NeuroImage 278 (2023) 120274

Contents lists available at ScienceDirect

NeuroImage

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

ConnSearch: A framework for functional connectivity analysis designed for interpretability and effectiveness at limited sample sizes

Paul C. Bogdan [a,b,*], Alexandru D. Iordan [c], Jonathan Shobrook [d], Florin Dolcos [a,b,e]

[a] Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, Urbana, IL, USA.
[b] Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA.
[c] Department of Psychiatry, University of Michigan, Ann Arbor, Michigan
[d] Department of Mathematics, University of Illinois at Urbana-Champaign, Champaign, IL, USA
[e] Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

ARTICLE INFO

Keywords:
fMRI
HCP
Predictive modeling
Fingerprinting
Supervised learning

ABSTRACT

Functional connectivity studies increasingly turn to machine learning methods, which typically involve fitting a connectome-wide classifier, then conducting post hoc interpretation analyses to identify the neural correlates that best predict a dependent variable. However, this traditional analytic paradigm suffers from two main limitations. First, even if classifiers are perfectly accurate, interpretation analyses may not identify all the patterns expressed by a dependent variable. Second, even if classifiers are generalizable, the patterns implicated via interpretation analyses may not replicate. In other words, this traditional approach can yield effective classifiers while falling short of most neuroscientists' goals: pinpointing the neural correlates of dependent variables. We propose a new framework for multivariate analysis, *ConnSearch*, which involves dividing the connectome into components (e.g., groups of highly connected regions) and fitting an independent model for each component (e.g., a support vector machine or a correlation-based model). Conclusions about the link between a dependent variable and the brain are based on which components yield predictive models rather than on interpretation analysis. We used working memory data from the Human Connectome Project ($N = 50$–250) to compare *ConnSearch* with four existing connectome-wide classification/interpretation methods. For each approach, the models attempted to classify examples as being from the high-load or low-load conditions (binary labels). Relative to traditional methods, *ConnSearch* identified neural correlates that were more comprehensive, had greater consistency with the WM literature, and better replicated across datasets. Hence, *ConnSearch* is well-positioned to be an effective tool for functional connectivity research.

1. Introduction

Multivariate analyses using machine learning are becoming increasingly common in functional connectivity research. The traditional machine learning analytic paradigm involves (a) fitting a connectome-wide classifier to predict a dependent variable, then (b) applying a post hoc interpretation tool to discern the features that best predict the dependent variable. The traditional paradigm is widely used and can produce accurate models (Chen et al., 2022; Eryilmaz et al., 2020; Li et al., 2021b; Shen et al., 2017; Yoo et al., 2018). However, the present research asks whether it is necessarily the most effective way of applying machine learning to study the brain, particularly for studies on task-fMRI. We do not doubt that classifiers are often highly predictive and that interpretation analyses are generally successful in explaining what factors drive the predictions. Rather, we investigate whether the traditional paradigm is suitable for the more basic aim of most neuroscientific research – namely, identifying the networks, connections, and regions sensitive to a task manipulation. The present research raises new questions about the traditional paradigm in this light and proposes a new tool that has the potential for broad usefulness in cognitive neuroscience research.

We argue that the traditional paradigm faces two key limitations that limit its efficacy for most neuroscience studies. First, the interpretations generated via the traditional paradigm may not be *comprehensive* as interpretation analyses may find only a subset of the neural correlates expressed by a dependent variable. Second, the interpretations may not

* Corresponding author at: SCoPE Neuroscience Laboratory, Beckman Institute for Advanced Science & Technology, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61801, USA.
E-mail address: pbogda2@illinois.edu (P.C. Bogdan).

https://doi.org/10.1016/j.neuroimage.2023.120274
Received 7 May 2023; Received in revised form 1 July 2023; Accepted 11 July 2023
Available online 12 July 2023

be *robust* as the neural correlates identified via interpretation analyses may not replicate across studies (Tian and Zalesky, 2021). Along with investigating these drawbacks, we propose a new framework for multivariate analysis of functional connectivity data. The framework is designed for general-purpose use in task-fMRI research where the goal is to pinpoint a manipulation's impact on the brain.

Rather than fitting a connectome-wide classifier and interpreting it, the proposed framework involves fitting many models on different subsets of the connectome. Neuroscientific conclusions are drawn based on which models are predictive, not interpretation analyses. Thus, this procedure deviates from the traditional paradigm. Consistent with the aim of creating a general-purpose framework, the proposed method is flexible. Depending on researchers' goals, subsets of the connectome can be modeled using classifiers or other techniques (e.g., correlational similarity). Our studies tested the efficacy of the proposed framework for these purposes and compared it to existing methods in terms of the two limitations raised above. We notably considered the framework's efficacy for not just large studies but also smaller ones ($N = 25$–50), where machine learning methods are regularly thought to be less viable but where they may prove particularly useful given their sensitivity

1.1. Existing multivariate analytic strategies

Multivariate fMRI tools generally fall under one of the following two categories: (1) fitting models at the group level or (2) fitting models within-subject (Gilron et al., 2017; Wang et al., 2020a). Analyzing data with both group-level and subject-specific techniques is beneficial as these strategies may yield distinct but complementary results (Gilron et al., 2017; Wang et al., 2020a). For instance, group-level analysis could reveal that high working memory (WM) load is associated with increased connectivity within the default mode network (DMN) across the vast majority of participants. In contrast, subject-specific analysis could show that high WM load is linked to increased frontoparietal control network (FPCN) connectivity in some participants but decreased FPCN connectivity in other participants. Taken together, such results would suggest that both the DMN and FPCN are relevant to WM, but high load elicits homogeneous effects on the DMN yet heterogeneous effects on the FPCN. The analyses together provide a more complete perspective of how WM impacts connectivity.

Group-level multivariate analysis often relies on machine learning. In the context of functional connectivity, this is typically done by fitting a classifier using the full connectome as input features. For instance, a task-based classifier could use connectome-spanning patterns to predict whether an example belongs to a high or low WM condition. Then, to interpret the classifier and identify the neural correlates supporting the prediction, post hoc analyses are performed that seek to find the most predictive connectivity edges and regions of interest (ROIs) (Eryilmaz et al., 2020; Li et al., 2021b; Shen et al., 2017; Yoo et al., 2018). This general paradigm can use different types of connectome-wide classifiers, such as support vector machines (SVMs) or ridge regressions. Interpretation analyses can also be done in different ways, such as with recursive feature elimination or post hoc frequentist tests.

In contrast, subject-specific techniques involve fitting separate models for each participant. Subject-specific analyses are primarily used in research on brain activation, often applying a technique named searchlight multivoxel pattern analysis (MVPA), which can be used for both group-level and subject-specific analysis (Weaverdyck et al., 2020). This voxelwise method involves defining a 5–10 mm sphere around every voxel and then submitting the levels of activation within each sphere to a multivariate test (e.g., testing correlational similarity between odd and even trials; Haxby et al., 2001; Walther et al., 2016). For functional connectivity research, subject-specific analyses may also prove insightful. Some connectome-wide models have been used for these purposes but primarily in the context of resting-state scans or for the analysis of data from a single task condition (see research on connectome fingerprints; Finn et al., 2015; Li et al., 2021a; Ravindra et al., 2021). Far fewer studies have investigated subject-specific patterns in the connectivity changes elicited by task manipulations, although as we mentioned above, doing so can provide a more thorough picture of how a task influences connectivity.

1.2. Limitations of existing approaches

Connectome-wide classification and interpretation methods come with well-known limitations. For example, these methods are typically used with large datasets, containing data from hundreds, if not thousands, of participants (e.g., Dubois et al., 2018; Wang et al., 2020b). However, their effectiveness for moderately sized datasets – on the order of tens of participants, which is common for task-fMRI research – remains unclear. These methods also come with a risk of over-interpretation: For instance, an SVM or regression assigning positive weight to a feature does not necessarily speak to the directionality of the feature's relationship with the dependent variable (Scheinost et al., 2019). Although some interpretation techniques solve this dilemma, like the weight-transformation procedure by Haufe et al. (2014), other issues remain.

The present report also puts forth two new possible concerns regarding traditional methods, specifically on the use of interpretation analysis. The first limitation is about *comprehensiveness* – i.e., whether the results of an interpretation analysis capture the complete range of neural correlates elicited by a given dependent variable. Traditional interpretation analyses may face challenges in this respect because they are designed to identify the features that best predict a dependent variable. However, this may lead them to miss moderately predictive effects. Moderately predictive features are not necessarily less meaningful, but their signals may simply be noisier than, or collinear with, the stronger predictors. For example, a hypothetical classifier trained to predict working memory load may achieve 100% testing accuracy, while interpretation analyses show that the classifier primarily uses occipital lobe features to predict load. Such results ignore regions (frontal & parietal) and connections (FPCN & DMN) that are likewise involved in working memory but have signals that do not match the stability seen in the occipital lobe (Grill-Spector et al., 1999). Available evidence suggests that this can occur across a wide variety of tasks and even with sophisticated neural network classifiers (Wang et al., 2020b).

The second limitation concerns the *robustness* of interpretations. Studies implementing connectome-wide classifiers generally use cross-validation and permutation-testing to evaluate the accuracy of classifiers and assess the statistical significance of accuracy. Significant accuracy establishes that the dependent variable is linked to functional connectivity in some way. However, significant accuracy does not imply that all of the features implicated by interpretation analysis genuinely predict the dependent variable. Consider a hypothetical dataset where only a single connectivity edge is genuinely related to the dependent variable. A classifier could, in principle, yield perfect accuracy via this predictive edge, while an interpretation analysis that seeks to identify the 1000 most predictive features would implicate 999 spurious edges. In real data, it is unlikely that only a single edge is a genuine predictor. Nonetheless, this concern is serious, given that the data associated with singular connectivity edges is often unreliable, which creates challenges in differentiating genuine vs. spurious predictors. The dilemma ultimately raises questions about the replicability of features implicated via interpretation analysis (Tian and Zalesky, 2021).

1.3. The proposed approach

The goal of this investigation was to test the efficacy of a novel framework for analyzing fMRI data: the "Connectome Searchlight" (*ConnSearch*) approach. The most similar existing method is the Connectivity-Based Psychometric Prediction approach by Wu et al. (2021). Their tool involved fitting a series of classifiers, wherein each classifier was trained using all connections to a given ROI as features.

The authors examined which classifiers best predicted a dependent variable. For example, the classifier based on connections to a supramarginal gyrus ROI best predicted fluid intelligence, so the authors concluded that this region is relevant to intelligence. The approach by Wu et al. (2021) generates neuroscientific conclusions based on whether a model trained using a subset of the connectome achieved a high fit, rather than via post hoc interpretation analyses. *ConnSearch* follows the same premise, fitting models on subsets of the connectome and assessing which models achieve high fit. However, as we will show below, *ConnSearch* incorporates innovations to increase sensitivity, permit usage even when datasets are limited, and open the door for subject-specific analyses.

