OXFORD

*Cerebral Cortex*, 2023, **33**, 1550–1565
https://doi.org/10.1093/cercor/bhac155
Advance access publication date 28 April 2022
Original Article

# Towards automated in vivo parcellation of the human cerebral cortex using supervised classification of magnetic resonance fingerprinting residuals

Shahrzad Moinian[1,2], Viktor Vegh[1,2], David Reutens[1,2,*]

[1]Centre for Advanced Imaging, The University of Queensland, Building 57, Research Road, St Lucia, QLD 4072, Australia,
[2]ARC Training Centre for Innovation in Biomedical Imaging Technology, The University of Queensland, Building 57, Research Road, St Lucia, QLD 4072, Australia
*Corresponding author: Centre for Advanced Imaging, The University of Queensland, Building 57, Research Road, St Lucia, QLD 4072, Australia.
Email: d.reutens@uq.edu.au

**Background:** Accurate parcellation of the cerebral cortex in an individual is a guide to its underlying organization. The most promising in vivo quantitative magnetic resonance (MR)-based microstructural cortical mapping methods are yet to achieve a level of parcellation accuracy comparable to quantitative histology.
**Methods:** We scanned 6 participants using a 3D echo-planar imaging MR fingerprinting (EPI-MRF) sequence on a 7T Siemens scanner. After projecting MRF signals to the individual-specific inflated model of the cortical surface, normalized autocorrelations of MRF residuals of vertices of 8 microstructurally distinct areas (BA1, BA2, BA4a, BA6, BA44, BA45, BA17, and BA18) from 3 cortical regions were used as feature vector inputs into linear support vector machine (SVM), radial basis function SVM (RBF-SVM), random forest, and *k*-nearest neighbors supervised classification algorithms. The algorithms' prediction performance was compared using: (i) features from each vertex or (ii) features from neighboring vertices.
**Results:** The neighborhood-based RBF-SVM classifier achieved the highest prediction score of 0.85 for classification of MRF residuals in the central region from a held-out participant.
**Conclusions:** We developed an automated method of cortical parcellation using a combination of MR fingerprinting residual analysis and machine learning classification. Our findings provide the basis for employing unsupervised learning algorithms for whole-cortex structural parcellation in individuals.

*Key words*: cerebral cortex parcellation; machine learning classification; magnetic resonance fingerprinting; residual.

## Introduction

Accurate subdivision of the human cerebral cortex into distinct structural areas in individuals enables the effects of aging and neurological disorders on different cortical regions to be examined (Lebel et al. 2008; Rose et al. 2008). Additionally, it is of value in evaluating the association between tissue microstructure and areal functions (Wahl et al. 2010). Furthermore, cortical parcellation in individuals at the voxel level can assist the delineation of abnormal tissue (Awad et al. 1991; Duffau 2012). This may be clinically important, for example in individuals with intractable epilepsy, in whom the identification of subtle malformations of cortical development associated with the seizure focus may enable curative resective surgery.

Recent advances in high-field magnetic resonance imaging (MRI) allow in vivo imaging at mesoscopic scale (0.1–0.5 mm), with different image contrasts. These images can be used in indirect quantitative methods that derive information about tissue microstructure (Weiskopf et al. 2015a). As different MR modalities are differentially sensitive to microstructural tissue properties, integration of complementary information from different modalities may increase the accuracy with which microarchitectonically distinct cortical areas can be distinguished (Cohen-Adad et al. 2012; Mangeat et al. 2015; Weiskopf et al. 2015b; Marques et al. 2017; Cercignani and Bouyagoub 2018; Edwards et al. 2018). Quantitative MR relaxometry parameters (e.g. $T_1$, $T_2$, $T_2$*) are sensitive to microarchitectonic features such as fiber orientation (Cohen-Adad et al. 2012), myelin density (Geyer et al. 2011; Tardif et al. 2013), and cortical depth (Tardif et al. 2013). However, the relationship between these parameters and microstructural features is nonspecific; for example both higher myelination and greater iron concentration reduce $T_1$ values (Weiskopf et al. 2015b; Cercignani and Bouyagoub 2018 ; Edwards et al. 2018). Hence, complementary information from other MR modalities is required for accurate parcellation of the cerebral cortex (Cercignani and Bouyagoub 2018; Edwards et al. 2018). For example, in areas such as the primary and secondary visual cortices where

**Received:** November 23, 2021. **Revised:** March 29, 2022. **Accepted:** March 30, 2022

iron and myelin are co-localized, opposing effects on susceptibility may be used to complement information from $T_1$ and $T_2$* measurements to allow the contribution of myelin to MR signals to be quantified more accurately (Stuber et al. 2014; Marques et al. 2017).

The use of multi-contrast MRI for in vivo microstructural mapping requires optimal data acquisition and data integration methods to be defined. These should account for potential collinearity between different MR parameters, thereby allowing the extraction of complementary information that can be combined to yield novel information that is greater than the sum of information from individual MR contrasts (Cercignani and Bouyagoub 2018). Currently, there is no consensus on the optimal data acquisition and integration methods (Cercignani and Bouyagoub 2018). Conventionally, multimodal MRI involves independent acquisition of each contrast, a process that is time-consuming and costly. Moreover, differences in image resolution and nonlinear geometric mismatches between image modalities acquired using different types of readouts may result in large errors during data integration and large biases in derived microstructural properties (Mohammadi et al. 2015; Cercignani and Bouyagoub 2018). To obviate this, specially designed joint acquisition MRI sequences (Yablonskiy and Haacke 1997; Marques et al. 2010; De Santis et al. 2016) have been developed to enable more efficient acquisition of multiple image weighting in a single scan. Complex analytical models are required to extract information from multiple MR contrasts from data acquired in this fashion (Cercignani and Bouyagoub 2018). Previous models were mostly developed for ex vivo experiments (e.g. Stuber et al. 2014) and performed inconsistently when used for in vivo whole-cortex microstructural characterization (e.g. Marques et al. 2017) due to field inhomogeneity, lower signal-to-noise ratio (SNR), and the effects of unmodeled biological factors on the MR signal (Edwards et al. 2018).

MR fingerprinting (MRF) has the potential to avoid the obstacles noted above. MRF facilitates simultaneous quantification of multiple tissue properties (e.g. $T_1$ and $T_2$) in a single scan by pseudorandom variation of acquisition parameters (e.g. repetition time, echo time, flip angle) (Ma et al. 2013). Pseudorandom changes in acquisition parameters result in unique signal evolution from voxels comprising structurally different tissues. Using the same pseudorandom acquisition parameters, Bloch equation simulations of MRF signals corresponding to a range of tissue properties of interest are used to precompute a dictionary and dictionary entries are matched to the acquired MRF signal in each voxel. The tissue properties used to generate the simulated signal which best matches the acquired signal are then assigned to the voxel (Ma et al. 2013). In a previous study, we introduced the use of MRF to dissociate microstructurally distinct regions of the human cerebral cortex (Moinian et al. 2022). We developed a MRF residual analysis framework in which each voxel's residual signal contained the collective unmodeled information in the acquired MRF signal after removing the effect of $T_1$ and $T_2$*. MRF residuals revealed region-specific patterns that allowed the identification of all 11 cortical areas evaluated in our previous study. In contrast, estimated $T_1$ and $T_2$* failed to dissociate all cortical areas in some participants. This finding suggests that for the purpose of cortical microstructural parcellation, MRF signals contain spatiotemporal information that is richer than the spatial information from static quantitative $T_1$ and $T_2$* maps. Our previous analysis was based on the average MRF residual signals from each target area and the utility of residual-based analyses at the voxel-level was not investigated (Moinian et al. 2022).

The MRF residual signal from each voxel contains information about an ensemble of tissue properties that is not explained by the Bloch equations, an established MR signal model that incorporates $T_1$ and $T_2$*. The combined effect of unmodeled MR properties on the residual signal is complex. The approach of using analytical models to describe these effects at the voxel level would face the same issues that impede modeling approaches to joint acquisition MRI sequences. Furthermore, to benefit from the spatiotemporal information in MRF signals, these models would need to be developed on high dimensional signals (because each MRF signal could contain over 1,000 data points) from a large number of voxels, adding to the complexity of identifying area-specific patterns. Machine learning (ML) classification algorithms provide a data-driven alternative. They facilitate the integration of information from high-dimensional data to develop a statistical model characterizing MRF residual signals so as to identify cortical region-based differences at the voxel level. In supervised ML classification, a model is trained using a set of data samples, labeled to describe the distinctions between categories present in the training dataset. A generalizable classification model would allow reliable, efficient, and automatic prediction of the class of future data samples with unknown labels (Magnin et al. 2009; Focke et al. 2011; Orru et al. 2012; Sacchet et al. 2015; Usman and Rajpoot 2017).

This research aims to establish the feasibility of automated in vivo parcellation of the cerebral cortex in individual subjects using data-driven ML-based algorithms to identify region-specific patterns in spatiotemporal MRF residuals.

## Materials and methods

### Subjects

Six volunteers (1 male and 5 females) participated in a 2-h-long MRI scan session. Participants were healthy individuals aged between 27 and 35 years with no history of neurological disease. Each participant underwent 2 scans separated by a 15-min break. A whole-brain 3D echo-planar imaging (EPI)-based MRF scan was performed in both sessions. We also performed a whole-brain 3D Saturation-prepared with 2 Rapid Gradient Echoes (SA2RAGE) and Magnetization-Prepared

2 Rapid Acquisition Gradient Echoes (MP2RAGE) scan. All participants were provided with an overview of the experiment and signed a written consent prior to the MR scans. The scan protocol was approved by the local ethics committee and was conducted at the Centre for Advanced Imaging, The University of Queensland.

