recent evidence from left and right interlimb coordination tasks have further confirmed the existence of abstract movement encoding in the parietal cortex, which in combination with the premotor cortex, form the highest level in the action representation hierarchy (Swinnen et al., 2010). In contrast, effector-dependent activation in IPL was more anterior and inferior and bordered the post-central gyrus. A similar pattern emerged in SPL with effector-dependent activity bordering the post-central gyrus in more anterior and superior regions compared to effector-independent activity. Somatotopic organization has also been shown in SMA, with a rostral-caudal face-to-foot organization (Chainay et al., 2004; Zeharia et al., 2012), and in the putamen, with a ventral-to-rostral foot-to-face organization (Zeharia et al., 2015). Five of our classification models identified SMA as effector-dependent; however, these supervoxels were large (>450 voxels), and extended posteriorly into

the paracentral lobule. Likewise, a single classification model identified the putamen as an effector-dependent region. The limited number of hits across classification models and the absence of voxelwise effector-dependent activation in these regions may be the result of large clusters and widespread activation across the visuomotor system associated with the implemented visuomotor task. Previous evidence using interlimb coordination tasks across the foot and the hand have identified many of these same regions (IPL, SPL, SMA, putamen) as encoding abstract movement representations and our findings are generally consistent with these data (Swinnen et al., 2010). A key difference between our findings and those of Swinnen et al. was the identification of effector-independent activation in the premotor cortex. We found some evidence of effector-dependence activation at the border of the dorsal and ventral premotor cortex, with the cluster of activation extending from its center of mass in the post-central gyrus anteriorly to the precentral gyrus. Importantly, follow-up contrasts showed greater activity for the mouth task relative to the foot and the hand, but no difference between the foot and hand tasks. A key difference, therefore, is the inclusion of the mouth task in the current study. Indeed, consistent with Swinnen et al, we did not find evidence of activation differences between the foot and hand tasks in the premotor cortex. The bilateral nature of the mouth task, coupled with its novelty compared to the foot and hand, may help explain the activation differences. Common activation was also evident in the middle and inferior occipital gyrus, consistent with other visuomotor studies that use real-time feedback (Coombes et al., 2010; Vaillancourt et al., 2003). While these regions are grouped with effector-independent regions, we further refine our interpretation of their activation in the current study as being more related to the general processing of online visual information and less to the abstract coding of movement representation. Indeed, our paradigm differed from many other somatotopy studies in the extent to which the visuomotor network was engaged, and this may have masked the

more fine-grained somatotopic organization that others have observed in these regions, while also accounting for the addition of other regions in the visual cortex that were commonly active across tasks.

Our observations reveal that the cerebellum contains both effector-dependent and effector-independent representations. The cerebellum was the third most commonly identified effector-dependent region. The majority of these regions were in the left hemisphere contralateral to the hand and foot, likely driven by the machine learning algorithm exploiting the bilateral activation associated with the mouth task (Figure 4B). However, consistent with prior literature (Boillat et al., 2020; Mottolese et al., 2013; Nitschke et al., 1996; Schlerf et al., 2010; Stoodley and Schmahmann, 2009), our voxelwise analysis did identify foot- and hand-related effector-dependent activity in right (ipsilateral) lobules IV-V, in addition to left lobule VI activity associated with the mouth task. Four different homunculi have been identified in the cerebellum (Manni and Petrosini, 2004), and our findings are consistent with one of these in the anterior lobe. Indeed, anterior to posterior, foot-to-head organization has previously been shown for lobules II-VI (Boillat et al., 2020; Grodd et al., 2001a; Nitschke et al., 1996; Schlerf et al., 2010), with activity related to the foot also more medial relative to the hand (Debaere et al., 2001), consistent with our data (see Figure 4D, row 2). Additionally, some evidence of effector-independent cerebellar regions also exists (Liu et al., 2020). We found effector-independent activity in left lobule VI (-25, -65, -26), close to the region of lobule VI that we previously identified as responding to multimodal pain and motor stimuli (-26, -65, -24: Coombes and Misra, 2016). Other evidence points to regions of Lobule VI being involved in early learning during explicit sequence learning tasks (Bernard and Seidler, 2013), suggesting more generalized effector and modality independent sensorimotor processing within this cerebellar region. Invasive and non-invasive stimulation

studies in rodents and humans have targeted the cerebellum (Grimaldi et al., 2014; Hardwick and Celnik, 2014; Miterko et al., 2021; Reis et al., 2014), and our findings add to the growing literature informing the stimulation of effector-dependent and effector-independent cerebellar regions.

The effector classification performance increased with an increasing number of supervoxels, reaching 94% accuracy with 10,000 specified supervoxels (1,755 actual supervoxels). We varied the total number of supervoxels allowed across models to identify the extent to which coarse to fine parcellation alters classification accuracy. As expected, coarser parcellation using fewer and larger supervoxels led to lower accuracy whereas finer parcellation using more and smaller supervoxels led to higher accuracy. Our approach outperformed standardized brain atlases in classifying effectors. It is interesting to note that while all three standardized brain atlases perform worse than the supervoxel-derived atlas, the Brainnetome atlas (Fan et al., 2016) achieved the highest accuracy among the three standardized atlases. The Brainnetome includes subdivisions in the precentral and post-central gyrus for the hand and face region, trunk, upper limb, as well as lower limb subdivisions in the paracentral lobule. The AAL does not subdivide the precentral gyrus but does include the paracentral lobule (Tzourio-Mazoyer et al., 2002). Finally, the Harvard atlas includes the entire pre- and post-central gyrus without any subdivisions (Desikan et al., 2006). It is therefore not surprising that classification accuracy improves with increasing granularity, especially within the post-central gyrus, which across our models, was the most frequently identified region that contributed to the classification accuracy. Our data demonstrate that predefined atlases lead to classification accuracies in the 70-80% range, whereas the current supervoxel approach reached 94%. A supervoxel approach has previously been used to segment and recognize 3D boundaries and shapes in mitochondria using electron microscopy (Lucchi et al., 2012), livers using volumetric CT (Wu et al., 2016), brain tumors in MRI (Soltaninejad et al.,

2018), and resting-state fMRI networks (Wang and Wang, 2016) with success. Here, for the first time, we implement a supervoxel approach to classify the movement of different effectors using fMRI data.

Both multivariate pattern analysis (MVPA) and classification approaches have been used to determine how well brain activity can distinguish different limbs performing similar movements (Bhattacharyya et al., 2010; Gallivan et al., 2013; Morash et al., 2008; Zhao et al., 2019), different movements performed by the same limb (Gallivan et al., 2013; Robinson et al., 2013; Shiman et al., 2017), and imagining as compared to executing movements (Morash et al., 2008; Sleight et al., 2009). The frontoparietal motor network has been consistently identified across these studies as having the ability to distinguish between different movements. This is evident despite variations in the movements being made, the neuroimaging technique being employed, and the analytical approach being taken. Indeed, the redundancy in the system, in terms of how many areas show somatotopic organization, was confirmed in our occlusion analysis which demonstrated that even when removing the most commonly identified brain region, classification accuracy remained ~80% or higher.

The SLIC algorithm has advantages over other clustering-based methods in our generalizable approach; however, there may be cases that SLIC may be unreliable. For instance, SLIC is computationally faster than other clustering algorithms. It accomplishes this by searching in a limited area surrounding the cluster. However, this may not account for joint activations between two separate regions. That is, although control of supervoxel compactness and size gives us the flexibility to conform to anatomical brain regions, there may be a tradeoff of flexibility and interpretability. Too much flexibility may allow supervoxels to associate unrelated activations together. However, we compensate for this by using a range of supervoxel counts. Finer

granularities and coarser parcellations are both generated to account for supervoxel variances. Furthermore, SLIC generates fixed initialized seeds with equidistant points arranged in a grid-like fashion. This ensures repeatability in experiments as long as the data remains consistent in size. However, this grid-based initialization may be a disadvantage if the clusters are tightly packed together.