Analogously to how searchlight MVPA defines a sphere around every voxel, *ConnSearch* defines a network component "around" every ROI. One way to do this is by treating each ROI as a "Core/Seed" and grouping it with the ROIs to which it is most strongly connected. This yields a large number of overlapping components that span the connectome (Fig. 1A). Then, *ConnSearch* models each component's data independently, to identify the components that significantly predict a dependent variable. Component analysis can be done in a variety of ways, much like how searchlight MVPA can be used for both group-level and subject-specific analysis. For instance, for each component, we can train a classifier based on data from one set of participants, then test classifier accuracy using data from other participants (group-level analysis; Fig. 1B). Alternatively, for each component, we can evaluate fit when modeling each participant's data independently, wherein models use a participant's data from one session to draw predictions about their data in another session (subject-specific analysis; Fig. 1C). In both cases, by examining which components give rise to statistically significant models, we can conclude which portions of the connectome are sensitive to the dependent variable.[1]

By fitting and analyzing multiple independent models, *ConnSearch* is poised to avoid both limitations of the classification/interpretation paradigm noted earlier. (1) Regarding comprehensiveness, traditional interpretation analyses aim to find the most predictive features, but this can result in missing moderately predictive features. In contrast, *ConnSearch* does not solely aim to discern which components are most predictive but also tests whether each component is predictive at all. Hence, *ConnSearch* may be better suited to identify both strongly and moderately predictive effects. (2) Regarding robustness, interpretation analyses may produce results that are a mix of genuine and spurious patterns because statistical significance is only tested for classifier accuracy and not for specific interpreted features. On the other hand, *ConnSearch* tests statistical significance at the level of the component, which allows robust conclusions on the link between a component and a dependent variable. In turn, this may enhance replicability.

However, along with these possible advantages, *ConnSearch* comes with a risk that classifiers based on small subsets of the connectome may simply not be significantly predictive because the number of features is too small. Furthermore, because *ConnSearch* fits many models, this risk is amplified as the threshold for statistically significant accuracy must be raised to account for multiple hypothesis testing. To overcome these challenges, we focus on task-fMRI and classifiers that dissociate task conditions. Compared to the effects of individual differences, task manipulations generally elicit much larger effects (Yarkoni, 2009), allowing more accurate classification. Additionally, our implementation of *ConnSearch* innovates in how cross-validation is performed with procedures that enhance sensitivity in identifying significant classifiers (see MVPA research on this topic, Valente et al., 2021).

[1] We refer to these subsets of the connectome as "components" for consistency with the Network-Based Statistic approach by Zalesky, Fornito, and Bullmore (2010).

1.4. The current study

To assess the suitability of *ConnSearch* for neuroimaging research and to compare it with existing techniques, we analyzed WM task-fMRI data from the Human Connectome Project (HCP; Barch et al., 2013; Van Essen et al., 2013). This task was selected because WM relies on distributed processes involving multiple brain networks (D'Esposito and Postle, 2015), and WM is among the most studied aspects of cognition (Wieczorek et al., 2021), so there is abundant literature with which the *ConnSearch* results can be compared. The HCP WM task has a block design and is a variant of the N-back paradigm: Participants completed two sessions of the task, with each session including eight N-back blocks (27.5 s each). Four blocks were 2-back (high-load) and four blocks were 0-back (low-load). To frame this dataset as a binary classification problem, each participant's data were organized into examples, such that each participant produced two 2-back examples and two 0-back examples. The task conditions were used as binary labels for supervised learning models.

To evaluate *ConnSearch's* utility as a general-purpose framework, we performed a series of analyses testing it for both group-level and subject-specific modeling. We focused on the comprehensiveness and robustness of its results. In terms of group-level modeling, we report several findings. First, we used *ConnSearch* to analyze the neural correlates of WM load using a limited dataset ($N = 50$) as well as a larger dataset ($N = 250$). Second, we compared the *ConnSearch* results to the findings generated by four different existing classification/interpretation methods, which consist of recursive feature elimination (Guyon et al., 2002), neighborhood component feature selection (Eryilmaz et al., 2020), connectome predictive modeling (Shen et al., 2017), and Haufe transformation (Chen et al., 2022; Haufe et al., 2014). Third, we assessed the replicability of the *ConnSearch* results across five 50-subject datasets and compared this to the replicability of the recursive feature elimination results. Finally, we also submitted the limited and large datasets to subject-specific *ConnSearch* to evaluate the flexibility of the proposed framework for multiple types of modeling.

2. Methods

2.1. Data and preprocessing

2.1.1. Participants and WM task design

Analyses were conducted on WM task data from the S1200 HCP dataset (57% female; $M_{age} = 28.7$ [22–37]; 11% left-handed). For a complete description of the task, see Barch et al. (2013). Analyses were done using three organizations of this dataset: (1) a sample of 50 participants, to evaluate the viability of *ConnSearch* when data are limited; (2) a sample of 250 participants, which was either analyzed in its entirety or split into five 50-participant groups to investigate replicability; (3) a sample of 45 test-retest participants, which was used for the analysis of reliability. This final sample constituted the complete HCP Test-Retest dataset, which was made up of all participants who completed the full WM protocols twice (four sessions total). This last sample was used for the reliability analysis because having more data per participant allowed us to better measure intraclass correlation. Across every dataset, N-back performance was generally high: For the 2-back condition, the mean accuracy was 84.4% (SD = 9.4%), and for the 0-back condition, the mean accuracy was 89.9% (SD = 10.8%). Among the five 50-participant groups used for the replicability analysis, no pair of groups showed significant differences in either 2-back or 0-back accuracy ($ps \geq 0.12$); tested using Wilcoxon rank-sum tests because the data were left-skewed. These datasets are publicly available, and this research did not require review by the University of Illinois Institutional Review Board.

2.1.2. HCP imaging protocol

Functional MRI data for the WM task were collected using a

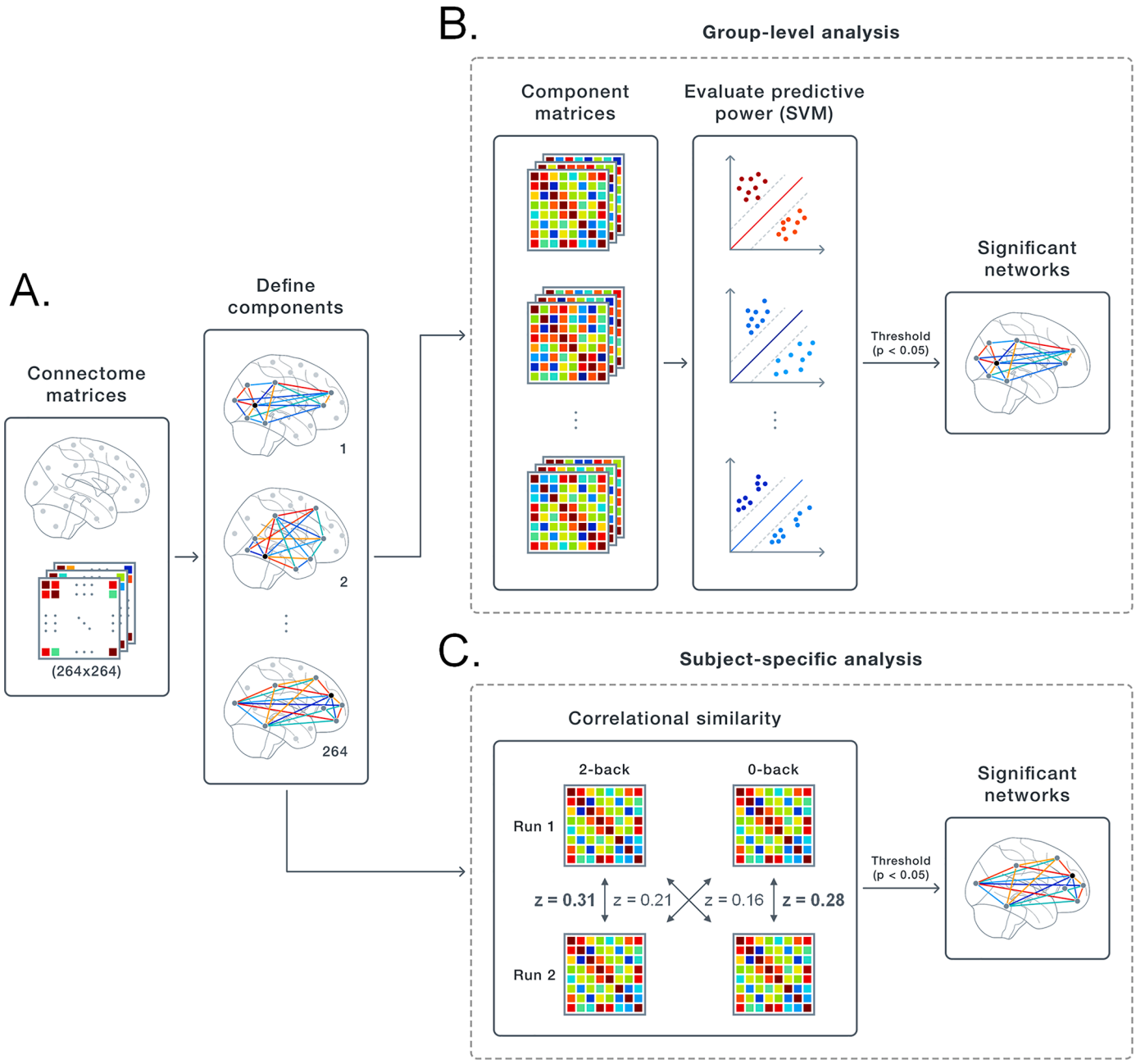

Fig. 1. *Diagram of ConnSearch.* **A.** Connectome matrices (264 × 264) are extracted to create examples, each with a binary label (e.g., 2-back vs. 0-back). Based on the average connectome, overlapping sets of ROIs ("network components") are defined. Here, 264 components are created by using each ROI as a "Core" to generate a component – the component includes said ROI and the N-1 other ROIs to which it is most strongly connected (N is a flexible parameter). Each component is used to extract $N \times N$ matrices from the connectome examples. Each component's data are then modeled individually. **B.** *Group-level ConnSearch.* For each component's $N \times N$-matrix examples, a classifier is trained to predict the labels (e.g., classify 2-back vs. 0-back), and its accuracy is tested. Permutation-testing is used to assess whether testing accuracy is statistically significant. **C.** *Subject-specific ConnSearch.* This analysis requires that each participant provides at least four examples, two for each label (Haxby et al., 2001). For each component and each participant, correlational similarity is calculated as the within-condition edge-by-edge correlation minus between-condition correlation (Fisher-z transformed Pearson correlations). The correlational similarity measurement for each participant is submitted to a one-sample *t*-test to assess whether the components show positive similarity across most people. Note that these two modeling branches are not meant to be exclusive, other types of component models are possible, but these are the two tested here.

customized 3T Siemens Skyra scanner with a 32-channel head coil (Van Essen et al., 2013) and a multiband pulse sequence that permitted high temporal and spatial resolution (multiband factor = 8; TR = 720 ms; TE = 33 ms; flip angle = 52°; 2 mm isotropic). High-resolution T_1-weighted images were collected before functional scanning (TR = 2400 ms, TE = 2.14 ms; flip angle = 8°; 0.7 mm isotropic). Before release, the HCP submitted the data to a "minimal preprocessing" pipeline, which removed spatial artifacts/distortions, produced segmentation data, registered the data within-subject, and normalized all participants' data to a common brain template (Glasser et al., 2013). Slice timing correction was not needed due to the short TR time. Further details on the scanning and preprocessing procedures are described by Van Essen et al. (2013) and Glasser et al. (2013).

