



Distributional independent component analysis for diverse neuroimaging modalities

Ben Wu¹ | Subhadip Pal² | Jian Kang³  | Ying Guo⁴ 

¹ Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China

² Department of Biostatistics and Bioinformatics, University of Louisville, Louisville, Kentucky, USA

³ Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA

⁴ Department of Biostatistics and Bioinformatics, Emory University, Atlanta Georgia, USA

Correspondence Ying Guo, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA.
Email: yguo2@emory.edu

Funding information

NIH, Grant/Award Numbers: R01MH105561, R01MH118771, R01DA048993, R01GM124061, R01CA249096

Abstract

Recent advances in neuroimaging technologies have provided opportunities to acquire brain images of different modalities for studying human brain organization from both functional and structural perspectives. Analysis of images derived from various modalities involves some common goals such as dimension reduction, denoising, and feature extraction. However, since these modalities have vastly different data characteristics, the current analysis is usually performed using distinct analytical tools that are only suitable for a specific imaging modality. In this paper, we present a Distributional Independent Component Analysis (DICA) that represents a new approach that performs decomposition on the distribution level, providing a unified framework for extracting features across imaging modalities with different scales and representations. When applying DICA to fMRI images, we successfully recover well-established brain functional networks in neuroscience literature, providing empirical validation that DICA delivers neurologically relevant findings. More importantly, we discover several structural network components when applying DICA to DTI images. Through fiber tracking, we find these DICA-derived structural components correspond to several major white fiber bundles. To the best of our knowledge, this is the first time these fiber bundles are successfully identified via blind source separation on single subject DTI images. We also evaluate the performance of DICA as compared with existing ICA methods through extensive simulation studies.

KEYWORDS

DTI, fMRI, independent component analysis, multimodality neuroimaging

1 | INTRODUCTION

With the advancement of neuroimaging techniques, it has become increasingly common for neuroscience studies to collect multimodal brain imaging data in order to obtain a more comprehensive understanding of brain function and structure. For example, functional magnetic resonance imaging (fMRI), which is currently the most prominent functional neuroimaging modality, measures the hemodynamic response related to neural activity in the brain; diffu-

sion tensor imaging (DTI), an increasingly important structural imaging modality, maps white fiber tract structure by measuring water diffusion in the brain. Data collected from these different imaging modalities offer complementary views on the brain organization and could potentially provide new insights into the relationship between brain function and structure (Guye *et al.*, 2008). For example, the structural connections revealed by DTI can help understand and verify the functional networks identified based on fMRI and also provide useful information for

constructing biologically plausible models for brain functional networks (Ramnani *et al.*, 2004). It has become increasingly clear that utilizing information from diverse imaging modalities leads to more effective neuroscience research.

Analyzing data derived from different imaging modalities is a challenging task. One of the major difficulties is that imaging measurements across these modalities have different data representations. For example, fMRI records a time series of blood-oxygen-level-dependent (BOLD) signals reflecting the hemodynamic responses at a voxel, that is, the volume element within fMRI data. DTI data are represented by a 3D tensor matrix at each voxel representing the local water diffusion pattern. Furthermore, these modalities have different noise levels and intensities. Given these issues, existing methods have to adopt different analytical approaches that are specific to each modality (Calhoun and Sui, 2016). This often causes challenges for integrating information and results across modalities. Therefore, it is desirable to develop a general analytical framework that can accommodate different types of imaging data.

Some common objectives in analyzing data from different imaging modalities include dimension reduction, denoising and extraction of latent features such as the functional networks from fMRI data and white matter structural networks from DTI data. A useful computational tool that can achieve these goals is independent component analysis (ICA). ICA is one of the most widely applied blind source separation methods for recovering latent features underlying multivariate observations. This rapidly evolving technique has found successful applications in a wide range of scientific areas such as biomedical imaging (e.g., neuroimaging and optical imaging), visual receptive fields, signal processing, and machine learning (Hyvärinen *et al.*, 2001; Bartlett *et al.*, 2002; Beckmann *et al.*, 2005). Essentially, ICA is a computational method for decomposing observed multivariate data into additive components that are statistically as independent as possible. The problem is typically expressed as follows. Let \mathbf{Y} be the observed multivariate data matrix with the dimension of $T \times J$ that contains T mixed signals of length J each. The classical noise-free ICA method decomposes the observed data into a linear combination of latent independent sources as

$$\mathbf{Y}_{T \times J} = \mathbf{A}_{T \times L} \mathbf{S}_{L \times J}, \quad (1)$$

where \mathbf{A} is a mixing matrix, \mathbf{S} represents non-Gaussian source signals, and L is the number of latent independent sources that is smaller than T . The L source signals are assumed to be statistically independent. ICA has been shown to be a highly effective tool for dimension

reduction, denoising and extraction of latent source signals (Hyvärinen *et al.*, 2001). ICA has several appealing properties that make it a popular tool in many fields. For example, compared with methods such as PCA and factor analysis that are based on the second-order statistics of the data, ICA exploits information from higher order statistics that are relevant for non-Gaussian data. Additionally, the key assumption of ICA, that is, statistical independence across components, is often supported by large-scale data with sparse signals (Beckmann *et al.*, 2005). Diverse algorithms have been developed for ICA over the past decades. For example, the well-known traditional ICA algorithms include Infomax (Bell and Sejnowski, 1995) and FastICA (Hyvärinen, 1999). More advanced methods were developed soon in engineering, statistical and computer science societies, such as maximum likelihood estimations (e.g., Hastie and Tibshirani (2003)), rank-based estimations (e.g., Hallin and Mehta (2015)), and estimations using deep learning techniques (e.g., Ngiam *et al.*, 2010; Le *et al.*, 2011). Of note, although the non-Gaussianity of sources is one of the assumptions used in many ICA algorithms, the ICA optimization can be achieved in various ways in addition to maximizing the non-Gaussianity. For example, non-Gaussianity and sample dependence have been considered together to achieve better performance (Adali *et al.*, 2014). Furthermore, recent work demonstrates that statistical independence and sparsity can be considered simultaneously in ICA in order to achieve sparse signal decomposition (Boukouvalas *et al.*, 2018).

In neuroscience research, ICA has become one of the most commonly applied tools for identifying brain functional networks based on fMRI data (McKeown *et al.*, 1998; Calhoun *et al.*, 2001; Beckmann and Smith, 2005; Guo, 2011; Guo and Tang, 2013; Shi and Guo, 2016; Wang and Guo, 2019). The classical ICA model in (1) can be readily applied to decompose a subject's fMRI data (McKeown *et al.*, 1998). Specifically, \mathbf{Y} is the $T \times J$ matrix of observed fMRI signals where T is the number of fMRI scans acquired, each row of \mathbf{Y} is a concatenated 3D brain image acquired during each scan and J is the number of voxels of an image. \mathbf{S} is the spatial source signal matrix where each row represents a concatenated 3D map of an independent source signal. \mathbf{A} represents the temporal mixing matrix that mixes the L spatial sources to generate the observed time series of fMRI images. For fMRI data, each independent component (IC) potentially corresponds to a brain functional network where each row of \mathbf{S} and the corresponding column of \mathbf{A} characterizes the spatial distribution and temporal dynamics for a functional network, respectively. ICA naturally has advantages applied to fMRI data. The spatial independence assumption of ICA corresponds well to the sparsity of fMRI signals, and thus ICA can identify brain functional networks without

constraining the temporal dynamics (Calhoun *et al.*, 2009). In addition, the non-Gaussianity assumption of ICA means we may capture the structured noise of fMRI data with noise components simultaneously, and thus denoise the data. Existing studies have found that ICA can be successfully applied to either resting-state or task-based fMRI analysis, single-subject or multi-subject studies.

