EVALUATING TANGENT SPACES, DISTANCES, AND DEEP LEARNING MODELS TO DEVELOP CLASSIFIERS FOR BRAIN CONNECTIVITY DATA

A Thesis

Submitted to the Faculty

of

Purdue University

by

Michael Wang

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Industrial Engineering

August 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF THESIS APPROVAL

Dr. Joaquín Goñi, Chair

School of Industrial Engineering

Dr. Mario Ventresca

School of Industrial Engineering

Dr. Juan Wachs

School of Industrial Engineering

Approved by:

Dr. Abhijit Deshmukh

Head of the School of Industrial Engineering

This thesis is dedicated to my wife Elizabeth whose unyielding love, enthusiasm, and encouragement have inspired me to pursue and complete this research.

ACKNOWLEDGMENTS

I would like to first express my deepest gratitude to my committee chair, Dr. Joaquín Goñi, whose excitement for discovery and knowledge greatly inspired me. His personal generosity and inclusiveness made my academic experience at Purdue University enjoyable. My appreciation also extends to my colleagues within the CONNplexity Lab. Dr. Kausar Abbas's mentoring and domain expertise were especially helpful throughout the research process. The inspiration to pursue many aspects of this work was due to Duy Duong-Tran's encouragement and advice. Uttara Tipnis also supported me with her expertise in fMRI twin studies. Lastly, the CONNplexity Lab served as a multi-cultural and welcoming research environment to me where we not only shared ideas but also different worldviews.

PREFACE

The basis for this research originally stemmed from my fascination with the brain and its hidden complexity. As publicly available neuroimaging data explode in size, complex algorithms such as deep neural networks are viable to uncover knowledge within this complexity. The potential of using fMRI and functional connectivity for predictive modeling is incredible. I could not have achieved the level of success that I have without a strong support system. First of all, my parents Daming and Jenny, who supported me with love and understanding. And secondly, my CONNplexity Lab colleagues, each of whom has provided patient advice and guidance throughout the research process. Thank you all for your unwavering support.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABSTRACT

Wang, Michael MS, Purdue University, August 2020. Evaluating Tangent Spaces, Distances, and Deep Learning Models to Develop Classifiers for Brain Connectivity Data. Major Professor: Joaquín Goñi.

A better, more optimized processing pipeline for functional connectivity (FC) data will likely accelerate practical advances within the field of neuroimaging. When using correlation-based measures of FC, researchers have recently employed a few data-driven methods to maximize its predictive power. In this study, we apply a few of these post-processing methods in both task, twin, and subject identification problems. First, we employ PCA reconstruction of the original dataset, which has been successfully used to maximize subject-level identifiability. We show there is dataset-dependent optimal PCA reconstruction for task and twin identification. Next, we analyze FCs in their native geometry using tangent space projection with various mean covariance reference matrices. We demonstrate that the tangent projection of the original FCs can drastically increase subject and twin identification rates. For example, the identification rate of 106 MZ twin pairs increased from 0.487 of the original FCs to 0.943 after tangent projection with the logarithmic Euclidean reference matrix. We also use Schaefer's variable parcellation sizes to show that increasing parcellation granularity in general increases twin and subject identification rates. Finally, we show that our custom convolutional neural network classifier achieves an average task identification rate of 0.986, surpassing state-of-the-art results. These post-processing methods are promising for future research in functional connectome predictive modeling and, if optimized further, can likely be extended into clinical applications.

1. INTRODUCTION

The use of functional magnetic resonance imaging (fMRI) has revolutionized our understanding of the human brain, allowing for the modeling of large-scale brain networks in a non-invasive manner [1] [2]. Its capability of high spatial resolution imaging of the brain yields detailed structural and functional connectivity networks. Applications of brain functional connectivity include discovering clinical biomarkers [3], identifying brain fingerprints that allow for the identification of individual subjects on test/retest acquisitions [4], and highlighting task-dependent reconfiguration of brain networks [5], among many others. In this study, we evaluate the effects of processing methods including PCA reconstruction and tangent space projection for classification of functional connectivity data. The processing methods evaluated are optimal reconstruction based on principal components analysis and projections to five different tangent spaces. Classification performance evaluated include test/retest subject identification for each fMRI task, twins identification for each fMRI task, and task identification across unrelated subjects.

1.1 Group vs. Subject Level Approaches in Functional Connectivity

With fMRI, researchers have analyzed functional brain connectivity patterns to identify group-level differences in brain networks among clinical and healthy populations [6]. These brain disorder studies often use statistical tests to uncover significance in the aggregated data. The performance of these tests is usually quantified by the p-values between groups with a certain significance threshold. Group-level analyses result in easier interpretation of the general networks that contribute to the differences between groups. However, by only analyzing at a group-average level, individual variability within-group is largely ignored [7]. Individual variability within functional

connectomes contains identifying features, or fingerprints, that can be used to identify an individual apart from the cohort [4] [8]. Nevertheless, there is a large effect of motion on functional connectivity data and one may argue that subject-specific motion could be partially driving the fingerprinting results. Finn et al. (2015) quelled this concern by developing a metric that captures the movement patterns for each subject per FC scan. These metrics were then used with the same identification pipeline that they used for the subject identification of FC matrices. In the end, identification rates based on movement alone were very low at 2.5% showing that motion noise is not sufficient to identify a subject's fingerprint. In contrast to group-level analyses, the goal of subject classification is to classify each subject into groups based on predetermined labels (e.g., clinical vs. healthy) with supervised learning. The success of this approach and classification problems in general is measured by the classification accuracy. Showing group differences within FCs and prediction with FCs are very different research questions, and in this study, we focus on the latter.

1.2 Traditional Machine Learning Algorithms and their Applications to Brain Connectivity Data

Within traditional connectome predictive modeling methods, proper feature selection of the model is essential to ensure models avoid overfitting and are generalizable to new data [11]. The majority of such studies that use machine learning for fMRI classification use traditional algorithms such as support vector machines (SVM) and regularized linear regression models. Specifically, a review of 77 papers using machine learning on fMRI showed that over half of the articles used SVMs [12]. Support Vector Machines create a line or hyperplane that separates the data into classes. If a clean line or hyperplane cannot be drawn within the dimensionality of the dataset, the data is mapped into a higher dimension, a process known as kernelling. SVMs perform well with smaller and cleaner datasets because they resist overfitting [13].

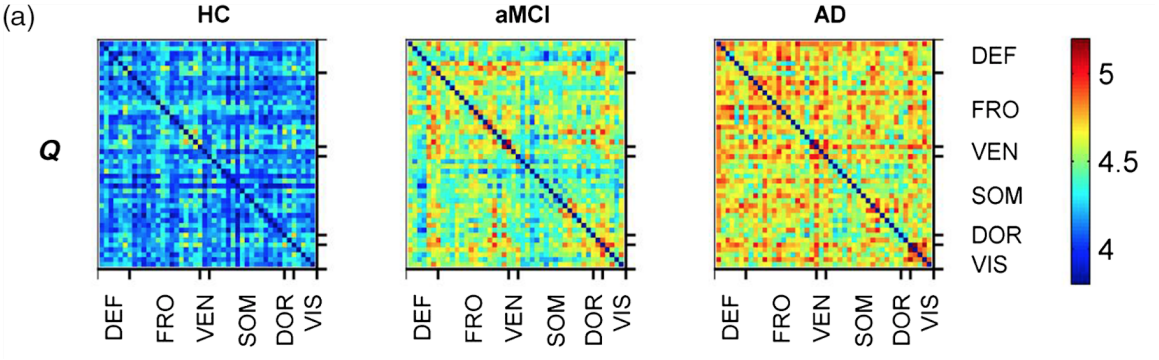


Fig. 1.1. Group level representations of whole-brain functional connectivity for three different cohorts. Namely, healthy controls (HC), amnestic mild cognitive impairment (aMCI), and Alzheimer's disease (AD). Figure adapted from [9].

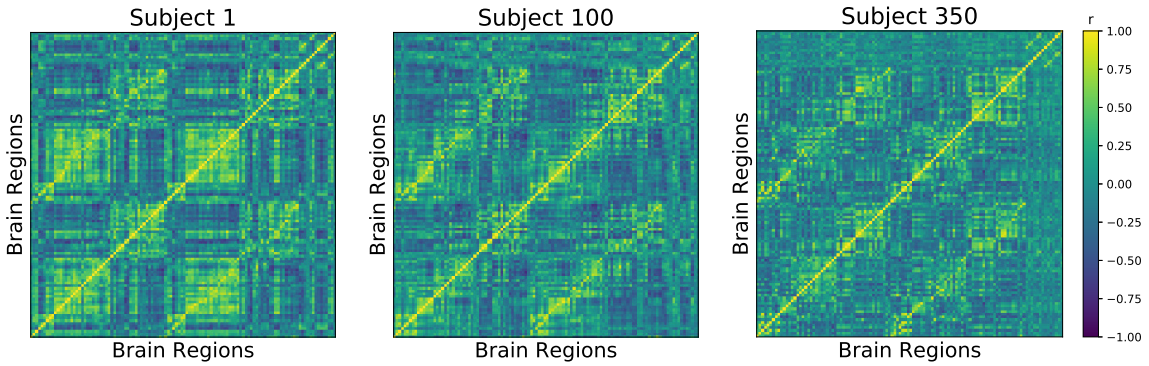


Fig. 1.2. Examples of three whole-brain resting-state functional connectomes (FC) of three individual subjects from HCP dataset using Schaefer's parcellation with 100 brain regions [10].

However, they require careful feature selection, which can prove to be a difficult task with high-dimensional data such as functional connectivity matrices.

Regularization is a method that simplifies a model to prevent overfitting and has been widely used in the field of machine learning [15]. It is especially important in the context of functional connectivity analysis due to the high dimensional space and proneness to overfitting. One of the most popular regularized linear regression models, elastic net, has yielded promising results [16]. Elastic net combines both L1

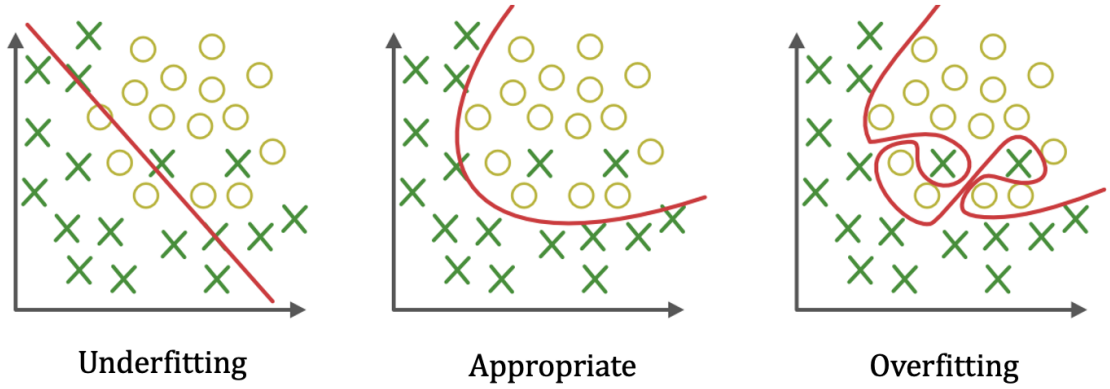


Fig. 1.3. Example of binary class separation with SVM with various levels of fit. Underfitting would produce a classifier that is too simple to model the relationships within the data, whereas overtiffing would produce a classifier that would learn relationships that are not generalizable to new data. Figure adapted from [14]

norm and L2 norm with a hyperparameter. The L1 norm penalizes larger weights more severely and prefers many weights close to 0 whereas the L2 norm encourages a smaller, simpler model. A model with more sparse weights or parameters has a few advantages in functional connectivity analysis. These include mitigation of overfitting in high-dimensional feature spaces and interpretability. In the context of deep learning models, regularization is necessary to prevent the overfitting of weights to the training dataset. However, too strong of a regularization coefficient will cause the model to be too simple and unable to learn the complex relationships within the data. We can apply L2 regularization with the weight decay parameter within the loss function.

1.3 Deep Learning Methods

Within the field of machine learning, a subset of classifiers that has multiple layers of nodes and weights are known as Deep Learning methods. Due to advancements in deep learning methods, it is more viable to train these models on high-dimensional datasets like fMRI than the traditional machine learning methods outlined above [17]. Recently, it has been shown that fully connected deep neural networks can be used successfully for connectome-based classification [18]. Furthermore, inspired by the success of convolutional neural network (CNN) architectures in the famous image classification challenge ImageNet [19], researchers have used CNNs to classify functional connectivity data to great success [20]. The convolutional layer(s) in a CNN apply a small, square filter to its input to perform a dot product calculation as an activation. Repeated convolutions across the input create what is called a feature map, which determines the presence of a particular feature in the input. This can be interpreted as a form of automated feature selection.

Functional connectivity matrices, as a data structure, are conveniently similar to input images in image classification. FC matrices are classically represented as square matrices of size $n \times n$. Here, n is the number of brain regions of a given parcellation - the segmentation of the brain according to an atlas. The value of the (i, j) location within the matrix represents the correlation between BOLD time series of brain regions i and j. As such, these matrices may be interpreted as grayscale images with pixel intensities between -1 and 1. One key difference between traditional images and functional connectomes is that the local features of traditional images do not smoothly translate to connectomes. For example, the clustering of a square 5×5 group of pixels may include an outline of a dog's ear in a traditional image whereas for a functional connectome, this clustering depends on the network structure and ordering of brain regions. Therefore, a strategic implementation of CNN architecture including filter sizes and depths, as well as the ordering of brain regions, is necessary.