2.1.3. *Further preprocessing*

Prior to analyses, the fMRI data were further preprocessed using the

Connectivity Toolbox (CONN; Whitfield-Gabrieli and Nieto-Castanon, 2012). Denoising was carried out with linear regression and removed physiological and other sources of noise via the inclusion of the following temporal covariates: First, using the anatomical CompCor method (Behzadi et al., 2007), as implemented in CONN, noise ROIs were created for each participant based on white matter and cerebrospinal fluid masks derived during segmentation and eroded by 1 voxel to minimize partial volume effects. For each ROI, five principal components representing its signal were added to the denoising regression model as temporal covariates. Second, twelve motion parameters (3 rotation, 3 translation, and first-order temporal derivatives) were added. Third, finite impulse response (FIR) covariates were also added to ensure that analyses targeted functional connectivity and not co-activation (Cole et al., 2019). One hundred twelve FIR covariates were added in total, which covered the length of each block (38 volumes) plus 18 additional volumes to account for the likely duration of the HRF. Separate covariates were used for each condition, hence $112 = (38 + 18) \times 2$. Fourth and finally, for each participant, a variable number of regressors were included to account for "outlier volumes". These volumes were identified using Artifact Detection Tools (www.nitrc.org/projects/_artifact_detect/), as scans exceeding 0.5 mm in composite motion (combination of translational and rotational displacements) and/or showing activation 3 standard deviations above or below the global mean signal. For each outlier volume, a "scrubbing" covariate was added, which consisted of 1 for the outlier volume and 0 s for all the other volumes. After regression, high-pass filtering was performed (> 0.01 Hz) to remove slow fluctuations while maintaining task-related changes.

After preprocessing, ROI-to-ROI connectivity was calculated using HRF-weighted Pearson correlations between time series extracted from ROIs. To ensure *ConnSearch* is effective regardless of the atlas employed, connectivity matrices were calculated using the 264-ROI atlas by Power et al. (2011) and the 1000-ROI variant of the atlas by Schaefer et al. (2018). The Power et al. (2011) atlas is specified by 264 cortical and subcortical coordinates, established from both resting-state and task-based data. For each coordinate, an ROI was created as a 5 mm radius sphere centered at the coordinate. The Power et al. (2011) atlas was used for the analyses of the 50-participant datasets and the 45-participant Test-Retest dataset. The Schaefer et al. (2018) atlas specifies 1000 ROIs and was developed based on resting-state data. This atlas was used for the analyses of the 250-participant dataset, where the increased density of ROIs, together with the larger sample size, allows fine-grain modeling of component data and precisely localizing effects of the task.

2.2. The ConnSearch approach

2.2.1. Overview

Here, we provide a broad overview of the analysis, and later subsections provide details: Examples were labeled 2-back or 0-back based on the task condition from which they originated. The task conditions served as binary labels (Y), which the models sought to predict. Each participant completed both conditions twice and thus generated two examples of each label. Each example is associated with a functional connectome matrix ($X_{M\times M}$), where M is 264 or 1000 depending on the atlas tested. Based on the M ROIs, *ConnSearch* defines M overlapping network components. Each network component (C_m) is a set of N ROIs, defined as ROI m and the $N-1$ ROIs most strongly connected to ROI m. Each network component was used to extract an $N \times N$ matrix ($X_{C_mC_m}$) from each example's connectome matrix.

Each component's data were submitted to independent models – i.e., one model was fit for component 0, one model for component 1, etc. Modeling fell under one of two distinct tracks, including group-level modeling and subject-specific modeling. (1) For group-level modeling, M SVMs were trained/tested, one per component. Each SVM predicted the labels (Y) based on data from a given component ($X_{C_mC_m}$). Group-level *ConnSearch* aims to identify which components give rise to significant classifiers. For each classifier, the accuracy threshold required for statistical significance was defined based on permutation-testing. (2) For subject-specific modeling, correlational similarity was used to test for subject-specific effects (Haxby et al., 2001). For a given participant, correlational similarity depends on the correlation between the participant's two examples of the same label minus the correlation between their examples of opposite labels. Applying this analysis separately for each component generates M measures of correlational similarity per participant. Subject-specific *ConnSearch* is to identify components that show positive correlational similarity across most participants. For each component, the significance of participants' correlational similarity was assessed by submitting every participant's value to a one-sample t-test.

2.2.2. Defining the network components

After extracting connectome matrices, the ROIs were organized into partially overlapping sets, which we refer to as *"network components."* A network component is defined for every ROI of the atlas. Each ROI's network component includes the ROI itself (the *"Core ROI"*) and the $N-1$ other ROIs with which it was most strongly connected (Fig. 2). Connectivity strength was assessed as the average connectivity matrix across both conditions and all participants. Defining components based on strong connectivity is motivated by network research defining communities of nodes based on their strong connections (Sporns, 2013). Although this strategy will exclude analysis of weak connections and thus may miss some neural correlates, inclusion of these connections is still possible by adjusting how components are defined (see Discussion 4.3).

Notably, pooling data from all participants when defining components technically means that the testing set "contaminates" the training. Nonetheless, it enables clearer results without compromising the analysis. Because group-level modeling involves cross-validation, doing the alternative – defining components solely on training data – would require defining a unique set of components for every cross-validation fold. This would create challenges in interpreting significant components. Additionally, *a priori* there is no clear reason to suspect why such contamination would inflate classifier accuracy, but even if it did, such inflation would not influence the results because significance was established using a permutation-testing approach. Accordingly, preliminary analyses using group-level *ConnSearch* and the limited ($N = 50$) dataset, showed that both approaches yield the same number of significant components (ten).

The number of ROIs in each component (N) is a flexible parameter that is specified before analysis. The main text focuses on results from where $N = 16$, but to test for consistency, the primary analyses were carried out multiple times, using 16, 24, or 32 ROIs. Overall, each ROI setting supported similar conclusions. However, as will be shown in the Results, larger sizes increase sensitivity but decrease localization – i.e., for larger components, more will be found to be significant but, for each, less can be said about specific ROIs and edges.

After defining network components, data for each component were extracted from the main connectome matrix. Analyses proceeded by either building group-level (Section 2.2.3) or subject-specific models (Section 2.2.4). To report the significant components, figures and tables are provided, which organize the components' ROIs in terms of canonical networks, as labeled by the Yeo et al. (2011) network atlas. Subcortical ROIs not assigned to any network were labeled "Uncertain." Supplemental Table S1 lists all Power et al. (2011) coordinates with their Yeo et al. (2011) labels.

2.2.3. Group-level ConnSearch

For group-level *ConnSearch*, an independent SVM was fit for each component, using its connectivity edges as features and the task conditions as labels (0-back vs. 2-back). Component data are symmetric matrices, and thus the bottom triangles were extracted and flattened into vectors before analysis. Analyses used SVMs with linear kernels

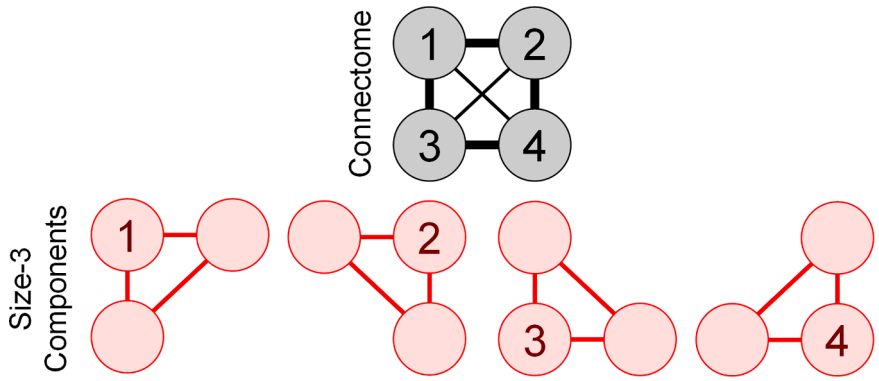

Fig. 2. *Procedure for defining overlapping components*. This diagram shows how a hypothetical 4-ROI connectome could be divided into four overlapping size-3 components. For the connectome graph, each edge's line width represents connectivity strength.

because available evidence shows their efficacy in connectivity-based classification (e.g., linear SVMs perform better than radial SVMs or Linear Discriminant Analyses; Plitt et al., 2015). The SVM regularization parameter was left at the default (C = 1), as no neuroscientific research has found support for a specific setting.

Each SVM's accuracy was assessed via 10-repeated 5-fold cross-validation. This procedure involves repeating 5-fold cross-validation ten times and averaging across the repetitions to generate an overall assessment of testing accuracy. Participants were organized by fold and, for each fold, no participants were split with some examples in the training set and some in the testing set. Notably, the use of repetitions for cross-validation is not yet standard in neuroimaging research, where unrepeated cross-validation is more common. Nonetheless, averaging across 10-repeats increases statistical power (see Supplemental Methods 1.1 and for a related demonstration, see Valente et al., 2021). For each SVM, accuracy was compared to a threshold of statistical significance, which was established via permutation-testing and which accounts for multiple comparisons at a family-wise error level (Supplemental Methods 1.2). For a 50-participant dataset with 264 components, $p < .05$ corresponds to over 63% accuracy, and for a 250-participant dataset with 1000 components, $p < .05$ corresponds to over 56% accuracy; note that although 56% accuracy may not intuitively seem far above chance, it corresponds to correct classification 1200 more times than incorrect classification.