Many studies have used EEG measurements to classify movement, given its logistical and practical advantages in real-world settings as compared to the confines of fMRI. Sleight et al, assayed cortical dynamics using 64 channel EEG while subjects imagined or executed unimanual or bimanual hand and foot flexion. Classification at the individual level reached 69% which dropped to 56% at the group level (Sleight et al., 2009). Individual-level classification accuracy reached 84% when differentiating left- and right-hand movements performed by one individual using EEG data collected from 2 electrodes (Bhattacharyya et al., 2010). As one would expect, classification accuracy drops when the movements being made are subtler. Morash et al. required subjects to execute or imagine a right-hand squeeze, left-hand squeeze, press of the tongue on the roof of the mouth, or a right foot toe curl while EEG data were collected (Morash et al., 2008). The classification accuracy was at or below 60% for executed movements and at or below 50% for imagined movements. The inclusion of the tongue press may have attenuated classification accuracy, with tongue/lip/mouth movements being harder to classify as compared to movements of the hand and foot (Zhao et al., 2019). Within-limb studies have also used EEG to show that forward reaching movements to different targets can be classified with between 50-80% accuracy depending on the number of targets being decoded and the number of subjects being tested (Robinson et al., 2013; Shiman et al., 2017). In each case, signals over the parietal cortex prominently contributed to classification accuracy, consistent with our observations. Frontal and

parietal brain areas have also been identified in fMRI studies that have classified eye and hand movements with ~55-65 accuracy (Gallivan et al., 2011), and the planning and execution of reaching and grasping movements with the left and right hand with 55-70% accuracy (Gallivan et al., 2013). Executing, observing, and imagining reaching movements of the same limb has been classified with 80% (Filimon et al., 2015) and within-subject mouth movements with 90% accuracy (Bleichner et al., 2015). Other fMRI studies have implemented MVPA approaches within specific regions such as M1 during hand and foot rotation (Shay et al., 2019). General linear modeling approaches have also been implemented to determine differences in brain activation during flexion of the right finger, elbow, and ankle (Cunningham et al., 2013). We extend the current literature by using fMRI data and a generalizable classifier to differentiate three different effectors performing the same task with the same relative amount of force. Our novel approach achieved 95% accuracy using 1755 supervoxels. This is notable because we implemented a whole-brain analysis approach that also allowed us to vary the number of supervoxels across different models, while also using a between- rather than within-subjects classification approach to increase the generalizability of our findings.

REFERENCES

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 34, Issue 11). <https://doi.org/10.1109/TPAMI.2012.120>
- Ai, L., Xiong, J., 2016. Temporal-spatial mean-shift clustering analysis to improve functional MRI activation detection. *Magn. Reson. Imaging* 34, 1283–1291.
<https://doi.org/10.1016/J.MRI.2016.07.009>
- Alkadhi, H., Crelier, G.R., Boendermaker, S.H., Golay, X., Hepp-Reymond, M.C., Kollias, S.S., 2002. Reproducibility of primary motor cortex somatotopy under controlled conditions. *Am. J. Neuroradiol.* 23, 1524–1532.
- Batson, M.A., Petridou, N., Klomp, D.W.J., Frens, M.A., Neggers, S.F.W., 2015. Single session imaging of cerebellum at 7 Tesla: obtaining structure and function of multiple motor subsystems in individual subjects. *PLoS One* 10, e0134933.
<https://doi.org/10.1371/journal.pone.0134933>
- Bernard, J.A., Seidler, R.D., 2013. Cerebellar contributions to visuomotor adaptation and motor sequence learning: An ALE meta-analysis. *Front. Hum. Neurosci.*
<https://doi.org/10.3389/fnhum.2013.00027>
- Bhattacharyya, S., Khasnobish, A., Chatterjee, S., Konar, A., Tibarewala, D.N., 2010. Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data, in: *International Conference on Systems in Medicine and Biology, ICSMB 2010 - Proceedings*. pp. 126–131.
<https://doi.org/10.1109/ICSMB.2010.5735358>
- Bleichner, M.G., Jansma, J.M., Salari, E., Freudenburg, Z. V., Raemaekers, M., Ramsey, N.F.,

2015. Classification of mouth movements using 7 T fMRI. *J. Neural Eng.* 12.
<https://doi.org/10.1088/1741-2560/12/6/066026>
- Boillat, Y., Bazin, P.L., van der Zwaag, W., 2020. Whole-body somatotopic maps in the cerebellum revealed with 7T fMRI. *Neuroimage* 211, 116624.
<https://doi.org/10.1016/j.neuroimage.2020.116624>
- Chainay, H., Krainik, A., Tanguy, M.-L., Gerardin, E., Le Bihan, D., Lehericy, S., 2004. Foot, face and hand representation in the human supplementary motor area. *Neuroreport* 15, 765–769. <https://doi.org/10.1097/00001756-200404090-00005>
- Chyzyk, D., Graña, M., 2014. Findings in resting-state fMRI by differences from K-means clustering. *Stud. Health Technol. Inform.* 207, 300–310. <https://doi.org/10.3233/978-1-61499-474-9-300>
- Coombes, S. a S.A., Corcos, D.M.D.M., Sprute, L., Vaillancourt, D.E.D.E., 2010. Selective regions of the visuomotor system are related to gain-induced changes in force error. *J. Neurophysiol.* 103, 2114–23. <https://doi.org/10.1152/jn.00920.2009>
- Coombes, S.A., Misra, G., 2016. Pain and motor processing in the human cerebellum. *Pain* 157, 117–127. <https://doi.org/10.1097/j.pain.0000000000000337>
- Cunningham, D.A., Machado, A., Yue, G.H., Carey, J.R., Plow, E.B., 2013. Functional somatotopy revealed across multiple cortical regions using a model of complex motor task. *Brain Res.* 1531, 25–36. <https://doi.org/10.1016/j.brainres.2013.07.050>
- Debaere, F., Swinnen, S.P., Béatse, E., Sunaert, S., Van Hecke, P., Duysens, J., 2001. Brain areas involved in interlimb coordination: A distributed network. *Neuroimage* 14, 947–958.
<https://doi.org/10.1006/nimg.2001.0892>
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L.,

- Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.
<https://doi.org/10.1016/j.neuroimage.2006.01.021>
- Diedrichsen, J., 2006. A spatially unbiased atlas template of the human cerebellum. *Neuroimage* 33, 127–138. [https://doi.org/S1053-8119\(06\)00641-0](https://doi.org/S1053-8119(06)00641-0)
[pii]10.1016/j.neuroimage.2006.05.056
- Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., Ramnani, N., 2009. A probabilistic MR atlas of the human cerebellum. *Neuroimage* 46, 39–46.
<https://doi.org/10.1016/j.neuroimage.2009.01.045>S1053-8119(09)00080-9 [pii]
- Diedrichsen, J., Maderwald, S., Küper, M., Thürling, M., Rabe, K., Gizewski, E.R., Ladd, M.E., Timmann, D., 2011. Imaging the deep cerebellar nuclei: A probabilistic atlas and normalization procedure. *Neuroimage* 54, 1786–1794.
<https://doi.org/10.1016/j.neuroimage.2010.10.035>
- Diedrichsen, J., Zotow, E., 2015. Surface-Based Display of Volume-Averaged Cerebellar Imaging Data. *PLoS One* 10, e0133402. <https://doi.org/10.1371/JOURNAL.PONE.0133402>
- Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2016. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex* 26, 3508–3526.
<https://doi.org/10.1093/cercor/bhw157>
- Filimon, F., Rieth, C.A., Sereno, M.I., Cottrell, G.W., 2015. Observed, Executed, and Imagined Action Representations can be Decoded From Ventral and Dorsal Areas. *Cereb. Cortex* 25, 3144–3158. <https://doi.org/10.1093/cercor/bhu110>

- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29, 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Gallivan, J.P., Adam McLean, D., Smith, F.W., Culham, J.C., 2011. Decoding effector-dependent and effector-independent movement intentions from human parieto-frontal brain activity. *J. Neurosci.* 31, 17149–17168. <https://doi.org/10.1523/JNEUROSCI.1058-11.2011>
- Gallivan, J.P., McLean, D.A., Flanagan, J.R., Culham, J.C., 2013. Where One Hand Meets the Other: Limb-Specific and Action-Dependent Movement Plans Decoded from Preparatory Signals in Single Human Frontoparietal Brain Areas. *J. Neurosci.* 33, 1991–2008. <https://doi.org/10.1523/JNEUROSCI.0541-12.2013>
- Gesslbauer, B., Hruby, L.A., Roche, A.D., Farina, D., Blumer, R., Aszmann, O.C., 2017. Axonal components of nerves innervating the human arm. *Ann. Neurol.* 82, 396–408. <https://doi.org/10.1002/ana.25018>
- Grimaldi, G., Argyropoulos, G.P., Boehringer, A., Celnik, P., Edwards, M.J., Ferrucci, R., Galea, J.M., Groiss, S.J., Hiraoka, K., Kassavetis, P., Lesage, E., Manto, M., Miall, R.C., Priori, A., Sadnicka, A., Ugawa, Y., Ziemann, U., 2014. Non-invasive cerebellar stimulation - A consensus paper. *Cerebellum*. <https://doi.org/10.1007/s12311-013-0514-7>
- Grodd, W., Hülsmann, E., Lotze, M., Wildgruber, D., Erb, M., 2001a. Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Hum. Brain Mapp.* 13, 55–73. <https://doi.org/10.1002/hbm.1025>
- Grodd, W., Hülsmann, E., Lotze, M., Wildgruber, D., Erb, M., Hülsmann, E., Lotze, M., Wildgruber, D., Erb, M., 2001b. Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Hum Brain Mapp* 13, 55–73. <https://doi.org/10.1002/hbm.1025>

- Hardwick, R.M., Celnik, P.A., 2014. Cerebellar direct current stimulation enhances motor learning in older adults. *Neurobiol. Aging* 35, 2217–2221.
<https://doi.org/10.1016/j.neurobiolaging.2014.03.030>
- Hardwick, R.M., Rottschy, C., Miall, R.C., Eickhoff, S.B., 2013. A quantitative meta-analysis and review of motor learning in the human brain. *Neuroimage* 67, 283–297.
<https://doi.org/10.1016/j.neuroimage.2012.11.020>
- Hlustík, P., Solodkin, A., Gullapalli, R.P., Noll, D.C., Small, S.L., 2001. Somatotopy in human primary motor and somatosensory hand representations revisited. *Cereb. Cortex* 11, 312–321. <https://doi.org/10.1093/cercor/11.4.312>
- Huntley, G.W., Jones, E.G., 1991. Relationship of intrinsic connections to forelimb movement representations in monkey motor cortex: a correlative anatomic and physiological study. *J. Neurophysiol.* 66, 390–413. <https://doi.org/10.1152/jn.1991.66.2.390>
- Liepmann, H., 1905. Die linke Hemisphere und das Handeln, *Munchener Medizinische Wochenschrift*. Munch. Med. Wochenschr. 49.
- Liu, Y., Vannuscorps, G., Caramazza, A., Striem-Amit, E., 2020. Evidence for an effector-independent action system from people born without hands. *Proc. Natl. Acad. Sci. U. S. A.* 117, 28433–28441. <https://doi.org/10.1073/pnas.2017789117>
- Lotze, M., Erb, M., Flor, H., Huelsmann, E., Godde, B., Grodd, W., 2000. fMRI evaluation of somatotopic representation in human primary motor cortex. *Neuroimage* 11, 473–481.
<https://doi.org/10.1006/nimg.2000.0556>
- Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P., 2012. Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. *IEEE Trans. Med. Imaging* 31, 474–486. <https://doi.org/10.1109/TMI.2011.2171705>

- Manni, E., Petrosini, L., 2004. A century of cerebellar somatotopy: a debated representation. *Nat. Rev. Neurosci.* 5, 241–249. <https://doi.org/10.1038/nrn1347>
- Mayer, A.R., Zimbelman, J.L., Watanabe, Y., Rao, S.M., 2001. Somatotopic organization of the medial wall of the cerebral hemispheres: a 3 Tesla fMRI study. *Neuroreport* 12, 3811–3814. <https://doi.org/10.1097/00001756-200112040-00042>
- Mayka, M.A., Corcos, D.M., Leurgans, S.E., Vaillancourt, D.E., 2006. Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: a meta-analysis. *Neuroimage* 31, 1453–1474. <https://doi.org/10.1016/j.neuroimage.2006.02.004>
- Meier, J.D.J.D., Aflalo, T.N.T.N., Kastner, S., Graziano, M.S.A., 2008. Complex organization of human primary motor cortex: a high-resolution fMRI study. *J Neurophysiol* 100, 1800–1812. <https://doi.org/10.1152/jn.90531.2008> [pii]
- Miterko, L.N., Lin, T., Zhou, J., van der Heijden, M.E., Beckinghausen, J., White, J.J., Sillitoe, R. V., 2021. Neuromodulation of the cerebellum rescues movement in a mouse model of ataxia. *Nat. Commun.* 12, 1–16. <https://doi.org/10.1038/s41467-021-21417-8>
- Morash, V., Bai, O., Furlani, S., Lin, P., Hallett, M., 2008. Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries. *Clin. Neurophysiol.* 119, 2570–2578. <https://doi.org/10.1016/j.clinph.2008.08.013>
- Mottollese, C., Richard, N., Harquel, S., Szathmari, A., Sirigu, A., Desmurget, M., 2013. Mapping motor representations in the human cerebellum. *Brain* 136, 330–342. <https://doi.org/10.1093/brain/aws186> [pii]
- Nitschke, M.F., Kleinschmidt, A., Wessel, K., Frahm, J., 1996. Somatotopic motor representation in the human anterior cerebellum. A high-resolution functional MRI study.

- Brain 119, 1023–1029. <https://doi.org/10.1093/brain/119.3.1023>
- Penfield, W., Boldrey, E., 1937. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain 60, 389–443.
<https://doi.org/10.1093/brain/60.4.389>
- Penfield, W., Rasmussen, T., 1950. The cerebral cortex of man; a clinical study of localization of function. Macmillan, Oxford.
- Reis, J., Schambra, H.M., Cohen, L.G., Buch, E.R., Fritsch, B., Zarahn, E., Celnik, P.A., Krakauer, J.W., Bizzi, E., Reisab, J., Schambraa, H.M., Cohen-, L.G., Buchar, E.R., Fritschab, B., Zarahns, E., Celnikd, P.A., 2014. enhances motor skill Noninvasive cortical stimulation an effect acquisition over multiple days through on consolidation.
- Robinson, N., Guan, C., Vinod, A.P., Keng Ang, K., Peng Tee, K., 2013. Multi-class EEG classification of voluntary hand movement directions. J. Neural Eng. 10, 056018.
<https://doi.org/10.1088/1741-2560/10/5/056018>
- Sanes, J.N., Schieber, M.H., 2001. Orderly somatotopy in primary motor cortex: does it exist? Neuroimage 13, 968–974. <https://doi.org/10.1006/nimg.2000.0733>
- Schellekens, W., Petridou, N., Ramsey, N.F., 2018. Detailed somatotopy in primary motor and somatosensory cortex revealed by Gaussian population receptive fields. Neuroimage 179, 337–347. <https://doi.org/10.1016/j.neuroimage.2018.06.062>
- Schlerf, J.E., Verstynen, T.D., Ivry, R.B., Spencer, R.M.C., 2010. Evidence of a novel somatopic map in the human neocerebellum during complex actions. J Neurophysiol 103, 3330–3336.
<https://doi.org/10.1152/jn.01117.2009> [pii]
- Shaprio, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>

- Shay, E.A., Chen, Q., Garcea, F.E., Mahon, B.Z., 2019. Decoding intransitive actions in primary motor cortex using fMRI: toward a componential theory of ‘action primitives’ in motor cortex. *Cogn. Neurosci.* 10, 13–19. <https://doi.org/10.1080/17588928.2018.1453491>
- Shiman, F., López-Larraz, E., Sarasola-Sanz, A., Irastorza-Landa, N., Spüler, M., Birbaumer, N., Ramos-Murguialday, A., 2017. Classification of different reaching movements from the same limb using EEG. *J. Neural Eng.* 14, 046018. <https://doi.org/10.1088/1741-2552/aa70d2>
- Sleight, J., Pillai, P., Mohan, S., 2009. Classification of Executed and Imagined Motor Movement EEG Signals.
- Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T.L., Barrick, T.R., Howe, F.A., Ye, X., 2018. Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels. *Comput. Methods Programs Biomed.* 157, 69–84. <https://doi.org/10.1016/j.cmpb.2018.01.003>
- Spraker, M.B., Corcos, D.M., Kurani, A.S., Prodoehl, J., Swinnen, S.P., Vaillancourt, D.E., 2012. Specific cerebellar regions are related to force amplitude and rate of force development. *Neuroimage* 59, 1647–1656. <https://doi.org/10.1016/j.neuroimage.2011.09.019>
- Spraker, M.B., Yu, H., Corcos, D.M., Vaillancourt, D.E., 2007. Role of individual basal ganglia nuclei in force amplitude generation. *J Neurophysiol* 98, 821–834. <https://doi.org/00239.2007> [pii]10.1152/jn.00239.2007
- Stoodley, C.J., Schmahmann, J.D., 2009. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* 44, 489–501. <https://doi.org/10.1016/j.neuroimage.2008.08.039>S1053-8119(08)00972-5 [pii]