## Regions of interest

We selected 8 cortical areas from 3 regions of the cerebral cortex extracted from an atlas of a subset of Brodmann areas (BA) integrated in FreeSurfer 7.1.1 image analysis suite (Fischl et al. 2008; Fischl 2012): primary somatosensory areas BA1 and BA2, primary motor area BA4a and premotor area BA6 from the central region, areas BA44 and BA45 from Broca's region, and areas V1 (BA17) and V2 (BA18) from visual cortical region. These cortical areas were chosen because they are microstructurally distinct on histology (Geyer et al. 1996, 1999; Amunts et al. 1999, 2000; Geyer 2012).

The FreeSurfer BA labels provide individual-specific surface-based probability maps for each cortical area, representing the interindividual variation in the spatial location of each area. A high probability value indicates that there is a high chance that the corresponding vertex number on an individual's cortical surface belongs to the area of interest. To avoid overlap between probability maps of selected areas, we used thresholded maps of regions of interest (ROIs) for each individual, created by FreeSurfer surface analysis.

## MRI acquisition

The scans were performed at the Centre for Advanced Imaging, The University of Queensland, on a 7T whole-body MRI research scanner (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil (Nova Medical, Wilmington, Massachusetts). We performed third-order $B_0$ shimming at the beginning of each scan session to reduce susceptibility effects.

## $B_1^+$ map

We utilized a 3D SA2RAGE sequence (Eggenschwiler et al. 2012) to obtain a whole-brain transmit magnetic field ($B_1^+$) map for $B_1^+$ inhomogeneity correction of the MRF signals (Ma et al. 2017). The SA2RAGE acquisition parameters were as follows: repetition time (TR) = 2,400 ms, echo time (TE) = 0.95 ms, flip angle (FA)1/2 = 6°/10°, voxel size = 4 × 4 × 4 mm, and matrix size = 48 × 58 × 64.

## $T_1$-weighted anatomical image

We acquired a 3D $T_1$-weighted anatomical scan of each participant, using a prototype MP2RAGE sequence (Marques et al. 2010) with TR = 6,000 ms, TE = 3.97 ms, inversion time TI1/2 = 800 ms/2,700 ms, FA1/2 = 4°/5°, voxel size = 1 × 1 × 1 mm, and matrix size = 222 × 240 × 144. $T_1$-weighted images were used to perform surface-based processing of individuals' MRF signals and to project MRF signals to vertices on an inflated model of the cortical surface of each individual subject. $T_1$-weighted images were also used to create gray matter mask of each individual.

## MRF acquisition

A 3D EPI-based (Poser et al. 2010) MRF sequence developed in-house was used to acquire 1,000 frames of 3D MRF images with the following parameters:

FA = 10–77° (Fig. 1A), TR = 41–99 ms (Fig. 1B), TE = 12–48 ms (Fig. 1B), partial Fourier phase = 6/8, voxel size = 1.4 × 1.4 × 1.4 mm, and matrix size = 142 × 142 × 88. We used GRAPPA parallel imaging (Griswold et al. 2002) in both phase encoding (with acceleration factor = 3 and reference lines = 36) and slice encoding directions (with acceleration factor = 2 and reference lines = 12). Chemical shift selective fat saturation technique (Frahm et al. 1985) was used to reduce common artifacts observed in EPI sequences at high-field scanners (Edelman et al. 1994).

The sinusoidal FA pattern used for the MRF acquisitions (Fig. 1A) was adopted from Buonincontri and Sawiak (2016). We added abrupt FA changes (0°-40°-50°-60°-0°) at the final MRF frames to increase the sensitivity of MRF signals to $B_1^+$ variations as previously suggested (Buonincontri et al. 2017).

We also used pseudo-randomized patterns of TE variation suggested by Rieger et al. (2017) (Fig. 1B), to improve SNR as a result of the larger number of short TE values. Shorter TE values also allowed the total MRF acquisition time to be optimized by utilizing the minimum possible value of TR for each TE at each MRF frame. The alternating TE pattern has also been shown to increase sensitivity of MRF signals to $T_1$ and $T_2$* variations (Rieger et al. 2017), potentially improving estimation accuracy for these tissue properties.

## MRF dictionary matching

An MRF dictionary was generated using Bloch equation simulation of MRF signal evolutions (in MATLAB R2019b, MathWorks, Natick, MA) for a range of $T_1$, $T_2$*, and $B_1^+$ values as follows:

$T_1$ = 0.5–5 s (in increments of 10 ms, 20 ms, 30 ms, 40 ms, and 50 ms for 0.5–1 s, 1–2 s, 2–3 s, 3–4 s, and 4–5 s, respectively); $T_2$* = 10–100 ms (in increments of 2 ms, 3 ms, and 5 ms for 10–20 ms, 20–40 ms, and 40–100 ms, respectively); $B_1^+$ = 0.3–1.4 in increments of 0.05. The $B_1^+$ value for each MRF dictionary entry was multiplied by the nominal FA values (Fig. 1A) to calculate the actual FA values for MRF signal simulation (Buonincontri and Sawiak 2016).

We performed MRF dictionary matching for each voxel of the 3D MRF volumes in MRF native space. To increase the efficiency of MRF dictionary matching, we used the SA2RAGE $B_1^+$ value of each voxel to only include the relevant subset of dictionary entries in the signature matching process (Chen et al. 2016). The dictionary signal profile which had the largest inner product with the acquired MRF signal from the voxel of interest was selected as the best fitting simulated MRF signal.

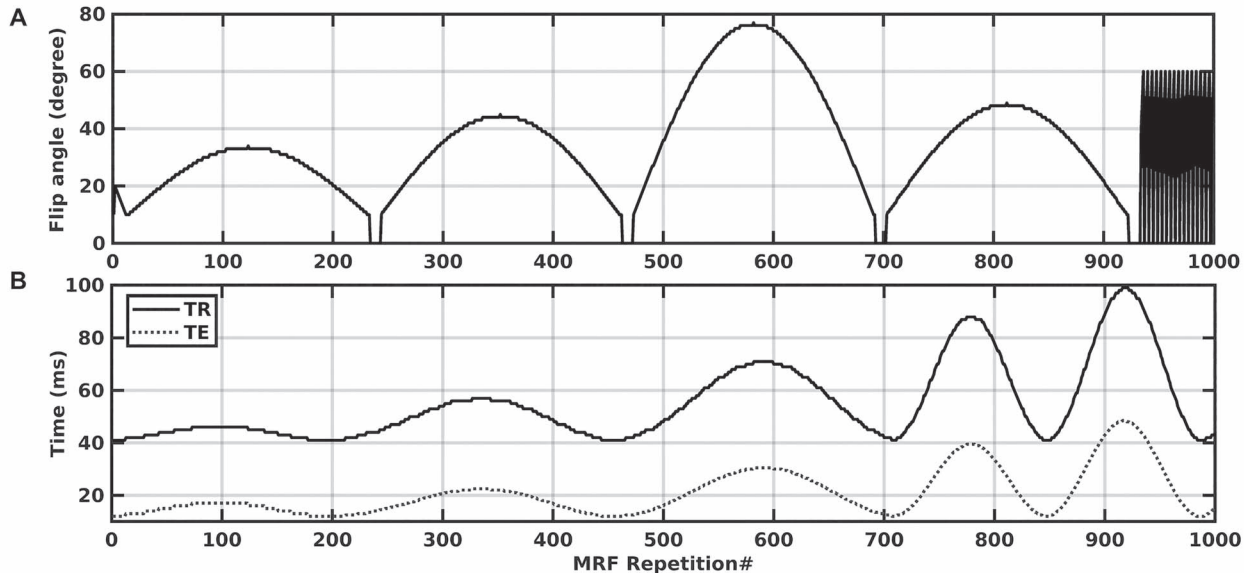


**Fig. 1.** A) The sinusoidal pseudorandomized pattern of flip angles and B) the pseudorandomized pattern of TR and TE, used to acquire the 3D MRF images.

## Image processing

We first extracted the brain tissue (using FSL-BET; Smith 2002; Smith et al. 2004) from all MRF, $T_1$-weighted, and $B_1^+$ images, to obtain a higher accuracy of brain tissue co-registration between these images. Using MCFLIRT (Jenkinson et al. 2002), we performed rigid body (6 parameter) co-registration of all 1,000 MRF 3D volumes of each MRF scan to the mean volume of the series. Next, we performed rigid body registration (using FSL-FLIRT; Jenkinson and Smith 2001; Jenkinson et al. 2002) of the MRF volumes of the second scan session to the MRF volumes of the first scan session and subsequently calculated the average MRF volume between the 2 co-registered scans for each MRF frame. We used FSL-FLIRT to perform rigid body registration of the MP2RAGE $T_1$-weighted image to the average MRF image of the subject.

We utilized FSL_anat tool (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/fsl_anat) to create the gray matter mask of each individual using the MP2RAGE $T_1$-weighted image co-registered to the average MRF volume. The FSL_anat tool performs enhanced bias-field correction, which improves tissue segmentation accuracy. Each individual's gray matter mask was then used to exclude voxels with gray matter probability of smaller than 90%.

The SA2RAGE $B_1^+$ map was first linearly co-registered to the MP2RAGE $T_1$-weighted image of the subject using FSL-FLIRT. Next, it was transformed to the MRF native space using the linear transformation matrix previously obtained by co-registering the MP2RAGE $T_1$-weighted image to the subject's average MRF volume.

## Surface-based processing

The MP2RAGE $T_1$-weighted image co-registered to the average MRF volume was processed using the FreeSurfer 7.1.1 *recon-all* pipeline to generate an inflated model of the cerebral cortex surface for each individual. We calculated the transformation between voxel coordinates in MRF native space and surface coordinates to project voxel-wise MRF signals to corresponding vertices on the cortical surface model of an individual.

The FSL_anat tool (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/fsl_anat) was used to perform enhanced bias-field correction of MP2RAGE images to improve FreeSurfer's white and pial surface estimations.

## Supervised classification model

We employed classical supervised ML classification algorithms to perform vertex-wise cortical parcellation on the 8 cortical areas of interest. Supervised classification involves learning a model based on a set of labeled training data samples (i.e. with known classes) characterized by a set of statistical features. The model should be able to predict the class label of unseen data samples (i.e. test set), which are described using the same set of features.