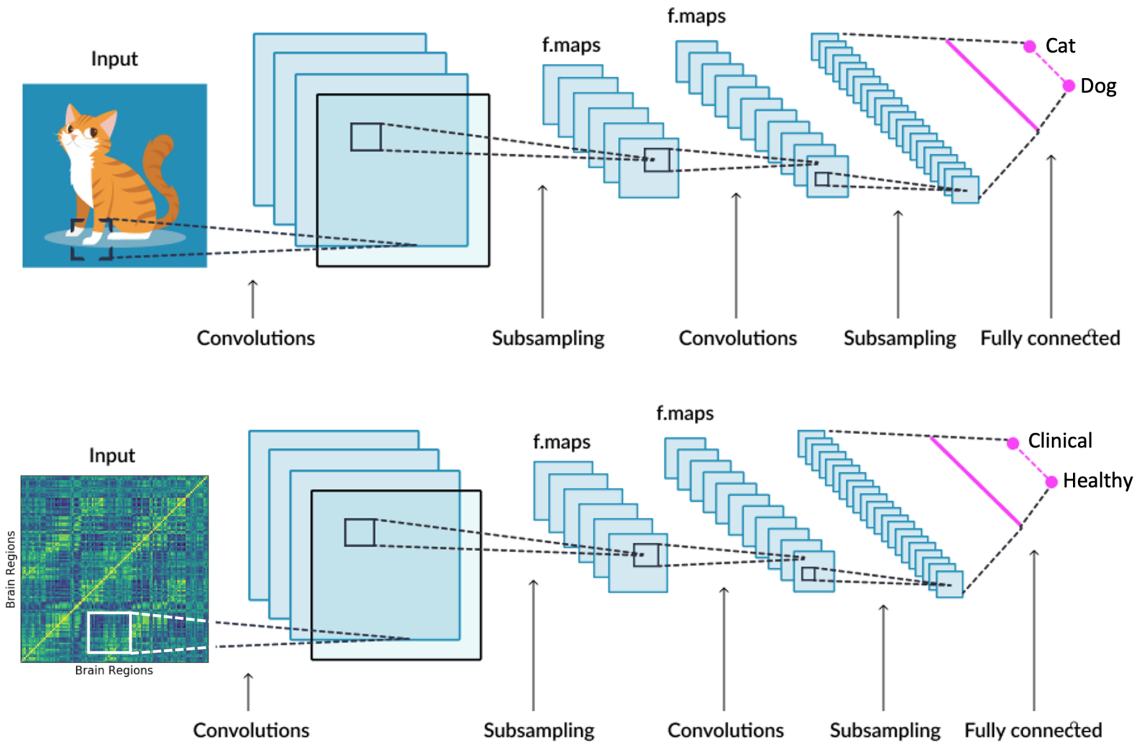


Fig. 1.4. Convolutional neural network (CNN) architecture examples for both binary image classification (top) and binary FC classification (bottom). The input matrix to both CNNs contain values of pixel intensities (top) or pearson correlation values (bottom). A convolution filter moves along the input matrix and creates additional channels based on the filter dimensions. Then, subsampling methods such as pooling are employed to shrink the data into more interpretable features. This process is repeated until the network merges the features into fully connected layer(s). Finally, the fully connected layer(s) produce an output layer the size of the number of output labels. In the image classification example, the two labels are Cat vs. Dog while in the FC classification example, the two labels are Clinical and Healthy. Figure adapted from [21].

The quantity, quality, and size of publicly available neuroimaging datasets have increased significantly in the past few years. Examples of these datasets include the 1000 Functional Connectome Project [22], ADNI2 and ADNI3 [23] with clinical

fMRI data for Alzheimer's disease, and the 1200 Subjects Release of the Human Connectome Project [24] [25] [26]. Although the additional data facilitates deep learning classification, these neuroimaging datasets still pale in comparison to the size of image recognition datasets. fMRI scans are expensive to collect and, consequently, fMRI studies have at most a few thousand training examples. Compared to training on the 1.2 million instances used in ImageNet, training on the smaller neuroimaging datasets can often lead to overfitting. Therefore, it is important to take rigorous steps for regularization and construct a relatively simple convolutional neural network architecture. Nevertheless, deep learning methods have already been used to classify autism spectrum disorder (ASD) [27], amnestic Mild Cognitive Impairment (aMCI) [20], Alzheimer's disease (AD) [28], cognitive impairment [29], and schizophrenia [30].

1.4 Clinical Applications of fMRI

Since its inception in the early 1990s, fMRI has reshaped the neuroimaging research community. It has allowed for in-vivo characterization of whole brain functional connectomes in humans [31], leading to the discovery of several critical brain networks implicated in schizophrenia, attention deficit hyperactivity disorder, autism, and Alzheimer's disease (AD) [6]. Despite its popularity in academia, fMRI has seen rather meager application in the clinical environment [32]. Abnormalities in brain networks identified in clinical research datasets have not translated into practical diagnostic tools for use on individual patients. One major reason for this disconnect is the low within-subject reliability and between-subject differentiability ('fingerprinting') for subject-level prediction or diagnosis [33] [4]. fMRI data is contaminated with noise including motion and scanner artifacts. Slight changes in subject positioning can easily contaminate the fMRI signals. Because of the high spatial resolution of fMRI, even the slightest of movements can trigger significant artifacts within the data. There has been a great deal of research done on correction methods in frequency fil-

ters to mitigate this issue [34]. Such artifacts lead to high inter-subject variability which hinders the performance of subject-level classification.

Recent work in differential identifiability has shown that individuals can be reasonably distinguished from each other using FC. The performance of distinguishing individuals is measured by identification rate [4], perfect separability rate [35], or differential identifiability [33]. Differential identifiability can be improved with increased scan length and multiple scan sessions [36]. Consequently, studies have shown that typical 6-minute fMRI acquisitions do not have adequate reliability, and likely causes significant issues in using clinical fMRI datasets for subject-level modeling. Further, a recent study has demonstrated that the use of multiple connectomes across sessions additional tasks improves predictive power [37]. However, extended fMRI acquisitions and the collection of task-based fMRI is often infeasible for clinical populations who often have trouble completing tasks and enduring a long acquisition session [11]. Additionally, it has also been shown that the presence of neurologic or psychiatric conditions makes differentiating between subjects more difficult [38]. To address these issues, Amico and Goni proposed the differential identifiability framework [33], which is a PCA-based denoising algorithm to uncover fingerprints in functional connectomes. By separating the data into linearly independent principal components by decreasing order of explained variance, researchers can then reconstruct the original dataset with a subset of these principal components. In this study, subject-level identifiability was maximized with a subset of the components that explain the most variance. Their results suggest that this cleaning method can remove scanner and motion artifacts while preserving the defining features on the individual level. They demonstrated improvements in FC fingerprinting beyond what could be achieved by increasing scan length [33]. Similar improvements have also been shown in inter-scanner and multi-site identifiability where subjects are scanned at different locations with different scanners [39]. The differential identifiability framework shows that improving across-session reliability of functional connectomes (FCs) also improves reliability of derived

network measures [8]. A similar improvement in FC fingerprinting was also found in data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [11].

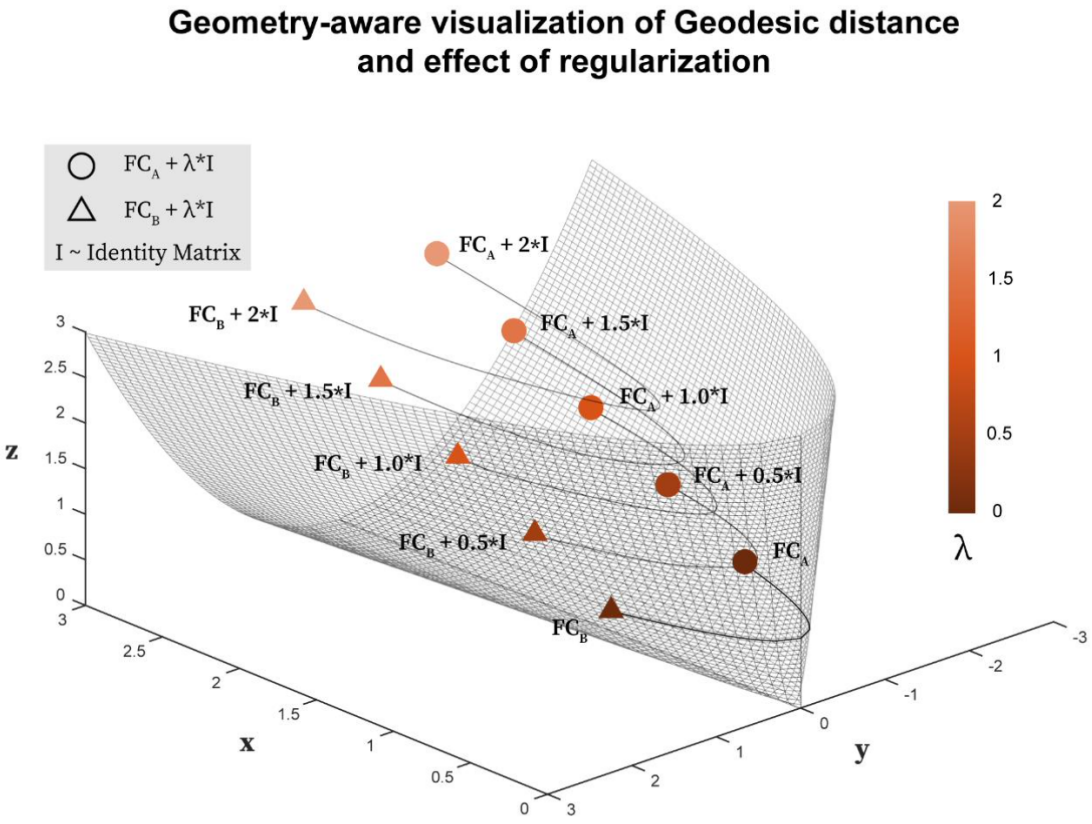


Fig. 1.5. Positive-definite (full rank) functional connectomes (FCs) reside in the interior of the positive semidefinite cone pictured above. Because they are correlation matrices, they do not naturally form a Euclidean space. The surface of this cone is comprised of all the rank-deficient positive semi-definite FCs (having at least one 0 eigenvalue). Abbas et al. showed that different magnitudes of λ for regularization offset the FCs within the positive semi-definite cone. [40]

There has also been a push to use geometry-aware analysis methods on functional connectivity data [41]. The reason is that non-geometry-aware similarities or distance measurements such as Person's correlation coefficient (r or $1 - r$) assume that the

underlying FC data lie in the Euclidean space. Instead, FC matrices computed by Pearson correlating BOLD time series data lie on a non-linear manifold called the positive semidefinite cone whose geometry is non-Euclidean [41]. Mathematical operations that are commonly used in a classification algorithm such as addition and multiplication do not translate to this space. By employing the geometry-aware projections outlined in Section 2.4.3, we can compare and classify functional connectomes properly in their natural space.

1.5 Twin Studies and Behavior Genetics

Phenotypic variation in humans can be physical (e.g., height or weight), physiological (e.g., blood pressure or brain volume), cognitive (e.g., memory), and psychological (e.g., susceptibility to depression). The debate over the source of this variation being from genetics, environment, or a mixture of both is known as the nature versus nurture debate. The classical twin study is a practical and powerful family design to uncover sources of phenotypic variation. There are two types of twins: identical (monozygotic; MZ) and non-identical (dizygotic; DZ) twins [42]. Monozygotic twins form when the zygote (fertilized egg) divides, usually within 2 weeks of fertilization. Both monogyzotic twin individuals originate from the same sperm and egg and share identical genetic information. Dizygotic twins form when more than one egg is released by the ovaries at the same time each of the eggs is fertilized by a separate sperm cell. Dizygotic twins, then, share the same amount of genetic information as if they were siblings - roughly 50% [42]. However, they do have shared prenatal environments, as they were conceived at the same time and shared the womb.

1.6 Aims and Hypotheses

The goal of this research is to investigate better data pipelines to process and project functional connectivity data for predictive modeling. To do so, we investigate the effects of two established data-based frameworks: the differential identifiability

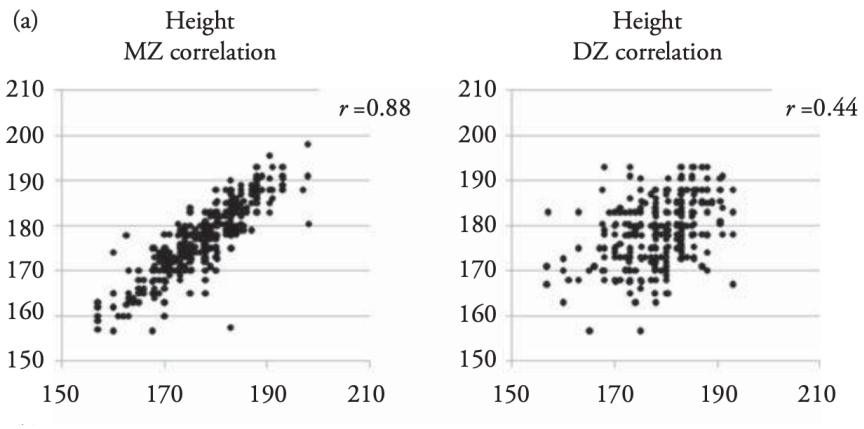


Fig. 1.6. Scatter plots of monozygotic (MZ) and digyzotic (DZ) twin pair for height (cm) in males. Altogether, correlations for MZ and DZ suggest a high heritability for this trait. Figure adapted from [42].

framework [33] and the use of geodesic distance [41]. It has been shown that there exist task-evoked network architectures of the human brain and that different areas of the brain are more or less involved in specific tasks [43]. Our hypothesis is that these ask-specific functional connectivity networks are identifiable within each subject's functional connectome depending on the task being performed. We hypothesize that aforementioned frameworks increase the power of the underlying networks and lead to higher performance. Additionally, Amico and Goñi and Finn demonstrated that there exists a subject-level fingerprint within functional connectivity data. We hypothesize that this fingerprint also extends to twin studies where monozygotic and dizygotic twins share portions of the subject-level fingerprint due to the sharing of genetic material. Again, we aim to adapt both the differential identifiability framework and the use of geodesic distance to maximize twin identifiability of functional connectivity data.