- Swinnen, S.P., Vangheluwe, S., Wagemans, J., Coxon, J.P., Goble, D.J., Van Impe, A., Sunaert, S., Peeters, R., Wenderoth, N., 2010. Shared neural resources between left and right interlimb coordination skills: The neural substrate of abstract motor representations. *Neuroimage* 49, 2570–2580. <https://doi.org/10.1016/j.neuroimage.2009.10.052>
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. <https://doi.org/10.1006/nimg.2001.0978>
- Vaillancourt, D.E., Thulborn, K.R., Corcos, D.M., 2003. Neural basis for the processes that underlie visually guided and internally guided force control in humans. *J. Neurophysiol.* 90, 3330–3340. <https://doi.org/10.1152/jn.00394.2003>
- van der Zwaag, W., Kusters, R., Magill, A., Gruetter, R., Martuzzi, R., Blanke, O., Marques, J.P., 2013. Digit somatotopy in the human cerebellum: a 7T fMRI study. *Neuroimage* 67, 354–362. <https://doi.org/10.1016/j.neuroimage.2012.11.041>
- Venkataraman, A., Van Dijk, K., Buckner, R., Golland, P., 2009. EXPLORING FUNCTIONAL CONNECTIVITY IN FMRI VIA CLUSTERING. *Proc. ... IEEE Int. Conf. Acoust. Speech, Signal Process. ICASSP 2009*, 441–444. <https://doi.org/10.1109/ICASSP.2009.4959615>
- Wang, J., Wang, H., 2016. A supervoxel-based method for groupwise whole brain parcellation with resting-state fMRI data. *Front. Hum. Neurosci.* 10, 659. <https://doi.org/10.3389/fnhum.2016.00659>
- Wu, C.W., Bichot, N.P., Kaas, J.H., 2000. Converging evidence from microstimulation, architecture, and connections for multiple motor areas in the frontal and cingulate cortex of prosimian primates. *J. Comp. Neurol.* 423, 140–177. <https://doi.org/10.1002/1096->

9861(20000717)423:1<140::aid-cne12>3.0.co;2-3

Wu, W., Zhou, Z., Wu, S., Zhang, Y., 2016. Automatic Liver Segmentation on Volumetric CT Images Using Supervoxel-Based Graph Cuts. *Comput. Math. Methods Med.* 2016.

<https://doi.org/10.1155/2016/9093721>

Zeharia, N., Hertz, U., Flash, T., Amedi, A., 2015. New Whole-Body Sensory-Motor Gradients Revealed Using Phase-Locked Analysis and Verified Using Multivoxel Pattern Analysis and Functional Connectivity. *J. Neurosci.* 35, 2845–2859.

<https://doi.org/10.1523/Jneurosci.4246-14.2015>

Zeharia, N., Hertz, U., Flash, T., Amedi, A., 2012. Negative blood oxygenation level dependent homunculus and somatotopic information in primary motor cortex and supplementary motor area. *Proc. Natl. Acad. Sci. U. S. A.* 109, 18565–18570.

<https://doi.org/10.1073/pnas.1119125109>

Zhao, M., Marino, M., Samogin, J., Swinnen, S.P., Mantini, D., 2019. Hand, foot and lip representations in primary sensorimotor cortex: a high-density electroencephalography study. *Sci. Rep.* 9, 1–12. <https://doi.org/10.1038/s41598-019-55369-3>

Table 1. List of the parameters and values considered. A randomized search was performed to sample multiple combinations of parameters to optimize performance.

Parameters	Value
Learning_rate	0.01, 0.05, 0.10, 0.20, 0.30, 0.40
n_estimators	50, 100, 500, 1000
max_depth	5, 10, 15, 20, 25
subsample	0.5, 0.6, 0.7, 0.8, 0.9, 1.0
colsample_bytree	0.5, 0.6, 0.7, 0.8, 0.9, 1.0
scale_pos_weight	30, 40, 50, 300, 400, 500, 600, 700
gamma	0.00, 0.05, 0.10, 0.15, 0.20, 1, 5

Table 2. Mean $\pm$ SD of force, force variability, and target force error for the hand, foot, and mouth effector.

Effector	Mean Force	Force Variability (SD)	Target Force Error (RMSE)
Hand	9.33 $\pm$ 0.64	0.40 $\pm$ 0.09	0.74 $\pm$ 0.70
Foot	9.53 $\pm$ 0.97	0.62 $\pm$ 0.44	0.70 $\pm$ 0.28
Mouth	9.86 $\pm$ 0.34	0.15 $\pm$ 0.11	0.59 $\pm$ 0.16

Table 3. Clusters showing significant ($p = 0.005$; $\alpha = 0.05$; cluster size = 24 voxels) effector-dependent activity based on the ANOVA, and the between effector differences for each t-test. MNI coordinates indicate the cluster center of mass. T-values for each effector comparison are taken from the voxel with the maximum difference between effectors within the cluster. Only significant t-values are shown. (R=right, L=left, F=foot, H=hand, M=mouth).

Cluster Size (voxels)	MNI Coordinates (x, y, z)	Anatomical Location	F vs H (t-value)	F vs M (t-value)	H vs M (t-value)
678	-48, -6, 39	L Postcentral Gyrus (Hand area)	-7.07 (F<H)	-3.97 (F<M)	-5.75 (H<M)
430	-6, -32, 67	L Paracentral Lobule (Leg area)	4.01 (F>H)	3.10 (F>M)	-
126	-35, -67, 30	L Middle Occipital Gyrus	-	-	-3.46 (H<M)
121	48, -20, 13	R Rolandic Operculum	3.90 (F>H)	-	-3.99 (H<M)
111	46, -61, 34	R Angular Gyrus	-	-3.72 (F<M)	-3.41 (H<M)
85	45, -16, 43	R Precentral Gyrus	-	-	-4.17 (H<M)
45	21, -77, 32	R Superior Occipital Gyrus	-	-	-3.57 (H<M)
36	-46, -28, 9	L Superior Temporal Gyrus	-	-	-4.11 (H<M)
36	-41, -72, 15	L Middle Occipital Gyrus	-	-	-3.77 (H<M)
26	-38, -18, 0	L Insula Lobe	-	-	-4.21 (H<M)
25	44, 1, -34	R Inferior Temporal Gyrus	-	-	5.43 (H>M)
82	17, -56, -20	R Cerebellum Lobules IV-V	-7.66 (F<H)	-	5.11 (H>M)
53	-14, -59, -20	L Cerebellum Lobule VI	-	-4.91 (F<M)	-4.15 (H<M)
27	27, -48, -38	R Cerebellum Lobule VI	-	-	-

Table 4. Cluster size, MNI coordinate, and anatomical location of common effector-independent brain activity common across all three effectors (foot, hand, and mouth).

Cluster Size (voxels)	MNI Coordinates (x, y, z)	Anatomical Location
196	1, -1, 61	Supplementary Motor Area
83	33, -88, -3	Right Inferior Occipital Gyrus
79	-26, 2, -1	Left Putamen
56	-30, -90, -4	Left Middle Occipital Gyrus
35	34, -50, 53	Right Inferior Parietal Lobule
27	-34, -53, 58	Left Superior Parietal Lobule
24	-13, -19, 6	Left Thalamus
66	-25, -65, -26	Left Cerebellum Lobule VI
44	-14, -72, -45	Left Cerebellum Lobule VIII
32	6, -74, -40	Right Cerebellum Lobule VII

Table 5. Evaluation metrics. Accuracy and macro-F1 score increase as the number of supervoxels increases. The maximum number of supervoxels represents the number input into the SLIC algorithm, whereas the number of actual supervoxels represents the actual supervoxels used for classification. Accuracy, precision, recall, and F1 score show an increase with an increase in spatial resolution (i.e., number of supervoxels). The F1 score paired with the accuracy indicates the model did not have any major issues with misclassified labels.