## Multi-class supervised classification

We formulated parcellation of cortical areas of interest as a regional multi-class supervised classification problem (Weston and Watkins 1998). All vertices of interest from the 8 cortical areas were uniquely labeled based on the FreeSurfer's thresholded BA probability maps of each subject.

## One-versus-one multi-class classification

In cortical regions with more than 2 ROIs (i.e. the central region), we employed a one-versus-one (OVO) multi-class classification approach, in which one classifier is trained for each pair of classes (Géron 2019). The class that receives the majority of votes at the time of prediction is considered to be the final class of the input sample.

Compared with one-vs-the-rest multi-class classification, OVO is potentially more robust to the class imbalance problem (Géron 2019), which we discuss further in Section 2.21.3.

### Feature representation

Classical ML classification algorithms require data samples that are characterized by statistical measures that distinguish between classes, with instances of the same class showing the highest similarity (Meyer-Baese and Schmid 2014). We previously demonstrated that the normalized autocorrelation of MRF residual signals can be used to characterize structural distinctions between different cortical areas (Moinian et al. 2022). Accordingly, here we used the normalized autocorrelation values (ranging from −1 to +1) of the MRF residual signals to form a feature vector for each vertex of interest. We examined the efficacy of 2 approaches to feature representation:

### Single vertex feature representation

Here, we used the normalized autocorrelation of the MRF residuals corresponding to the vertex of interest and created a 999-dimensional feature vector per vertex number (i.e. autocorrelations at all lags other than lag 0 at which the normalized autocorrelation is always 1).

### Neighborhood-based feature representation

An established approach in quantitative histological studies requiring delineation of boundaries between cortical regions is to exploit neighborhood information when characterizing a particular location in the cortex (Geyer et al. 1996; Amunts et al. 1999; Geyer et al. 1999; Amunts et al. 2000). This allows the local statistical characteristics of tissue properties of interest (e.g. neuronal cell density) to be derived from neighboring spatial locations in the cortex (Eickhoff et al. 2018). We followed a similar approach by incorporating information from a neighborhood of vertices immediately adjacent to the vertex of interest. A feature vector was created for each vertex by calculating the autocorrelations of the average MRF residuals in a vertex neighborhood centered on the target vertex. The neighborhood vertices add confirmatory information to the description of the vertex, which, in particular, may improve the differentiation of vertices with a large degree of similarity from 2 different areas (e.g. vertices near the border between areas).

### Classical supervised classification algorithms

We employed the following supervised classification methods: linear SVM (L-SVM), radial basis function kernel SVM (RBF-SVM), random forest (RF), and *k*-nearest neighbors (KNN). To implement these classifiers, we used scikit-learn (v0.21.3) (Pedregosa et al. 2011), which provides an open-source ML library for Python (v3.5). Model selection (described in Section 2.22) was performed to fine-tune the key parameters for each algorithm.

### Support vector machine

Generally, in a high-dimensional feature space, SVM aims to find a separating hyperplane (i.e. decision boundary), which splits the data samples into distinct classes, such that the distance from the hyperplane to the data samples is maximal (Boser et al. 1992). SVM-based classifiers are robust to model overfitting when dealing with a high-dimensional feature space in a variety of neuroimaging applications (Fan et al. 2005; Zhang et al. 2011). Key parameters of the SVM model are as follows.

#### *Regularization parameter (C)*

The regularization parameter defines the penalty applied to a misclassified data sample when the SVM algorithm is searching for the optimal decision boundary. More regularization (i.e. smaller values of C) creates smoother decision boundaries, leading to greater tolerance to noise or outliers in the data samples.

#### *Kernel function*

The measure of similarity between data samples in feature space is based on a distance function, the SVM kernel. We used both a linear (L-SVM) and a nonlinear (RBF-SVM) kernel. In the former, the decision boundary between data samples in feature space is linear, whereas in the latter, data samples are transformed into a new feature space that permits a linear decision boundary to be found. The linear decision boundary in the new feature space corresponds to a nonlinear decision boundary in the original feature space. We chose the RBF as the kernel function, in light of previous reports of better prediction performance and more time-efficient training for high-dimensional data classification, compared to other SVM kernel types (e.g. polynomial) (Yekkehkhany et al. 2014). The kernel function of the RBF-SVM algorithm is defined as:

$$K(x_1, x_2) = e^{-\gamma \cdot \|x_1 - x_2\|^2} \quad (1)$$

where $x_1$ and $x_2$ are the feature vectors of 2 data samples and $\|x_1 - x_2\|^2$ is the distance between the 2 samples in the original feature space.

#### *Gamma parameter ($\gamma$)*

This parameter specifies the similarity radius of the kernel function. As seen in Equation (1), the similarity between 2 data samples in the transformed feature space (K) is a function of their distance in the original feature space ($\|x_1 - x_2\|^2$). Smaller $\gamma$ values lead to slower decay of the exponential similarity function K. This in turn forms a larger similarity radius around each sample data point and thus the samples at greater distances in the feature space will be considered as similar, creating smoother decision boundaries.

### Random forest

A RF classifier consists of an ensemble of decision tree predictors, each of which is built using a random number of training data samples and a random subset of features

at each splitting node (Breiman 2001). The majority voting method is used to predict the class of unseen data samples, reducing the risk of model overfitting, often seen with single decision tree models (Han et al. 2012). RF models are a theoretically appropriate choice for the current study because they are reportedly robust against outliers (which may occur due to misregistration or the probabilistic ROI extraction here) and can handle multiclass classification problems efficiently (Han et al. 2012).

To avoid model overfitting and improve the generalizability of the RF model, we used bootstrapping so that each tree in the model was only trained on a random number of data samples. The key RF algorithm parameters tuned during model selection are as follows.

*n-estimators*

This parameter defines the total number of decision trees built for the RF model. A very large value may lead to overfitting while a small value may cause underfitting.

*max-depth*

This parameter controls the maximum depth to which each individual decision tree is allowed to grow (i.e. split further). If not specified, or if set too high, it may lead to full-grown trees where all the leaf nodes are pure (i.e. all the data samples in the leaf nodes belong to a single class), causing overfitting to the training data.

*max-features*

This parameter specifies the maximum number of features that could randomly be selected at each node of the tree to find the best split (i.e. the split that will create more pure child nodes).

*criterion*

This is an impurity measure that is used to decide the features that create the best split. We evaluated the use of entropy and the Gini index:

$$\text{Entropy} = -\sum_j P_j \log_2 P_j \tag{2}$$

$$\text{Gini} = 1 - \sum_j P_j^2 \tag{3}$$

where $P_j$ is the probability of class $j$ in the child node.

## K-nearest neighbors

The KNN algorithm classifies each data sample according to the class of its K nearest samples in the feature space (Cunningham and Delany 2007). This is a simple and effective classification algorithm, which usually performs well in a wide range of applications dealing with low-dimensional data sets (Cunningham and Delany 2007). However, its prediction performance is competitive for high-dimensional data when combined with neighborhood component analysis (NCA) as a distance metric learning algorithm (Cover and Hart 1967). Additionally, KNN performs well in data sets with irregular decision boundaries (Cunningham and Delany 2007). We tuned the following parameters during model selection.

*n-neighbors (K)*

This parameter denotes the number of nearest neighbors to consider for each data sample. K influences the smoothness of the decision boundaries (i.e. model complexity). Decision boundaries are less smooth for small $K$ values, increasing the chance of model overfitting.

*weights*

This parameter defines a weighting function for the distance between a data sample and its KNN. We evaluated uniform weighting, in which all K neighbors are equally weighted, and distance weighting, in which the weights of neighbors are inversely proportional to their distance from the data sample.

## Data preparation

Our data preparation pipeline is illustrated in Fig. 2.

## Feature preprocessing

*Feature scaling*

To give the values at all autocorrelation lags equal importance in terms of descriptive information provided to the classification models, we calculated normalized autocorrelations. The normalization scales the values of each feature (i.e. the autocorrelation value at each lag number) to the range of 0 to 1 in all data samples. The scaling prevents the distance measures from being biased towards data samples of higher magnitude but maintains the original relationship between features (Han et al. 2012).

*Dimensionality reduction*

If the autocorrelations of the MRF residual signals were to be used as feature vectors, the feature space would have 999 dimensions. High-dimensional feature space may not always benefit classification algorithms (Jain et al. 2000) because training and prediction time increase and prediction accuracy may drop (Jain et al. 2000). Feature reduction is usually performed to find the most effective subset of features for the classification problem at hand (i.e. to remove features that do not contribute much to sample discrimination in the training set). This is especially important when the number of features is large relative to the number of samples available in the training set. We performed dimensionality reduction using:

(a) Feature selection

This eliminates the features that contribute little to the separability of different classes (Meyer-Baese and Schmid 2014). We employed a MRF residual signal subsampling scheme as a feature selection method (see Supplementary Figs. S1 and S2).

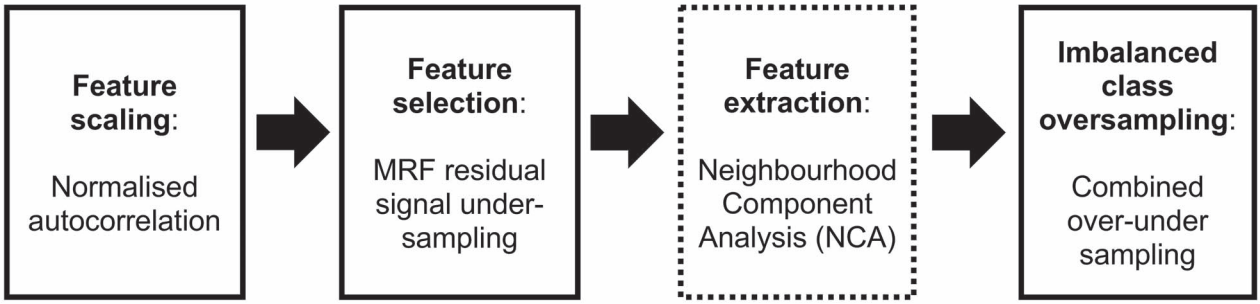


**Fig. 2.** The data preparation pipeline performed in this study. Note that the feature extraction step was not applied to all classification algorithms.