Although ICA has been widely applied to fMRI data analysis, its applications to other imaging modalities, especially DTI data, have been very limited. One main reason is that the classical ICA was mainly developed to decompose multivariate observations, such as fMRI time series, but is not applicable to the diffusion tensor matrices in DTI. In a few cases where ICA was applied to DTI data (Li *et al.*, 2012; Ouyang *et al.*, 2015), investigators first obtained at each voxel the fractional anisotropy (FA) (Basser *et al.*, 1994; Koay *et al.*, 2006) that is a scalar summary statistic derived from the estimated tensor matrix and then applied the classical ICA model to decompose multi-subject FA data. This FA-based ICA application has several major limitations. The diffusion tensor matrix from DTI data characterizes both the shape and orientation of the diffusion ellipsoid at each voxel where the eigenvectors of the tensor matrix represent the diffusion directions and the eigenvalues are associated with the speed of diffusion in these directions. The FA, which is a scalar summary statistic derived based on the differences among the eigenvalues, only captures the degree of anisotropy in the diffusion ellipsoid but does not use information from the eigenvectors. Consequently, the FA-based ICA does not take into account the orientation of the diffusion tensors, which is a crucial piece of information in DTI data. Furthermore, the FA-based ICA method requires multi-subject FA values and is not applicable to single subject DTI analysis.

To help address these challenges, we propose a new Distributional Independent Component Analysis (DICA) method to provide a unified framework for extracting source signals from diverse types of data from different imaging modalities including but not restricted to fMRI and DTI. We note that the aforementioned limitations of the classical ICA mainly relate to the fact that the classical ICA aims to decompose the observed data. Hence the ICA model and associated estimation methods need to be tailored to the specific imaging modality. The proposed DICA represents a fundamentally different approach of source separation. Instead of separating the observed data, DICA aims to separate the characterizing parameters in the distribution function of the observed data as a mixture of source signals. Specifically, in the DICA framework, we first model the observed data with a mixture distribution model where the component distributions are chosen appropriately based on the data representations from the specific imaging modality. The mixture distribution model

is characterized by the component distributions and a set of weight parameters that is a probability vector representing the loading on each component distribution. At the second stage of DICA, we perform the ICA decomposition on the posterior weights by separating them as a linear mixture of independent latent source signals. In this new framework, the component distributions of the mixture distribution model at the first stage of DICA can be viewed as a set of bases in the distribution space of the imaging data. The posterior weights of the mixture distribution can be viewed as the set of coordinates of the observed data on the bases of component distributions. For diverse types of data from different modalities, the basis component distributions are specific to each imaging modality while the weights are comparable across modalities. For example, in DTI analysis, the first stage of DICA models the whole tensor using the mixture of Wishart distributions, allowing us to take into account information from both eigenvalues and eigenvectors of the tensor matrices. The voxel-specific posterior weights derived from the first-stage DICA contain both the shape and direction information of a representative diffusion tensor. DICA is also able to decompose single-subject DTI data. Therefore, the proposed DICA addresses the major limitations in existing ICA analysis of DTI data. Our goal of the proposed DICA is to provide a unified platform to decompose diverse types of data derived from different imaging modalities. Building on this unified platform, future research can be conducted to extend the DICA method to performing joint decomposition across multiple imaging modalities.

We applied DICA to decompose fMRI data and DTI data from the Philadelphia Neurodevelopmental Cohort (PNC) study (Satterthwaite *et al.*, 2014). For fMRI data, DICA successfully recovered well-known resting state brain function networks that have been consistently identified in fMRI studies. This demonstrates the validity of the proposed DICA for fMRI data. For DTI data, the DICA extracted independent components that are composed of white matter regions that correspond to major fiber bundles in the brain. The results demonstrate the applicability of DICA for revealing structural networks related to white matter pathways. To the best of our knowledge, the proposed DICA is the first ICA framework that can achieve this goal for single subject DTI data.

The rest of this paper is organized as follows. In Section 2, we introduce the DICA framework and its estimation procedure, and discuss the connection and distinction between DICA and the classical ICA model. In Section 3, we include the results obtained by applying DICA to the fMRI and DTI imaging data from the PNC study. Section 4 presents simulation studies to evaluate the performance of DICA for separating data generated from different underlying mechanisms, compared with

two popular classical ICA algorithms. We conclude with a discussion in Section 5.

2 | THE DISTRIBUTIONAL ICA

The DICA method is a general decomposition method that can be applied to data collected from diverse imaging modalities such as fMRI and DTI. These modalities measure different aspects of brain function and structure. For instance, fMRI measures the blood-oxygenation-level-dependent (BOLD) signal as a correlate of neural activity. Series of BOLD signals are captured across time to investigate the temporal dynamics of the neural processing in response to experimental stimuli or during resting state. From fMRI data, we can infer brain functional networks consisting of brain regions demonstrating coherent BOLD temporal dynamics. DTI is another MRI modality that maps white matter fiber tracts in the brain by measuring the diffusion of water molecules within brain tissues. Specifically, DTI models local water diffusion using a zero-mean 3D Gaussian distribution. The 3×3 covariance matrix of the Gaussian distribution, known as the diffusion tensor matrix, characterizes the water diffusion pattern in 3D space. From diffusion tensors, we can infer local white fiber orientation and subsequently construct brain structural networks consisting of major white matter fiber bundles.

2.1 | A two-stage method

Let \mathbb{I} represent the imaging data space where we collect data from. Denote by $\mathbf{Y}_j \in \mathbb{I}$ the imaging measurement obtained at voxel j ($j = 1, \dots, J$). The imaging data space \mathbb{I} is specific to an imaging modality. For fMRI, \mathbf{Y}_j represents the BOLD signal series measured over T time points at voxel j . Hence, $\mathbb{I} = \mathbb{R}^T$ is a T -dimensional Euclidean space. For DTI, \mathbf{Y}_j represents the diffusion tensor matrix at voxel j that is visualized as an ellipsoid characterizing the diffusion pattern of water molecules in local brain tissue. Hence, $\mathbb{I} = PG_3$ represents the space of 3×3 symmetric positive definite matrices.

The proposed DICA consists of two stages. At stage one, DICA models the observed data using a mixture distribution model and projects the data onto a space of probability vectors containing posterior weights of the mixture distribution. At stage two, DICA decomposes the posterior probability weights to extract latent source signals. Please see Figure 1 for a schematic illustration of the DICA. We present some detailed introduction in the following.

At stage one, we model the probability distribution of imaging measurements using a mixture distribution of K

components:

$$\mathbf{Y}_j \sim \sum_{k=1}^K \pi_k \mathcal{D}(\Theta_k), \quad \text{for } j = 1, \dots, J,$$

where π_k is the weight of the component k in the mixture distribution and $\mathcal{D}(\Theta_k)$ specifies the distribution of component k , which is chosen according to the specific imaging modality. For fMRI data, $\mathcal{D}(\Theta_k)$ can be specified as a multivariate Gaussian distribution with $\Theta_k = (\theta_k^{fMRI}, \tau_k^{fMRI})$ where θ_k^{fMRI} and τ_k^{fMRI} are the mean and covariance of the multivariate Gaussian. For DTI data, $\mathcal{D}(\Theta_k)$ can be specified as a Wishart distribution with $\Theta_k = (\theta_k^{DTI}, \tau_k^{DTI})$, where θ_k^{DTI} represents the expected tensor shape and τ_k^{DTI} is the degree of freedom. From the definition of Wishart distribution on the space PG_3 , we require that $\tau_k^{DTI} > 2$. To simplify presentation, for the rest of the paper, we use the generic notation $\Theta_k = (\theta_k, \tau_k)$ when related computations are generally applicable to both modalities. When we discuss scenarios that are specific to a modality, we use the notation $\Theta_k^{fMRI} = (\theta_k^{fMRI}, \tau_k^{fMRI})$ for fMRI data and $\Theta_k^{DTI} = (\theta_k^{DTI}, \tau_k^{DTI})$ for DTI data.

For each voxel, we obtain the posterior probability $\mathbf{w}_j = (w_{1j}, \dots, w_{Kj})^T$ as follows:

$$w_{kj} = \frac{\pi_k \mathcal{D}(\mathbf{Y}_j; \Theta_k)}{\sum_{k'=1}^K \pi_{k'} \mathcal{D}(\mathbf{Y}_j; \Theta_{k'})}, \quad k = 1, \dots, K, \quad (2)$$

where $\mathcal{D}(\cdot; \Theta_k)$ is the probability density function for component k . Given the component distributions, the probability vector \mathbf{w}_j provides a re-representation of the measurements \mathbf{Y}_j . The mixture model has been widely used for T1 brain image segmentation (Greenspan *et al.*, 2006) where the component distributions model voxels from various tissue types and the posterior probability characterizes which tissue type a voxel is associated with. Similarly, at stage one of DICA, the component distributions model brain voxels with various types of characteristics within the specific imaging modality (e.g., fMRI BOLD series patterns or DTI diffusion ellipsoid patterns), and the posterior probability characterizes a voxel's association with the different types of characteristics (Figure 1).