Max # of Supervoxels	# of Actual Supervoxels	Balanced Accuracy	Macro Precision	Macro Recall	Macro F1 Score
1	1	0.42	0.37	0.42	0.38
10	6	0.49	0.48	0.49	0.46
100	53	0.85	0.90	0.85	0.86
1000	285	0.84	0.89	0.84	0.84
10000	1755	0.94	0.95	0.94	0.94

Figure 1. Experimental setup and paradigm. (A) Custom-designed foot force device (left). Force was generated by dorsiflexion of the right foot and measured by the force transducer. Exemplary force time series (right) shows the 10% MVC target force (green), as well as the force produced (black). (B) Hand force device which measured right-hand pinch force. (C) The custom-designed bite device. Force was generated against the bite plates held in the subject's mouth and was measured by the force transducer. (D) Depiction of the task paradigm. After a 30s rest period, the red force bar turned green to cue force production. The subject then produced force to match the white target force bar for 4s. When the bar turned red, subjects rested for 2s. This pattern of 4s force and 2s rest, repeated 5 times during the 30s force block.

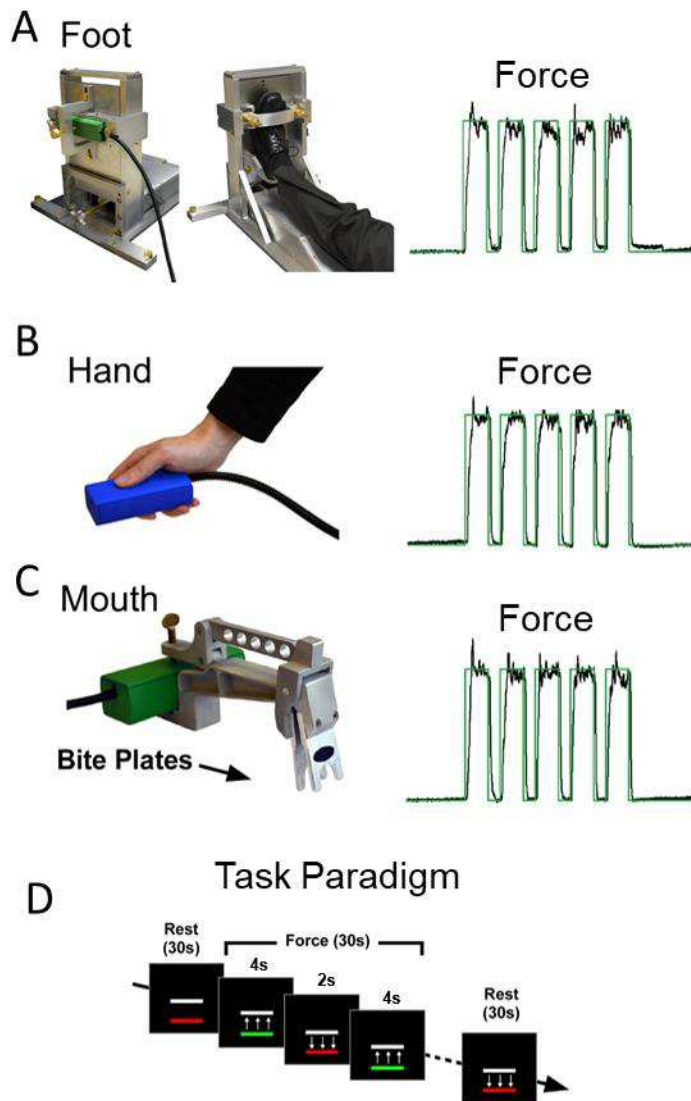


Figure 2. The pipeline serves to generate a uniform supervoxel map which is used for all subjects. All subject data (a) are averaged to produce a single whole-brain map (b). This ensures consistency between subjects. c) Supervoxel maps are generated using the single whole-brain map. **This step onwards is repeated for every resolution of supervoxels from the same single whole-brain map.* d) Supervoxels are masked using an activation mask to limit the number of voxels included in the analysis. This will affect generated supervoxels by removing all or a portion of voxels. e) Supervoxels are organized into a tabular format for machine learning where the supervoxels are represented as feature inputs. Supervoxels are used as features in the machine learning model (f) to produce weights (g). h) Weights are visualized to determine anatomical location.

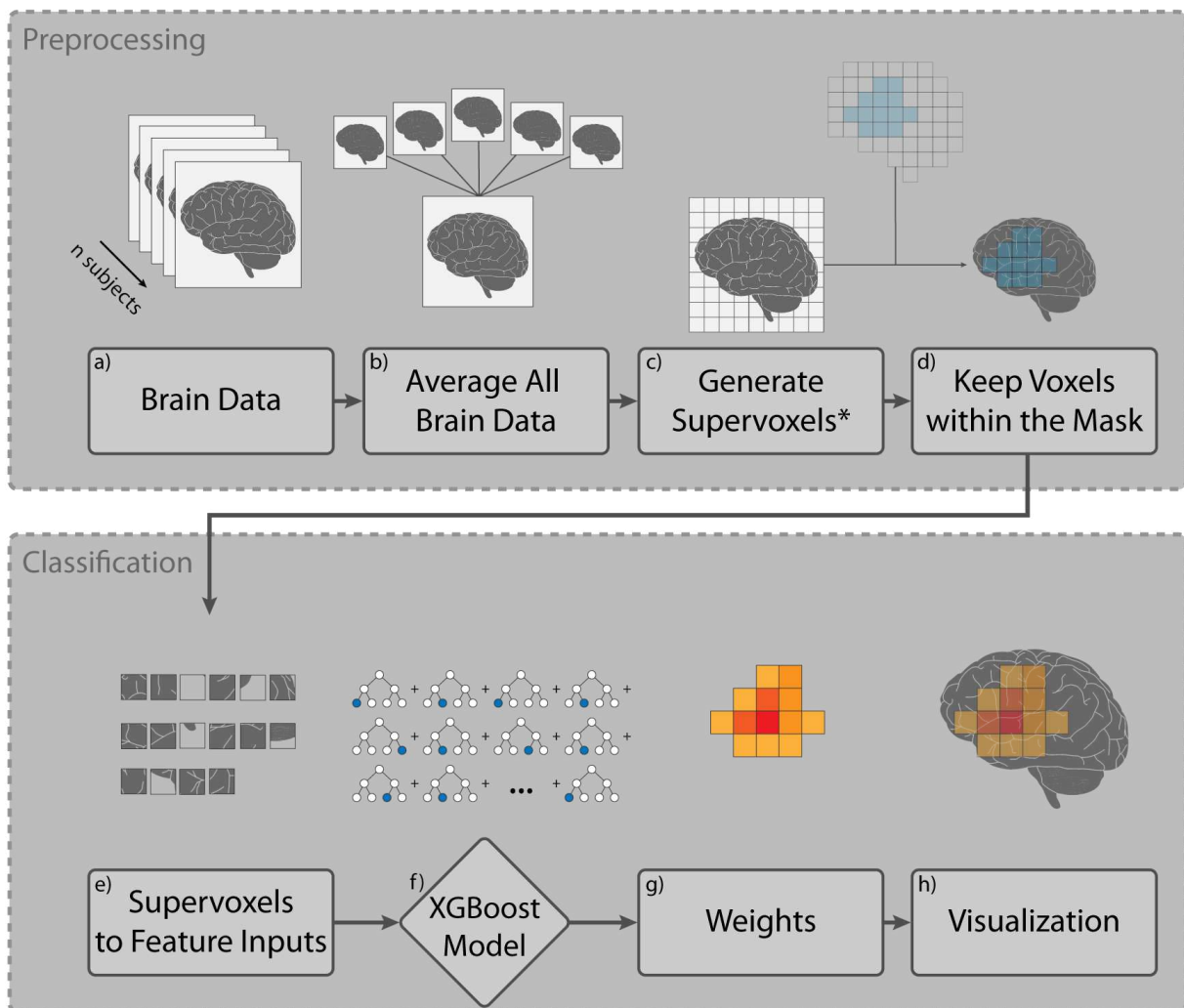


Figure 3. Example of 5-fold nested cross-validation. A) There are a total of 51 task-based data sets or beta coefficient voxel maps across 17 subjects. Each subject has a hand, foot, and mouth voxel map. The full dataset is split and stratified into 5-folds where each fold has 10-11 voxel maps and roughly the same amount of each task. B) Each fold is iteratively used as a test fold where the rest of the data is used for training. The cross-validation outer loop controls error estimation and provides an overall performance measure from this dataset. C) The cross-validation inner loop performs cross-validation using the training subsets from the outer loop to tune hyperparameters. The nested cross-validations estimate an unbiased performance of a model.