2.2.4. Subject-specific ConnSearch

Subject-specific *ConnSearch* is based on correlational similarity (Haxby et al., 2001; Kriegeskorte et al., 2008; Weaverdyck et al., 2020). To measure correlational similarity, component data matrices are flattened into vectors, as with group-level *ConnSearch*. As noted, each participant provides four examples for a given component (2 conditions x 2 sessions), x_1^{0bk}, x_2^{0bk}, x_1^{2bk}, and x_2^{2bk}. Pearson correlations were assessed between pairs of examples from different sessions, and each correlation was Fisher Z-transformed. Correlational similarity (f) was calculated as within-condition similarity minus between-condition similarity,

$$f_{within} = Z\left(x_1^{0bk}, x_2^{0bk}\right) + Z\left(x_1^{2bk}, x_2^{2bk}\right)$$

$$f_{between} = Z\left(x_1^{0bk}, x_2^{2bk}\right) + Z\left(x_1^{2bk}, x_2^{0bk}\right)$$

$$f = f_{within} - f_{between},$$

where Z represents a Fisher-Z-transformed correlation. The above equations were applied independently for each component and participant. High correlational similarity for a participant's data indicates that, across the two sessions, connectivity data were similar within-condition and distinct between-condition. For each component, one-sample t-tests were used to assess whether most participants showed a substantial correlational similarity effect for that component. T-tests were corrected for multiple hypotheses using Holm's method with Šidák correction (Abdi, 2007) – a family-wise error approach similar to Bonferroni correction. To confirm that significant results were rooted in subject-specific patterns, correlational similarity was also measured after subtracting group-average 2-back vs. 0-back effects on an edge-by-edge basis (Supplemental Methods 1.3). This technique, which is analogous to "cocktail-blank removal" in MVPA research (Bonhoeffer and Grinvald, 1993; Walther et al., 2016), omits all group-level patterns from the data prior to measuring correlations, meaning that any subsequent results must be rooted in neural patterns unique to each participant.

2.3. Comparisons with other analytic methods

The group-level *ConnSearch* results were compared to those of four alternative strategies, applied to the same 50-participant and 250-participant datasets. Each alternative method followed the traditional machine learning approach, whereby a classifier was fit using the whole connectome as features, then a post hoc interpretation analysis was conducted to identify the most predictive edges. The four strategies used different feature selection techniques for finding predictive features; note, feature selection was used here solely for post hoc interpretation and not to enhance classifier accuracy. Supplemental Methods 1.4 also details frequentist approaches that were implemented for comparison, including t-tests and Network-Based Statistic analysis (results for a 50-participant dataset provided in Supplemental Results 2.1; Zalesky et al., 2010).

The first connectome-wide classification and interpretation approach used *recursive feature elimination* (Guyon et al., 2002), which involved (a) fitting a linear SVM using the whole connectome as features, (b) identifying the 0.1% of features with the lowest magnitude weights, and (c) eliminating those edges. This was performed recursively until 1000 edges were achieved; preliminary analyses tested alternative selection thresholds (500–4000 edges), but these all yielded similar interpretations (no shifts in significance for patterns reported in the Results, Table 2). The second method used *neighborhood component feature selection* (Yang et al., 2012), which involved identifying the edges that showed the smallest distance among examples within-condition and the largest distance among examples between-condition. Using this strategy, 397 features were identified as most predictive. This 397-feature cutoff was used because it provides high and stable accuracy (following the procedure by Eryilmaz et al., 2020). This second method could only be used for the analysis of the 50-participant dataset, as it was too computationally intensive for the 250-participant dataset. The third method was *connectome predictive modeling* (Shen et al., 2017), which involved submitting every edge to a frequentist test (t-test comparing 0-back vs. 2-back), then filtering edges to only those where uncorrected

$p < .05$; preliminary analyses tested other cutoffs ($p < .001$–0.10), but these all yielded similar interpretations. The fourth method used *Haufe transformation* (Chen et al., 2022; Haufe et al., 2014), which involved fitting a kernel ridge regression, then transforming the coefficient weights with the technique by Haufe et al. (2014). Unlike the first three methods, this final technique yields continuous measures of predictiveness for each edge.

For each method, its sets of most predictive edges were organized in terms of which Yeo et al. (2011) network their corresponding ROIs belonged to (e.g., an edge could be labeled FPCN-DMN). These effects are plotted for each method, showing how some pairs of networks had a greater number of edges deemed highly predictive. Contrasting most studies using these methods, the present research also tested the statistical significance of interpretations: Analyses examined whether the number of highly predictive edges between a given pair of networks significantly surpassed what would be expected by chance if the edges were uniformly distributed; p-values were assessed via permutation-testing. The Haufe results could not be used for testing significance, as its results are continuous and do not lay out a specific set of highly predictive edges.

2.4. Reliability and replicability analyses

ConnSearch and the other methods were compared in terms of both reliability and the replicability of the brain regions they implicate (i.e., the replicability of interpretations). Paralleling earlier HCP reliability research (Elliott et al., 2020), reliability was quantified for significant group-level components as the mixed-effects intraclass correlation (ICC) – i.e., the extent that a participant's connectivity within the first run predicted their connectivity in the second run, while task-related effects were controlled. Component ICC was compared to the ICC associated with individual edges. Further details on measuring reliability are provided in Supplemental Methods 1.5. The results and conclusions on reliability apply to both group-level and subject-specific *ConnSearch*.

Replicability was assessed by first organizing the HCP WM data into five independent datasets ($N = 50$ each). Then, *ConnSearch* was performed using each dataset, and overlap was measured among each dataset's results. Overlap was measured using the Dice Similarity Coefficient (DSC) in terms of the ROIs that made up each dataset's significant components (Dice, 1945; Maitra, 2010; Sorensen, 1948). Substantial overlap would mean that the neuroscientific conclusions based on one dataset would similarly be identified in another dataset. Overlap was measured between each pair of datasets and then averaged across all possible pairs. Replicability was independently measured for group-level and subject-specific *ConnSearch*.

For comparison purposes, the replicability was also assessed based on interpretation using the first connectome-wide method (recursive feature elimination). The method was likewise used to analyze five datasets. For each dataset, recursive feature elimination identified 1000 edges as the most predictive. Based on these edges, ROIs were defined as "significant" if they were associated with eight or more edges included in the pool of 1000. Replicability was measured based on the overlap of "significant" ROIs between datasets, again calculated using the Dice Similarity Coefficient. The eight-edge threshold was chosen because it drives the number of "significant" recursive feature elimination ROIs (105 on average) to roughly equal the number of "significant" *ConnSearch* ROIs (102 on average). The eight-edge threshold is not meaningful beyond this specific comparison.

Note that neither *ConnSearch* nor recursive feature elimination formally tests statistical significance in terms of individual ROIs. Instead, *ConnSearch* tests statistical significance for components' models, and our connectome-wide analyses test significance in terms of edges. By defining "significant" ROIs, this approach created a "common currency", which allowed the different methods' replicabilities to be compared even though their outputs are fundamentally different.

2.5. Software

The *nilearn* package was used for the glass brain visualizations (Abraham et al., 2014). The new code has been released so that other researchers can apply *ConnSearch* to their own datasets or modify it (https://github.com/paulcbogdan/ConnSearch). The code notably contains tools for alternative component definitions (e.g., proximity), permutation-testing, and creating tables. The visualization code used for the chord diagrams was released as a separate package (https://github.com/paulcbogdan/NiChord).

3. Results

3.1. Limited sample size

3.1.1. Group-level ConnSearch

Consistent with our aim to develop a tool that is sensitive and provides a comprehensive perspective on the effects of a task, group-level *ConnSearch* identified ten significant network components, even under a limited sample size ($N = 50$). These ten significant components span canonical resting-state networks and provide a look into multiple aspects regarding the effects of working memory load (Fig. 3 & Table 1). For example, two significant components were primarily composed of parietal and occipital ROIs, which are linked to visual and attentional processing networks (Entries #1 & #2). Significant components also arose linked to prefrontal ROIs and higher-order networks, such as the FPCN and DMN (e.g., Entries #4 & #10). Finally, several components spanned both visual/attentional and frontal areas (e.g., Entries #8 & #9), pointing to the collaboration between these processes. For robustness, this analysis was also performed using 24-sized and 32-sized components, which yielded significant results of similar topography (Supplemental Tables S2 & S3). Overall, these findings demonstrate that *ConnSearch* can generate significant results, even when (1) sample sizes are limited,[2] (2) classifiers use only small fractions of the connectome, fewer than 0.4% of all edges, and (3) strict accuracy thresholds that correct for multiple hypotheses are employed.

3.1.2. Comparison to connectome-wide interpretation methods

Testing alternative methods showed they were less comprehensive in identifying the breadth of neural correlates elicited WM load compared to *ConnSearch*. For this analysis, we used the same 50-participant dataset as above and first confirmed that the SVM and ridge regression were successful classifiers. Indeed, both yielded high accuracy (over 80%). Then, four approaches to post hoc interpretation were tested, and the four methods yielded mostly similar results: Every method heavily implicated edges linked to the Visual Network and Dorsal Attention Network (DAN), which were associated with posterior ROIs (Fig. 4). However, the alternative methods were more limited in identifying patterns beyond these networks. Statistically, only edges linked to the Visual Network and DAN were deemed most predictive at frequencies that significantly surpassed chance level (Table 2). This contrasts *ConnSearch*, which delineated not just posterior components but also FPCN-FPCN, FPCN-DMN, and DMN-DMN components that included frontal ROIs. These results provide evidence for our hypothesis on the comprehensiveness of *ConnSearch*, in terms of more fully capturing the effects of a dependent variable. These different paradigms will be further

[2] Our primary focus was on 50 participant datasets, as after accounting for the short HCP scan times (10 minutes per participant), they are most representative of typical fMRI studies, usually involving 20-25 participants. However, we also tested *ConnSearch* using yet smaller HCP datasets (N = 25). This analysis used large component sizes (32 ROIs) to compensate for the limited amount of data. Even at this very small sample size, *ConnSearch* identified 14 significant components, which included ones that can be considered occipital, occipitoparietal, or frontoparietal.

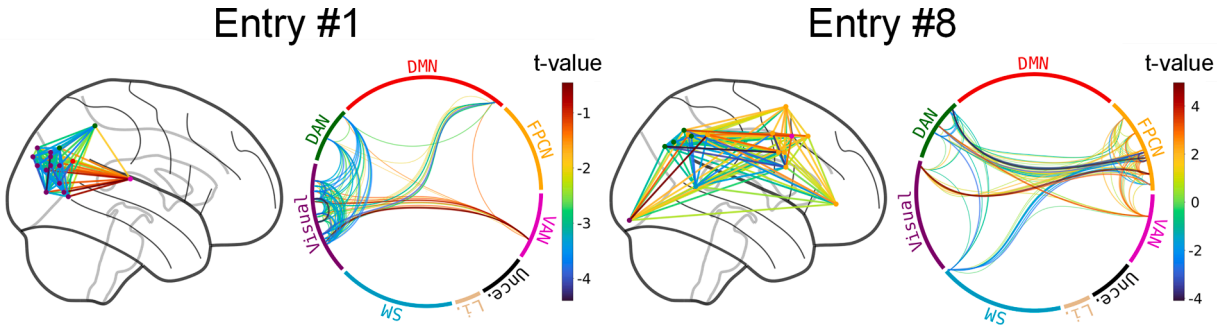

Fig. 3. *Significant network components identified by group-level ConnSearch.* Two components that are significant at $N = 50$ are shown here and titled by their corresponding entry number in Table 1. Edges are colored based on the results of paired t-tests, rather than SVM weights, as such weights cannot be compared across models. The network names were assigned based on the labels of the Yeo et al. (2011) network atlas. VAN, Ventral attention network; DAN, Dorsal attention network; SM, Somatomotor network; Li., Limbic network; Unce., Uncertain (subcortical).