1.7 Optimizing the Processing of Functional Connectomes Based on Classification Performance

Analysis of fMRI data is held back by the lack of a common cleaning and analysis pipeline within the neuroimaging community. Predictive models that use functional connectivity typically follow three main steps: parcellating the brain into regions, estimating the interactions between these defined regions, and feeding the data into a classifier for prediction [16]. In this study, we attempt to address all three steps of this pipeline by determining parcellation effects, data cleaning transformations, and a robust deep learning classifier. The goal of this research is to apply these data pipeline optimizations to maximize classification accuracy of functional connectivity data.

With respect to parcellations, finer-grain segmentations yield higher spatial resolution and more functional connectivity information. However, the dimensionality of the features scales exponentially and can easily lead to overfitting. We test the various parcellations granularities of Schaefer's brain atlas [10].

Next, to delineate the effects of data cleaning and processing, we use the minimally processed functional connectomes as the control. We also apply PCA reconstruction with varied levels of principal components (PCs). In within-task subject identifiability, it has been shown that PCA reconstruction with slightly less than 50% of PCs results in optimal performance [33]. It is thought that the functional connectivity included in base function are represented by the first few principal components that explain the most variance, and as the explained variance decreases, task and subject-level variability are introduced. Finally, in the last 50% of PCs, scanner artifacts and motion noise are likely introduced. For the other data cleaning method, we test tangent space projection of the functional connectomes with various reference matrices. These reference matrices can be as simple as the group mean of the entire dataset and serve as an anchor into the tangent space. Since PCA reconstruction is a linear combination of principal components, it cannot be used simultaneously with tangent

space projection. Finally, for classification, we use both a distance-based K-Nearest Neighbors classifier and a custom convolutional neural network (CNN) classifier. The purpose of the K-Nearest Neighbors classifier is to serve as a proxy model in its simplicity to be able to test many hyperparameters in both our proposed data cleaning methods. Once the best parameters are identified, the input data from all three pipelines are tested with the CNN model. The CNN creates a more complex model with nonlinearities throughout the training of many weights within the hidden layers. Because of the smaller size of the HCP dataset, we employ a simpler CNN model architecture with proper regularization methods. We demonstrate that fine-grain parcellations boost performance in the CNN. We also show that both PCA reconstruction and tangent space projection clean the data for task classification, but with some caveats in the latter data cleaning method. Finally, we present state-of-the-art results with the optimized data pipeline in the classification of functional connectome data.

2. METHODS

2.1 Human Connectome Project Dataset

The main dataset analyzed during this study is the Human Connectome Project 1200 Subjects Release and is available online in the HCP repository established by the National Institute of Health [44]. Per HCP protocol, all subjects gave written, informed consent to the Human Connectome Project consortium. Of the 1200 subjects, we use a subset of 424 unrelated subjects. This subset, as specified by the HCP, ensures that each individual is not related to any other subject within the subset. This criterion was included in the early stages of this study to eliminate any genetic confounding factors within the study. All subjects gave written informed consent to the Human Connectome Project consortium.

The HCP dataset consists of scans of the subjects at rest (resting state) and while performing seven separate tasks: gambling, relational, social, working memory, motor, language, and emotion. These 7 tasks and resting state represent 8 total output labels in our predictive models. For each subject and task, there were two fMRI data acquisition sessions, which function as replicates and are denoted by test and retest. To avoid confounds due to the directionality of acquisition, the RL and LR scans were randomly assigned to test and retest [45]. The working memory, gambling, and motor tasks were completed on the first day of acquisition while the others were completed on the second day. The HCP scanning protocol was approved by the local Institutional Review Board at Washington University in St. Louis. All experiments were performed in accordance with relevant guidelines and regulations.

2.1.1 Twin Subset

The HCP 1200 Subject Release also includes 116 monozygotic twin pairs, who share 100% of their genetic structure and a common environment, and 76 dizygotic twin pairs, who share 50% of their genetic structure and a common environment. The data structure is identical to the HCP dataset as described above as it is a subset of the 1200 subjects. There are test and retest scans for each of the 8 tasks and for each of the twin pair individuals.

2.1.2 Data Preprocessing

The HCP functional minimal preprocessing pipeline was used [24] which includes artifact removal, motion correction, and registration to the standard space. Further processing was done for both resting state and task fMRI data outlined by Amico, Arenas, and Goñi [46]. This pipeline includes spatial preprocessing, in both volumetric and grayordinate forms with motion correction [47], weak high-pass temporal filtering for slow drift removal, MELODIC ICA [48] applied to volumetric data, and scanner artifacts identified and removed with FIX [49]. For the resting-state fMRI data, global gray matter signal was regressed out of the time series [50]. A band-pass first-order Butterworth filter in forward and reverse directions [0.001 Hz, 0.08 Hz] was applied [50], and the voxel time courses were z-scored and averaged per brain region to exclude outlier time points outside of 3 standard deviations from the mean [25]. For task fMRI data, we applied the same above mentioned steps with a less restrictive range for the band-pass filter [0.001 Hz, 0.25 Hz].

2.1.3 Estimation of Subject-Level Functional Connectomes

For each task fMRI session of each subject, a functional connectivity matrix (i.e. the functional connectome) was obtained by computing Pearson's correlation coefficients between pairs of time courses of each brain region. These functional connec-

tomes are symmetric of size $n \times n$ where n is the number of brain regions in the given parcellation. The functional connectomes were neither thresholded nor binarized and kept as correlation coefficients of values between -1 and 1. A value close to 1 signifies strong functional connectivity between the brain regions as their BOLD signal activity are closely correlated with each other. A value close to 0 means that there is very little functional connectivity between the two brain regions. Finally, a value close to -1 shows a strong negative correlation, but that one region tends to be activated while the other one is at rest, and vice versa. The construction of functional connectomes was done for all variable sized Schaefer parcellations with subcortical regions. An example of an individual functional connectome from the Schaefer 100 brain parcellation is shown in Figure 2.1.3. The resulting individual functional connectivity matrices are automatically ordered (rows and columns) according to seven resting-state cortical subnetworks (RSNs) as proposed by Yeo et al. (2011). For completeness, an eighth subnetwork including the 14 HCP subcortical regions was added (as analogously done in a recent paper; Amico et al., 2018) [45].

2.2 Task Classification

Task-based fMRI is regularly used to identify brain regions that are functionally involved in the execution of a specific task, while resting state fMRI is more often used to highlight the underlying brain networks. From Task-based fMRI, we can use the information about which brain regions activate during which tasks to help understand how the brain is organized in resting state fMRI [51]. We aim to classify functional connectomes into each of the 8 different task labels (7 tasks and resting state) from the Human Connectome Project dataset. Examples of each of the 7 task FCs and resting state FC are shown in Figure 2.1.3. These example FCs all belong to Subject 1 with Schaefer's 100 brain region parcellation plus 14 subcortical regions. Previous work has shown that there is much subject-level variation in functional connectivity [46] [4]. To consistently classify individual FCs into the correct task

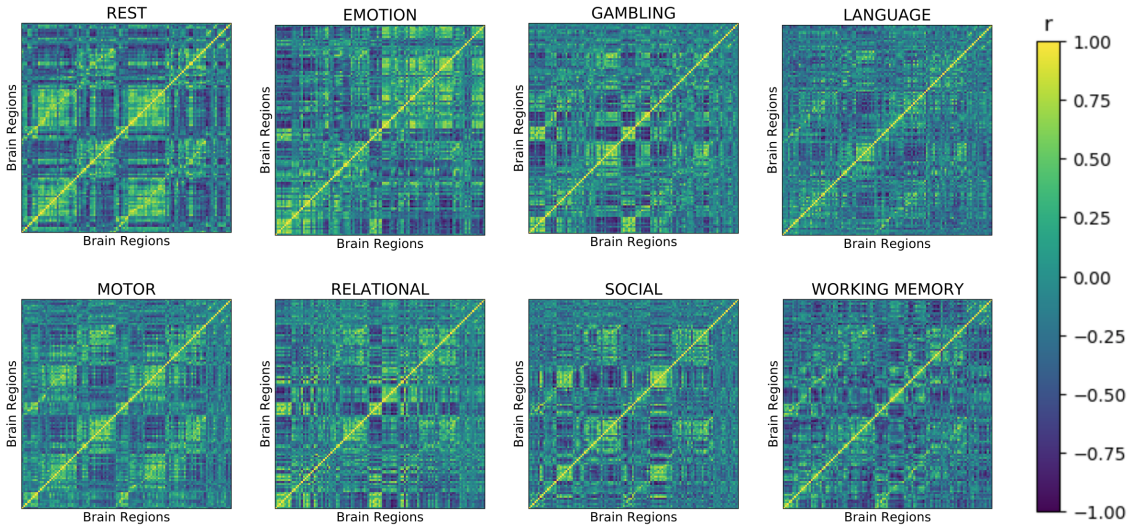


Fig. 2.1. Example of 8 functional connectomes of Subject 1, one per fMRI task with Schaefer's parcellation with 100 brain regions and 14 subcortical regions.

labels, the classifier must overcome subject-level differences and instead focus on task-specific variability in the functional connectivity data. Previous studies have used the HCP dataset to differentiate gambling and relational tasks (92% accuracy) [52], motor and working memory tasks (95.9% accuracy) [53], and classify into all 7 HCP tasks (93.7% accuracy) [18]. They use a variety of classifiers including SVMs [52], random forests [53], and deep neural networks [18]. In this study, we will use both a distance-based K-Nearest Neighbor classifier and a Convolutional Neural Network classifier. Details on these algorithms will be discussed in the following sections.

2.3 Twin Identification

There is a strong individual 'fingerprint' within a subject's functional connectome [4]. Whether this fingerprint is determined by genetics, environment, or a mixture of both is a question of nature versus nurture. A recent study has classified zygocity

(monozygotic vs. dizygotic) in twin pairs with 92.23% accuracy [54]. Another study has attempted to identify twin pairs with resting state fMRI data with 64% and 22% accuracy for monozygotic ($n = 25$) and dizygotic ($n = 25$) twins pairs. We also assess the identification problem on twin pairs with the Human Connectome Project's subset of monozygotic and dizygotic twins in an attempt to uncover the role of genetics in subject-level identifiability. In addition to resting state fMRI, we also examine the predictive power of task-based fMRI for twin pair identification. Each individual in this dataset has a test and retest scan for each of the 7 tasks and resting state. We employ post-processing techniques that are detailed in the following section to optimize performance of our predictions with a distance-based classifier.

2.4 Functional Connectivity Parcellations and Post-Processing Methods

To best solve the aforementioned classification problems, we employ a variety of post-processing methods in an attempt to increase performance. These methods include increasing or decreasing granularity of brain parcellations, adapting the differential identifiability framework to reconstruct functional connectomes (FCs), and projecting FCs into the tangent space for geometry-aware analysis. We implement these post-processing methods for all analyses with the goal to recommend best-practice post-processing methods for functional connectivity data.

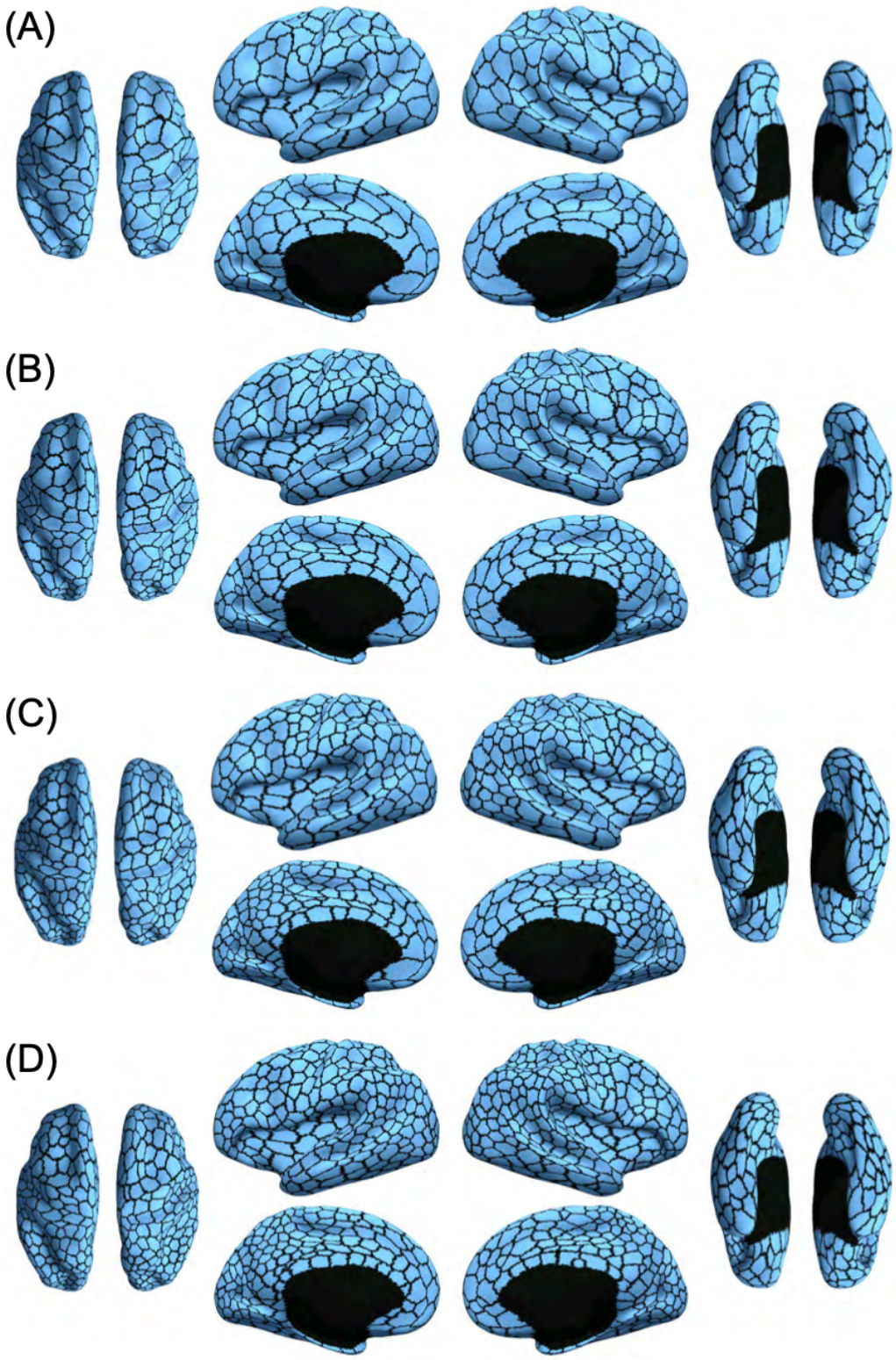


Fig. 2.2. Schaefer's cerebral cortex parcellations of the human brain with (A) 400 regions (B) 600 regions (C) 800 regions and (D) 1000 regions.