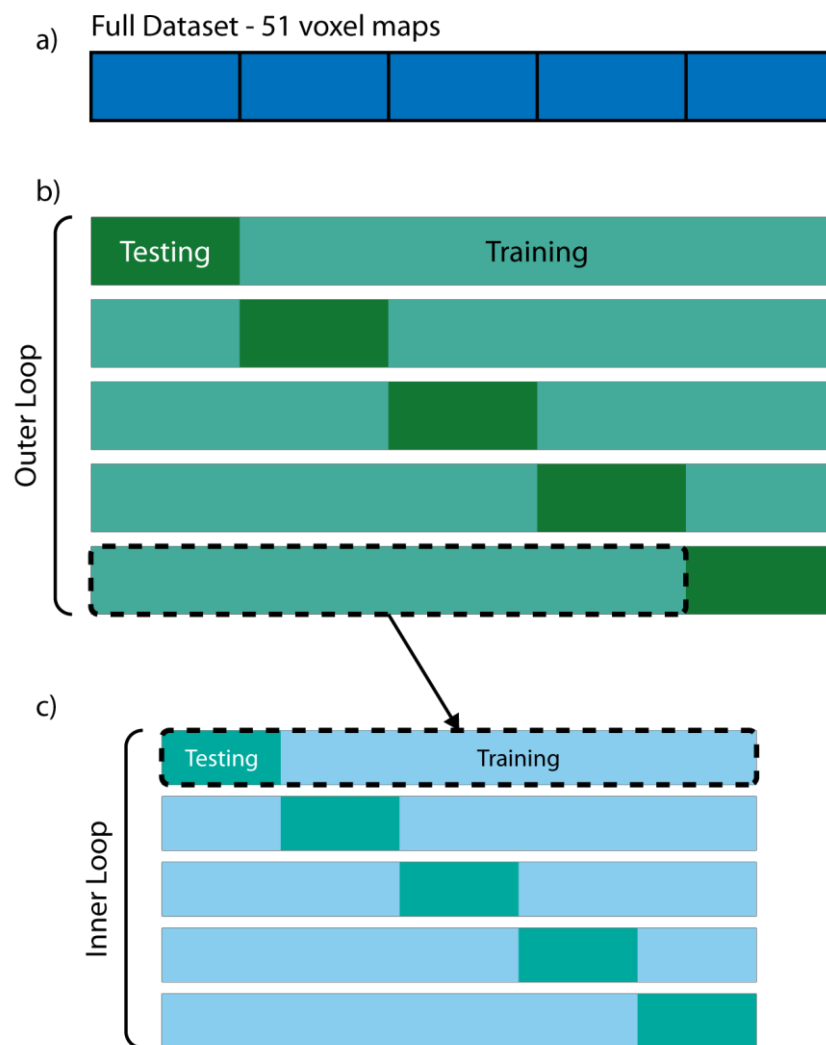


Figure 4. **A)** Whole-brain activation maps showing t-values for each of the three effectors vs. rest. All three effectors show activation of the supplementary motor area. The foot (top) shows activation in the medial sensorimotor cortex. The hand (middle) shows activation laterally in the left hemisphere, consistent with the right hand being used. The mouth (bottom) shows more widespread, lateral activation bilaterally. **B)** Cerebellum activation maps showing t-values for each of the three effectors vs. rest. The foot (top) shows activity in the right anterior lobules, as well as bilateral activity in the posterior lobules of the cerebellum. The hand (middle) activity is localized to the right anterior and posterior lobules. The mouth (bottom) shows bilateral activation of the posterior lobules. **C)** Whole-brain 1-way ANOVA (top) identifying areas showing significant effector-dependent activity. Follow-up t-tests (bottom) determined which effector(s) differed at each region. **D)** Cerebellum 1-way ANOVA (top) identifying areas showing significant effector-dependent activity. Follow-up t-tests (bottom) determined which effector(s) differed at each region. Images were threshold at $p = 0.005$, $\alpha = 0.05$, and cluster size = 24 voxels.

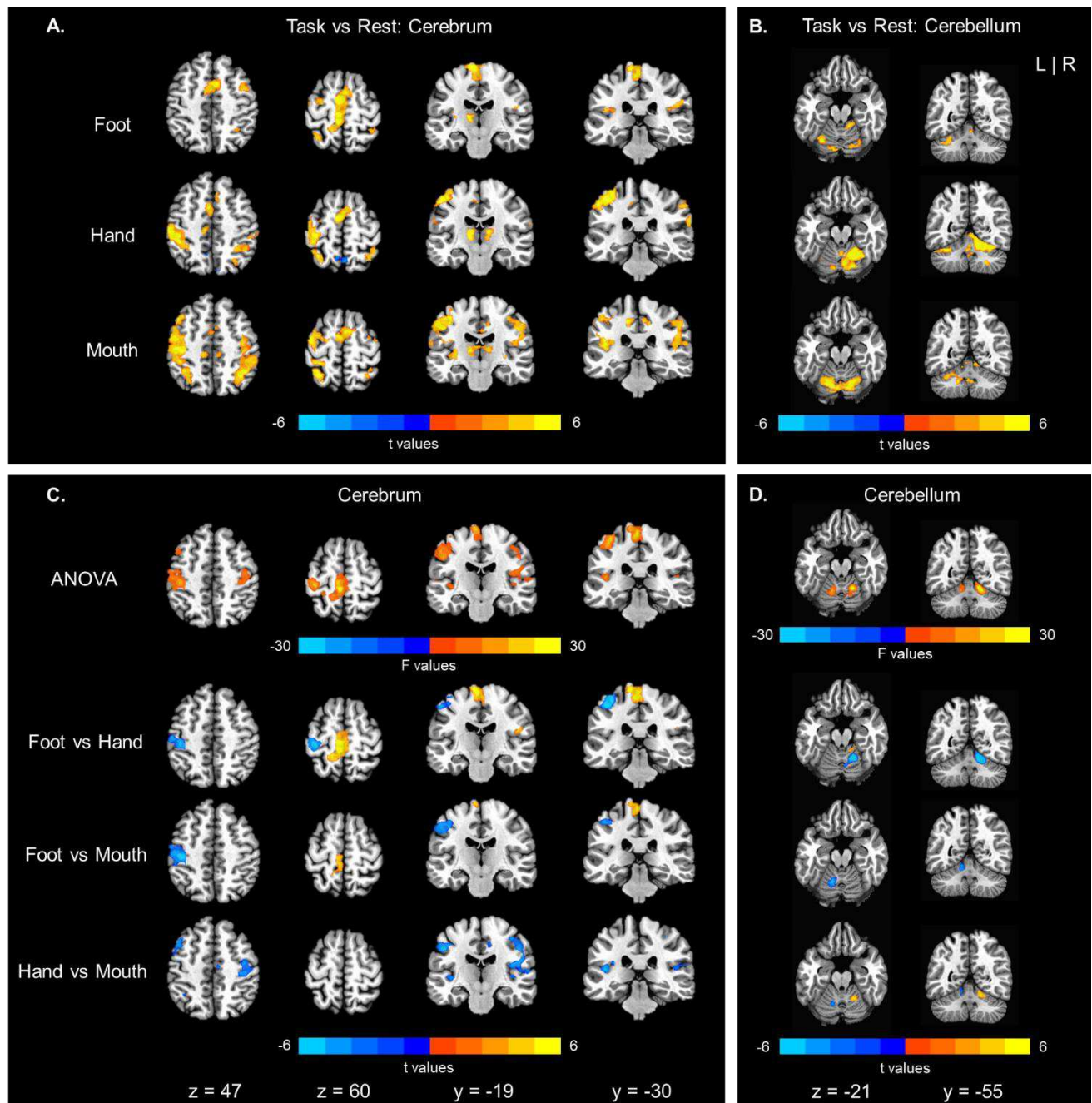


Figure 5. Whole-brain (top) and cerebellum (bottom) overlap maps showing common, effector-independent areas of activation across all three effectors. Images were threshold at $p = 0.005$, $\alpha = 0.05$, and cluster size = 24 voxels.