Table 1

Significant network components identified by group-level ConnSearch. Each row corresponds to a significant component. The three columns under "Core ROI Details" describe the properties of the Core ROI used to create the component, including its region, Brodmann Area (BA), and its "ROI #" (see Supplemental Table S1 for details on each ROI #). The "Component Network Composition" columns represent the breakdown of each significant component's ROIs. Rows are ordered from posterior (#1) to anterior (#10) in terms of the location of their Core ROI. Due to rounding, some rows sum to 99% rather than 100%. L, Left; R, Right; Cun, Cuneus; SPL, Superior parietal lobule; ITG, Inferior temporal gyrus; STG, Superior temporal gyrus; PCu, Precuneus; FuG, Fusiform gyrus; PrG, Precentral gyrus; MFG, Middle frontal gyrus.

	Core ROI Details				**Network Component Composition**						
Entry #	Region	BA	ROI #	Accuracy	DAN	DMN	FPCN	Limbic	SM	VAN	Visual
1	L Cun	7	165	64.3%	12%	6%	0%	0%	0%	6%	75%
2	R SPL	7	255	63.4%	44%	0%	12%	0%	0%	0%	44%
3	L ITG	20	3	64.2%	12%	19%	25%	0%	6%	0%	38%
4	R STG	40	234	63.9%	12%	25%	31%	0%	12%	19%	0%
5	R PCu	7	202	64.2%	6%	0%	0%	0%	50%	19%	12%
6	L FuG	20	247	66.9%	25%	6%	25%	6%	0%	0%	38%
7	R PrG	6	263	65.8%	50%	0%	31%	0%	0%	12%	6%
8	L PrG	6	173	66.0%	25%	6%	56%	0%	6%	6%	6%
9	L MFG	9	186	66.1%	25%	0%	56%	0%	0%	0%	19%
10	R MFG	6	192	64.8%	19%	12%	69%	0%	0%	0%	0%

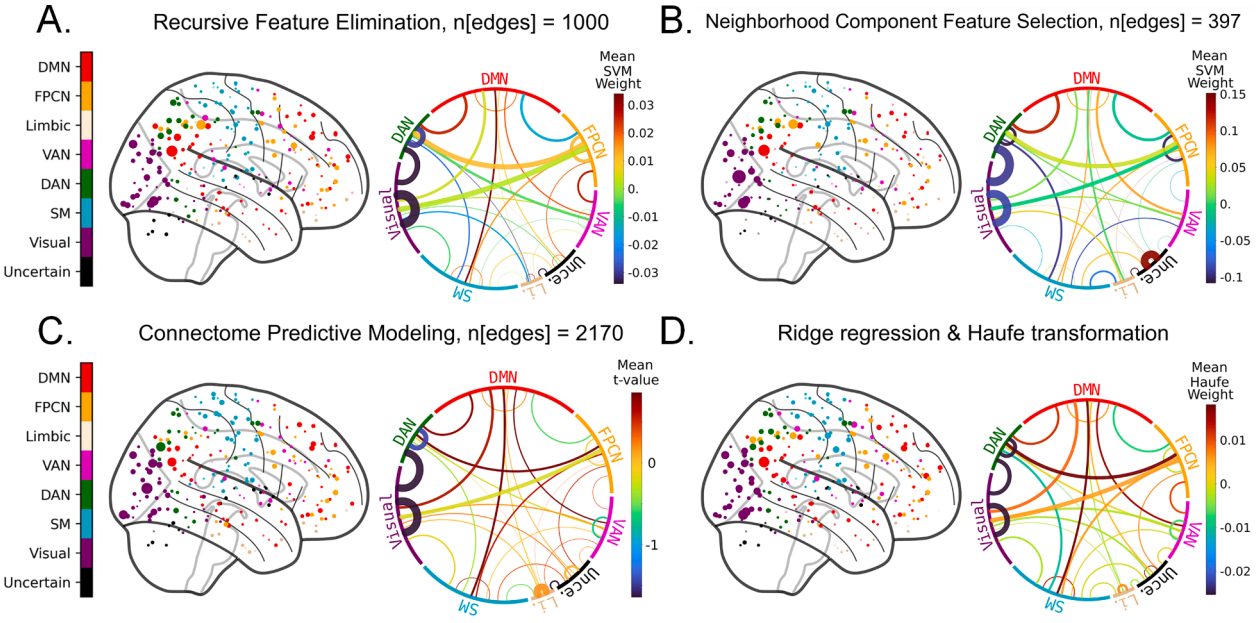

Fig. 4. *Results of four connectome-wide classification and feature selection methods.* Analysis used the limited ($N = 50$) dataset parecellated using the Power et al. (2011) atlas. For each method's results, the glass brains indicate which ROIs were linked to the most predictive edges, whereby larger-sized dots indicate that a given ROI was associated with more predictive edges. The chord diagrams indicate which networks were associated with the most predictive edges, whereby thicker arcs indicate that a given pair of networks is associated with many predictive edges.

Table 2
Prevalence of each network pair among the top-1000 edges implicated by recursive feature elimination. The table details the recursive feature elimination results visualized in Fig. 4A. These results are representative of the results for the other connectome-wide methods, as each other method likewise only identified significant frequencies for edges connected to the Visual network or DAN. This table follows the reporting by Eryilmaz et al. (2020) but with added significance testing. For a given pair of networks, "#Top 1000″ specifies the number of associated edges implicated by feature selection, and "#Features" indicates the total number of edges between a given network pair regardless of the feature selection results. The latter count was used to calculate the expected frequency of top edges for a given network pair if top edges were uniformly distributed. "Weighted prevalence" is the number of top edges divided by expected frequency. The p-values correspond to whether the weighted prevalence significantly surpassed 1.0. Only pairs of networks where weighted prevalence surpassed 1.0 are reported.

Within Network				
Network Pair	#Features	#Top 1000	Weighted Prevalence	p-value
Visual-Visual	703	40	1.98	.001
DAN-DAN	276	13	1.64	.042
FPCN-FPCN	595	22	1.28	.189
Between Network				
Network Pair	#Features	#Top 1000	Weighted Prevalence	p-value
Visual-FPCN	1330	65	1.70	.002
FPCN-DAN	840	39	1.61	.002
Visual-DAN	912	41	1.56	.011
VAN-DAN	576	22	1.33	.066
DMN-DAN	1392	51	1.27	.099
VAN-FPCN	840	29	1.20	.141
FPCN-DMN	2030	70	1.20	.148
Visual-DMN	2204	74	1.17	.187
Visual-Limbic	418	13	1.08	.293
SM-DMN	2552	76	1.03	.384
VAN-DMN	1392	41	1.02	.378

compared, next, using a larger dataset.

3.2. Large sample size

3.2.1. Group-level ConnSearch

Submitting the larger dataset (N = 250) parcellated with the Schaefer et al. (2018) atlas to *ConnSearch* allowed it to map the topography of connectivity patterns elicited by WM load with precision (Fig. 5). This precision is evident in two notable patterns that arose. First, in the posterior cortex, the 250-participant analysis pointed to two clusters: one early visual cluster and one posterior parietal cluster, which may reflect dissociable visual and attentional processing, respectively (Weber et al., 2017). Second, in the prefrontal cortex (PFC), a dorsolateral PFC cluster emerged and showed an anterior-posterior gradient, which may reflect the hierarchical organization of processing (Jung et al., 2022; Nee and D'Esposito, 2016). Neither of these specific patterns emerged in the analysis of the limited dataset. Thus, at a larger sample size, *ConnSearch* yielded sharper localizations on the effects of WM load. Although it cannot be said for certain that these findings are "correct" as the "ground truth" is unknown, the results are consistent with existing WM research (see Discussion 4.1). It is also worth noting that networks and regions typically thought to be less involved in WM, such as the Limbic network or orbitofrontal cortex, continue to not be identified, which suggests that *ConnSearch* does not produce false positives.

3.2.2. Comparison to connectome-wide interpretation methods

Comparisons to existing methods using the larger dataset provided further evidence that the *ConnSearch* results more comprehensively represented the patterns elicited by WM load. For example, the occipital vs. parietal dissociation identified via *ConnSearch* was less evident in the results from the connectome-wide classification/interpretation methods (Fig. 6). Specifically, recursive feature elimination only showed the involvement of edges connected to the early visual cortex, while connectome predictive modeling and Haufe transformation both only pointed to a single broad occipitoparietal cluster without any dissociation. In addition, *ConnSearch* identified particular PFC effects, primarily implicating posterior portions of the lateral PFC. On the other hand, recursive feature elimination showed no clear frontal involvement, and the other two techniques primarily implicated medial and relatively anterior portions of the PFC. As noted above, we have no "ground truth" about what regions are more relevant, but meta-analyses on BOLD activation elicited by WM generally show more prominent lateral and posterior PFC effects than medial and anterior PFC effects (de Voogd and Hermans, 2022; Wang et al., 2019). These results overall complement our initial findings using the smaller dataset and add to the evidence that *ConnSearch* more precisely and comprehensively localizes the neural correlates of a task manipulation than these alternative methods.

3.3. Group-level reliability and replicability

Our reliability analyses revealed that network components provide more reliable measurements than individual edges. Reliability – the extent to which participants' measurements in one run predict their measurements in other sessions – was assessed using the test-retest HCP dataset. This revealed that medium-to-highly reliable component-level measurements (ICC ≈ 0.60) require six sessions of data (30 min of scan time), whereas similarly reliable edge-level measurements require nine sessions of data (45 min; for more details, see Supplemental Results

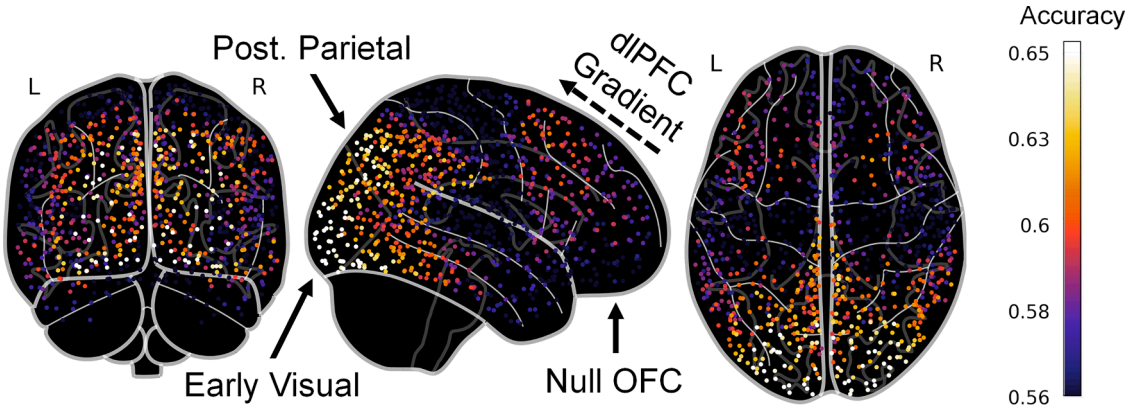

Fig. 5. *Network component accuracies at a large sample size.* Analysis used connectomes derived from parcellation with the Schaefer et al. (2018) atlas (N = 250 dataset). Each ROI's color indicates the accuracy of the component classifier for which it is the Core ROI. For example, the white ROIs in the early visual cortex indicate that components based on early visual ROIs yielded highly predictive classifiers. Post., Posterior; dlPFC, Dorsolateral prefrontal cortex; OFC, Orbitofrontal cortex.