2.4.1 Brain Parcellations

In this study, we use Schaefer's parcellation with variable brain regions. These were each appended with an additional 14 subcortical regions for completeness [45]. Schaefer's parcellation is a gradient-weighted Markov Random Field (gwMRF) model that integrates both local gradient and global similarity approaches. One key advantage to Schaefer's proposed parcellation is that it has variable degrees of segmentation, from 100-1000 brain region parcellations in intervals of 100 regions. In this study, we use the 100-500 brain region parcellations due to the exponential increase in feature space dimensionality of high resolution parcellations.

2.4.2 Differential Identifiabilty Framework Adapted for Classification

After segmenting the brain into regions of interest based on the aforementioned parcellations and computing the minimally processed functional connectomes, we adapt the differential identifiability framework to clean the data from noise and artifacts. Specifically, we reconstruct the original dataset with a subset of its principal components (PCs). Principal component analysis computes the eigenvectors, or principal axes, of the dataset and sorts them by their eigenvalues in decreasing order of explained variance. Each principal component is a linear combination of the inputs (in this case, of functional connectomes). The data is centered by subtracting its mean and is then projected onto these principal axes to yield principal components (PCs). While PCA is typically used as a technique for dimensionality reduction and/or feature selection, the PCs can be linearly combined to reconstruct the dataset. The purpose of PCA reconstruction is to clean the data from noise that is contained within the principal components with little explained variance. This method has been used in optimizing subject-level differential identifiability [33] to great success. Differential identifiability assesses the strength of the individual fingerprint of a subject's connectome. Each entry in the identifiability matrix i, j represents the correlation between connectome of subject i test and subject j retest. Thus, the correlation

coefficients between replicates of the same subject is denoted as I_{self} and reside on the main diagonal. The non-diagonal elements are the correlations between a run of a subject i and subject j where i and j are different subjects (I_{others}). Differential identifiability is then defined as

$$I_{diff} = (I_{self} - I_{others}) \times 100 \tag{2.1}$$

The differential identifiability framework and our adaptation requires at least two samples of each label to calculate identification rate. In the context of subject identification, these are designated as test and retest scans. For twin identification, there are four total samples for each twin pair (test and retest for each individual). To maximize differential identifiability, there exists an optimal number of PCs to include in reconstruction. Amico and Goñi. (2018) found that including approximately 40-50% of the total principal components yielded the maximum differential subject identifiability [33]. In this study, we use this finding as inspiration to maximize task identifiability; that is, to differentiate one task functional connectome from other tasks. We vectorize the upper triangular portion of each individual FC and aggregate them into one large matrix of size $m \times n$ where m is the number of edges in the flattened FC and n is the total number of FCs. PCA decomposition on this large matrix is performed and separated into orthogonal eigenvectors W sorted by decreasing order of explained variance. We subtract the mean and project the centered data onto eigenvectors to create the principal components Z. Finally, we reconstruct the large matrix of size $m \times n$ with a subset of k principal components and adding the mean back into the matrix as shown in equation 2.2. Please refer to a visualization of this process shown in Figure 2.4.2 adapted from Amico and Goñi.

$$\underset{m\times n}{\mathrm{X_r}} = \underset{m\times k}{\mathrm{W}} \times \underset{k\times n}{\mathrm{Z'}} + \mu \tag{2.2}$$

We incrementally test the proportion of PCs to include in reconstruction to maximize classification accuracy. Due to the large search space of optimal reconstruction, using an expensive model (i.e. convolutional neural network) to determine the best

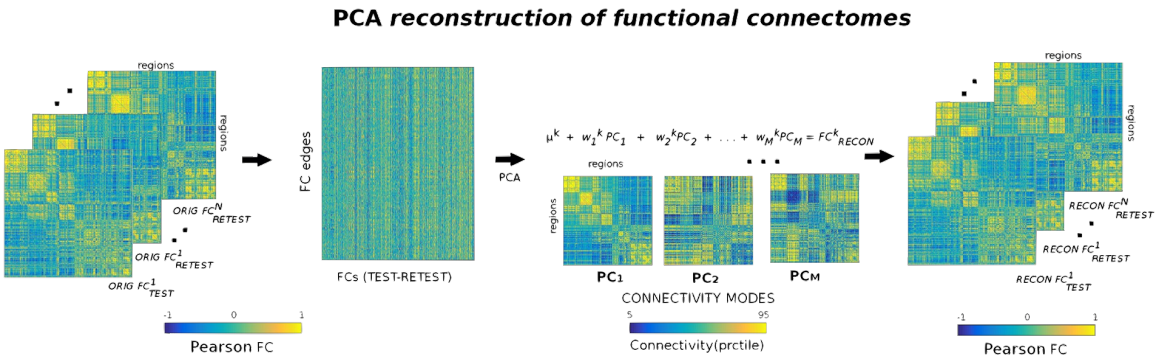


Fig. 2.3. Example of PCA reconstruction of the original dataset of functional connectivity (FC) matrices. The upper triangular portion of the original FCs are first vectorized and aggregated into a single large matrix. Then, the eigenvalues and eigenvectors of this matrix are produced through PCA decomposition. Finally, the original dataset is reconstructed with a subset of principal components and reshaped into square, individual FC matrices.

proportion of PCs is infeasible. Rather, we use a simple distance-based classifier as a stand-in and use the best configuration in the more complex and expensive CNN model.

2.4.3 Tangent Space Projection

In many fMRI analyses, it is common to encounter the concept of quantifying similarity between functional connectomes. The classic and intuitive approach is to vectorize the upper triangular of the FC matrix excluding the main diagonal into a one-dimensional vector and compute the correlation distance between the vectors. Then, if two connectomes are similar their flattened matrices would yield a relatively high correlation coefficient. The correlation distance approach has produced impressive results in applications such as participant identification through 'fingerprinting' [4] [33]. However, there exists another way to analyze functional connectomes that, interestingly, preserves their native geometry.

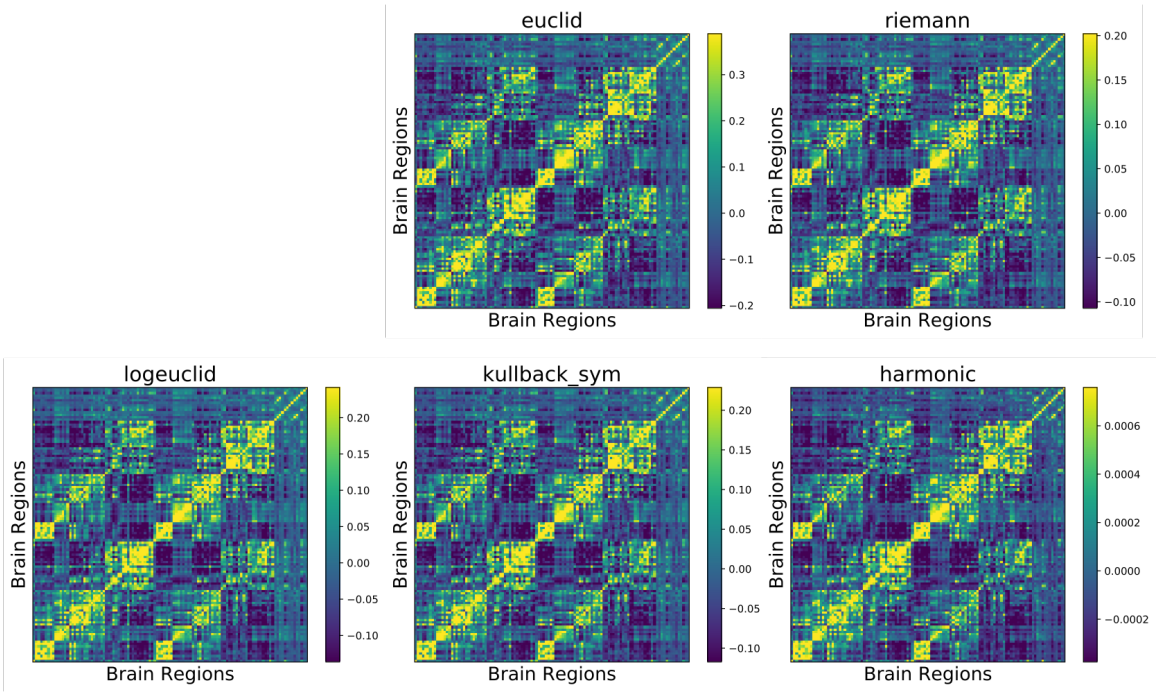


Fig. 2.4. Reference matrices calculated from the Human Connectome Project ($n = 424$) with Schaefer's 100 brain region parcellation. Note that the scales are different for each reference matrix.

Table 2.1.
Reference matrices C_g for tangent space projection. Table adapted from [16].

Reference	**Equation**
Euclidean	$\frac{1}{N}\sum_i C_i$
Harmonic	$(\frac{1}{N}\sum_i C_i^{-1})^{-1}$
LogEuclid	$\exp_m(\frac{1}{N}\sum_i \log_m C_i)$
Kullback	$C_e^{\frac{1}{2}}(C_e^{-\frac{1}{2}} C_h C_e^{-\frac{1}{2}})^{\alpha} C_e^{\frac{1}{2}}$
Riemmanian	$\arg\min(\sum_i \delta_R(C_e C_i)^2)$

Matrices computed by correlating time series data are positive definite and lie within a non-linear surface called the positive semidefinite cone. Suppose we are given a simple example of an FC matrix with two brain regions:

$$\begin{bmatrix} x & z \\ z & y \end{bmatrix}$$

FC matrices always satisfy $x \geq 0$ and $y \geq 0$ because their self-correlation coefficient is always 1 and $xy - z^2 \geq 0$; therefore, they are positive definite. Because of this, their geometry is non-Euclidean and classic distance measures such as correlation distance or euclidean distance cannot be used if geometry is to be preserved. The operations that classification algorithms use such as addition, subtraction, multiplication, and division operate within the Euclidean space. To preserve geometry and use such classifiers, the Pearson correlated FC matrices must be first projected into the tangent space [16]. Geodesic distance, a non-Euclidean distance metric that accounts for the manifold on which the data lies, improves participant identification compared to the Pearson correlation distance metric [41]. Geodesic distance, as defined by this study, is the shortest path between two FC matrices along the semi-positive definite manifold. As a result, a small euclidean distance does not always imply a small geodesic distance, and vice versa.

For most state-of-the-art machine learning algorithms such as deep neural networks and SVMs, it is infeasible to operate on the positive semidefinite cone. Instead of converting all Euclidean operations within these classifiers to geodesic distance-based operations, we can project the FC matrices into the tangent space. From there, the projected FC matrices can be treated as Euclidean objects and classification algorithms can be simply applied as usual in the tangent space. Given a covariance matrix C, in our case an FC matrix, we can project it into the tangent space with a reference matrix C_g using the following equation:

$$\hat{C} = \log_m \left(C_g^{-\frac{1}{2}} C C_g^{-\frac{1}{2}} \right) \tag{2.3}$$

Reference matrix C_g serves as the group anchoring point in the tangent space. There are many different methods to calculate a reference matrix. It is unclear which group mean estimate produces the best results, so we test all the reference matrices shown in Table 2.1. Visualizations of these reference matrices in respect to the Human Connectome Project dataset with Schaefer's 100 region parcellation can be found in Figure 2.4.3. Examples of the functional connectomes after tangent space projection of a constant subject with resting state FC are shown in figure 2.4.3. Tangent projection potentially aids in classification by transforming the data into a more representative space. All tangent space projections were performed with the Python package `pyriemann` developed by Dr. Alexandre Barachant [55].