(b) Feature extraction

Feature extraction methods use all the information provided in the original feature space to transform the data into a new lower-dimensional feature space. In most feature extraction methods, each feature in the new space combines the most discriminative features from the original space (Meyer-Baese and Schmid 2014).

Performing both feature extraction and feature selection can improve the efficiency and predictive performance of a KNN classifier in a high-dimensional feature space (Goldberger et al. 2005). We used NCA, a supervised Mahalanobis distance metric learning algorithm that aims to maximize the prediction performance of KNN by transforming the data samples into a new feature space in which instances from the same class have higher similarity (Goldberger et al. 2005). The main advantage of NCA over other feature extraction methods is that it preserves information during the dimensionality reduction process (Raghu and Sriraam 2018). Euclidean distance was used to measure the similarity of data samples in the transformed space.

We also used NCA as a feature extraction method to train our RF classifier, as it has also been shown to improve the efficiency and performance of other classification algorithms (Raghu and Sriraam 2018). However, as SVM models can deal with high-dimensional data effectively, we did not apply NCA to SVM classifiers.

### *Imbalanced class oversampling*

Selected cortical areas in each region of interest in this study differed in size. The difference in the total number of instances of each class (i.e. vertices in each FreeSurfer BA probability map) in the training dataset leads to a class imbalance problem, which can affect classifier performance. One remedy is to balance the training set by increasing the number of the minority class instances (i.e. oversampling) or decreasing the number of the majority class instances (i.e. undersampling). To avoid overfitting caused by random oversampling or to lose important information by random undersampling, we performed combined over- and undersampling on all classes. The combination of synthetic minority oversampling technique (SMOTE) and edited nearest neighbor (ENN), called SMOTEENN, has been shown to outperform other combined sampling methods (Batista et al. 2004).

## Model selection and evaluation

For each classification algorithm, we used a grid search to evaluate the prediction performance of models built using different combinations of parameter values (Hsu et al. 2003). The set of parameter values that resulted in the model with the best prediction performance (i.e. with the highest evaluation score) on the unseen testing data samples was selected as the best parameter set and the resultant model as the best classifier. Grid search model selection is efficient for applications in medium-sized parameter spaces (Hsu et al. 2003), as is the case in this study.

## Evaluation metric

We used the area under the receiver operating characteristic (ROC) curve (ROC-AUC) (Bradley 1997) to evaluate each classifier's prediction performance. AUC-ROC allowed us to identify the best probability threshold for each classifier and to select the best model when comparing different classifiers.

In a binary classification problem with so-called positive and negative classes, the classification algorithms often return the probability of each data sample being from the positive class. A probability threshold is then applied to define the class of the data sample. The proper discrimination threshold for a classifier varies with application. The ROC curve identifies the best threshold by plotting the true positive rate (TPR) against the false positive rate (FPR) of predictions at different prediction probability thresholds. This is a means of cost–benefit analysis, representing how much sensitivity (i.e. TPR) is gained at each threshold at the cost of a decline in specificity (i.e. 1 − FPR) (Bradley 1997). Additionally, for model selection on our imbalanced dataset, as suggested by Tang et al. (2009), we needed a single metric combining sensitivity and specificity measures when comparing the effectiveness of different models. ROC-AUC is a reliable performance measure for imbalanced datasets (Provost and Fawcett 1997) and indicates the probability of all the data samples being classified correctly. An AUC of 1 represents a model that always distinguishes data samples correctly, while an AUC of 0.5 occurs

when a classifier randomly guesses the class of the samples. Note that we used the average of the evaluation score across all classes (macro-average) to give equal importance to accurate classification of all classes.

### Repeated K-fold cross validation

We performed 10-fold cross validation (Wong 2015) on each set of model parameter values in the grid search procedure to avoid overfitting the classifiers to the limited data samples. This method randomly subsampled the whole dataset into 10 mutually exclusive folds and repeated the grid search 10 times for each set of model parameters. At each repetition, 9 partitions of the data samples were used as the actual training set and 1 partition was held out as the validation set. The validation set was then used to evaluate the model performance on unseen data, to ensure that the model was reliably generalized. The average of the evaluation metric over the 10 repetitions of the grid search was then taken as the overall performance score of the classifier for the model parameters of interest. To follow the paired experimental design approach for comparisons of performance between different models, we used the same randomly subsampled data partitions for all classifiers.

To keep the portion of data samples of each class in each fold the same as in the initial dataset, we used a stratified method for dividing the data samples into 10 folds. Additionally, we used repeated K-fold cross-validation with 10 repetitions to test the statistical significance of classifier prediction performance during the model selection process.

### Testing the selected model

We used a leave-one-subject-out model evaluation approach. We held out the data from one participant as the test set for the final model evaluation and used the data samples from the other 5 participants for training as explained above. Evaluating the model performance on a separate test set improves model generalization power, by minimizing the risk of information leakage (Kaufman et al. 2012) during model training.

### Data and code availability

As no consent was obtained from the participants to share their data in a public repository, the anonymized data can be made available upon request to the corresponding author. We also provide all the code for data analysis and ML model development in this study at GitHub.

## Results

### Model selection using single vertex approach

The results of feature selection and model selection for the single vertex approach in the central region (areas 1, 2, 4a, and 6) are presented in Fig. 3. The steps of the horizontal axis correspond to the MRF residual signal subset that was included in the model selection process. RBF-SVM (C = [BA1:6, BA2:4, BA4a:1, BA6:2] and $\gamma = 0.01$) was found to perform best with a macro-average ROC-AUC of 0.8 for class predictions for the held-out test set.

Overall, RBF-SVM also outperformed the other 3 algorithms when different subsets of MRF residuals were selected. When sampling the MRF residual at every third timepoint, the macro-average ROC-AUC for RBF-SVM was 0.79. However, in return for a 1% drop ($P$-value $<0.05$) in ROC-AUC, there was a 54% and 61% reduction ($P$-value $<0.05$) in the model fitting time and prediction time, respectively.

### Model selection using neighborhood-based approach

The results of model selection process using the neighborhood-based feature representation approach in the central region (areas 1, 2, 4a and 6) are detailed in Table 1. The RBF-SVM model (with C = [BA1:30, BA2:4, BA4a:1, BA6:2 and $\gamma = 0.01$) outperformed all other models investigated in this study with macro-average ROC-AUC = 0.85 (sensitivity = 0.77, specificity = 0.92).

### Comparing single vertex and neighborhood-based approaches

In Fig. 4, the normalized confusion matrices of the best performer of both feature representation approaches are provided for cortical areas from 3 cortical ROIs. We found the ROC-AUC of predictions on the held-out test set improved in all regions with the use of neighborhood-based feature representation. ROC-AUC of predictions increased by 5% (Fig. 4A and B), 8% (Fig. 4C and D), and 4% (Fig. 4E and F) ($P$-value $<0.05$) in the central region, Broca's region, and visual cortical region, respectively.

The prediction sensitivity scores for the cortical areas of interest are represented along the diagonal of the confusion matrices. Comparing these scores between the 2 matrices in each region shows an overall increase in sensitivity values by using neighborhood-based feature representation for all cortical areas. In average, the sensitivity of predictions rose by 7%, 8%, and 4% ($P$-value $<0.05$) in the central region, Broca's region, and visual cortices, respectively. We observed the largest increase in the sensitivity score for area BA6 with an increase of 19%.

Comparison of the similarity between autocorrelation profiles of cortical areas of interest (Supplementary Fig. S3) confirms that false positive predictions for each pair of areas (Fig. 4B, D, and F) reflect the degree of similarity between autocorrelation profiles of cortical areas. It is more challenging for the classification algorithm to separate samples with more similar autocorrelation profiles.

### Spatial distribution of model predictions

Figure 5 depicts the spatial distribution of vertex-wise label predictions for the unseen data samples of the held-out participant using the best RBF-SVM classifier trained with the neighborhood-based feature vectors. The confusion matrices of these predictions are presented in Fig. 4B, D, and F. The cortical area boundaries defined by

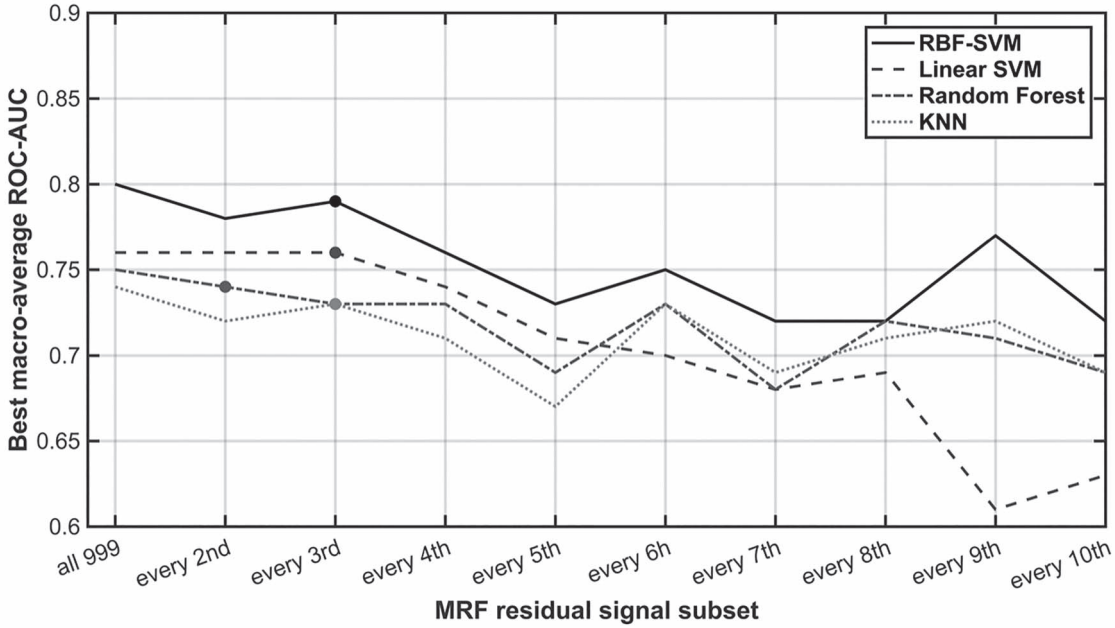


**Fig. 3.** The best macro-average ROC-AUC scores of 4 supervised classification algorithms are compared when trained with autocorrelation of different subset of MRF residual signals as the feature vectors. The classification methods used here are L-SVM, RBF-SVM, RF, and KNN. The solid circle on each plot represents the subset of MRF residual signals at which the classification algorithm showed its best performance, compared to the initial case of including the whole MRF residual signal.