At stage two, we decompose \mathbf{w}_j into L independent components as follows:

$$g(\mathbf{w}_j) = \sum_{l=1}^L \mathbf{a}_l S_{lj}.$$

Here, $g(\cdot)$ is a link function for the probability vector that provides a mapping from \mathbb{P}^K to \mathbb{R}^{K-1} ,

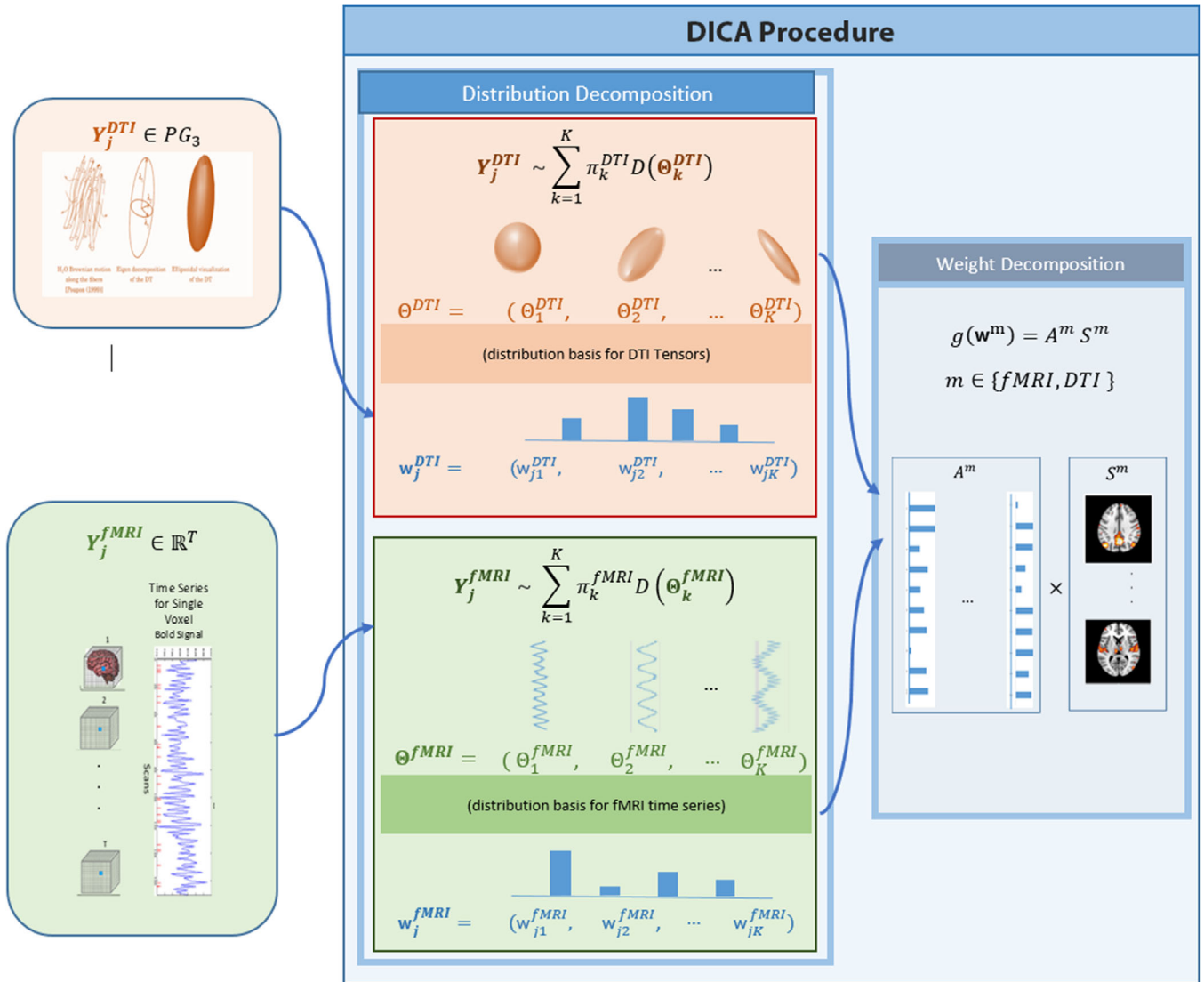


FIGURE 1 The schematic representation of the DICA method

where $\mathbb{P}^K = \{\mathbf{x} \in \mathbb{R}^K : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} > 0\}$. For example, we can specify the *mlogit* link function with $g(\mathbf{x}) = (\log\{x_1/x_K\}, \dots, \log\{x_{K-1}/x_K\})$. The term $S_{lj} \in \mathbb{R}$ represents the latent source signal of the l th independent component at voxel j . The parameters $\mathbf{a}_l = (a_{l1}, \dots, a_{lK})$ are the mixing coefficients that mix the independent source signals to generate the mixture weights \mathbf{w}_j that are related to the observed images. At stage two, we have the same assumptions as the classical ICA. That is, the source signals $S_{lj} (l = 1, \dots, L)$ are statistically independent; the number of sources L is less than the number of mixtures K ; and the source signals S_{lj} follow a non-Gaussian distribution. Our ICA model is spatial ICA with statistical independence assumed in the spatial domain. Previous work has shown that the spatial independence assumption is well suited to the sparse distributed nature of the

spatial pattern for brain functional networks (McKeown and Sejnowski, 1998). Similar observations have also been made for structural networks derived from DTI data. Therefore, ICA applications with fMRI and DTI data are predominantly performed as spatial ICA (Calhoun *et al.*, 2001; Beckmann and Smith, 2005; Guo and Pagnoni, 2008; Li *et al.*, 2012; Ouyang *et al.*, 2015; Shi and Guo, 2016).

Here we provide an intuitive explanation of the rationale for the DICA. In the distribution model, the collection of the component distributions $\{D(\Theta_k)\}_{k=1}^K$ may be viewed as a set of bases characterizing the distributional space of the mixture model for \mathbf{Y}_j ; and the posterior weights \mathbf{w}_j represent the coordinates of the observed data \mathbf{Y}_j on this set of distributional bases. It is worth noting that only the distributional bases, that is, $\{D(\Theta_k)\}_{k=1}^K$, depend on the data characteristics of specific imaging modalities while

the posterior weights \mathbf{w}_j have common representations across different modalities. DICA uses the mixture distribution model to capture modality-specific data characteristics, extracts posterior weights \mathbf{w}_j as re-representations of the imaging measurements and conducts ICA source separation on \mathbf{w}_j that are free of specific modality characteristics. The ICs extracted from the posterior weights are assumed to be spatially independent with each other based on the fact that different brain networks usually have limited spatial overlaps. Under such an assumption, each IC potentially corresponds to a brain functional or structural network.

In neuroimaging literature, the number of ICs L can be chosen either using quantitative methods such as Laplace approximation (Minka, 2000) or based on the biological interpretations of brain networks. Previous work (Smith *et al.*, 2013) has shown that the extracted brain networks are largely robust to the selection of the number of ICs within a range. For L corresponding to qualitatively different model orders, results show that the low model orders lead to large ICs representing networks responsible for very broad sets of similar functions, while high model orders give rise to small ICs representing sub-networks with more specific functions under the same umbrella. Our experiments show that the choice of L has similar effects on DICA. In practice, the choice of L depends on the established knowledge in neuroimaging literature and also the scale of brain networks investigators are interested in identifying.

2.2 | Computation

In this section, we present computation details of the DICA method. At stage one, we fit the mixture distribution model using the Expectation-Maximization (EM) algorithm. Specifically, for fMRI data, we fit the mixture of Gaussian distributions using the standard EM algorithm. For DTI data, we fit the mixture of Wishart distributions using k-MLE proposed by Nielsen (2012), which can be regarded as a hard membership clustering version of the EM algorithm with a higher efficiency. We obtain the MLE of each Wishart component using the method proposed by Saint-Jean and Nielsen (2014). The number of components in the mixture model is chosen according to the Bayesian information criteria (BIC). At stage two, we transform the estimated posterior mixture weights $\hat{\mathbf{w}}_j$ using the *mlogit* function, that is, $mlogit(\hat{\mathbf{w}}_j)$. A standard ICA algorithm such as Infomax (Bell and Sejnowski, 1995) is then applied to decompose $mlogit(\hat{\mathbf{w}}_j)$ to estimate the mixing matrix $\mathbf{A} = \{a_{ki}\}$ and the source signal $\mathbf{S} = \{S_{lj}\}$.