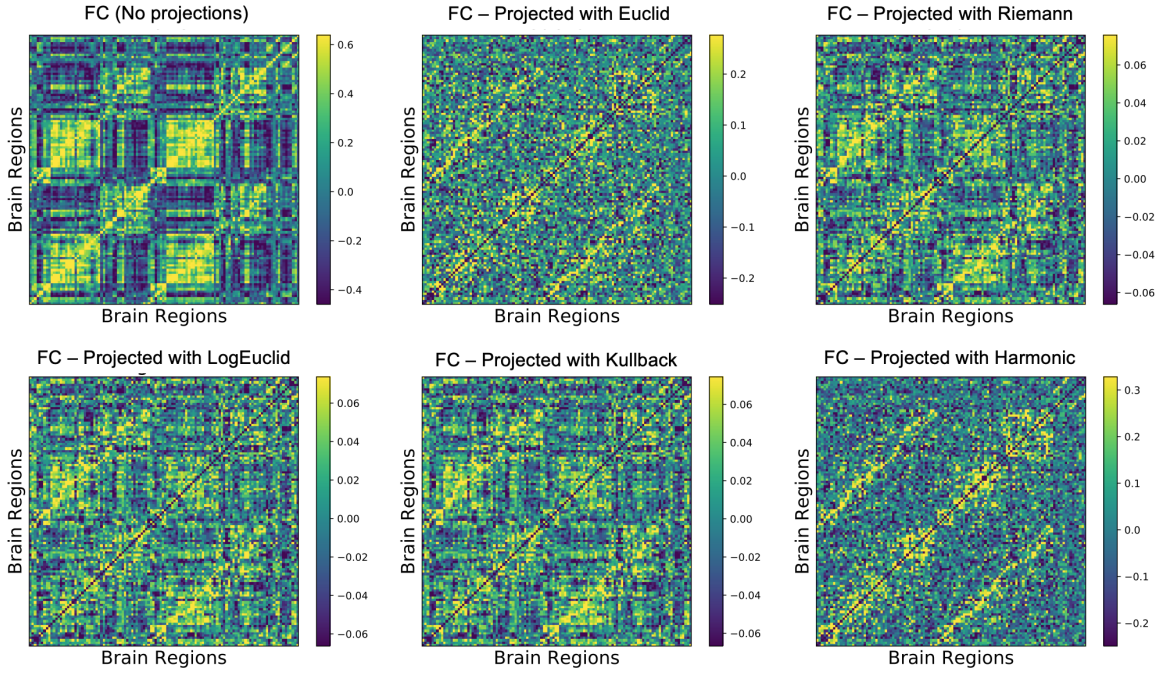


Fig. 2.5. Tangent projected FCs (by using different C_g) matrices) of Subject 1 resting state functional connectome with and without the five reference matrices.

2.5 K-Nearest Neighbors Classifier

K-Nearest Neighbors is a simple machine learning algorithm that compares the distances between the test sample and instances within the training dataset for classification. The output is a class label. The test sample is classified by a plurality vote decided by the class most common among its k nearest neighbors. The value of k is typically a small, positive integer. In the case where $k = 1$, the predicted label is simply determined by the label of the closest instance within the training set. This classification method is a type of instance-based learning where there is no training of a model and the computational expense is deferred to test sample evaluation. The best choice of the parameter k largely depends on the dataset. In general, a larger value of k reduces the effect of noise on the classifier, but at the expense of computation and less distinct boundaries between classes.

In many fMRI analyses, the common method of comparing similarity between FCs is to flatten each FC matrix into one-dimensional vectors and compute the correlation distance between the vectors. Two connectomes that are similar in the underlying brain activity would yield a relatively lower distance. This can be seen in studies that achieve impressive results in applications such as participant identification [4] [33]. Here, we do the same flattening procedure and distance analysis with the K-nearest neighbor classifier but with with some adjustments. These are various types of similarity metrics and the value of k. For the similarity metrics, we first test euclidean distance, cosine similarity, and correlation coefficients with Python's `scikit-learn` package and its KNN function.

Of all similarity measures, Euclidean distance is the most common and often serves as a homonym of 'distance'; it is the basis of many measures of similarity. Euclidean distance calculates the square root of the sum of squared differences between two vectors. Given two vectors x and y, euclidean distance d is calculated by

$$d(x, y) = \sqrt{\sum_{i}^{n} (x_i - y_i)^2} \tag{2.4}$$

Cosine similarity is the cosine of the angle between two n-dimensional vectors in an n-dimensional space [56]. It is computed as the dot product of two vectors divided by the product their Euclidean norms. Let x and y be two vectors for comparison, then the cosine similarity S is given by

$$S(x, y) = \frac{xy}{\|x\|\|y\|} \tag{2.5}$$

The last similarity metric used is the correlation between two vectors. Again, let x and y be two vectors for comparison. The correlation coefficient r between two vectors x and y is given by

$$r(x, y) = \frac{\frac{1}{n}\sum_i x_i y_i - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{2.6}$$

where values for μ are the means of the respective vectors and σ represents the standard deviations of the vectors x and y. The covariance of x and y is represented by the expression in the numerator. From there, we must standardize the covariance to unit variance and therefore we divide by the product of the two vectors' standard deviation. A higher correlation value means that the two vectors are more similar to each other.

2.5.1 Selection of k IN KNN

In KNN, determining a value for k is not very straightforward. It is highly dependent on the dataset and the number of features. In general, a small value of k means that noise will more strongly affect the classification but computation is efficient. This is because low values of k mean less comparisons to the training set are necessary. A larger value of k reduces the noise in the dataset but can make the classifier very computationally expensive. A general rule of thumb in the machine learning community is to try a value of $k = \sqrt{n}$ where n is the number of training samples. In the domain of functional connectivity, the dimensionality is quite large as each correlation between brain regions is a single feature. For task classification,

we tested values of $k = \{1, 5, 30\}$. For twin identification, because there were only 2 examples of each twin individual per task, a value of $k = 1$ was used. All analyses were conducted with Python and scikit-learn class `KNeighborsClassifier`.

2.6 Deep Learning Classifier

For datasets with many features and dimensions, deep learning models can learn a much more complex output function than can traditional machine learning methods such as the KNN outlined above [17]. Deep learning is a subset of machine learning which learns complex relationships of labeled data incrementally through its hidden layer architecture. The hidden layers contain adjustable weights in the training process. Deep neural networks contain an input layer, an output layer, and at least one hidden layer in between. They employ nonlinear activation functions to learn complex relationships between the input and output layers. Convolutional Neural Networks (CNN) are one of the variants of deep neural networks and are commonly used computer vision applications. It gets its name from its hidden layers that typically consist of convolutional layers, pooling layers, and fully connected layers. Instead of using the normal activation functions defined above, convolution and pooling functions are also used as activation functions. The success of convolutional neural network (CNN) architectures in the famous image classification challenge ImageNet [19], researchers have used CNNs to classify functional connectivity data to great success [20]. A convolutional layer in a CNN applies a small, square filter to the input that performs the dot product as an activation. Repeated convolutions across the input create what is called a feature map, which determines the presence of a particular feature in the input. Benefits of convolution layers include automatic feature learning, weight sharing (decreasing the number of learnable weights significantly), and performance on image recognition tasks. CNNs are sophisticated enough to learn the complex processes of the brain, with the convolution layers, nonlinear transformations, and fully connected layers. However, because of the large number of weights, deep learning methods re-

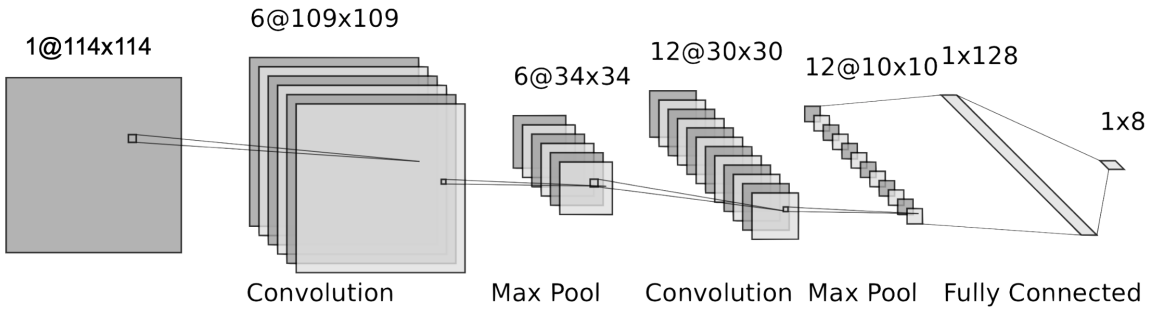


Fig. 2.6. Example of the custom Convolutional Neural Network (CNN) architecture applied to functional connectivity data based on Schaefer's 100 region parcellation. Input matrices of size 114x114 and the output are the 8 labels consisting of resting state and 7 HCP tasks. The CNN architecture includes two convolution layers, two max pooling layers, and two fully connected layers.

quire a large training set to reap the benefits of the algorithm's complexity. For this reason, we designed a relatively simple CNN architecture for classification of functional connectivity matrices.

2.6.1 Neural Network Architecture

Inputs to the convolutional neural network (CNN) vary in size depending on the brain parcellation used in data preprocessing. For example, the Schaefer's 100 region parcellation has a total of 114 brain regions including subcortical regions. This produces a functional connectivity matrix of size 114×114 as input into the CNN. The CNN treats each of these square matrices as one training or test sample. Each FC has a ground-truth label of the task that subject was performing in the scanner. For the purpose of this example, we will show the architecture for the Schaefer 100 region parcellation as depicted in Figure 2.6.1. Other parcellations have slightly different dimensions for each hidden layer but have the same general steps.

The input FCs of size 114×114 convolve into 6 channels with filter size 5, and stride of 1. In the field of computer vision, a filter sizes are commonly either 5×5 or 3×3. An odd filter size allows us to center the filter on a specific pixel and mitigates problems with input and output dimensions. We chose a the larger filter size of 5×5 because a larger filter has a larger receptive field and can mitigate overfitting. Then, a max pooling layer with a 3×3 filter is used to condense each 3×3 block into a single value. Max pooling finds the maximum value within the filter's receptive field, essentially passing on only the most important feature. Then, a rectified linear unit (ReLU) is applied as an activation layer. The ReLU activation function is superior to other common activation functions such as sigmoid and tanh because it mitigates the vanishing gradient problem. We then repeat this process by following up with a second convolution layer with 12 output channels, a second max pooling layer, again with filter size 3×3, and finally another ReLU activation. Between these layers, batch normalization was applied to regularize the data and promote learning. After convolution, pooling, and activation, we have 12 channels of a 10×10 matrix each. The 3-dimensional layer is then flattened into a one-dimensional vector - in this case, the size is 1×1200. This flattened layer is fully connected to another hidden layer of size 1×128. Finally, the hidden layer is linked to the final output layer of 8 classes. Recall that these 8 classes correspond to the task performed by the subject in the fMRI scanner. Softmax probability assigns a likelihood estimate on each class; the class with the highest softmax probability is the classifier's prediction.

2.6.2 Training of the Classifier

The data was split into an 80-20 train-test split followed by another 80-20 split of the training set into training and validation sets. The purpose of the validation set is to provide feedback during the training loop to quantify overfitting or underfitting. The 20% of data not included in training is set aside purely for testing. That is, the neural network does not see this data until it is completely finished with train-

ing. Then, we normalized the entire dataset by subtracting the training mean and dividing by standard deviation. This helps us avoid slow or unstable learning during training because the features are all similarly scaled and the gradients can be updated uniformly. For the optimizer, we test stochastic gradient descent and Adam as an additional hyperparameter to optimize. Stochastic gradient descent (SGD) samples only one sample at a time to calculate the derivative of the loss function. It is computationally efficient compared to batch optimizers but can be swayed by excessive noise in the dataset since it only operates on one sample at a time. The Adam optimizer features an adaptive per-parameter learning rate that improves performance on problems with sparse gradients. Adam has found to be very adaptive and easy to use without much hyperparameter tuning [57]. During our trials, however, we discovered that SGD generally outperforms Adam in the domain of FC classification.

With respect to learning rates, we performed a rough hyperparameter optimization by testing rates between $1e-2$ to $1e-5$ on a logarithmic scale. The standard learning rate of $1e-3$ seemed to converge at a reasonable rate and did not suffer from the problem of overshooting the global optimum that larger learning rates may encounter. Small learning rates such as $1-e5$ did not converge in a reasonable amount of time in our application.

The model was built in PyTorch and employed CUDA for parallel computation on an NVIDIA GTX960 GPU with 6GB of vRAM. GPU computing drastically shortens computation time versus a multi-core CPU. Each instance of the model was trained for 200 epochs with an early stopping parameter of 5 epochs. That is, whenever there are 5 epochs in a row where the validation loss does not decrease, the training ends. This also mitigates overfitting the model to the training data.

2.6.3 Evaluation of Post-processing Methods

We use the optimal configuration for PCA reconstruction found in Figure 3.2 at 80 principal components. We also evaluate the five tangent reference matrices from the

previous KNN classifier results. The convolutional neural network architecture and optimizing function introduce inherent randomness into the model. For this reason, we tested 20 iterations of each pipeline to not only examine average performance but also the variance of the classifier. Sources of variance include the stochastic gradient descent optimizer where a different sample is used to calculate the gradient at each step of the function. There is also a random test and train split of the data for each iteration. An ideal classifier would not only have high performance but also low variance for consistent results. We repeated this procedure for both the 100- and 300-region Schaefer brain parcellation to determine the effect of parcellation granularity on classifier performance.

3. RESULTS

3.1 Task Classification: K-Nearest Neighbors

The K-Nearest Neighbor classifier was run for different values of k on 424 unrelated subjects from the Human Connectome Project for all 7 tasks and resting state functional connectivity. Values for k tested included $1, 5, 10, 15$, and 20. Distance measures include correlation, cosine similarity, and Euclidean distance. The variable for classification was the task that the subject was performing in each scan. Figure 3.1, shows the accuracy of K-Nearest Neighbor classifier in classifying 8 HCP tasks with Schaefer's 100 brain region parcellation. Correlation and cosine similarity outperformed Euclidean distance for all values of k. As seen in Figure 3.1, within the different values of k, a greater value generally resulted in greater performance. However, when increasing the value of k for the K-Nearest Neighbor classifier, computation time increases significantly. Since performance gains are minimal with k values greater than 10, we decide to use $k = 10$ for the remaining experiments in task classification with this classifier.