**Table 1.** The ROC-AUC scores of 4 supervised classification methods are compared, when the neighborhood-based approach was used for feature representation of each vertex.

| Classification algorithm | Best macro-average ROC-AUC | Best performer parameters |
|---|---|---|
| RBF-SVM | **0.85** | C = [BA1:30, BA2:4, BA4a:1, BA6:2], $\gamma = 0.01$ |
| L-SVM | 0.75 | C = 100 (balanced class weights) |
| KNN | 0.71 | n-neighbors = 25, weights: distance, distance metric: Minkowski distance with $P = 3$ on 6 NCA components |
| RF | 0.71 | n-estimators = 200, criterion = entropy, max-depth = 14, max-features = 6 NCA components |

The value in bold indicates the classification algorithm which achieved the highest ROC-AUC when the neighborhood-based feature representation approach was used.

the FreeSurfer's thresholded BA probability maps are also shown (solid cyan lines surrounding each area). The vertex-wise class labels are overlaid on an inflated model of the cortical surface of the right hemisphere of the participant generated by FreeSurfer.

## Discussion

We demonstrated the feasibility of developing an automated in vivo method of vertex-wise cortical parcellation in 3 cortical ROIs, using a combination of MRF residual signals and supervised ML. Using a RBF-SVM classifier and neighborhood-based feature representation, an average prediction score (ROC-AUC) of 0.85, 0.78, and 0.62 was achieved for vertex-wise parcellation of 8 cortical areas in the central region, Broca's region, and visual cortical region, respectively.

The present vertex-wise cortical parcellation method aimed at improving the parcellation performance based on the structural tissue properties. In contrast, previous ML-based cortical parcellation methods mostly utilize multiple neurobiological properties (i.e. functional, connectional, and topographical features) (Glasser et al. 2016; Cucurull et al. 2018). Those studies note that extended applicability of their parcellation method to other lesion segmentation applications (e.g. multiple sclerosis) would be conditioned on preparation of a proper training dataset (Cucurull et al. 2018). Identifying and extracting suitable feature representations are likely to be key challenges to the application of these methods. An important advantage of the parcellation framework presented here is more efficient data acquisition and feature extraction, compared to parcellation methods that rely on multimodality information on multiple neurobiological properties. The characterization of each voxel in the present study does not require acquisition of separate complementary MR modalities, unlike conventional multimodal MRI microarchitectural mapping methods. Moreover, the feature extraction process here

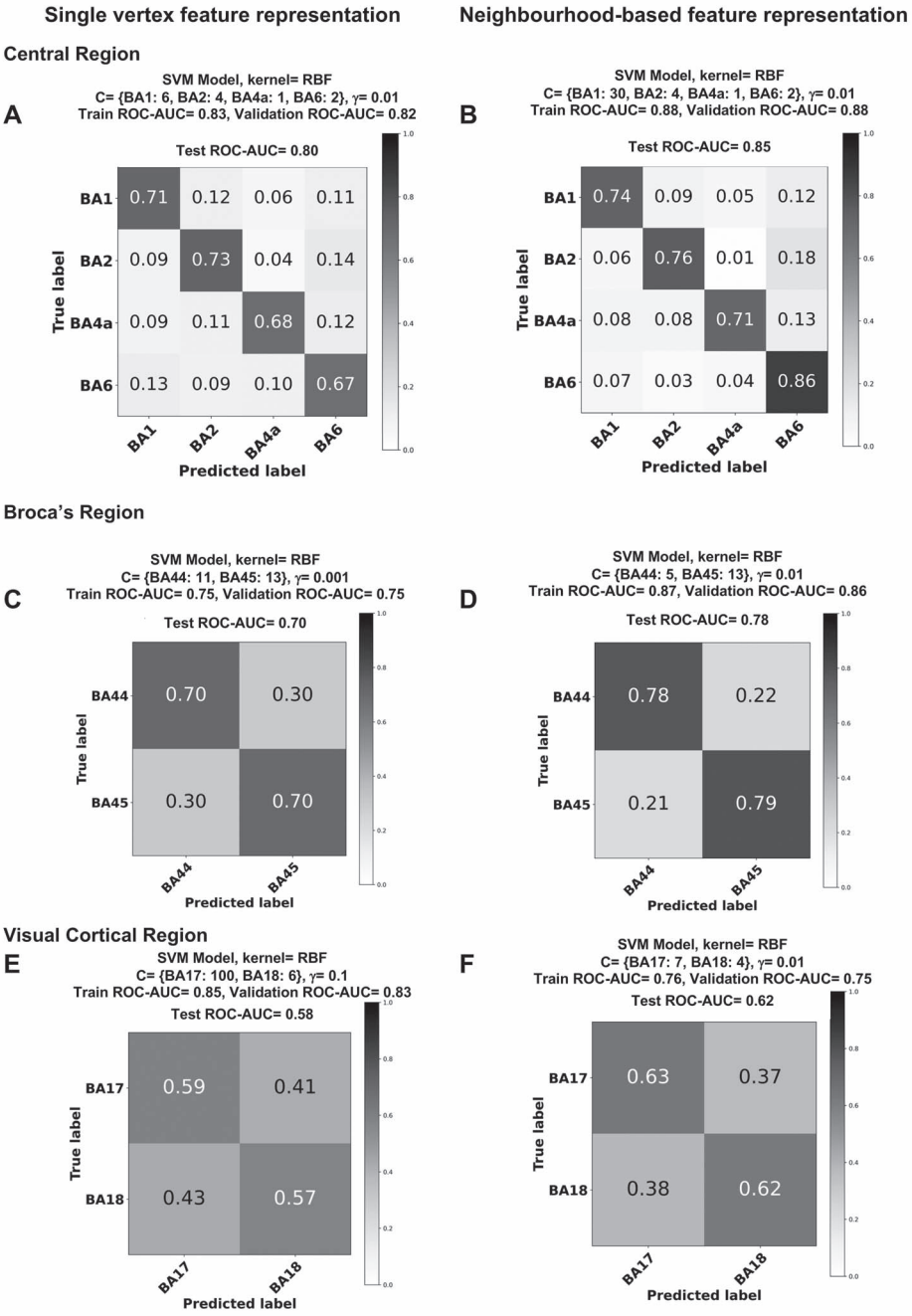


**Fig. 4.** Normalized confusion matrices for the best performer (i.e. RBF-SVM classifier) trained with (A, C, E) single vertex feature representation and (B, D, F) neighborhood-based feature representation in the central region (A, B), Broca's region (C, D), and visual cortical region (E, F). The model parameters, validation, and test ROC-AUC scores of the best classifier are stated at the top of each matrix. Cell values along the diagonal of each matrix represent the sensitivity of the model predictions on the held-out test set. The values were rounded up to 2 decimal places.

involves efficient computation of the MRF residuals and autocorrelation of the residuals, eliminating the need for complicated data integration from multiple modalities. Consequently, the efficient microstructural feature extraction method of the present study facilitates the application of our ML-based parcellation method to other voxel-wise tissue segmentation applications such as classification of brain tumor types.

We found the feature representation approach based on a patch of neighboring surface vertices led to higher prediction accuracy in comparison with the single vertex approach (Fig. 4). The neighborhood-based feature representation used in this study could be seen as a combined local–global brain partitioning method. Similar to local partitioning methods, the inclusion of features from the patch of vertices surrounding the vertex of interest provided the similarity functions of classification algorithms with descriptive information about the vertex neighborhood. At the global level, the similarity of each neighborhood was measured against all other