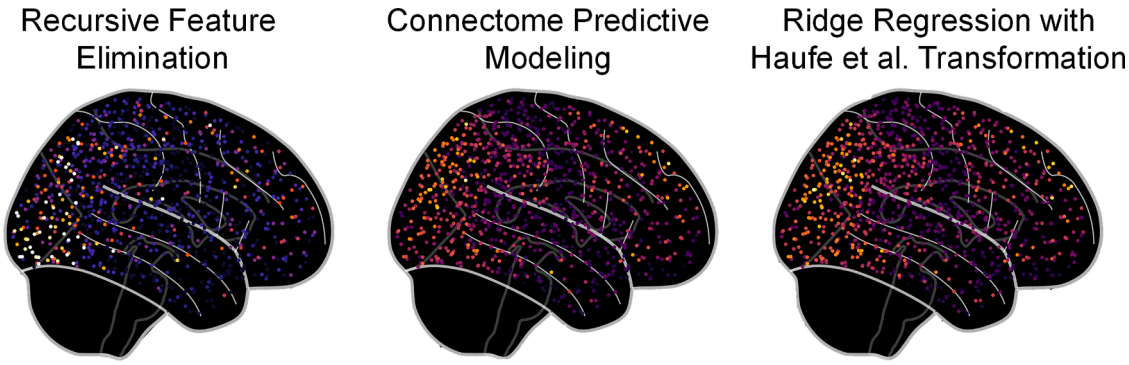

Fig. 6. *Results of alternative methods applied to the large dataset.* Analysis used connectomes derived from parcellation with the Schaefer et al. (2018) atlas ($N = 250$ dataset). Hot colors indicate ROIs whose associated edges were most frequently implicated as highly predictive by a given interpretation method. Neighborhood component feature selection was not computationally feasible for this larger dataset.

2.2). Hence, there is an advantage to analyzing network components, rather than single edges.

Our final group-level results showed that *ConnSearch* yielded more replicable findings than interpretation via recursive feature elimination. This was assessed by submitting five independent datasets ($N = 50$ each) to *ConnSearch* and then doing the same using the connectome-wide approach. This revealed that the ROIs implicated as relevant by *ConnSearch* showed greater overlap among the different datasets (DSC = 0.663 [.628, 0.698]) than the ROIs implicated by the connectome-wide method (DSC = 0.583 [.554, 0.612]).

3.4. Subject-specific ConnSearch

Beyond just group-level analyses, *ConnSearch* components can be also modeled in other fashions. Subject-specific *ConnSearch* models each component's data using correlational similarity rather than classification. This analysis attempts to identify components where the connectivity patterns elicited by a condition in the first session, for a given participant, predict the participant's connectivity patterns in the second session. As expected, subject-specific *ConnSearch* found several significant network components sensitive to WM load, even when using the limited dataset ($N = 50$; Fig. 7 top). For example, the FPCN, which was implicated by the group-level analyses, was also identified by this subject-specific analysis (Table 3). On the other hand, some networks that showed clear group-level effects did not show major subject-specific patterns (e.g., Visual network & DMN), suggesting that these networks' task-related responses are more homogeneous across the population. Such dissociations between networks highlight how subject-specific analyses provide unique insights that complement group-level analyses and how *ConnSearch* localizes these effects. For confirmatory purposes, subject-specific *ConnSearch* was also performed after removing group-level effects by subtracting group means for each condition on an edge-by-edge basis. Three components remained significant (Entries #1, #6, #7), meaning that, for at least these three components, WM load elicits configurations of connectivity unique to each participant.

Further analyses add to the robustness of the FPCN findings. For instance, similar results arose when using 24-sized or 32-sized components (Supplemental Tables S4 & S5). Moving to a larger dataset ($N = 250$) parcellated with the Schaefer et al. (2018) atlas also showed FPCN patterns and now with enhanced precision (Fig. 7 bottom). Analyses of replicability moreover revealed that the patterns found using 50

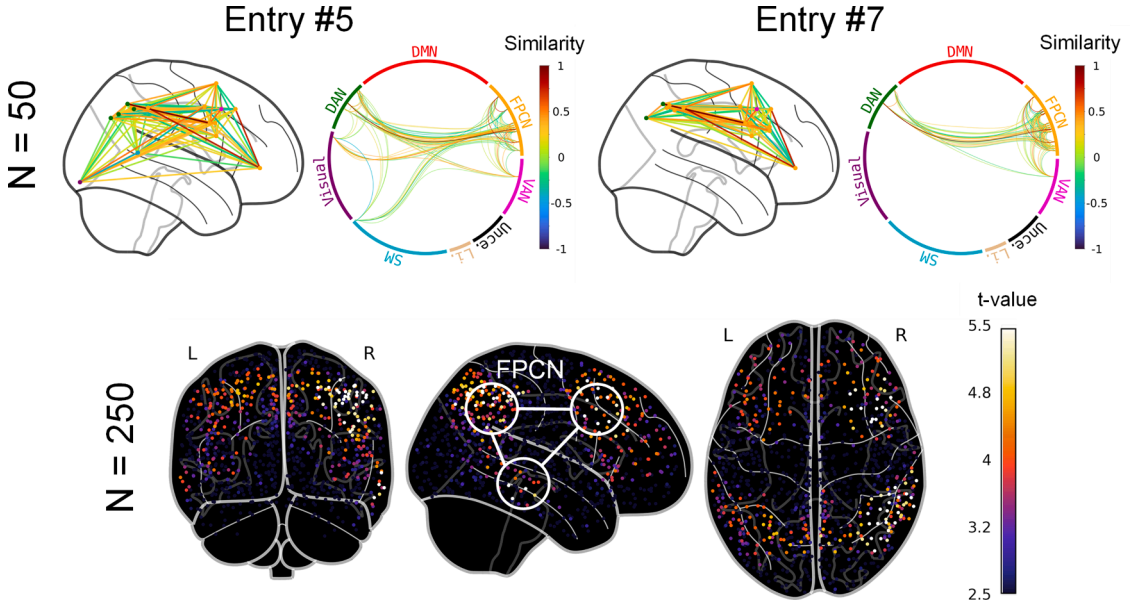

Fig. 7. *Significant network components identified by subject-specific ConnSearch.* **Top.** Two significant components are shown from the subject-specific analysis of the limited dataset ($N = 50$), parcellated using the Power et al. (2011) atlas. The components are titled by their entry number in Table 3. Edge colors correspond to similarity in terms of Euclidean distance. **Bottom.** Results reflect the analysis of the large dataset ($N = 250$), parcellated using the Schaefer et al. (2018) atlas and while subtracting group-level 2-back vs. 0-back differences. Each ROI's color indicates the t-value of the component model for which it is the Core ROI.

Table 3
Significant network components identified by subject-specific ConnSearch. Asterisks indicate components that remain significant ($p_{FWE} < 0.05$) after regressing out the group average. See the Table 1 caption for details on the table organization. AnG, Angular gyrus; FuG, Fusiform gyrus; SPL, Superior parietal lobule; PrG, Precentral gyrus; MFG, Middle frontal gyrus.

	Core ROI Details				Network Component Composition					
Entry #	Region	BA	ROI #	t-value	DAN	DMN	FPCN	SM	VAN	Visual
1	L AnG*	40	194	4.42	19%	12%	69%	0%	0%	0%
2	L FuG	37	243	3.90	12%	0%	0%	6%	0%	81%
3	R SPL	40	198	3.85	31%	0%	38%	0%	0%	31%
4	R PrG	6	204	4.64	25%	0%	31%	0%	6%	38%
5	L PrG	6	173	4.07	25%	0%	56%	6%	6%	6%
6	R MFG*	46	174	4.54	25%	12%	56%	0%	6%	0%
7	L MFG*	9	200	4.73	12%	0%	81%	0%	6%	0%

participants converge across multiple datasets more than what would be expected by chance (Supplemental Results 2.3). Overall, these results show that task manipulations elicit connectivity effects unique to each participant that go above and beyond group-level patterns. Furthermore, these results show that *ConnSearch* is an effective means of localizing these effects to specific areas of the connectome, showing the framework's flexibility and positioning it as a potential tool for the emerging literature on between-person heterogeneity in brain networks (Finn et al., 2015; Tobyne et al., 2018).

4. Discussion

The present research introduces a novel framework applicable to functional connectivity analysis, referred to as "*ConnSearch.*" The study yielded four main findings. Regarding the group-level analyses, *first, ConnSearch* is viable for task-fMRI research, even when sample sizes are limited ($N = 50$). *Second, ConnSearch* identifies a more extensive and detailed array of neural correlates than the four existing classification and interpretation methods tested. *Third, ConnSearch* operates with more reliable data and yields results that are more replicable compared to traditional methods. *Finally*, regarding the subject-specific analyses, *ConnSearch* can identify network components expressing subject-specific (heterogeneous) effects. This final result, more broadly, shows that *ConnSearch* is viable with different modeling strategies and shows how subject-specific analyses can shed new light on specific networks (e.g., only the FPCN shows subject-specific effects of WM load). Below, we describe how the neuroscientific patterns identified here are consistent with extant research, and we also discuss the methodological implications for *ConnSearch* and multivariate fMRI analysis in general.

4.1. Group-level results and comparisons

By isolating and separately analyzing network components, *ConnSearch* was able to detect the WM roles of both perceptual/attentional networks (Visual Network and DAN) along with executive processing networks (FPCN and DMN), which operate both independently and together with one another. Speaking to the sensitivity of ConnSearch, these results were first demonstrated using a limited ($N = 50$) sample, which is representative of typical sample sizes used in fMRI studies. While we generally advocate for large sample sizes in fMRI research (Durnez et al., 2016; Marek et al., 2022), there may be cases in which smaller sample sizes are unavoidable, such as when working with special populations, where the sensitivity provided by *ConnSearch* and similar approaches may be needed.

Applying *ConnSearch* to a larger sample ($N = 250$) yielded more precise results. *ConnSearch* localized early visual cortex vs. posterior parietal cortex processing. In addition, *ConnSearch* pinpointed the involvement of the posterior dorsolateral PFC as the most relevant PFC region, while showing that connections to other PFC areas, like the orbitofrontal cortex, are less sensitive to WM load. The observed patterns are consistent with earlier research showing specific roles of different perceptual processes during picture-based WM tasks (Makovski and Lavidor, 2014; Yu and Shim, 2017). The results also speak to evidence on the general roles of the FPCN and DMN across a range of WM paradigms (Eryilmaz et al., 2020; Kim, 2019). More broadly, these results point to how WM is a network phenomenon and emerges from the coordination between "general purpose" frontal control systems and more "specialized" posterior systems engaged according to the particular type of processing that they support (e.g., sensory-perceptual, verbal-semantic, and sensorimotor; Christophel et al., 2017; Lorenc and Sreenivasan, 2021; Reuter-Lorenz and Iordan, 2021). Hence, *ConnSearch* is poised to particularly benefit this type of research by assisting in distributed-oriented interpretations about brain networks rather than specific regions, which can thus validate the distributed nature of the tasks.