Adapted from the differential identifiability framework proposed by Amico and Goñi [33], the entire vectorized (flattened) matrix of functional connectomes was reconstructed with a specific percentage of the principal components. The task identification rates corresponding to each level of PCs used in PCA reconstruction are shown in Figure 3.2. The input for PCA consists here of a large matrix of functional connectomes (FCs) that cover a total of 6,874 scans, corresponding to a total dimensionality of the data of 6,874 principal components. When we include 100% of principal components in reconstruction, the output is identical to the original dataset because the cumulative sum of all explained variance from PCs equals 100%. Since 0 principal components included in reconstruction results in a dataset of all zeros, we

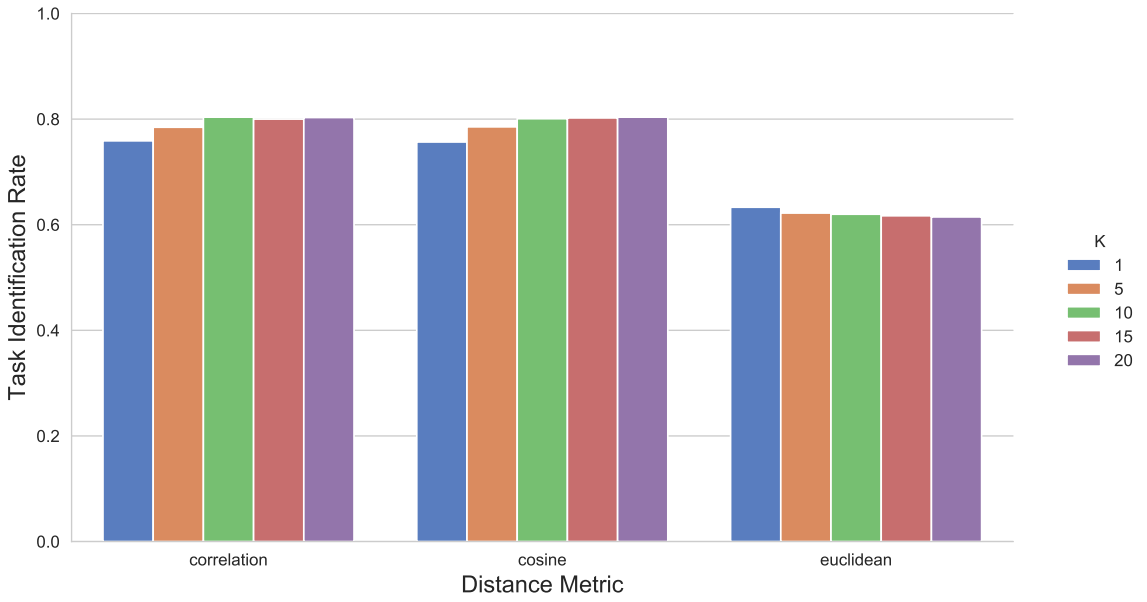
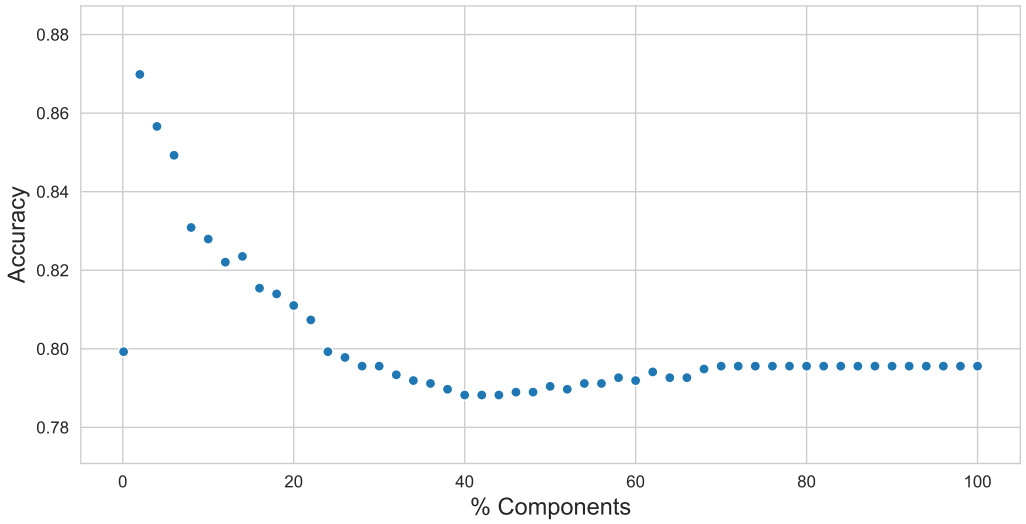


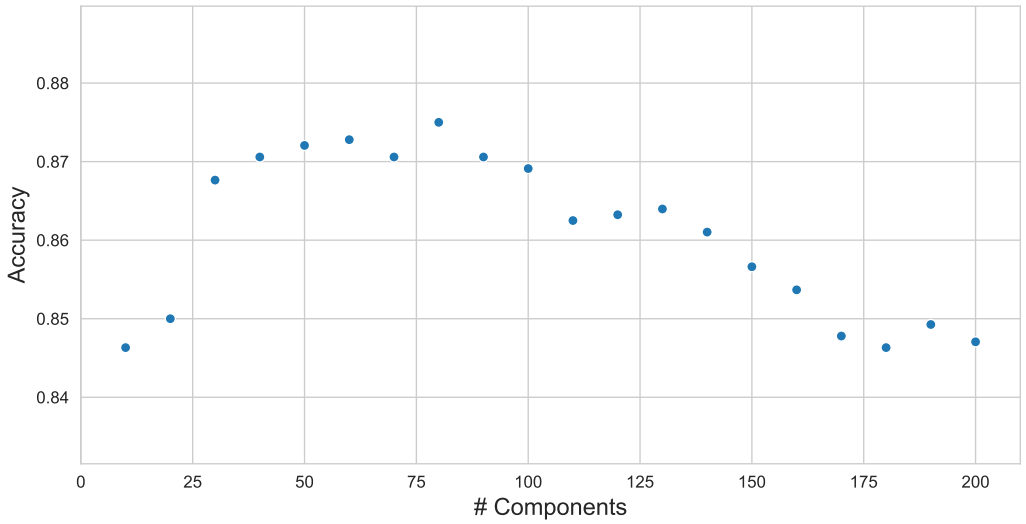
Fig. 3.1. Task identification rate of K-Nearest Neighbor classifier with correlation, cosine, and Euclidean distance metrics and various levels of k with Schaefer's 100 brain region parcellation. A greater value of k increases task identification rate but increases computational expense. Correlation distance and cosine similarity consistently outperformed Euclidean distance.

chose to start the first data point at one PC. Reconstruction was performed on Schaefer's 100 brain region parcellation. The task identification rate increases sharply at just 2% of PCs and then decreases until a minimum of 40% of PCs. From there, the accuracy slowly but steadily climbs to the accuracy of the untouched original dataset. At 2% principal components, we achieve a task identification rate on the test dataset of approximately 0.87. Without PCA reconstruction, the task identification rate is at 0.797 accuracy with the original FCs of the Schaefer 100 brain region parcellation.

Since there was significant spike between including just one PC and including 2% (approximately 120 PCs) in accuracy, we then took a more fine-grain approach by zooming in on this area of interest. Figure 3.2 also shows the accuracy between 10 and 200 components in intervals of 10 PCs. This appears to be a roughly convex function



(a) All proportions of principal components (PCs)



(b) Fine-grain proportions of PCs in reconstruction

Fig. 3.2. (a) Task identification rates of 7 HCP tasks and resting state of the K-Nearest Neighbor classifier ($k = 10$) with variable percentages of principal components (PCs) included in reconstruction. The total number of PCs is 6,784. In (b) we repeated the experiment with a PC range of 10-200 components, approximately 0-2.5% of all PCs of the dataset.

with a maximum at 80 principal components. Reconstruction at 80 PCs results in a task identification rate of approximately 0.875%. Compared to the original FCs, this is an improvement of over 0.07. Given the optimal performance seen in this particular reconstruction, we will refer to reconstruction of the HCP dataset with 80 PCs in the task identification problem as PCA - Optimal.

Figure 3.3 shows the KNN classifier's task identification rates on the original functional connectomes (FCs), PCA reconstructed dataset, and FCs projected into the tangent plane with various tangent reference matrices with Schaefer's 100 brain region parcellation. Both PCA-reconstructed and a few of the best tangent projected FCs produced higher performance than the original FCs. Of the reference matrices in tangent projection, the logarithmic Euclidean and Kullback means outperformed the others with task identification rates of 0.899 and 0.886, respectively, with the correlation distance metric. Similar to our findings in Figure 3.1, we observe that the correlation and cosine similarity measures performed the best within the K-Nearest Neighbor classifier across post-processing methods.

3.2 Task Classification: Convolutional Neural Network

Figure 3.4 shows the validation and training loss of the convolutional neural network (CNN) classifier in one example for each post-processing pipeline with Schaefer's 100 brain region parcellation. These curves were obtained over the course of 200 training epochs. The post-processing pipelines include the original, minimally processed functional connectomes (FCs), the optimally reconstructed FC matrices at 80 principal components, and tangent projected FC matrices with each of five reference matrices. There are slight differences in the training curves of the CNN trained on post-processed FCs versus the original FCs. First, in the case of the PCA optimally reconstructed FCs, and tangent projected Euclidean and harmonic FCs, the training and validation loss curves were more steep than that of the original FCs. Furthermore, the tangent projected FCs produced curves with miminal separation between

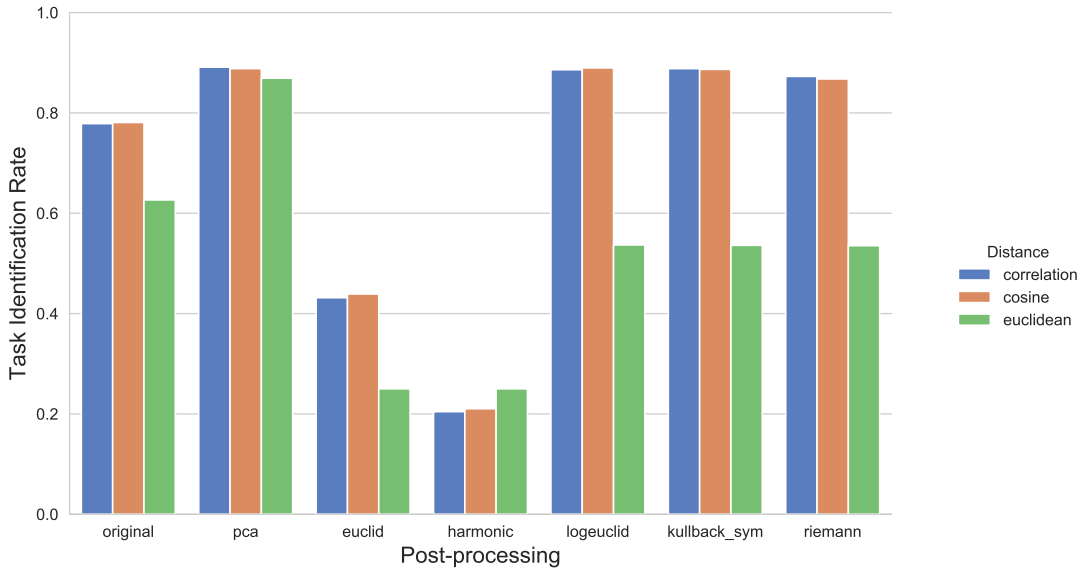


Fig. 3.3. KNN ($k = 10$) test accuracy with tangent reference matrices including Euclidean, Harmonic, logarithmic Euclidean, Kullback, and Riemman means. The raw data was also directly intput to the KNN classifier, denoted by `none`.

the training and validation curves. Furthermore, since the validation loss did not ever increase for more than 5 epochs in a row, early stopping did not occur. Hence these two observations suggest that the mitigation of overfitting was largely successful. In the case of the tangent projected FCs with Riemann and logarithmic Euclidean FCs, the loss curve decreased slower at first before sharply decreasing around epoch 15. In the end, both the training and validation loss resulted in values lower than that of the original FCs.