2.3 | Connection between the DICA and the classical ICA

The proposed DICA method is closely connected to the classical ICA model. The posterior probability of mixture component \mathbf{w}_j is a re-representation of the imaging measurements \mathbf{Y}_j , that is, a mapping $\mathbf{p}(\mathbf{y}) = \{p_1(\mathbf{y}), \dots, p_{K-1}(\mathbf{y})\}$ defined by Equation (2) with $\mathbf{p}(\mathbf{y}) : \mathbb{I} \rightarrow \mathbb{P}^K$. Here, the $\mathbf{p}(\mathbf{y})$ maps the imaging measurements \mathbf{Y}_j from their modality-specific space \mathbb{I} to a probability vector space \mathbb{P}^K that is free of modality characteristics. Then a link function $g(\cdot)$ is applied to $\mathbf{w}_j = \mathbf{p}(\mathbf{Y}_j)$ for the ICA decomposition, that is, $g(\mathbf{w}_j) = \sum_{l=1}^L \mathbf{a}_l S_{lj}$. The classical ICA model can be represented using a similar procedure with a different mapping and link function. Suppose \mathbf{Y}_j is an R dimensional vector. The classical ICA can be viewed as a two-stage procedure that applies the identity mapping $\mathbf{p}^*(\mathbf{y}) = \mathbf{y}$ at stage one and the identity link $g^*(\mathbf{y}) = \mathbf{y}$ at stage two followed by ICA decomposition, that is, $g^*(\mathbf{p}^*(\mathbf{Y}_j)) = \sum_{l=1}^L \mathbf{a}_l S_{lj}$. Some dimensional reduction steps prior to ICA such as PCA also can be considered as special mappings at stage one of classical ICA. The main distinction between DICA and classical ICA is that at stage one, DICA performs a mapping $\mathbf{p}(\mathbf{y}) : \mathbb{I} \rightarrow \mathbb{P}^K$ that transforms imaging measurements with various types of representations into a probability vector space that is free of modality characteristics. This makes DICA a unified source separation framework that can be applied to decompose diverse imaging modalities. Furthermore, unlike classical ICA that mainly separates linear mixtures, DICA is potentially capable of separating nonlinear mixtures on the original scale of the data because of the nonlinearity of the mapping, which we demonstrate in Section 4.

3 | APPLICATION TO REAL IMAGING DATA

3.1 | Data description

We applied the DICA to the rs-fMRI and DTI data obtained from a subject from the Philadelphia Neurodevelopmental Cohort (PNC) study. All images from the PNC study were acquired on a Siemens Tim Trio 3 Tesla scanner. The rs-fMRI scans were acquired with 124 volumes each containing a 3D brain image of $64 \times 64 \times 46$ voxels with the resolution of $3.0 \times 3.0 \times 3.0$ mm. DTI data were derived based on diffusion-weighted imaging (DWI) scans using a twice-refocused spin echo (TRSE) single-shot EPI sequence. The sequence consisted of 64 diffusion-weighted directions with $b = 1000s/mm^2$, and 7 scans with $b = 0s/mm^2$. More

details about image acquisition can be found in Satterthwaite *et al.* (2014).

3.2 | Analysis of rs-fMRI data

To preprocess the rs-fMRI data, skull stripping was performed on the T1 images to remove extra-cranial material, then the first four volumes of the functional time series were removed to stabilize the signal, leaving 120 volumes for subsequent preprocessing. The anatomical image was registered to the eight volume of the functional image and subsequently spatially normalized to the MNI standard brain space. These normalization parameters from MNI space were used for the functional images, which were smoothed with a 6 mm FWHM Gaussian kernel. Motion corrections were applied on the functional images. A validated confound regression procedure was performed on each subject's time series data to remove confounding factors including motions, global effects, white matter (WM) and cerebrospinal fluid (CSF) nuisance signals. The confound regression contained nine standard confounding signals (six motion parameters plus global / WM / CSF) as well as the temporal derivative, quadratic term, and temporal derivative of the quadratic of each. Furthermore, motion-related spike regressors were included to bound the observed displacement. Finally, the functional time series data were band-pass filtered to retain frequencies between 0.01 and 0.1 Hz that is the relevant frequency range for rs-fMRI. Prior to ICA analysis, we performed additional preprocessing steps including centering and dimension reduction. Specifically, we performed a PCA and reduced the dimension of the fMRI BOLD response at each voxel by projecting the fMRI time series onto the first 30 principal component (PC) directions, where the number of PCs was chosen based on the scree plot.

We then applied DICA to the preprocessed fMRI data. We considered a mixture of Gaussian (MoG) distribution with 20 Gaussian components for the first stage of DICA. In our experiments, we also considered MoG distribution models with a different number of Gaussian components ranging from 18 to 30. We found that the results from the second stage of DICA are fairly robust to the selection of the number of mixtures at stage one. At stage two of DICA, we extracted $L = 14$ ICs from the rs-fMRI data. The choice of L was motivated by the selection of the number of ICs in the existing neuroimaging ICA (Shi and Guo, 2016; Guo and Tang, 2013) that leads to neurobiologically interpretable ICs. Among the extracted components, we identified several well-known resting state functional networks such as the occipital pole visual network, lateral visual, default mode network, sensorimotor network,

auditory network, and executive control network. Figure 2 presents the DICA estimates of these networks. The maps were thresholded showing activated voxels whose z-scores based on the estimated source signal exceeding 95th percentile. These DICA-based rs-fMRI networks have been consistently identified in large sample rs-fMRI studies in the literature using the standard ICA methods (Smith *et al.*, 2009). The fact that DICA successfully recovered these networks using a single subject fMRI data demonstrates the applicability of DICA for extracting brain functional networks from fMRI. To evaluate the reliability of the results, we performed the DICA analysis for additional subjects from the PNC study. Web Appendix A in the Supporting Information presents results for the additional subjects. We obtained largely consistent brain functional networks across subjects, which are similar as the results reported in Figures 2.

3.3 | Analysis of DTI data

The DWI scans were preprocessed using the DWI pipeline in FSL. The procedure includes brain extraction to remove nonbrain regions, phase reversal distortion correction, and aligning diffusion-weighted images to the average nondiffusion weighted image by a rigid body affine transformation to remove motion artifact and also correction of Eddy current distortions. We then estimated the directional diffusion at each voxel based on a diffusion tensor model implemented via the Diffusion Toolbox (FDT) in FSL.

At stage one of DICA, we considered a mixture of Wishart distributions with 20 component distributions. We also considered distributions with a different number of components and the findings of the second stage DICA remained similar. At the second stage of DICA, we extracted 14 ICs. Among the extracted components, we identified several ICs that are found to correspond to major white matter fiber pathways via fiber tracking. These ICs were also identified when we extracted different number of ICs. Figure 3 presents the DICA estimates of the spatial source signals of the extracted ICs. The maps were thresholded showing significant voxels whose z-scores based on the estimated source signal posterior distribution exceeding 95th percentile. We also required a minimum activation cluster size of at least of 3000 voxels to improve the reliability of the voxels selected into the thresholded map. To demonstrate the white fiber pathways that each IC is associated with, we applied the DTI tractography to reconstruct the fiber tracts passing through the spatial regions of each IC (Figure 3). In addition, we illustrate the main DTI diffusion ellipsoid associated with each IC in Figure 3. The first IC corresponds to the corpus callosum and contains a major fiber bundle that connects left and