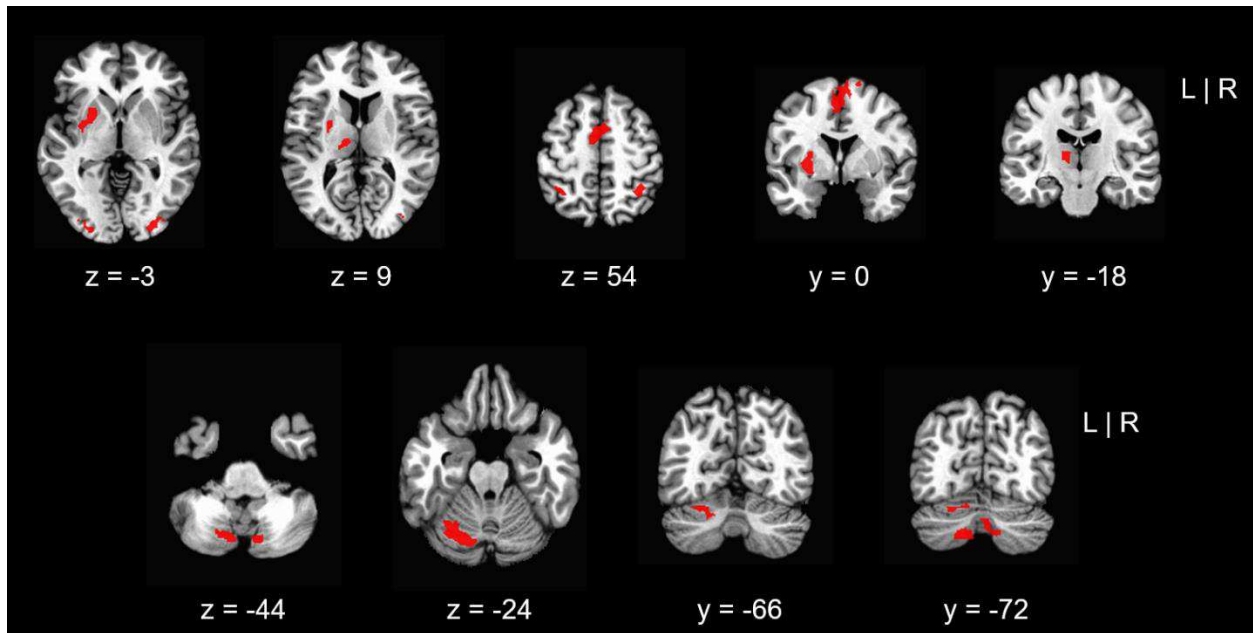


Figure 6A) Accuracy scores across supervoxels. An increase in balanced accuracy is seen with an increase in the number of generated supervoxels. Generated supervoxels are the outputs of the SLIC algorithm. The number of supervoxels is on a logarithmic scale. A 2nd-degree polynomial trend line was fit to show the steady increase from 1 to 1755 supervoxels. Randomly generated supervoxels and brain atlases have lower accuracy than the segmented supervoxels. **B)** F1 scores across labels. F1 scores increase at different rates across effectors. The F1 scores focus on the misclassification of effectors and show the effectors perform differently. The foot effector has the worst performance when a lower number of supervoxels is used (i.e., less spatial resolution). F1 scores greatly increase until around 1000 supervoxels where the foot scores begin to plateau. All three effectors benefit from an increased number of supervoxels or greater spatial resolution.

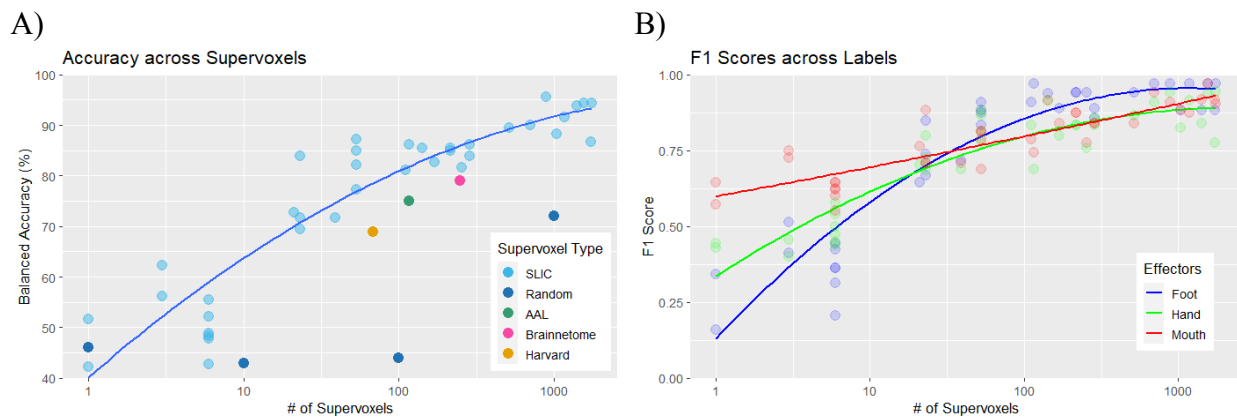


Figure 7. Confusion matrices. Confusion matrices are shown for different supervoxel models from 1 to 10,000. In each matrix, the main diagonal represents a correct prediction. Smaller supervoxel numbers tend to reflect the lower accuracies and metrics with a seemingly normal distribution of misclassifications. All confusion matrices were averaged together to generate an overall model. The overall model shows an equal distribution of correctly predicted labels. There is generally at least one misclassified label and no majorly difficult labels seen across all models.