Figure 3.1 shows the task identification rates of the CNN classifier trained on seven different sets of FCs. The violin plot visualizes the distribution of the 20 identification rates per all seven transformations. The original FCs resulted in a mean task identification rate of 0.926 with a standard deviation of 0.006. All 6 of the other post-processing methods produced FCs that resulted in increased average performance and decreased variance. As shown above in Figure 3.2, PCA reconstruction with 80 prin-

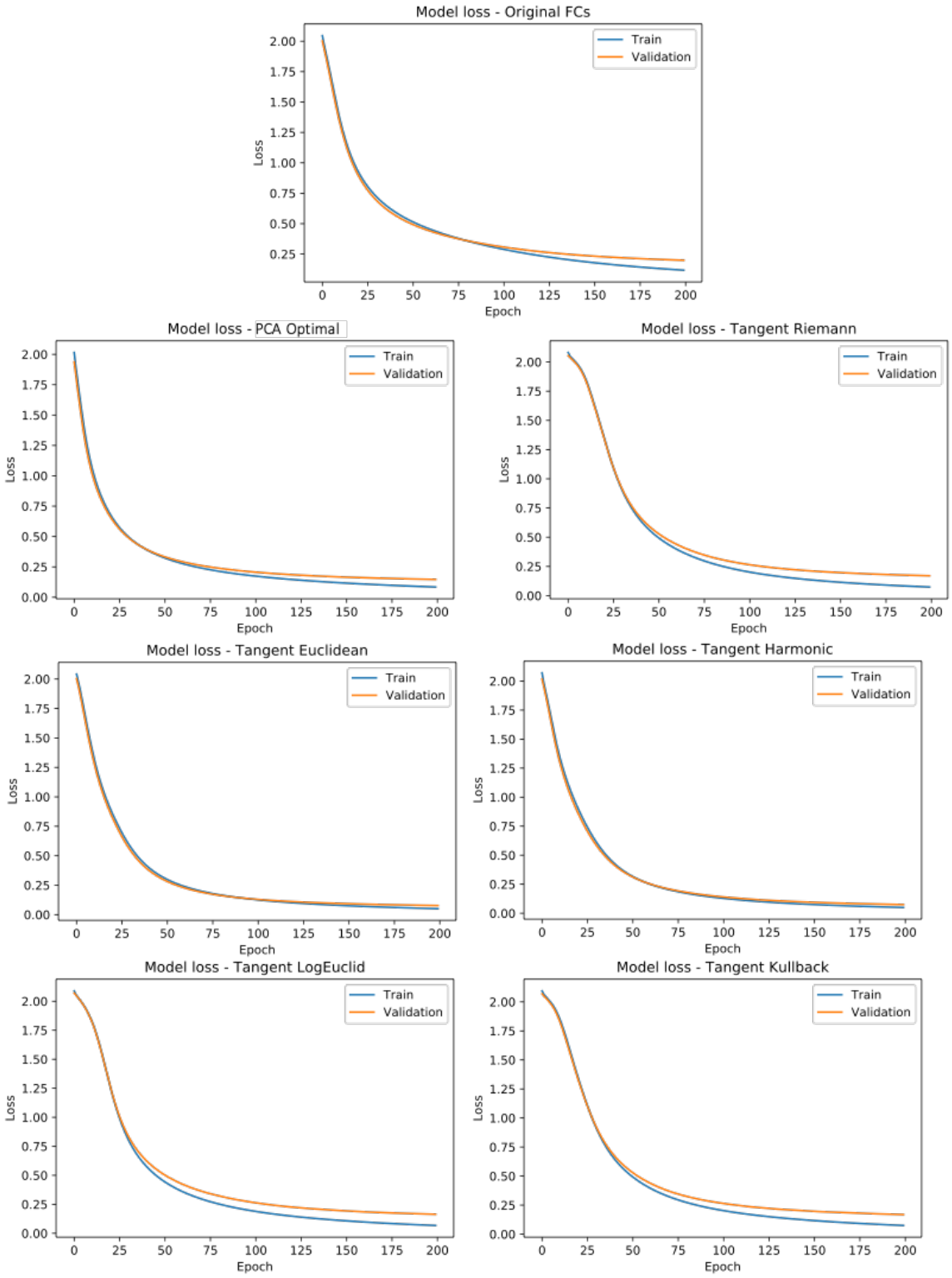


Fig. 3.4. Learning curve examples of CNN classifier on original functional connectomes (FCs) and post-processed FCs over 200 epochs of training.

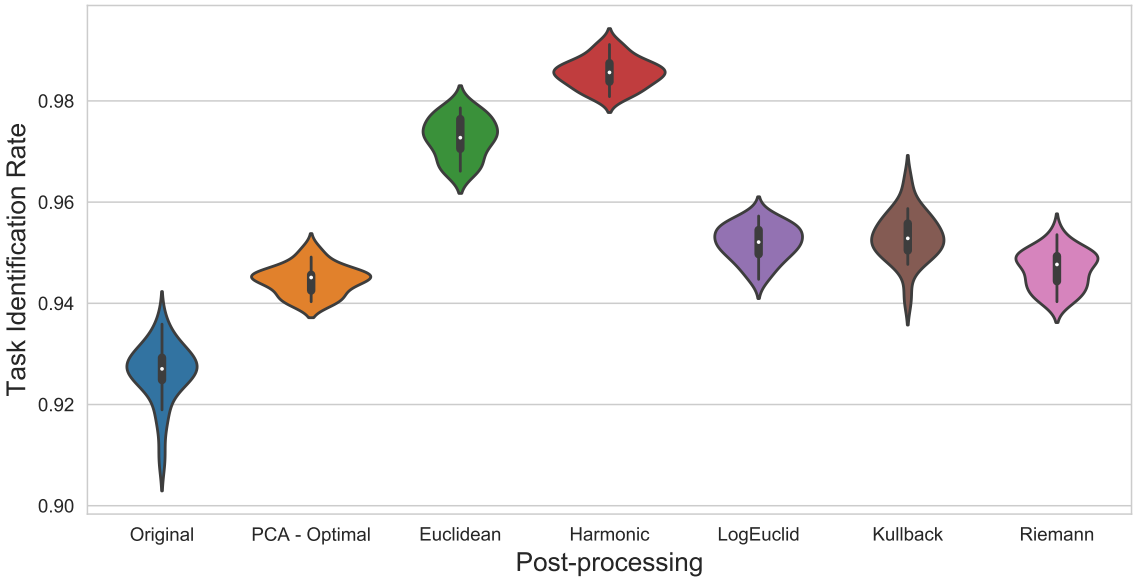


Fig. 3.5. Task identification rates of the convolutional neural network (CNN) classifier on original functional connectomes (FCs), optimally PCA reconstructed FCs, and tangent projected FCs with various reference matrices with Schaefer100 parcellation. The CNN model was trained 20 times per reference matrix to show within-classifier variability. Tangent projected FCs with the harmonic mean reference resulted in the highest mean task identification rate at 0.986 with a standard deviation of 0.0028.

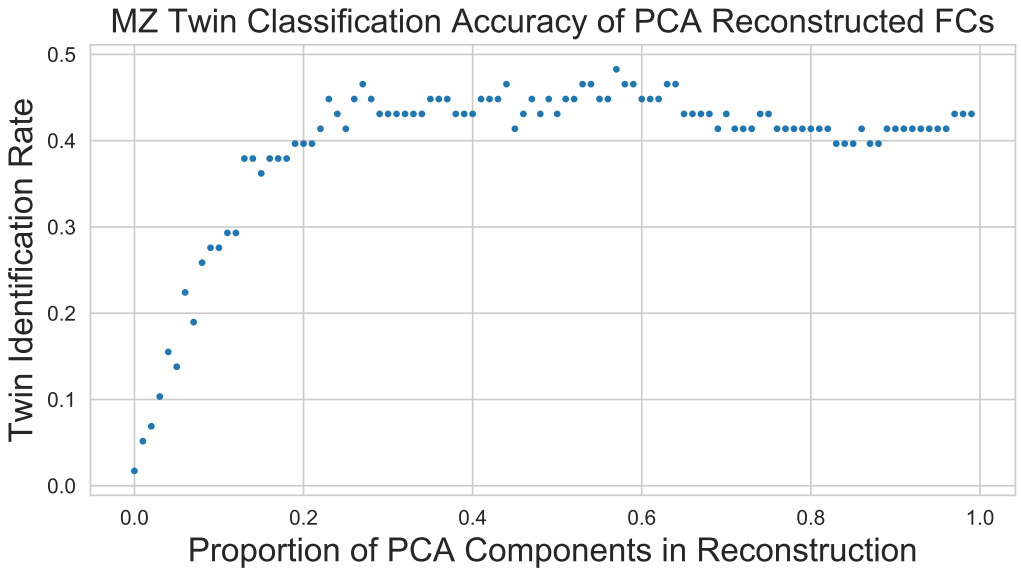
Table 3.1.
Average task identification rates and standard deviations for all post-processing methods

Post-processing	**Task Identification Rate**	**SD** (σ)
Original FCs	0.926	0.006
PCA - Optimal	0.945	0.003
Tan - Euclidean	0.973	0.004
Tan - Harmonic	0.986	0.003
Tan - LogEuclid	0.952	0.003
Tan - Kullback	0.953	0.005
Tan - Riemann	0.947	0.004

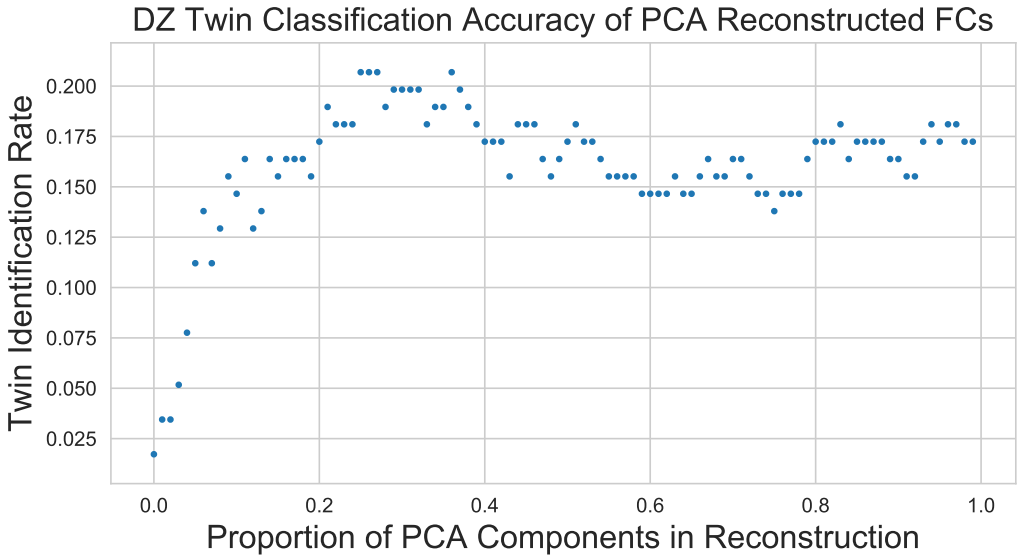
cipal components resulted in the optimal task identification rate in the KNN classifier. Due the expense of the CNN classifier, it was infeasible to again test the levels of PCs used in reconstruction. Therefore, we used the same 80 PCs in reconstruction for the CNN classifier. The average task identification rate of the 20 CNNs ran on PCA optimally reconstructed FCs was 0.945 with a standard deviation of 0.003. Out of the five tangent reference matrices, the harmonic mean outperformed the rest with an average task identification rate of 0.986 and a standard deviation of 0.003. Interestingly, the logarithmic Euclidean, Kullback, and Riemann means performed the worst out of the tangent projected FCs, opposite of the findings with the KNN classifier. These results were obtained using Schaefer's 100-region parcellation. Schaefer's 300-region parcellation was also tested, but due to the exponential increase in computation time, it was infeasible to finish the experiment. However, preliminary findings suggest that increase in parcellation granularity does not increase performance.

3.3 Twin Identification

As with task identification, we also experiment with various levels of PCs included in PCA reconstruction for twin datasets using Schaefer's 100-region parcellation. Here, however, the classification problem is to match twin pairs. Therefore, PCA reconstruction was done for each group of task FCs. In Figure 3.3, we show the twin identification rates of the KNN classifier for resting state fMRI of monozygotic (MZ) and dizygotic (DZ) twin pairs with 0 to 100% of the total PCs used in reconstruction. Within the HCP dataset, there were 106 pairs of MZ twins and 58 pairs of DZ twins with data across all 7 tasks and resting state FCs. There is a peak at roughly 50% of total PCs for MZ twins and at roughly 30% of total PCs for DZ twins. The peak of the DZ twins occurs earlier than that of the MZ twins in terms of number of PCs included in reconstruction. Also, it is important to note that the twin identification rate of the MZ twin pairs is significantly higher than that of the DZ twin pairs.



(a) Monozygotic twin pairs



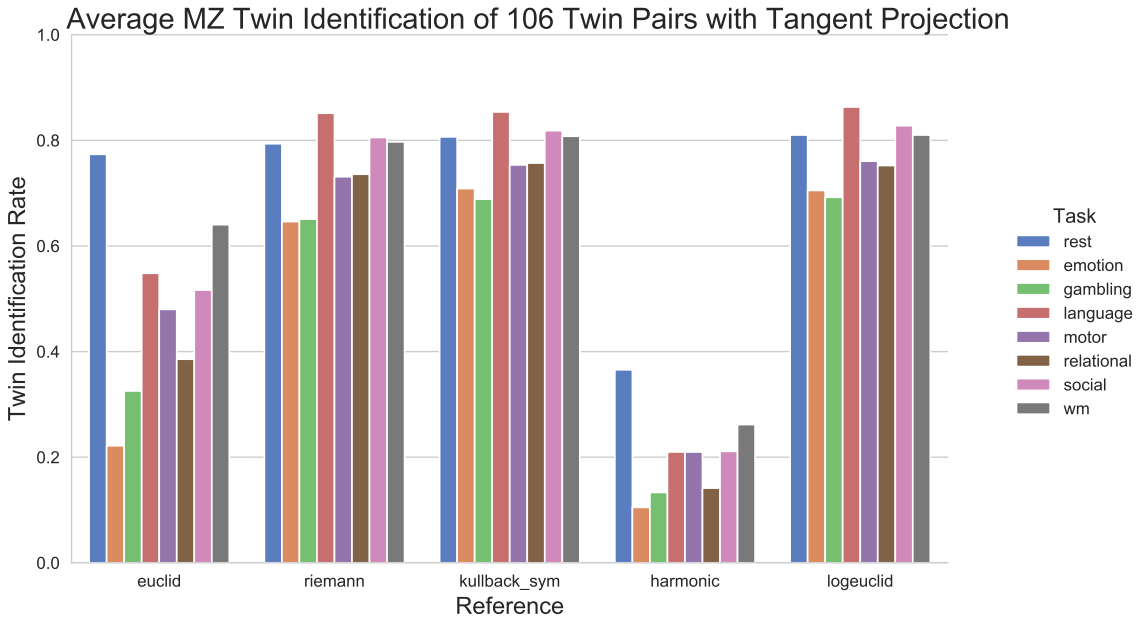
(b) Digyzotic twin pairs

Fig. 3.6. Twin identification rates of 106 monozygotic (MZ) and 58 dizygotic (DZ) twin pairs with variable proportions of principal components (PCs) used in reconstruction of resting state FCs. Analysis uses Schaefer's 100 brain region parcellation. There is an optimal reconstruction at approximately 50% of PCs for MZ twins and at 30% of PCs for DZ twins.