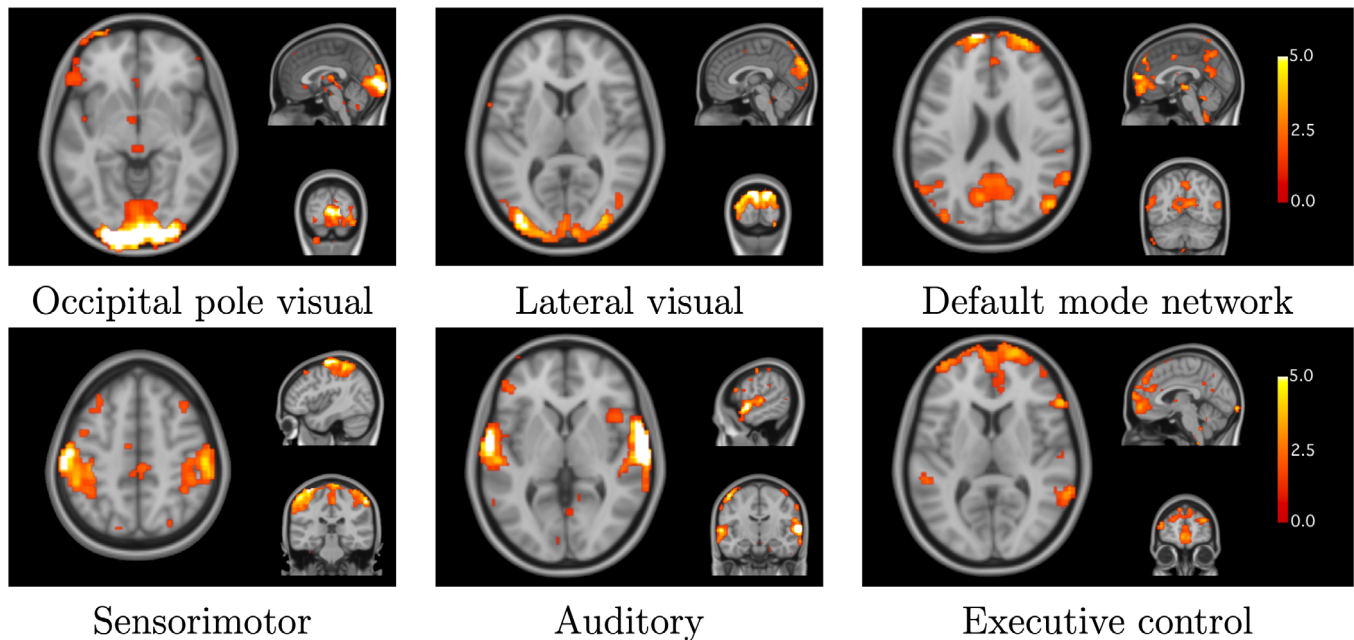


FIGURE 2 Brain functional networks estimated from a subject's resting state fMRI data from the PNC study using DICA. The IC maps are thresholded showing significant voxels with z-scores derived from the estimated source signal posterior distribution exceeding 95th percentile. The DICA successfully recovered well-known brain functional networks that have been consistently identified in fMRI studies

right cerebral hemispheres (Catani and Thiebaut de Schotten, 2008). We can see the main diffusion ellipsoid associated with this IC demonstrates diffusion between left and right in the brain that is consistent with the major fiber tract direction of the corpus callosum. The second IC is mostly located symmetrically in both hemispheres of the brain and mainly contains projection fiber tracts passing through the internal capsule and corona radiata that contains ascending fibers from the thalamus to the cerebral cortex and descending fibers from the frontoparietal cortex to subcortical nuclei (Catani and Thiebaut de Schotten, 2008). The main diffusion ellipsoid associated with this IC demonstrates inferior–superior diffusion pattern that is consistent with the direction of the projection tracts. The third IC represents cingulum, a medial associative fiber bundle that runs an antero-posterior course within the cingulate gyrus around the corpus callosum. Correspondingly, the main diffusion ellipsoid of this IC demonstrates antero-posterior diffusion pattern. The fourth IC contains corticoponto-cerebellar tracts that create communications between the cerebellum and the contralateral cerebral hemisphere (Catani and Thiebaut de Schotten, 2008). The diffusion ellipsoid of this IC demonstrates inferior–superior diffusion pattern that corresponds to the projection tracts that pass through this IC. Similar as in the analysis of fMRI data, we also present results for additional subjects in the Web Appendix A in the Supporting Information. We obtained largely consistent brain struc-

tural networks across subjects, which are similar as the results reported in Figures 3.

4 | SIMULATION STUDIES

We conducted several simulation studies to evaluate the performance of the proposed DICA method. We generated data from three distinct underlying models including a model that favors DICA where observed data are linear mixtures of source signals on the distribution level, the classical linear ICA model, and a nonlinear ICA model where observed data are nonlinear mixtures of latent source signals. Among the three settings, the first case is based on the DICA model while the latter two represent models that deviate from the DICA. In particular, the classical linear ICA is the underlying model of infomax and FastICA. Following previous work (Beckmann and Smith, 2005; Guo and Pagnoni, 2008), we evaluated the performance of each ICA method based on the correlation between the true and estimated source signals. Also, we evaluated the AUC of the estimated source signals for Simulations II and III where there were categorical definition of activated/nonactivated voxel in generating the sources. Since ICA recovery is invariant to permutations, each estimated IC was matched with the original source with which it had the highest spatial correlation (Beckmann and Smith, 2005). In comparison to DICA, we considered

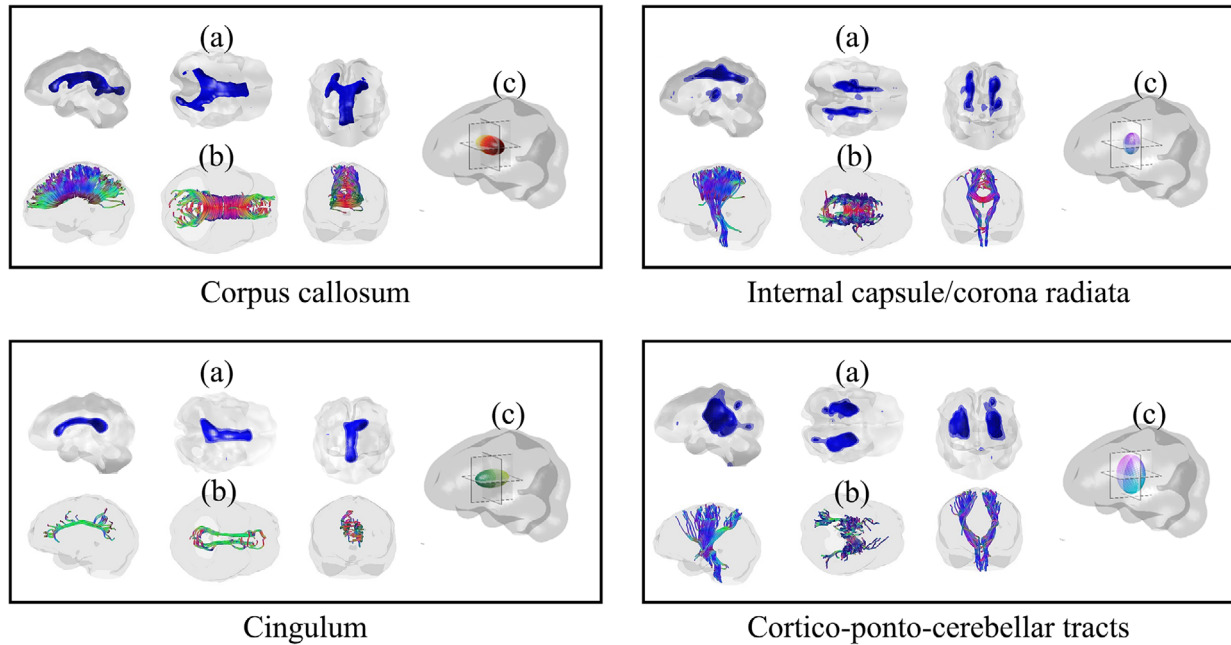


FIGURE 3 Brain structural networks estimated from a subject's DTI data from the PNC study using DICA. DICA discovered components corresponding to major white fiber pathways in the brain. (a) The estimated IC spatial maps, which are thresholded based on the z-scores derived from the estimated source signal posterior distribution, (b) tractography-reconstructed fiber tracts passing through the spatial regions of each IC, and (c) the main DTI diffusion ellipsoid associated with each IC

two commonly applied ICA algorithms: Infomax (Bell and Sejnowski, 1995) and FastICA (Hyvarinen, 1999). In addition, we considered two advanced and more recent ICA developments: SparseICA (Boukouvalas *et al.*, 2018) and nonlinear ICA (Almeida, 2003). The SparseICA was implemented with the SparseICA-EBM algorithm (Boukouvalas *et al.*, 2018). The nonlinear ICA was implemented with a MATLAB toolbox "MISEP" for Simulation III.