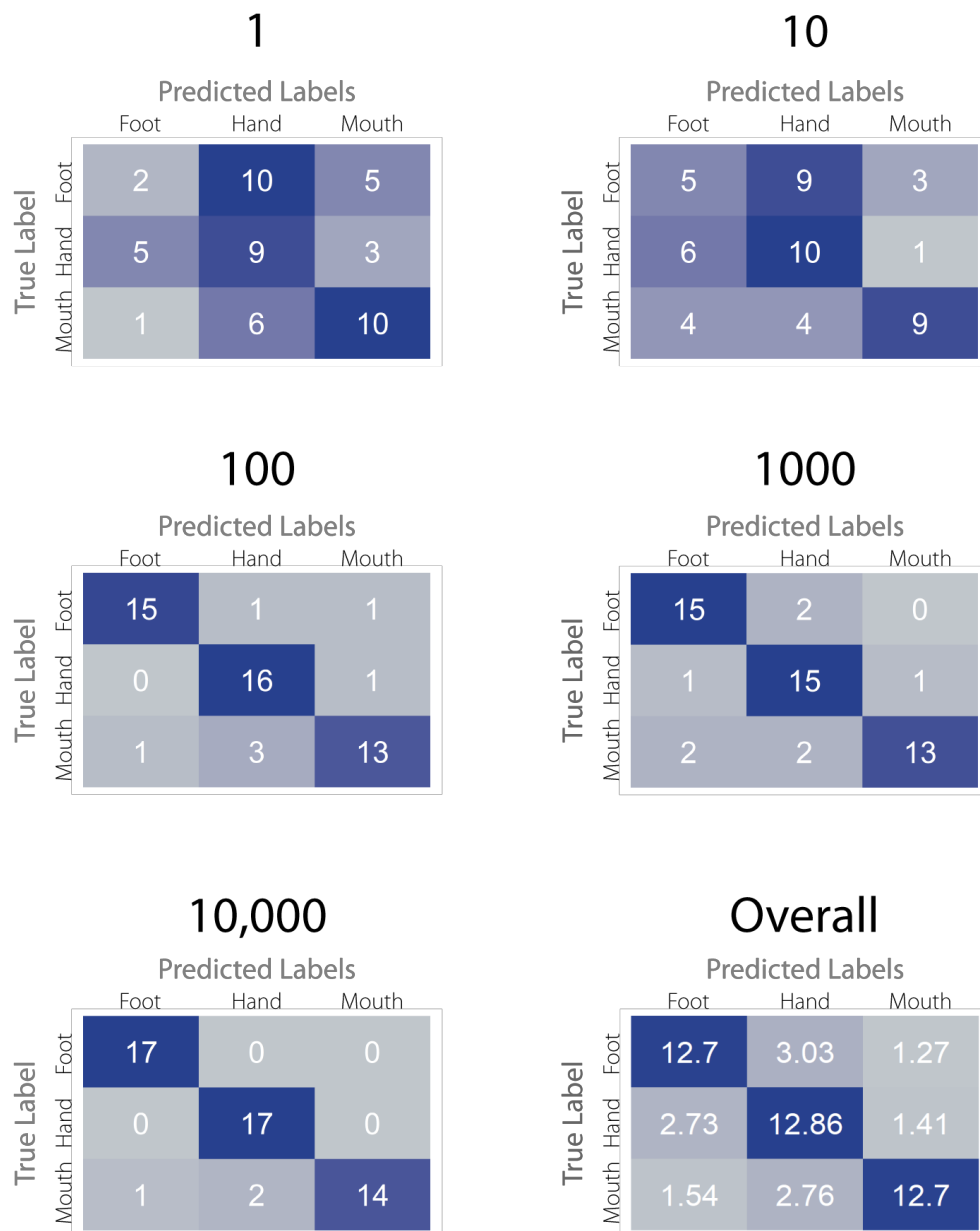


Figure 8. **A)** Model overlap counts of supervoxels in the post-central gyrus overlaid on an anatomical image and then on the HMAT for Z-slice = 47. **B).** Model overlap counts of supervoxels in the para-central lobule overlaid on an anatomical image and then on the HMAT for Z-slice = 73. **C)** Subset of regions identified by the classification analysis and the number of times the region was identified for the right and left hemispheres. The majority of regions were localized to the left hemisphere consistent with the movement of the right effector. Note that the left (contralateral) cerebellum was identified more often than the expected right (ipsilateral) cerebellum, likely because the classification algorithm could more easily exploit the bilateral activity associated with the mouth task. **D)** Each identified brain region was independently removed or occluded from the classification analysis. Regions were removed at the same time the conjunction map is applied. Each model uses a pre-specified 5000 supervoxels, leading to a parcellation model of 1038 regions. The accuracy shows the performance of a model without the respective brain region. Lower accuracies indicate the relative importance of a region in classification.

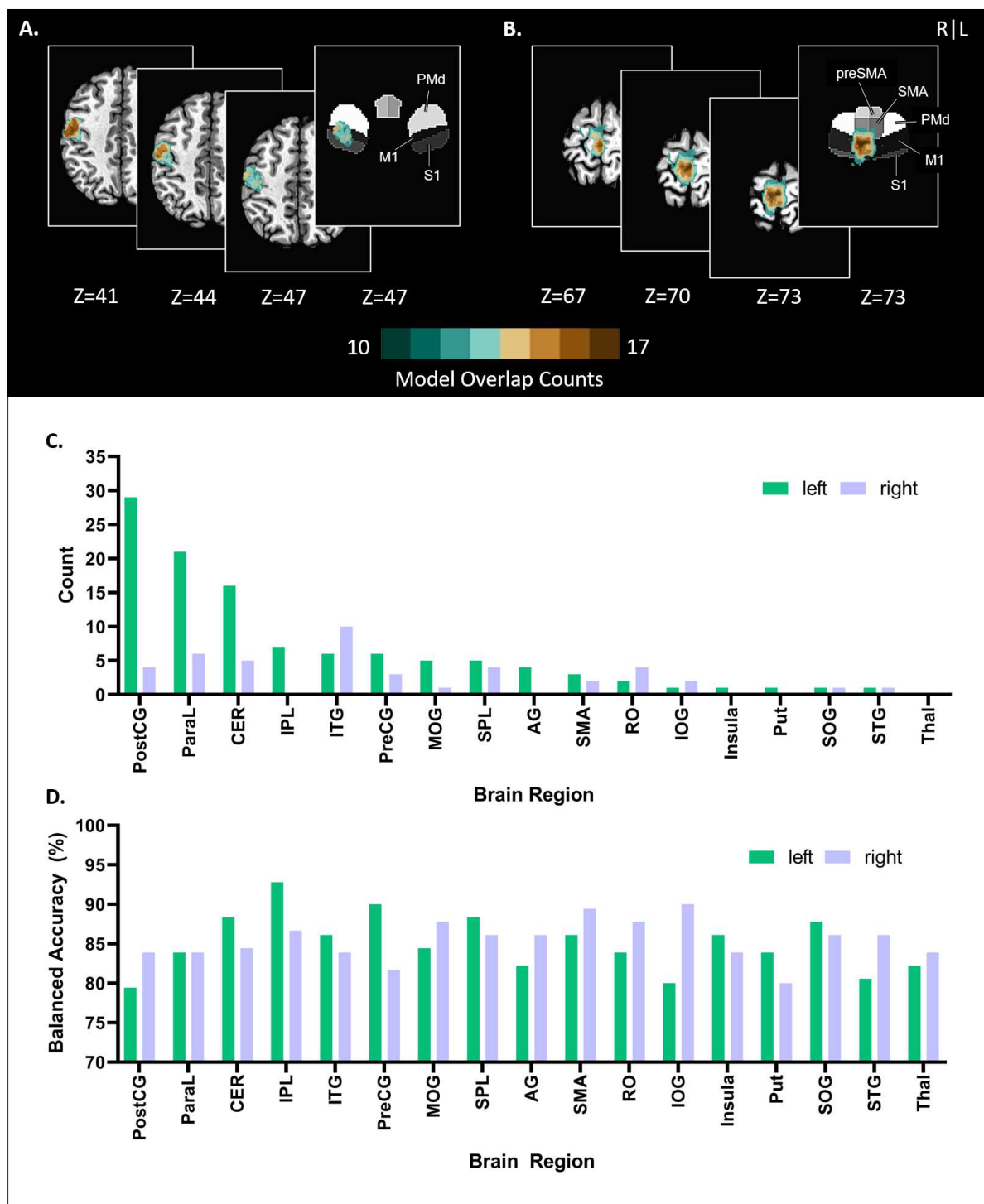


Table 6 (Supplementary Table). The classification # of occurrences columns show the total number of times a brain region was identified within the top 10 highest weights across all models. ‘L’ and ‘R’ indicate left and right hemispheres, respectively. The effector-dependent columns show an ‘X’ for the brain regions that were commonly active across all effectors based on the conjunction analysis outlined in section 2.7.2. The effector-independent columns show an ‘X’ for the brain regions whose activity significantly differed between effectors based on the between effector t-tests outlined in section 2.7.3.

	Classification # of occurrences		Effector- Dependent		Effector-Independent	
	L	R	L	R	L	R
Postcentral Gyrus	29	4	X			
Paracentral Lobule	21	6	X			
Cerebellum	16	5	X	X	X	X
Precuneus	9	3				
Inferior Parietal Lobule	7	0				X
Middle Frontal Gyrus	7	4	X			
Inferior Temporal Gyrus	6	10		X		
Precentral Gyrus	6	3		X		
Fusiform Gyrus	5	2				
Lingual Gyrus	5	2				
Middle Occipital Gyrus	5	1			X	
Superior Parietal Lobule	5	4			X	
Angular Gyrus	4	0		X		
Inferior Frontal Gyrus (p. Opercularis)	4	3				
Mid Orbital Gyrus	4	1				
Anterior Cingulate Cortex	3	2				
Caudate Nucleus	3	0				
Middle Temporal Gyrus	3	5				
Rectal Gyrus	3	4				
SMA	3	2			X	
ParaHippocampal Gyrus	2	2				
Rolandic Operculum	2	4		X		
Superior Frontal Gyrus	2	3				
Temporal Pole	2	0				
Hippocampus	1	1				
Inferior Occipital Gyrus	1	2				X
Insula Lobe	1	0	X			
Middle Cingulate Cortex	1	4				
Middle Orbital Gyrus	1	0	X			
Putamen	0	0			X	
Superior Orbital Gyrus	1	1				
Superior Temporal Gyrus	1	1	X			
SupraMarginal Gyrus	1	4				
Medial Temporal Pole	0	3				

Olfactory cortex	0	1		
Superior Occipital Gyrus			X	
Thalamus				X