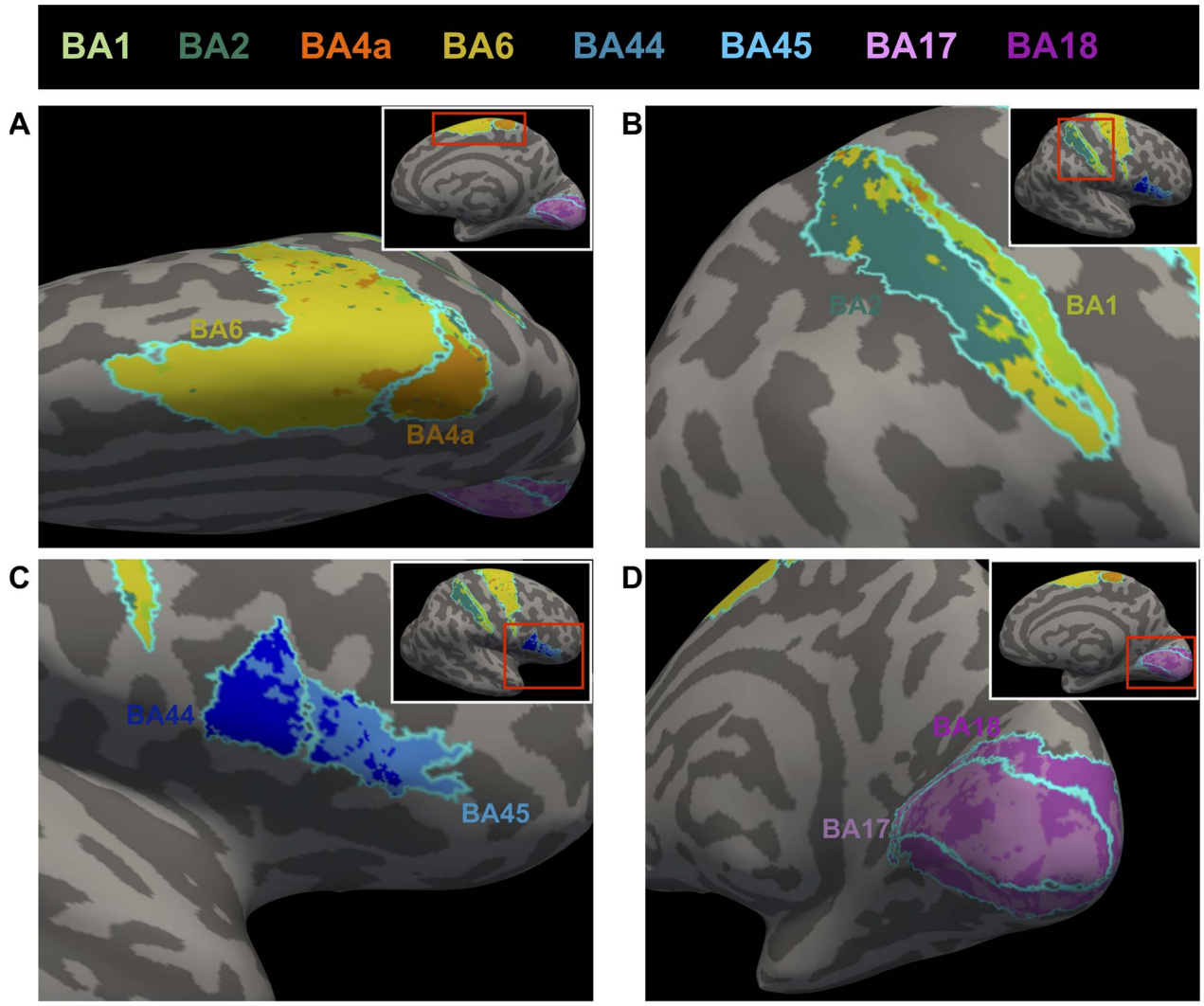


**Fig. 5.** Cortical parcellation results for (A) primary motor area BA4a (orange) and premotor area BA6 (yellow) from the precentral gyrus region, (B) primary somatosensory areas BA1 (light green) and BA2 (dark green) from the postcentral gyrus region, (C) areas BA44 (dark blue) and BA45 (light blue) from Broca's region, and (D) areas BA17 (light pink) and BA18 (magenta) from the visual cortical region of the held-out participant, using the best RBF-SVM classifier trained with the neighborhood-based feature representation. The cyan solid lines around each area show the true class borders. The predicted vertex-wise class labels are overlaid on an inflated model of the right hemisphere cortical surface of the participant, generated by FreeSurfer.

neighborhoods regardless of their spatial location on the cortical surface. Hybrid local–global brain partitioning methods have demonstrated improved parcellation performance, compared with local (e.g. classical histological mapping) or global (e.g. connectivity-based parcellation) partitioning approaches (Eickhoff et al. 2018; Schaefer et al. 2018). Higher prediction accuracy of neighborhood-based feature representation is likely to stem from its hybrid parcellation approach, in line with similar improvements observed in other hybrid methods of cortical mapping (Schaefer et al. 2018).

The pattern of false prediction rates along each row of the confusion matrix (excluding the elements on the diagonal) of the best classifier (Fig. 4B) may imply microstructural similarities between the cortical areas of interest. For example, when comparing the similarity of BA2, BA4a, and BA6, false BA4a predictions (1%) were significantly less frequent than false BA6 predictions (18%) for BA2 data samples. Additionally, false BA6 predictions (13%) were larger than false BA2 predictions (8%) for BA4a data samples. Further, BA6 data samples were more frequently predicted as BA4a (4%) than BA2 (3%). This comparison between these 3 cortical areas suggests that BA4a was more similar to BA6 than to BA2, causing the classification model to make less accurate predictions when distinguishing the MRF residual signals from BA4a and BA6. Additionally, the larger false BA6 predictions for BA2 data samples suggest higher similarity between BA2 and BA6, compared to the similarity between BA2 and BA4a. These findings are in line with the similarity measurements between these 3 areas reported in Moinian et al. (2022). The residual-based areal dissimilarity between these 3 areas in the current study is in agreement with the pattern of microstructural dissimilarity on histology. For example, larger microarchitectonic distinctions have

been observed between areas 2 and 4 than between areas 4 and 6 (Geyer et al. 1996, 1997; Geyer 2012; Nieuwenhuys 2013; Mangeat et al. 2015). Additionally, we observed the smallest false positive BA4a predictions for data samples of other 3 cortical areas in the central region using both feature representation methods (false BA4a prediction of 10% when using the neighborhood-based approach), which may correspond to the distinctive microstructural characteristics of this area as reported on histology. The presence of giant pyramidal cells (i.e. Betz cells) and lack of layer IV are reportedly the unique microstructural features of area 4 (Geyer et al. 1997; Amunts and Zilles 2015). Therefore, it is plausible to infer that our feature vectors (i.e. autocorrelation of MRF residuals as illustrated in Supplementary Fig. S3) could reflect microarchitectonic characteristics of the areas of interest at the voxel level.

Nonetheless, it should be noted that histological studies of cortical microstructure have mostly employed local partitioning methods (Eickhoff et al. 2018), making further microstructural inferences about the other areas of interest in this study challenging. Further validation of the microstructural similarity findings of the present study would require experiments that compare a combination of microarchitectonic characteristics of areas globally, instead of focusing on the local microarchitectural transitions at the borders between adjacent areas.

### The effect of non-microstructural factors on classification

Here we excluded MRF residual signals from voxels with gray matter probability less than 90% to reduce the contribution of neighboring white matter and cerebrospinal fluid to gray matter MRF signals (partial volume effect). However, due to the large voxel size of MRF images, complete mitigation of partial volume effect is not feasible. We also examined the effects of 3 non-microstructural features, which may influence the degree of partial volume effect in the cortical areas of interest: average cortical thickness, local cortical curvature, and vertex displacement from the mid-surface. Variation in cortical thickness (Fig. 6A), local cortical curvature (Fig. 6B), and displacement of vertices from the mid-surface (Fig. 6C) were not correlated with classification error defined as the percentage of misclassified test data samples from each area.

### Spatial distribution of classification errors

The spatial distribution of classification predictions on an inflated model of the cortical surface of the held-out participant (Fig. 5) shows that classification errors for BA1, BA2, BA4a, BA44, and BA45 are mostly located at the borders of these areas. However, the classification errors for BA6, BA17, and BA18 are centrally distributed in some regions, e.g. anterior part of BA17 and inferior part of BA6. In BA6, this may be explained by high within-area microarchitectonic variation. For example, Glasser and Van Essen (2011) observed higher myelination superiorly in BA6. High microarchitectonic heterogeneity in a large area like BA6 could affect the prediction performance of the present ML-based parcellation method. Increasing the number of subjects in the training set in future work may improve the classification performance in such cases.

Additionally, due to the large MRF image voxel size, cortical regions separated from the cortical area of interest by a sulcus can contribute to the MRF signal of the area of interest, especially in regions with highly folded, relatively thin gray matter (Supplementary Fig. S4) as in BA17 and BA18. This effect may have resulted in greater similarity between the MRF signals of BA17 and BA18, thus posing a greater challenge for the classification algorithm in differentiating between these 2 areas. Furthermore, tissue segmentation of visual cortical regions using FreeSurfer may be degraded by partial volume effect, thereby impairing surface reconstruction in visual areas (Glasser and Van Essen 2011). Greater inaccuracy in the gray matter masks of visual areas compared to other areas of interest may, in turn, lead to inclusion of voxels with larger partial volume effect in the training set. In the application of supervised ML to classification, errors in the class labels of the training dataset may affect the resultant model. In future work, higher MRF image resolution and improved bias field correction of $T_1$-weighted structural images may improve classification performance in visual areas.