4.1 | Simulation study I: Data from a model that aligns with DICA

In the first simulation study, we considered data derived from a model that aligns with the DICA procedure. Specifically, we generated \mathbf{Y}_j as 10×1 vectors for $j = 1, \dots, J$, $J = 6400$:

$$\mathbf{Y}_j \sim \sum_{k=1}^K w_{kj}^0 \mathcal{N}(\theta_k, \tau_k), j = 1, \dots, J. \quad (3)$$

Here, data were generated from a mixture of multivariate Gaussian distribution with the dimension of $d = 10$ and the number of mixture $K = 6$. The mean and covariance for each component were specified as $\theta_k = (k - (K + 1)/2)\mathbf{1}_d$ and $\tau = 0.1\mathbf{I}_{d \times d}$. The weights w_{kj}^0 were the posterior weights derived from a mixture of multivariate

Gaussian with the same mean and covariance as (3) and prior weights $w_{kj} = 1/6, k = 1, \dots, K, j = 1, \dots, J$. The true source signals S_{lj} and the mixing matrix $a_{kl}, l = 1, \dots, L$ with $L = 3$ were obtained by decomposing w_{kj}^0 as at stage two of DICA.

We applied the proposed DICA, FastICA, Infomax, and SparseICA to decompose the simulated data \mathbf{Y}_j . Table 1 provides the means (standard deviations) of the correlations between the true and estimated source signals across 100 simulation runs based on each of the three ICA methods. The results show the proposed DICA demonstrated a much better accuracy in recovering the true source signals as compared with the other methods when the data are generated in this process that align with the DICA model. FastICA, Infomax and SparseICA only captured one of the true signals while DICA successfully recovered all of them. The significant difference in the performance between the methods demonstrates that when data are mixed on the distribution level, the standard ICA that decomposes the observed data may yield poor results.

4.2 | Simulation study II: Data from the classical linear ICA

In the second simulation study, we generated data from the classical linear ICA model where the observed data

TABLE 1 (A) Mean(standard deviation) of the correlations between the true and estimated source signals based on different ICA methods with 100 simulation runs in the three simulation studies

Simulation I: Distributional ICA					
	DICA	FastICA	Infomax	SparseICA	
Source 1	0.993(0.000)	0.800(0.001)	0.800(0.001)	0.793(0.034)	
Source 2	0.991(0.000)	0.019(0.013)	0.018(0.012)	0.043(0.050)	
Source 3	0.948(0.001)	0.021(0.014)	0.021(0.012)	0.035(0.008)	
Simulation II: Linear ICA					
	DICA	FastICA	Infomax	SparseICA	
Source 1	0.817(0.085)	0.859(0.025)	0.859(0.025)	0.862(0.025)	
Source 2	0.593(0.107)	0.683(0.050)	0.683(0.050)	0.689(0.040)	
Source 2	0.654(0.109)	0.736(0.040)	0.736(0.040)	0.738(0.046)	
Simulation III: Nonlinear ICA					
	DICA	FastICA	Infomax	SparseICA	nonlinear ICA
Source 1	0.979(0.008)	0.869(0.010)	0.869(0.009)	0.807(0.043)	0.377(0.349)
Source 2	0.963(0.014)	0.523(0.011)	0.523(0.011)	0.529(0.034)	0.164(0.219)

TABLE 1 (B) Mean(standard deviation) of the AUCs of the estimated source signals based on different ICA methods with 100 simulation runs in Simulations II and III

Simulation II: Linear ICA					
	DICA	FastICA	Infomax	SparseICA	
Source 1	0.898(0.039)	0.912(0.012)	0.912(0.012)	0.917(0.013)	
Source 2	0.862(0.047)	0.893(0.019)	0.893(0.019)	0.900(0.016)	
Source 3	0.880(0.049)	0.907(0.016)	0.907(0.016)	0.912(0.016)	
Simulation III: Nonlinear ICA					
	DICA	FastICA	Infomax	SparseICA	non-linear ICA
Source 1	1.000(0.000)	1.000(0.000)	1.000(0.000)	0.978(0.014)	0.718(0.228)
Source 2	1.000(0.000)	0.853(0.008)	0.853(0.008)	0.862(0.024)	0.585(0.173)

are a linear mixture of the source signals via the mixing coefficients plus a noise term, that is, $\mathbf{Y}_j = \sum_{l=1}^L \mathbf{a}_l S_{lj} + \epsilon_j$. Specifically, we generated image data from three ICs. For each IC, we generated a 2D spatial map with the dimension of 100×100 and the activated signals in the ICs are displayed in Figure 4. This spatial maps were generated using MATLAB toolbox simTB (Erhardt *et al.*, 2012), which simulates spatial maps that mimic brain networks. The mixing coefficient vector \mathbf{a}_l was a time series of the length $T = 40$ based on i.i.d standard Gaussian random variables. We considered a Gaussian noise term added to the linear mixtures with zero mean and standard deviation of 0.8. We tried multiple noise levels with different standard deviations and obtained similar conclusions.

We applied the DICA, Infomax, FastICA, and SparseICA to decompose the data. We specified $K = 10$ and $L = 6$ for the DICA. Other specifications of K and L were also investigated and the results were found to be similar. We evaluated the methods with the AUCs of the estimated source signals as well as the correlations between the true and

the estimated signals. As shown in Table 1, the Infomax, FastICA, and SparseICA performed better than DICA in both AUCs and correlations for this data. This is expected given that the data were generated from the classical linear ICA that is the underlying model of the three methods. The accuracy of DICA was acceptable as compared with the three linear ICA methods. Results from Figure 4 show that DICA successfully recovered the true underlying source signals. The fact that DICA demonstrated reasonable accuracy indicates that the DICA can be an effective decomposition method for data that conform to the classical ICA model assumption.

4.3 | Simulation study III: Data from nonlinear mixtures

In the third simulation study, we compared the performance of the methods for recovering sources when the observed data are nonlinear mixtures of source signals,

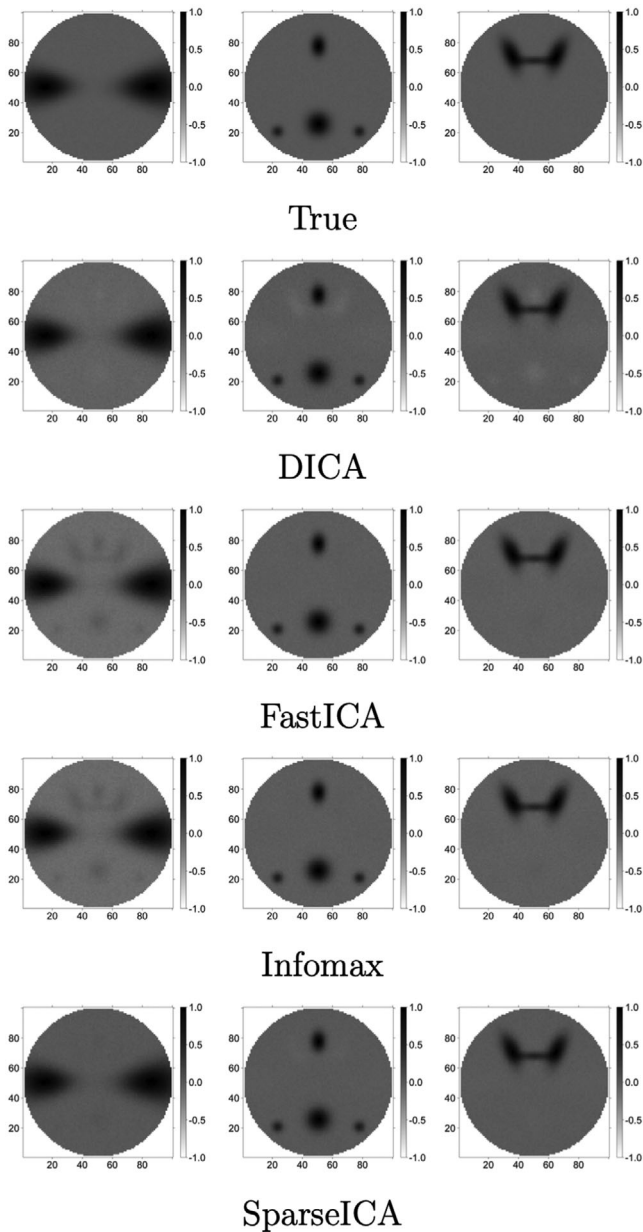


FIGURE 4 True source signals and the estimated ICs based on different ICA methods in simulation study II for data simulated from the classical ICA model (averaged over 100 runs)

which represents a scenario that deviates from both the DICA and classical linear ICA models. We considered two source signals each with a 50×50 spatial map. The true maps are presented in Figure 5(B), which were generated from the following functional forms added Gaussian noise with the standard deviation of 0.1 to each pixel, $s_1(x_1, x_2) = 0.95 I[(x_1 - 0.3)^2 + (x_2 - 0.3)^2 \leq 0.3^2]$ and $s_2(x_1, x_2) = 0.95 I[|x_1 - 0.7| + |x_2 - 0.7| \leq 0.3]$, where $(x_1, x_2) \in [0, 1]^2$ represents the location. We have considered different noise levels, which led to consistent conclusions. The observed data \mathbf{Y} were then generated by performing a nonlinear mixing of the source signals

via a nonlinear mixing function, that is, $\mathbf{Y} = f(\mathbf{s})$ with $f(\mathbf{s}) = f_1(\mathbf{s}) + f_2(\mathbf{s}) + f_3(\mathbf{s})$ (Figure 5A), where the functions were specified as