On the other hand, the comparison methods yielded less comprehensive results, which only supported conclusions about a subset of the WM correlates noted above. Specifically, the four connectome-wide classification and interpretation methods tested – recursive feature elimination, neighborhood component feature selection, connectome predictive modeling, and Haufe transformation – all found that visual/attentional edges best predicted WM load. However, the four methods were more limited in identifying relevant frontal areas and connections linked to the FPCN and DMN. These challenges were most prominent when the sample size was limited, but even when using the larger dataset, no method pinpointed the dorsolateral PFC's role in WM. For the larger dataset, one interesting pattern that arose was a difference between the methods in what posterior regions they most implicated. Recursive feature elimination found connections to the early visual cortex to be most predictive, whereas connectome predictive modeling and Haufe transformation pointed to connections with the posterior parietal cortex. Hence, the different interpretation methods diverge in what they deem most predictive, which can lead to patterns being missed.

Earlier studies using whole-connectome/whole-brain classifiers to predict task conditions have likewise found that the occipital lobe is among the most predictive areas of the brain whenever virtually any picture-based task is used (Xu et al., 2020). This focus on the occipital lobe even emerges in research employing deep neural networks (Wang et al., 2020b). It is probably due to the ample and stable signal in the occipital areas (Grill-Spector et al., 1999), which virtually any picture-based task will modulate in some way. Consistent with this explanation, overemphasis on occipital lobe patterns occurs less in studies regressing individual differences on resting-state data (Dubois et al., 2018; Eryilmaz et al., 2020; Shen et al., 2017). These patterns speak to our premise that this limitation of interpretation analyses may be a particular disadvantage for research on task classification. *ConnSearch* overcomes this pitfall by analyzing components individually, which precludes patterns within one component from influencing analyses of other components.

The usefulness of *ConnSearch* under sample size constraints is strengthened by our assessments of replicability and reliability. Replicability is receiving ever-growing attention in fMRI research, although relatively less so in the context of machine learning methods. Instead,

machine learning studies largely have focused on the *generalizability* of models, examining whether trained models can produce accurate predictions when applied to different datasets (e.g., Fountain-Zaragoza et al., 2019; Rosenberg et al., 2016). Yet, generalizability is distinct from *replicability* – whether different datasets lead to the same neuroscientific inferences. Our research provides initial evidence on this topic and points to its importance. We also provide evidence regarding *reliability*. We found that 30 min of scan time is sufficient for "good reliability" when analyzing components (ICC ≈ 0.60; Noble et al., 2021), whereas 45 min is required when analyzing edges. As fMRI-based machine learning approaches continue to develop, these topics will likely garner further interest.

To be clear, our goal was not to develop a classifier that surpasses others in classification accuracy. Deep learning techniques generally perform best in this regard. For functional connectivity data, graph neural networks achieve nearly perfect accuracy in dissociating task conditions (Li et al., 2021b). Likewise, for analyses of 4D activation maps, convolutional neural networks yield nearly perfect accuracy (Wang et al., 2020b). Our focus was solely on how multivariate methods can be leveraged to localize the neural correlates of a dependent variable. For these purposes, *ConnSearch* proved effective.

4.2. Sensitivity and cross-validation

In the Introduction, we noted that a challenge to analyzing components individually and fitting multiple classifiers is that models may not have sufficient data for accurate predictions. This is difficult because models tend to benefit from having access to as much information as possible – e.g., note how using 32-ROI components yielded many more significant components (Supplemental Table S3). Adding to this challenge, fitting many separate classifiers (264 or 1000 in this case) requires stricter thresholds for statistical significance as multiple hypothesis correction. Nonetheless, our results showed that these challenges can be overcome, even when limited to moderate sample sizes ($N = 50$ or less).

Much of this success hinged on the use of repeated k-fold cross-validation. That is, for cross-validation, the data were split into fifty overlapping 80/20 train-test sets. This type of cross-validation deviates from the more popular method, where just five 80/20 folds are used. However, as described in Supplemental Methods 1.1, repeated k-fold cross-validation will lower the accuracy threshold derived from permutation-testing (lowering Type II error) and increase the precision of the test accuracy (lowering Type I error). These benefits come from the fact that test accuracy is essentially measured using a pool that contains 1000% of the original examples (10 repetitions × 100% testing data = 1000%). Permutation-testing and Supplemental Fig. S1 show how, if the traditional 80/20 scheme were used, the permutation-testing significance threshold would be much higher, while not conferring any advantages. These points were explored in detail by Valente et al. (2021), who investigated and confirmed these aspects in the context of MVPA. Our findings demonstrate applicability to connectivity research and how, by leveraging these principles, new methods become available.

4.3. Subject-Specific ConnSearch

Subject-specific *ConnSearch* proved to be a similarly effective tool, and these analyses revealed that WM load elicits unique configurations of FPCN connectivity that vary across individuals but are stable within individuals. This pattern was first shown in the limited dataset ($N = 50$), which speaks to the sensitivity of *ConnSearch* and demonstrates how such analyses need not be exclusive to studies with huge samples. However, having more participants imparts benefits. Analyzing the larger dataset ($N = 250$), *ConnSearch* precisely mapped the FPCN's involvement, showing highly predictive components linked to ROIs from the parietal lobe, dorsolateral PFC, and lateral temporal lobe. The subject-specific FPCN patterns notably emerged above and beyond the trends common across the group. In contrast, heterogeneous effects like these were not seen for visual/attentional network components, suggesting that the effect of WM load on these networks is more homogeneous across the population. Hence, WM load elicits a mix of group-level and subject-specific effects, which can be found when combining multiple modeling strategies.

This general perspective on the heterogeneous vs. homogeneous nature of functional connectivity has seen increased attention recently, with fingerprinting research demonstrating that individuals can be distinguished from the rest of the population based on their connectomes (Finn et al., 2015; Li et al., 2021a; Ravindra et al., 2021). However, thus far, most fingerprinting studies have focused on resting-state data (Li et al., 2021a) or on participants' overall connectivity patterns during a task (Finn et al., 2015; Ravindra et al., 2021). Our study was relatively unique in that it examined 2-back vs. 0-back differences (Tobyne et al., 2018), and hence permits insight into how task-related cognitive processes manifest differently among different people. Examining cognitive processes in subject-specific manners may allow unique insights into complex processes such as intelligence, which, to a degree, rely on disparate mechanisms across the population (Iuculano et al., 2020). Altogether, the results showcase how *ConnSearch*, along with its underlying principles, can support investigations into mostly unexplored questions.

4.4. Future directions

The *ConnSearch* approach is modular and can be modified in two key manners: First, how the connectome is divided into subsets can be changed. Second, how each component is modeled can also be changed. Both are important degrees of freedom. Concerning the component definition, the present report used a definition that we thought to be the most straightforward and intuitive: One component was created for each ROI, using it and the $N - 1$ ROIs with which it is most strongly connected. However, this definition may carry limitations in some contexts, such as when weak connections are relevant (Santarnecchi et al., 2014). The released software can use other component definitions. For example, components can be organized in terms of the nodes that are most proximal to the Core ROIs. Our preliminary analyses (not reported) using the HCP motor task (dissociating hand vs. foot movement; Barch et al., 2013) showed that when components are defined in terms of proximity, *ConnSearch* pinpoints significant motor cortex components. For subject-specific analyses, researchers may also want to consider defining components differently for each participant, given that a uniform definition may not accurately reflect every person's functional network. Indeed, evidence shows that even the organization of the canonical networks varies somewhat across individuals (Gordon et al., 2017). Although defining unique components has disadvantages – e.g., components may be less reliable because data are not pooled from multiple people – this may be a fruitful direction.

Beyond just using components, it is also worth considering alternative ways of dividing the connectome, such as the technique by Wu et al. (2021) mentioned in the Introduction, which fits a classifier based on all edges connected to a given ROI. Comparatively, *ConnSearch* may be more effective especially when dealing with smaller samples, potentially because component data are less collinear than edges connected to a single ROI. Yet, their technique also has advantages, such as in pinpointing the relevance of specific ROIs. The present research broadly supports this emerging paradigm and encourages the development of further analytic strategies of this nature, which will likely come with other advantages and trade-offs.

Regarding the component-level modeling procedure, SVMs and correlational similarity were used, as earlier research has demonstrated the effectiveness of these techniques (Walther et al., 2016). However, any type of modeling can be applied, and more sophisticated algorithms may provide enhanced accuracy and sensitivity. The use of machine

learning methods itself is not a strict requirement. For example, Hotelling's multivariate *t*-tests can be used for group-level and/or subject-specific analyses (requires trial-based designs or a large number of blocks; Gilron et al., 2017). The released code permits different options and, as with the component definitions, researchers can specify custom classifiers.

5. Conclusion

Overall, the present report provides evidence for *ConnSearch* as a novel framework for group-level and subject-specific functional connectivity analyses. More broadly, our group-level findings demonstrate that, with the right procedures, machine learning tools can find effective roles within traditional neuroimaging studies, even when using moderate sample sizes. Additionally, our subject-specific findings illustrate heterogeneity regarding how different individuals' brains will respond to the same task, and how these patterns can be localized to specific networks. To support future investigations of this nature and the adoption of machine learning by cognitive neuroscientists, all the current code is made available publicly with detailed documentation and instructions.

Ethics statement

The present research uses an already collected public dataset, and it is considered by the University of Illinois Institutional Review Board to not require review.

Code and data availability statement

The code needed to reproduce the present results and the main preprocessed functional connectivity dataset used for the analyses have been released publicly on GitHub (https://github.com/paulcbogdan/ConnSearch). The functional connectivity dataset is based on the task-fMRI data by the Human Connectome Project, which is also publicly available online.

Conflict of interest statement

The authors declare no conflicts of interest.

CRediT authorship contribution statement

Paul C. Bogdan: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualization. **Alexandru D. Iordan:** Conceptualization, Methodology, Writing – original draft. **Jonathan Shobrook:** Software, Formal analysis, Methodology. **Florin Dolcos:** Conceptualization, Supervision, Project administration, Funding acquisition, Methodology, Resources, Writing – review & editing.

Declaration of Competing Interest

Declarations of interest: none.

Data availability

The code and the preprocessed functional connectivity data needed to reproduce our results have been released publicly on GitHub.