### The effect of MRF acquisition parameters on classification

We explored the effect of MRF acquisition parameters on our MRF-based cortical parcellation method by using the signal subsampling scheme introduced in the Supplementary Information (Supplementary Figs. S1 and S2). The model selection process in Fig. 3 showed that when we used every third MRF residual timepoint, the RBF-SVM model prediction score only dropped by 1%. We believe that this is because the shape of the residual signals was preserved when subsampled at every third timepoint. We also trained an RBF-SVM classifier using the first 500 MRF residual signal timepoints. The prediction score (ROC-AUC = 0.72) of this RBF-SVM classifier was 6% ($P$-value $<0.05$) lower than that of a RBF-SVM model trained using every second residual timepoint, which had the same feature vector length but a different pattern of acquisition parameters. These observations suggest that the pattern of MRF acquisition parameters affects our MRF-based cortical parcellation method, a subject for systematic investigation in future work. We used the autocorrelation profile to characterize MRF residual signals because it allowed us to incorporate the spatiotemporal information of MRF signals in the classification. Feature importance analysis methods would identify the most discriminative autocorrelation lag numbers but not the most important signal timepoints for cortical parcellation. Future studies may consider classification

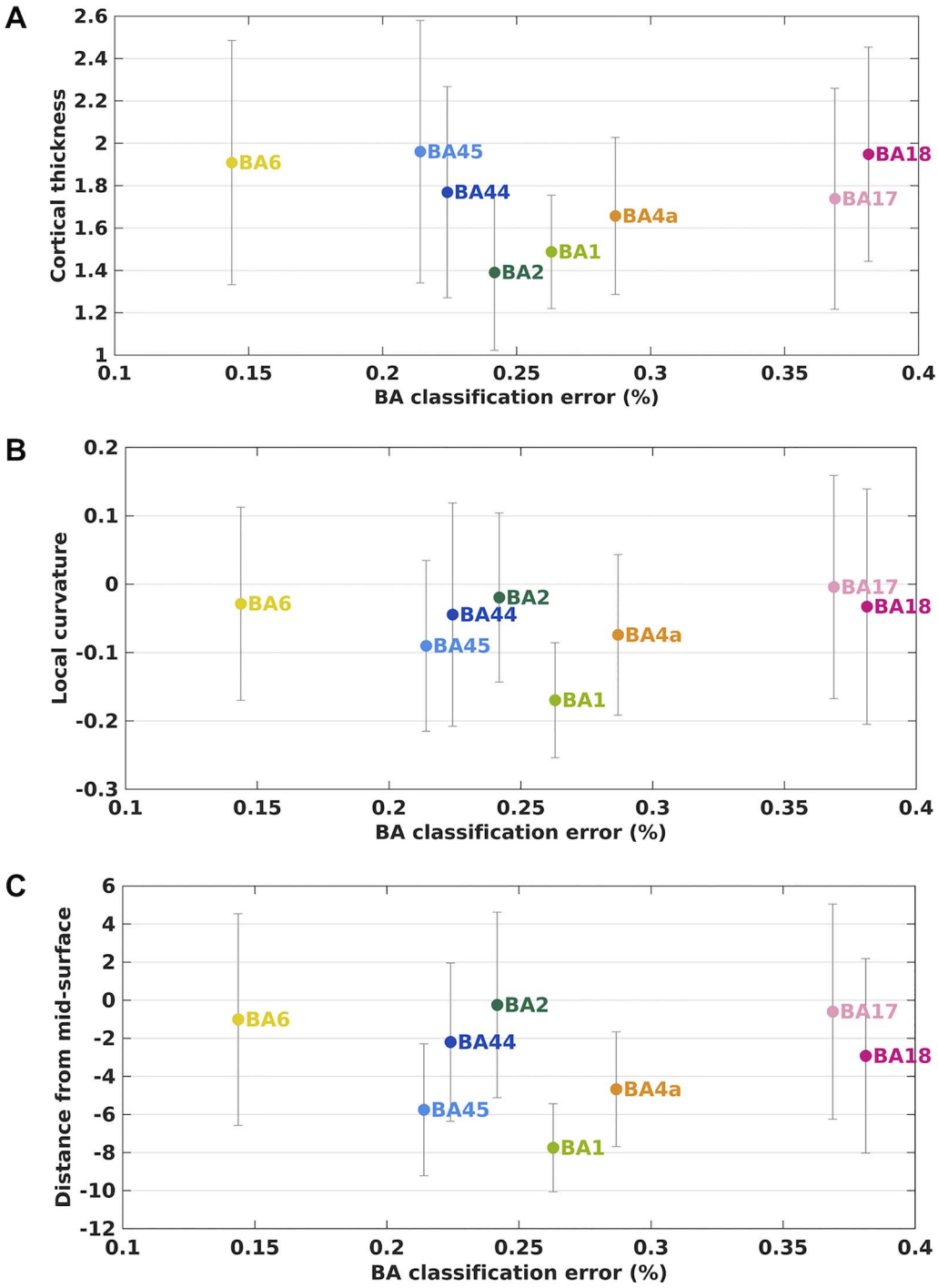


**Fig. 6.** A) Mean cortical thickness, (B) mean local cortical curvature, and (C) mean vertex distance from the mid-surface against the percentage of classification error (using neighborhood-based RBF-SVM classifier) for 8 cortical areas of interest. Solid circles indicate the mean of the non-microstructural feature and the error bars represent the standard deviation of that feature in each cortical area.

algorithms that use MRF residual signals as the input. In this case, feature importance analysis may provide more direct information on the effect of MRF acquisition parameters.

## Considerations for future work

Our voxel-wise parcellation method may be integrated with the image reconstruction procedure on the MR scanner, enabling real-time cortical parcellation. To achieve this, an essential step is to improve the time efficiency of the parcellation approach. The dimensionality reduction results illustrated in Fig. 3 demonstrate the feasibility of increasing the time efficiency of our parcellation approach. Including every third MRF residual signal timepoints (out of all 1,000 timepoints) for the RBF-SVM classifier resulted in a 54% and 61%

decrease ($P$-value $< 0.05$) in the model fitting and prediction time, respectively. Additionally, this would allow a reduction in total MRF data acquisition time from about 42 min to about 13 min in future work. This increased time efficiency increases the feasibility of real-time applications of parcellation of the whole cerebral cortex. However, the gains in time efficiency were accompanied by a 1% drop in the prediction score (ROC-AUC) of the classifier. In the context of voxel-wise cortical parcellation, reductions in accuracy could be clinically important. Thus, in future work, the trade-offs between time efficiency, prediction accuracy, and dimensionality reduction should be examined further. Moreover, the generalizability of the classification model needs to be investigated in future work by acquiring data from a greater number of individuals. Generalization may drop for the RBF-SVM method as more test subjects are processed.

There are many avenues for improving the accuracy of the present parcellation method in future work. Here, we only used autocorrelation to create feature vectors. Adding other statistical measures enhanced prediction accuracy in other ML studies (Wang et al. 2006). Moreover, deep learning methods are effective in learning features from time series data (Langkvist et al. 2014). In addition to enhanced efficiency, deep learning-based classification may improve accuracy by extracting more discriminative features from the MRF residual signals.

## Conclusion

Our results demonstrate the feasibility of developing a ML classification approach based on MRF residual signals for automated in vivo parcellation of the human cerebral cortex. We also showed that a neighborhood-based feature extraction approach improved the accuracy of predictions. In future, we expect to improve the accuracy of predictions by (i) using more statistical measures to characterize the MRF residual signals and (ii) collecting a larger training set.

## Acknowledgments

The authors acknowledge the facilities and scientific and technical assistance of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at the Centre for Advanced Imaging, The University of Queensland. We also thank Nicole Atcheson, Sarah Daniel, and Aiman Al Najjar from the Centre for Advanced Imaging, The University of Queensland, for helping with MR data acquisition.

## Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

## Funding

This work was conducted by the Australian Research Council Training Centre for Innovation in Biomedical Imaging Technology research and funded, in part, by the Australian Government through the Australian Research Council (DP140103593 and IC170100035). SM holds a post-doctoral position as part of the Australian Research Council Training Centre for Innovation in Biomedical Imaging Technology (chief investigators include DR and VV), grant number IC170100035.

*Conflict of interest statement*: The authors have no conflicts of interest to declare that are relevant to the content of this article.

## Ethics approval

The study was approved by the local ethics committee of the Centre for Advanced Imaging, The University of Queensland.

## Consent to participate

All participants were provided with an overview of the experiment and signed a written consent to participate and filled out an MRI safety form prior to the MR scans.

## References

Amunts K, Malikovic A, Mohlberg H, Schormann T, Zilles K. Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable? *NeuroImage*. 2000:11(1):66–84.

Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HBM, Zilles K. Broca's region revisited: cytoarchitecture and intersubject variability. *J Comp Neurol*. 1999:412(2):319–341.

Amunts K, Zilles K. Architectonic mapping of the human brain beyond Brodmann. *Neuron*. 2015:88(6):1086–1107.

Awad IA, Rosenfeld J, Ahl J, Hahn JF, Luders H. Intractable epilepsy and structural lesions of the brain: mapping, resection strategies, and seizure outcome. *Epilepsia*. 1991:32(2):179–186.

Batista GE, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor Newslett*. 2004:6(1):20–29.

Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on computational learning theory*. New York: ACM; 1992. p. 144–152.

Bradley AP. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recogn*. 1997:30(7):1145–1159.

Breiman L. Random forests. *Mach Learn*. 2001:45(1):5–32.

Buonincontri G, Sawiak SJ. MR fingerprinting with simultaneous B1 estimation. *Magn Reson Med*. 2016:76(4):1127–1135.

Buonincontri G, Schulte RF, Cosottini M, Tosetti M. Spiral MR fingerprinting at 7 T with simultaneous B1 estimation. *Magn Reson Imaging*. 2017:41:1–6.

Cercignani M, Bouyagoub S. Brain microstructure by multi-modal MRI: Is the whole greater than the sum of its parts? *NeuroImage*. 2018:182:117–127.

Chen Y, Jiang Y, Pahwa S, Ma D, Lu L, Twieg MD, Wright KL, Seiberlich N, Griswold MA, Gulani V. MR fingerprinting for rapid quantitative abdominal imaging. *Radiology*. 2016:279(1):278–286.

Cohen-Adad J, Polimeni JR, Helmer KG, Benner T, McNab JA, Wald LL, Rosen BR, Mainero C. T2* mapping and B0 orientation-dependence at 7 T reveal cyto-and myeloarchitecture organization of the human cortex. *NeuroImage*. 2012:60(2):1006–1014.

Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. 1967:13(1):21–27.

Cucurull G, Wagstyl K, Casanova A, Veličković P, Jakobsen E, Drozdzal M, Romero A, Evans A, Bengio Y. Convolutional neural networks for mesh-based parcellation of the cerebral cortex. In: *1st Conference on Medical Imaging with Deep Learning (MIDL)*. Amsterdam, The Netherlands; 2018.

Cunningham P, Delany SJ. k-nearest neighbour classifiers. *Multiple Classifier Syst*. 2007:34(8):1–17.

De Santis S, Assaf Y, Jeurissen B, Jones DK, Roebroeck A. T1 relaxometry of crossing fibres in the human brain. *NeuroImage*. 2016:141:133–142.