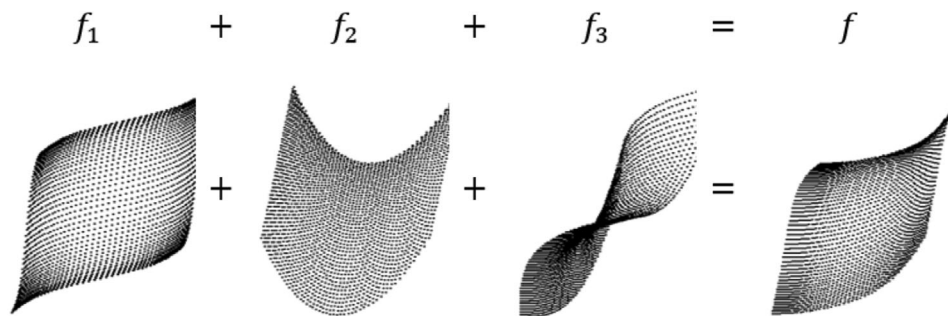
$$f_1(\mathbf{s}) = \begin{bmatrix} \tanh(4s_1 - 2) + \frac{2s_1 + s_2}{2} \\ \tanh(4s_1 - 2) + \frac{2s_1 + s_2}{2} \end{bmatrix},$$

$$f_2(\mathbf{s}) = \begin{bmatrix} \tanh\left(\frac{s_2}{2}\right) + \frac{2s_1 + s_2^2}{2} \\ s_1^3 - s_1 + \tanh(s_2) \end{bmatrix}, f_3(\mathbf{s}) = \begin{bmatrix} s_2^3 + s_1 \\ \tanh(s_2) + s_1^3 \end{bmatrix}.$$

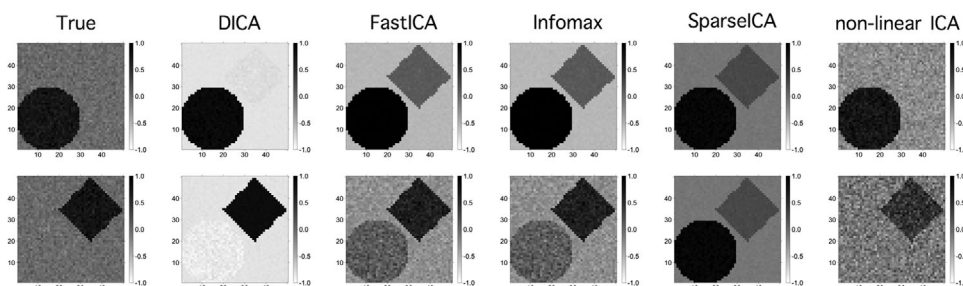
Figure 5(B) presents the true source signals and the reconstructed sources using DICA, FastICA, Infomax, SparseICA, and nonlinear ICA, and the estimation results were evaluated as shown in Table 1. In DICA, we specified $K = 12$ and $L = 4$, where other specifications yielded similar results. We observe that the proposed DICA showed satisfactory performance in recovering the sources from the nonlinear mixtures. In comparison, the three linear ICA methods, that is, FastICA, Infomax, and SparseICA, failed to separate the two underlying source signals. The nonlinear ICA implemented with MISEP did not perform well either in this case and failed to identify the true source patterns in many simulated data sets. Specifically, the nonlinear ICA only successfully identified the true source 1 and 2 with correlation greater than 0.5 in 41 and 15 simulation runs out of 100 runs, respectively. In most of the simulation runs, the nonlinear ICA could only identify one of the two sources. We also compared the nonlinear ICA and DICA in other simulation settings by generating nonlinear mixtures following the example provided by the MISEP toolbox itself. The performance of the proposed DICA is comparable or better than the nonlinear ICA in that scenario as well. This result indicates that DICA can potentially provide a useful decomposition tool for data derived from mixing mechanisms that are more complicated than the classical linear mixture model.

5 | DISCUSSION

In this paper, we propose a distributional ICA procedure to provide a general analysis tool for decomposing diverse types of imaging modalities. Compared with the classical ICA, the proposed DICA represents a new approach that aims to separate the observed data on the distribution level. We develop a two-stage estimation for the DICA procedure. We apply the DICA model to real-world fMRI and DTI data. Our analyses have generated scientifically meaningful and insightful findings. For fMRI data, DICA successfully identified well-known resting-state functional



(A) Nonlinear mixing function $f(s)$ and its three component functions



(B) True and estimated sources

FIGURE 5 Simulation III: (a) the nonlinear mixing function and (b) the true and estimated source signals (averaged over 100 runs) using DICA, FastICA, Infomax, SparseICA, and nonlinear ICA to separate the data from nonlinear mixtures

networks. For DTI data, DICA discovered components that correspond to major white fiber pathways in the brain. To the best of our knowledge, the proposed DICA is the first ICA method that can perform source separations for diffusion tensors based on single subject DWI scans and our findings are not obtainable using the existing ICA methods. The extensive simulation studies conducted in the paper further demonstrate satisfactory performance of the DICA for recovering source signals from data that were generated from different underlying mixture models.

In recent neuroimaging studies, investigators often need to deal with multimodal neuroimaging where measurements have different forms and dimensions. Methods have been proposed to jointly model features extracted from multimodality data. For example, some ICA methodologies (Eichele *et al.*, 2008; Franco *et al.*, 2008) were developed to jointly analyze multi-subject and multi-dimensional features extracted from fMRI, ERP, and genetic data. Methods based on the graph Laplacian have been proposed for joint modeling of multi-subject multimodal network features extracted from fMRI and DTI (Doderio *et al.*, 2014; Abdelnour *et al.*, 2018). Deep learning techniques such as auto-encoder, U-net, and GAN are also widely employed in feature extraction and image synthesis (Wang *et al.*, 2020). Compared to the existing methods, our DICA has a different goal that is to provide a unified ICA method to decompose images obtained using differ-

ent imaging modalities to identify underlying source signals. DICA is applicable to both single subject as well as multi-subject imaging data. The term “unified” is in the sense that DICA can be applied to diverse modalities, such as fMRI or DTI. In future research, this unified decomposition framework could be extended to performing joint decomposition across multiple modalities. Another useful application of the DICA is that it can potentially separate both linear and nonlinear mixtures on the original scale of the data, while the classical ICA mainly separates linear mixtures. DICA can be used as an alternative blind source separation tool that may provide new findings that complement the outputs from the standard ICA algorithms.

For future research, we plan to extend the current DICA model to decompose multi-subject imaging data. For fMRI data, a straightforward extension of DICA is to follow the temporal concatenation ICA (TC-GICA) framework (Calhoun *et al.*, 2001; Guo and Pagnoni, 2008; Guo, 2011) by concatenating the fMRI data on the temporal domain across subjects and then apply DICA. An alternative way for multi-subject extension is to adopt the hierarchical ICA framework (Guo and Tang, 2013; Shi and Guo, 2016) and develop multilevel DICA models. As another future research topic, we plan to investigate on other distribution models such as the matrix Langevin distributions on the Stiefel manifold to model the DTI data and multivariate Laplacian to model fMRI data.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute of Health under Award Number R01MH105561, R01MH118771, R01DA048993, R01GM124061, and R01CA249096. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

DATA AVAILABILITY STATEMENT

The PNC study data are publicly available to download from the database of Genotypes and Phenotypes (dbGaP) via Authorized Access. To request data access, investigators can login the dbGaP controlled-access portal at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v3.p2 and submit a Data Access Request.