Acknowledgments

This research was carried out in part at the University of Illinois' Beckman Institute for Advanced Science & Technology. During the preparation of this manuscript, P.C.B. was supported by a Predoctoral Fellowship provided by the Beckman Foundation and a Dissertation Completion Fellowship provided by the University of Illinois, and F.D. was supported by an Emanuel Donchin Professorial Scholarship in Psychology from the University of Illinois. The authors wish to thank Dr. Humphrey Shi, Dr. Aron K. Barbey, and Dr. Sanda Dolcos for their feedback on an earlier version of this manuscript. The authors also wish to thank members of the Dolcos Lab for feedback on the development of the tool. The authors would additionally like to thank the three anonymous reviewers for their constructive feedback

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120274.

References

Abdi, H., 2007. Bonferroni and Šidák corrections for multiple comparisons. Encycl. Meas. Stat. 3, 103–107.
Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8, 14.
Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Feldt, C., 2013. Function in the human connectome: task-fMRI and individual differences in behavior. Neuroimage 80, 169–189.
Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37 (1), 90–101.
Bonhoeffer, T., Grinvald, A., 1993. The layout of iso-orientation domains in area 18 of cat visual cortex: optical imaging reveals a pinwheel-like organization. J. Neurosci. 13 (10), 4157–4180.
Chen, J., Tam, A., Kebets, V., Orban, C., Ooi, L.Q.R., Asplund, C.L., Bzdok, D., 2022. Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. Nat. Commun. 13 (1), 1–17.
Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., Haynes, J.D., 2017. The distributed nature of working memory. Trends Cogn. Sci. (Regul. Ed.) 21 (2), 111–124.
Cole, M.W., Ito, T., Schultz, D., Mill, R., Chen, R., Cocuzza, C., 2019. Task activations produce spurious but systematic inflation of task functional connectivity estimates. Neuroimage 189, 1–18.
D'Esposito, M., Postle, B.R., 2015. The cognitive neuroscience of working memory. Annu. Rev. Psychol. 66, 115–142.
de Voogd, L.D., Hermans, E.J., 2022. Meta-analytic evidence for downregulation of the amygdala during working memory maintenance. Hum. Brain Mapp. 43 (9), 2951–2971.
Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.
Dubois, J., Galdi, P., Paul, L.K., Adolphs, R., 2018. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. Philos. Trans. R. Soc. B Biol. Sci. 373 (1756), 20170284.
Durnez, J., Degryse, J., Moerkerke, B., Seurinck, R., Sochat, V., Poldrack, R.A., Nichols, T.E., 2016. Power and sample size calculations for fMRI studies based on the prevalence of active peaks. Biorxiv, 049429.
Elliott, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Hariri, A. R., 2020. What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. Psychol. Sci. 31 (7), 792–806.
Eryilmaz, H., Dowling, K.F., Hughes, D.E., Rodriguez-Thompson, A., Tanner, A., Huntington, C., Roffman, J.L., 2020. Working memory load-dependent changes in cortical network connectivity estimated by machine learning. Neuroimage 217, 116895.
Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Constable, R. T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18 (11), 1664–1671.
Fountain-Zaragoza, S., Samimy, S., Rosenberg, M.D., Prakash, R.S., 2019. Connectome-based models predict attentional control in aging adults. Neuroimage 186, 1–13.
Gilron, R., Rosenblatt, J., Koyejo, O., Poldrack, R.A., Mukamel, R., 2017. What's in a pattern? Examining the type of signal multivariate analysis uncovers at the group level. Neuroimage 146, 113–120.
Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Polimeni, J.R., 2013. The minimal preprocessing pipelines for the human connectome project. Neuroimage 80, 105–124.
Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Sun, H., 2017. Precision functional mapping of individual human brains. Neuron 95 (4), 791–807 e797.
Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24 (1), 187–203.
Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (1), 389–422.
Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430.
Iuculano, T., Padmanabhan, A., Chen, L., Nicholas, J., Mitsven, S., de los Angeles, C., Menon, V., 2020. Neural correlates of cognitive variability in childhood autism and relation to heterogeneity in decision-making dynamics. Dev. Cogn. Neurosci. 42, 100754.
Jung, J., Ralph, M.A.L., Jackson, R.L., 2022. Subregions of DLPFC display graded yet distinct structural and functional connectivity. J. Neurosci. 42 (15), 3241–3252.
Kim, H., 2019. Neural activity during working memory encoding, maintenance, and retrieval: a network-based model and meta-analysis. Hum. Brain Mapp. 40 (17), 4912–4933.
Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis-connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4.
Li, K., Wisner, K., Atluri, G., 2021a. Feature selection framework for functional connectome fingerprinting. Hum. Brain Mapp. 42 (12), 3717–3732.
Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Duncan, J.S., 2021b. Braingnn: interpretable brain graph neural network for fmri analysis. Med. Image Anal. 74, 102233.
Lorenc, E.S., Sreenivasan, K.K., 2021. Reframing the debate: the distributed systems view of working memory. Vis. Cogn. 29 (7), 416–424.
Maitra, R., 2010. A re-defined and generalized percent-overlap-of-activation measure for studies of fMRI reproducibility and its use in identifying outlier activation maps. Neuroimage 50 (1), 124–135.
Makovski, T., Lavidor, M., 2014. Stimulating occipital cortex enhances visual working memory consolidation. Behav. Brain Res. 275, 84–87.
Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Hendrickson, T.J., 2022. Reproducible brain-wide association studies require thousands of individuals. Nature 603 (7902), 654–660.
Nee, D.E., D'Esposito, M, 2016. The hierarchical organization of the lateral prefrontal cortex. Elife 5, e12112.
Noble, S., Scheinost, D., Constable, R.T., 2021. A guide to the measurement and interpretation of fMRI test-retest reliability. Curr. Opin. Behav. Sci. 40, 27–32.
Plitt, M., Barnes, K.A., Martin, A., 2015. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. NeuroImage Clin. 7, 359–366.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Schlaggar, B.L., 2011. Functional network organization of the human brain. Neuron 72 (4), 665–678.
Ravindra, V., Drineas, P., Grama, A., 2021. Constructing compact signatures for individual fingerprinting of brain connectomes. Front. Neurosci. 15, 344.
Reuter-Lorenz, P.A., Iordan, A.D., 2021. Remembering over the short and long term: empirical continuities and theoretical implications. In: Logie, R.H., Camos, V., Cowan, N. (Eds.), Working Memory: State of the Science, 282-310. Oxford University Press.
Rosenberg, M.D., Finn, E.S., Scheinost, D., Papademetris, X., Shen, X., Constable, R.T., Chun, M.M., 2016. A neuromarker of sustained attention from whole-brain functional connectivity. Nat. Neurosci. 19 (1), 165–171.
Santarnecchi, E., Galli, G., Polizzotto, N.R., Rossi, A., Rossi, S., 2014. Efficiency of weak brain connections support general cognitive functioning. Hum. Brain Mapp. 35 (9), 4566–4582.
Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Yeo, B.T., 2018. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cereb. Cortex 28 (9), 3095–3114.
Scheinost, D., Noble, S., Horien, C., Greene, A.S., Lake, E.M., Salehi, M., Barron, D.S., 2019. Ten simple rules for predictive modeling of individual differences in neuroimaging. Neuroimage 193, 35–45.
Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., Constable, R.T., 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat. Protoc. 12 (3), 506–518.
Sorensen, T.A., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biologiske Skrifter 5, 1–34.
Sporns, O., 2013. Network attributes for segregation and integration in the human brain. Curr. Opin. Neurobiol. 23 (2), 162–171.
Tian, Y., Zalesky, A., 2021. Machine learning prediction of cognition from functional connectivity: are feature weights reliable? Neuroimage 245, 118648.
Tobyne, S.M., Somers, D.C., Brissenden, J.A., Michalka, S.W., Noyce, A.L., Osher, D.E., 2018. Prediction of individualized task activation in sensory modality-selective frontal cortex with 'connectome fingerprinting. Neuroimage 183, 173–185.
Valente, G., Castellanos, A.L., Hausfeld, L., De Martino, F., Formisano, E., 2021. Cross-validation and permutations in MVPA: validity of permutation strategies and power of cross-validation schemes. Neuroimage 238, 118145.
Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., 2013. The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79.
Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., Diedrichsen, J., 2016. Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage 137, 188–200.
Wang, H., He, W., Wu, J., Zhang, J., Jin, Z., Li, L., 2019. A coordinate-based meta-analysis of the n-back working memory paradigm using activation likelihood estimation. Brain Cogn. 132, 1–12.
Wang, Q., Cagna, B., Chaminade, T., Takerkart, S., 2020a. Inter-subject pattern analysis: a straightforward and powerful scheme for group-level MVPA. Neuroimage 204, 116205.
Wang, X., Liang, X., Jiang, Z., Nguchu, B.A., Zhou, Y., Wang, Y., Wu, F., 2020b. Decoding and mapping task states of the human brain via deep learning. Hum. Brain Mapp. 41 (6), 1505–1519.
Weaverdyck, M.E., Lieberman, M.D., Parkinson, C., 2020. Tools of the Trade Multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. Soc. Cogn. Affect. Neurosci. 15 (4), 487–509.
Weber, E.M.G., Hahn, T., Hilger, K., Fiebach, C.J., 2017. Distributed patterns of occipito-parietal functional connectivity predict the precision of visual working memory. Neuroimage 146, 404–418.
Whitfield-Gabrieli, S., Nieto-Castanon, A., 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect 2 (3), 125–141.
Wieczorek, O., Unger, S., Riebling, J., Erhard, L., Koß, C., Heiberger, R., 2021. Mapping the field of psychology: trends in research topics 1995–2015. Scientometrics 1–33.
Wu, J., Eickhoff, S.B., Hoffstaedter, F., Patil, K.R., Schwender, H., Yeo, B.T., Genon, S., 2021. A connectivity-based psychometric prediction framework for brain–behavior relationship studies. Cereb. Cortex 31 (8), 3732–3751.
Xu, W., Li, Q., Liu, X., Zhen, Z., Wu, X., 2020. Comparison of feature selection methods based on discrimination and reliability for fMRI decoding analysis. J. Neurosci. Methods 335, 108567.
Yang, W., Wang, K., Zuo, W., 2012. Neighborhood component feature selection for high-dimensional data. J. Comput. 7 (1), 161–168.
Yarkoni, T., 2009. Big correlations in little studies: inflated fMRI correlations reflect low statistical power—Commentary on Vul et al.(2009). Perspect. Psychol. Sci. 4 (3), 294–298.
Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkari, D., Hollinshead, M., Polimeni, J., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165.
Yoo, K., Rosenberg, M.D., Hsu, W.T., Zhang, S., Li, C.S.R., Scheinost, D., Chun, M.M., 2018. Connectome-based predictive modeling of attention: comparing different functional connectivity features and prediction methods across datasets. Neuroimage 167, 11–22.
Yu, Q., Shim, W.M., 2017. Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. Neuroimage 157, 97–107.
Zalesky, A., Fornito, A., Bullmore, E.T., 2010. Network-based statistic: identifying differences in brain networks. Neuroimage 53 (4), 1197–1207.