Duffau H. *Brain Mapping: From Neural Basis of Cognition to Surgical Applications*. Vienna: Springer; 2012.

Edelman RR, Wielopolski P, Schmitt F. Echo-planar MR imaging. *Radiology*. 1994:192(3):600–612.

Edwards LJ, Kirilina E, Mohammadi S, Weiskopf N. Microstructural imaging of human neocortex in vivo. *NeuroImage*. 2018:182:184–206.

Eggenschwiler F, Kober T, Magill AW, Gruetter R, Marques JP. SA2RAGE: a new sequence for fast B1+ −mapping. *Magn Reson Med*. 2012:67(6):1609–1619.

Eickhoff SB, Yeo BTT, Genon S. Imaging-based parcellations of the human brain. *Nat Rev Neurosci*. 2018:19(11):672–686.

Fan Y, Shen D, Davatzikos C. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. *Med Image Comput Comput Assist Interv*. 2005:8(Pt 1):1–8.

Fischl B. Free Surfer. *NeuroImage*. 2012:62(2):774–781.

Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo BTT, Mohlberg H, Amunts K, Zilles K. Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex*. 2008:18(8):1973–1980.

Focke NK, Helms G, Scheewe S, Pantel PM, Bachmann CG, Dechent P, Ebentheuer J, Mohr A, Paulus W, Trenkwalder C. Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic Parkinson syndrome and healthy controls. *Hum Brain Mapp*. 2011:32(11):1905–1915.

Frahm J, Haase A, Hanicke W, Matthaei D, Bomsdorf H, Helzel T. Chemical-shift selective MR imaging using a whole-body magnet. *Radiology*. 1985:156(2):441–444.

Géron A. *Hands-on machine learning with Scikit-Learn, Keras, and Tensor Flow: concepts, tools, and techniques to build intelligent systems*. Sebastopol, California: O'Reilly Media; 2019.

Geyer S. *The microstructural border between the motor and the cognitive domain in the human cerebral cortex*. Berlin Heidelberg: Springer Science & Business Media; 2012.

Geyer S, Ledberg A, Schleicher A, Kinomura S, Schormann T, Burgel U, Klingberg T, Larsson J, Zilles K, Roland PE. Two different areas within the primary motor cortex of man. *Nature*. 1996:382(6594):805–807.

Geyer S, Schleicher A, Zilles K. The somatosensory cortex of human: cytoarchitecture and regional distributions of receptor-binding sites. *NeuroImage*. 1997:6(1):27–45.

Geyer S, Schleicher A, Zilles K. Areas 3a, 3b, and 1 of human primary somatosensory cortex 1. Microstructural organization and interindividual variability. *NeuroImage*. 1999:10(1):63–83.

Geyer S, Weiss M, Reimann K, Lohmann G, Turner R. Microstructural parcellation of the human cerebral cortex - from Brodmann's post-mortem map to in vivo mapping with high-field magnetic resonance imaging. *Front Hum Neurosci*. 2011:5:19.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, et al. A multi-modal parcellation of human cerebral cortex. *Nature*. 2016:536(7615):171.

Glasser MF, Van Essen DC. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J Neurosci*. 2011:31(32):11597–11616.

Goldberger J, Hinton GE, Roweis ST, Salakhutdinov RR. Neighbourhood components analysis. *Advances in neural information processing systems*. San Mateo, California: Morgan Kaufmann Publishers; 2005. p. 513–520.

Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002:47(6):1202–1210.

Han J, Kamber M, Pei J. *Data mining: concepts and techniques*. 3rd ed. Burlington Massachusetts: Elsevier; 2012. p. 1–703.

Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification. Taipei: Department of Computer Science and Information Engineering, University of National Taiwan; 2003. Technical Report. p. 1–12.

Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell*. 2000:22(1):4–37.

Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*. 2002:17(2):825–841.

Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal*. 2001:5(2):143–156.

Kaufman S, Rosset S, Perlich C, Stitelman O. Leakage in data mining: formulation, detection, and avoidance. *ACM Trans Knowl Discov Data*. 2012:6(4):1–21.

Langkvist M, Karlsson L, Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recogn Lett*. 2014:42:11–24.

Lebel C, Walker L, Leemans A, Phillips L, Beaulieu C. Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*. 2008:40(3):1044–1055.

Ma D, Coppo S, Chen Y, McGivney DF, Jiang Y, Pahwa S, Gulani V, Griswold MA. Slice profile and B1 corrections in 2D magnetic resonance fingerprinting. *Magn Reson Med*. 2017:78(5):1781–1789.

Ma D, Gulani V, Seiberlich N, Liu KC, Sunshine JL, Duerk JL, Griswold MA. Magnetic resonance fingerprinting. *Nature*. 2013:495(7440):187–192.

Magnin B, Mesrob L, Kinkingnehun S, Pelegrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehericy S, Benali H. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*. 2009:51(2):73–83.

Mangeat G, Govindarajan ST, Maineroc C, Cohen-Adad J. Multivariate combination of magnetization transfer, T-2∗ and B0 orientation to study the myelo-architecture of the in vivo human cortex. *NeuroImage*. 2015:119:89–102.

Marques JP, Khabipova D, Gruetter R. Studying cyto and myeloarchitecture of the human cortex at ultra-high field with quantitative imaging: R1, R2∗ and magnetic susceptibility. *NeuroImage*. 2017:147:152–163.

Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele PF, Gruetter R. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T-1-mapping at high field. *NeuroImage*. 2010:49(2):1271–1281.

Meyer-Baese A, Schmid VJ. *Pattern recognition and signal analysis in medical imaging*. San Diego: Elsevier; 2014.

Mohammadi S, Carey D, Dick F, Diedrichsen J, Sereno MI, Reisert M, Callaghan MF, Weiskopf N. Whole-brain in-vivo measurements of the axonal G-ratio in a group of 37 healthy volunteers. *Front Neurosci*. 2015:9:441.

Moinian S, Vegh V, O'Brien K, Reutens D. Magnetic resonance fingerprinting residual signals can disassociate human grey matter regions. *Brain Struct Funct*. 2022:227(1):313–329.

Nieuwenhuys R. The myeloarchitectonic studies on the human cerebral cortex of the Vogt-Vogt school, and their significance for the interpretation of functional neuroimaging data. *Brain Struct Funct*. 2013:218(2):303–352.

Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*. 2012:36(4):1140–1152.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011:12:2825–2830.

Poser BA, Koopmans PJ, Witzel T, Wald LL, Barth M. Three dimensional echo-planar imaging at 7 Tesla. *NeuroImage*. 2010: 51(1):261–266.

Provost FJ, Fawcett T. *Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions*. Menlo Park, California: KDD; 1997. p. 43–48.

Raghu S, Sriraam N. Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. *Expert Syst Appl*. 2018:113:18–32.

Rieger B, Zimmer F, Zapp J, Weingartner S, Schad LR. Magnetic resonance fingerprinting using echo-planar imaging: joint quantification of T-1 and T-2∗ relaxation times. *Magn Reson Med*. 2017:78(5):1724–1733.

Rose SE, Janke Phd AL, Chalk JB. Gray and white matter changes in Alzheimer's disease: a diffusion tensor imaging study. *J Magn Reson Imag*. 2008:27(1):20–26.

Sacchet MD, Prasad G, Foland-Ross LC, Thompson PM, Gotlib IH. Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. *Front Psych*. 2015:6:21.

Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo XN, Holmes AJ, Eickhoff SB, Yeo BTT. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb Cortex*. 2018:28(9):3095–3114.

Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp*. 2002:17(3):143–155.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. 2004:23:S208–S219.

Stuber C, Morawski M, Schafer A, Labadie C, Wahnert M, Leuze C, Streicher M, Barapatre N, Reimann K, Geyer S, et al. Myelin and iron concentration in the human brain: a quantitative study of MRI contrast. *NeuroImage*. 2014:93:95–106.

Tang YC, Zhang YQ, Chawla NV, Krasser S. SVMs modeling for highly imbalanced classification. *IEEE Trans Syst Man Cybern B Cybern*. 2009:39(1):281–288.

Tardif CL, Dinse J, Schäfer A, Turner R, Bazin P-L. Multi-modal surface-based alignment of cortical areas using intra-cortical T1 contrast. In: *International Workshop on Multimodal Brain Image Analysis*. Cham, Switzerland: Springer; 2013. p. 222–232.

Usman K, Rajpoot K. Brain tumor classification from multi-modality MRI using wavelets and machine learning. *Pattern Anal Applic*. 2017:20(3):871–881.

Wahl M, Li YO, Ng J, LaHue SC, Cooper SR, Sherr EH, Mukherjee P. Microstructural correlations of white matter tracts in the human brain. *NeuroImage*. 2010:51(2):531–541.

Wang XZ, Smith K, Hyndman R. Characteristic-based clustering for time series data. *Data Min Knowl Disc*. 2006:13(3):335–364.

Weiskopf N, Mohammadi S, Lutti A, Callaghan MF. Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology. *Curr Opin Neurol*. 2015a:28(4):313–322.

Weiskopf N, Mohammadi S, Lutti A, Callaghan MF. Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology (vol 28, pg 313, 2015). *Curr Opin Neurol*. 2015b: 28(5):547–547.

Weston J, Watkins C. Multi-class support vector machines. Department of Computer Science, Royal Holloway, University of London; 1998. Technical Report CSD-TR-98-04. p. 1–9.

Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recogn*. 2015:48(9):2839–2846.

Yablonskiy DA, Haacke EM. An MRI method for measuring T2 in the presence of static and RF magnetic field inhomogeneities. *Magn Reson Med*. 1997:37(6):872–876.

Yekkehkhany B, Safari A, Homayouni S, Hasanlou M. A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data. *1st Isprs International Conference on Geospatial Information Research*. Hannover, Germany: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; 2014:40(2/W3): 281–285.

Zhang D, Wang Y, Zhou L, Yuan H, Shen D, Neuroimaging A's D, I. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*. 2011:55(3):856–867.