ORCID

Jian Kang  <https://orcid.org/0000-0002-5643-2668>

Ying Guo  <https://orcid.org/0000-0003-3934-3097>

REFERENCES

- Abdelnour, F., Dayan, M., Devinsky, O., Thesen, T. & Raj, A. (2018) Functional brain connectivity is predictable from anatomic network's Laplacian eigen-structure. *NeuroImage*, 172, 728–739.
- Adali, T., Anderson, M. & Fu, G.-S. (2014) Diversity in independent component and vector analyses: identifiability, algorithms, and applications in medical imaging. *IEEE Signal Processing Magazine*, 31, 18–33.
- Almeida, L.B. (2003) Misp-linear and nonlinear ICA based on mutual information. *Journal of Machine Learning Research*, 4, 1297–1318.
- Bartlett, M.S., Movellan, J.R. & Sejnowski, T.J. (2002) Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13, 1450–1464.
- Basser, P.J., Mattiello, J. & LeBihan, D. (1994) Mr diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66, 259–267.
- Beckmann, C.F., DeLuca, M., Devlin, J.T. & Smith, S.M. (2005) Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360, 1001–1013.
- Beckmann, C.F. & Smith, S.M. (2005) Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage*, 25, 294–311.
- Bell, A.J. & Sejnowski, T.J. (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129–1159.
- Boukouvalas, Z., Levin-Schwartz, Y., Calhoun, V.D. & Adali, T. (2018) Sparsity and independence: balancing two objectives in optimization for source separation with application to fMRI analysis. *Journal of the Franklin Institute*, 355, 1873–1887.
- Calhoun, V.D., Adali, T., Pearlson, G.D. & Pekar, J. (2001) A method for making group inferences from fMRI data using independent component analysis. *Human Brain Mapping*, 14, 140–151.
- Calhoun, V.D., Liu, J. & Adali, T. (2009) A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, 45, S163–S172.
- Calhoun, V.D. & Sui, J. (2016) Multimodal fusion of brain imaging data: a key to finding the missing link(s) in complex mental illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1, 230–244.
- Catani, M. & Thiebaut de Schotten, M. (2008) A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 44, 1105–1132.
- Dodero, L., Murino, V. & Sona, D. (2014) Joint laplacian diagonalization for multi-modal brain community detection. *2014 International Workshop on Pattern Recognition in Neuroimaging*. Piscataway, NJ: IEEE. pp. 1–4.
- Eichele, T., Calhoun, V.D., Moosmann, M., Specht, K., Jongsma, M.L., Quiroga, R.Q. et al. (2008) Unmixing concurrent eeg-fMRI with parallel independent component analysis. *International Journal of Psychophysiology*, 67, 222–234.
- Erhardt, E.B., Allen, E.A., Wei, Y., Eichele, T. & Calhoun, V.D. (2012) SIMTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. *NeuroImage*, 59, 4160–4167.
- Franco, A.R., Ling, J., Caprihan, A., Calhoun, V.D., Jung, R.E., Heileman, G.L. et al. (2008) Multimodal and multi-tissue measures of connectivity revealed by joint independent component analysis. *IEEE Journal of Selected Topics in Signal Processing*, 2, 986–997.
- Greenspan, H., Ruf, A. & Goldberger, J. (2006) Constrained gaussian mixture model framework for automatic segmentation of mr brain images. *IEEE Transactions on Medical Imaging*, 25, 1233–1245.
- Guo, Y. (2011) A general probabilistic model for group independent component analysis and its estimation methods. *Biometrics*, 67, 1532–1542.
- Guo, Y. & Pagnoni, G. (2008) A unified framework for group independent component analysis for multi-subject fMRI data. *NeuroImage*, 42, 1078–1093.
- Guo, Y. & Tang, L. (2013) A hierarchical model for probabilistic independent component analysis of multi-subject fMRI studies. *Biometrics*, 69, 970–981.
- Guye, M., Bartolomei, F. & Ranjeva, J.P. (2008) Imaging structural and functional connectivity: towards a unified definition of brain organization? *Current Opinion in Neurology*, 21, 393–403.
- Hallin, M. & Mehta, C. (2015) R-estimation for asymmetric independent component analysis. *Journal of the American Statistical Association*, 110, 218–232.
- Hastie, T. & Tibshirani, R. (2003) Independent components analysis through product density estimation. *Advances in Neural Information Processing Systems*, 15, 665–672.
- Hyvarinen, A. (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10, 626–634.
- Hyvärinen, A., Karhunen, J. & Oja, E. (2001) *Independent Component Analysis*. Hoboken, NJ: John Wiley & Sons.
- Koay, C.G., Chang, L.-C., Carew, J.D., Pierpaoli, C. & Basser, P.J. (2006) A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *Journal of Magnetic Resonance*, 182, 115–125.
- Le, Q., Karpenko, A., Ngiam, J. & Ng, A. (2011) ICA with reconstruction cost for efficient overcomplete feature learning. *Advances in Neural Information Processing Systems*, 24, 1017–1025.

- Li, Y.O., Yang, F.G., Nguyen, C.T., Cooper, S.R., LaHue, S.C., Venugopal, S. et al. (2012) Independent component analysis of DTI reveals multivariate microstructural correlations of white matter in the human brain. *Human Brain Mapping*, 33, 1431–1451.
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J. et al. (1998) Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6, 160–188.
- McKeown, M.J. & Sejnowski, T.J. (1998) Independent component analysis of fMRI data: examining the assumptions. *Human Brain Mapping*, 6, 368–372.
- Minka, T.P. (2000) Automatic choice of dimensionality for PCA. *NIPS*, 13, 598–604.
- Ngiam, J., Chen, Z., Chia, D., Koh, P., Le, Q. & Ng, A. (2010) Tiled convolutional neural networks. *Advances in Neural Information Processing Systems*, 23, 1279–1287.
- Nielsen, F. (2012) k-MLE: A fast algorithm for learning statistical mixture models. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 869–872.
- Ouyang, X., Chen, K., Yao, L., Wu, X., Zhang, J., Li, K. et al. (2015) Independent component analysis-based identification of covariance patterns of microstructural white matter damage in Alzheimer's disease. *PLoS ONE*, 10, e0119714.
- Ramnani, N., Behrens, T.E., Penny, W. & Matthews, P.M. (2004) New approaches for exploring anatomical and functional connectivity in the human brain. *Biological Psychiatry*, 56, 613–619.
- Saint-Jean, C. & Nielsen, F. (2014) Hartigan's method for k-MLE: Mixture modeling with wishart distributions and its application to motion retrieval. In Nielsen, F. (ed.), *Geometric Theory of Information*. Berlin: Springer. pp. 301–330.
- Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Loughhead, J., Prabhakaran, K., Calkins, M.E. et al. (2014) Neuroimaging of the philadelphia neurodevelopmental cohort. *NeuroImage*, 86, 544–553.
- Shi, R. & Guo, Y. (2016) Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis. *The Annals of Applied Statistics*, 10, 1930–1957.
- Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E. et al. (2009) Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106, 13040–13045.
- Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L. et al. (2013) Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, 17, 666–682.
- Wang, T., Lei, Y., Fu, Y., Wynne, J.F., Curran, W.J., Liu, T. et al. (2020) A review on medical imaging synthesis using deep learning and its clinical applications. *Journal of Applied Clinical Medical Physics*, 22, 11–36.
- Wang, Y. & Guo, Y. (2019) A hierarchical independent component analysis model for longitudinal neuroimaging studies. *NeuroImage*, 189, 380–400.

SUPPORTING INFORMATION

Web Appendices, Tables and Figures referenced in Section 3, example data, and R code implementing the proposed method are available with this paper at the Biometrics web-site on Wiley Online Library. The R package for DICA implementation is available on the Github at <https://github.com/benwu233/DICA>.

How to cite this article: Wu B, Pal S, Kang J, Guo Y. (2022) Distributional independent component analysis for diverse neuroimaging modalities. *Biometrics*, 78, 1092–1105.
<https://doi.org/10.1111/biom.13594>