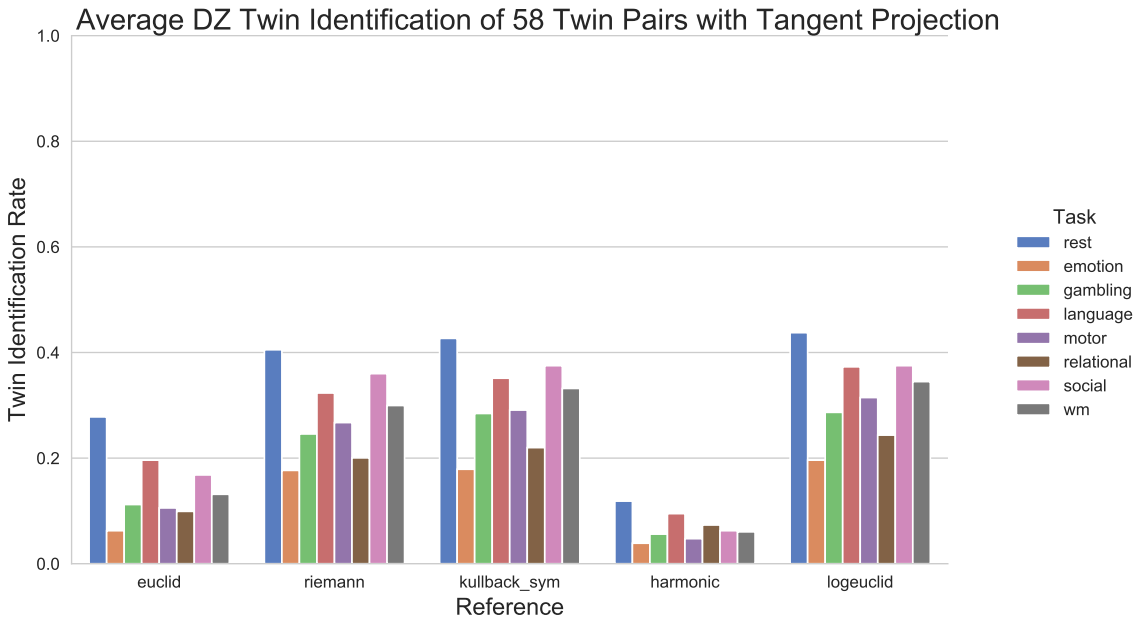
In Figure 3.7, we show the effects of the reference matrices used in tangent projection of the original FCs. Results are averaged from Schaefer's 100, 200, 300 and 400-region parcellations. Parcellation effects will be shown and discussed in the next figures. The KNN classifier was employed for each of the 7 tasks and resting state FCs after tangent projection with each of the five reference matrices. The logarithmic Euclidean, Kullback, and Riemann mean reference matrices performed the best across tasks. The harmonic mean showed meager results with an average twin identification rate across tasks of only 0.2. Out of the tasks, resting state functional connectivity was the best predictor of twin pairs as represented by the blue bars. These results were consistent overall for both MZ and DZ twin pairs.

In Figure 3.3, results of twin pair identification for monozygotic (MZ) and dizygotic (DZ) twin pairs are shown. Each subplot within this figure shows a different post-processing method applied to each of the 7 tasks and resting state FCs of the twin pairs. Further, the color bars represent the various Schaefer parcellation granularities used in the analyses. The first visual conclusion from these plots is that increasing parcellation granularity results in increased twin identification rates in the original FCs, PCA with optimal reconstruction, and tangent projected FCs with logarithmic Euclidean reference. These two references were chosen based on the former's high performance and the latter's interesting dichotomy between resting state and task FC identification rates. However, for the tangent Euclidean reference, a lower twin identification rate was observed for the gambling, language, motor, relational, and social tasks for parcellations of sizes 300 and 400. These findings were consistent for both MZ and DZ twins. Due to some inconsistencies in the HCP dataset of the 500-region parcellation, we were only able to obtain corresponding results for the original and PCA reconstructed FCs of MZ twins.

In the original FCs, resting state FCs significantly outperformed the other tasks with an average twin identification rate across parcellations of 0.487. Out of the seven HCP tasks, social, language, and working memory FCs produced higher twin identification rates than the other four tasks. However, after post-processing, the task



(a) Monozygotic twin pairs

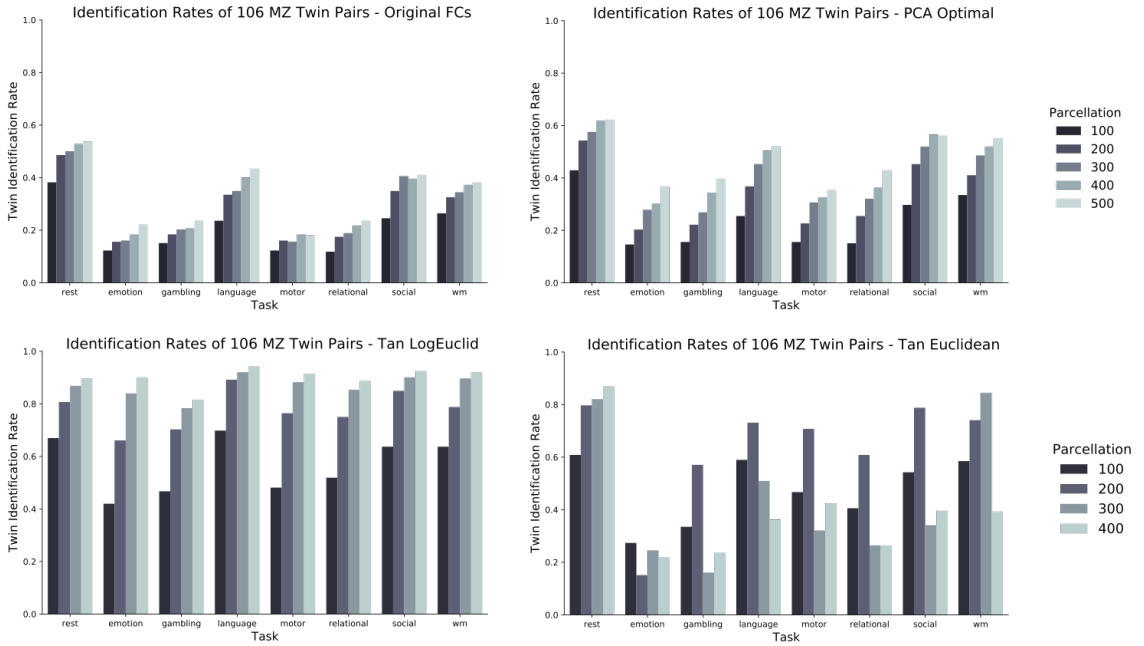


(b) Digyzotic twin pairs

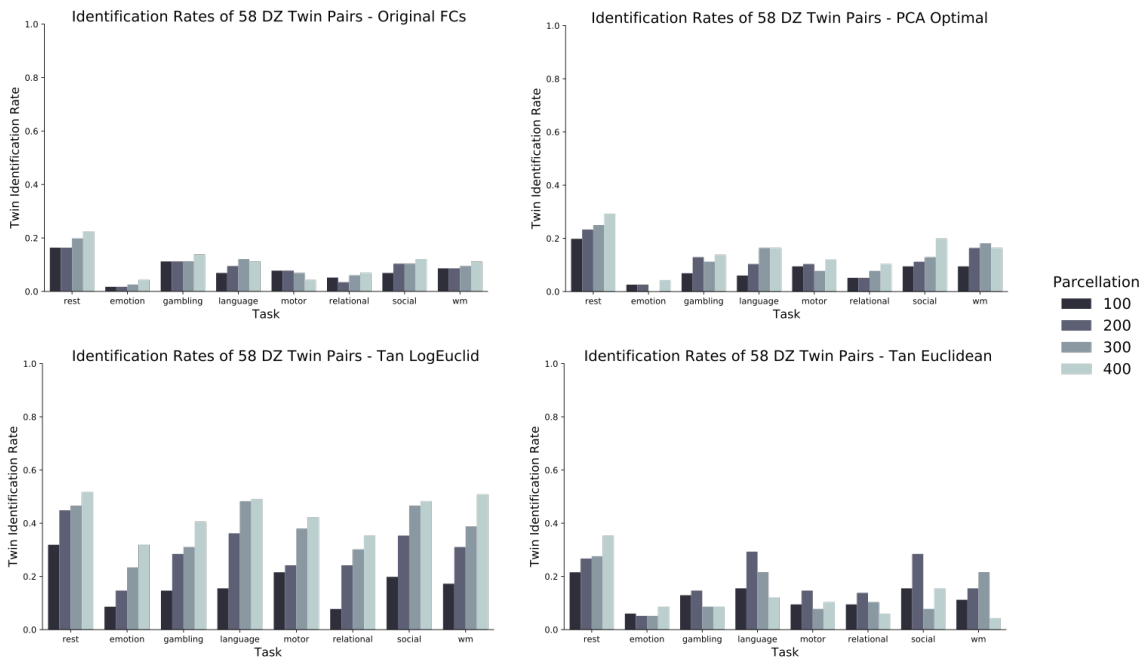
Fig. 3.7. Averaged twin identification rates of 106 monozygotic (MZ) and 58 dizygotic (DZ) twin pairs across Schaefer 100-400 parcellations with tangent projected FCs using different reference matrices. The logarithmic Euclidean and Kullback mean references resulted in the highest identification rates.

FCs matched or even in some cases surpassed the results obtained from resting state FCs. Using the results obtained from Figure 3.3, for optimal PCA reconstruction, we use 50% of PCs in MZ reconstruction and 30% of PCs in DZ reconstruction for each task. After reconstructing the dataset, the KNN classifier was employed on each task to produce the results shown in the upper-right subplot of both figures. Performance increased due to PCA reconstruction for all 7 tasks and resting state fMRI.

In Figure 3.9, we show the results of the two optimal configurations of PCA reconstruction and tangent space projection of functional connectomes (FCs) alongside the results of the original FCs. Again, both monozygotic (MZ) and dizyogtic (DZ) twin pairs are shown in this figure. Across all parcellation sizes and types of twin pairs, both PCA reconstruction and tangent space projection with the logarithmic Euclidean reference significantly increase the task identification rates. For DZ twins, tangent projection with logarithmic Euclidean reference produces task identification rates higher than that of the original FCs of MZ twins. The highest task identification rate was observed with the language FCs of Schaefer's 400-region parcellation with a rate of 0.943. For reference, the task identification rate of a random classifier correctly matching a twin to one of 106 pairs is less than 0.01. In Figure 3.3, we show the subject identification rates of 106 individuals. The identification rates of the original FCs and PCA optimally reconstructed FCs are greater than the twin identification rates. However, in the tangent projected FCs, the rest, gambling, language, and working memory task FCs resulted in decreased subject identification rates. Parcellation granularity again improves performance across all post-processing methods with the exception of tangent projection with the Euclidean mean reference matrix.

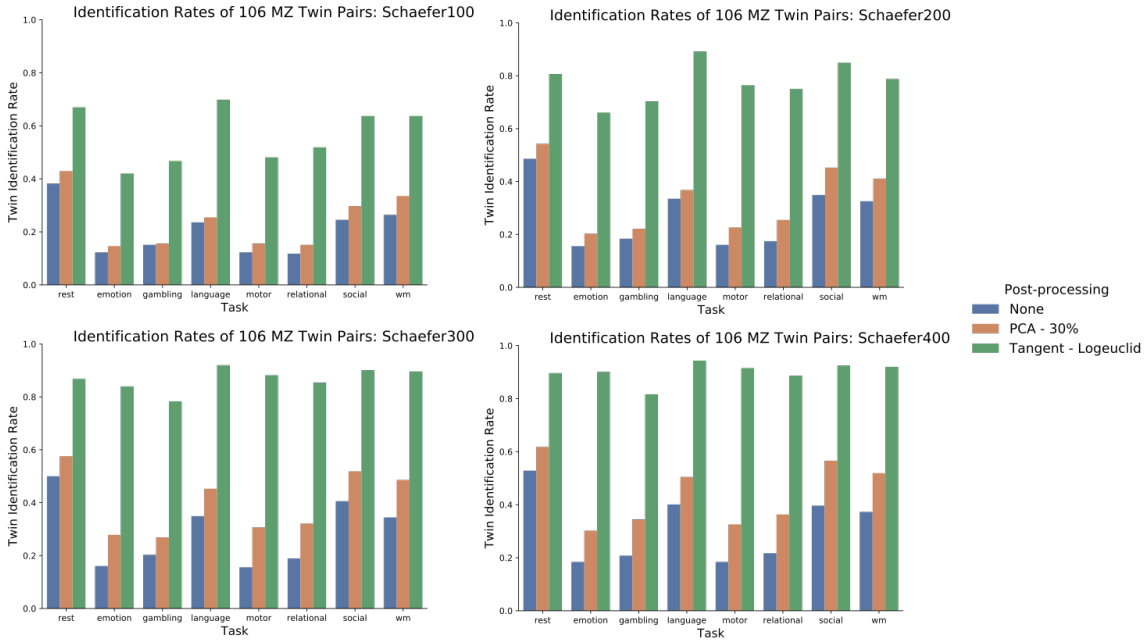


(a) Monozygotic twin pairs

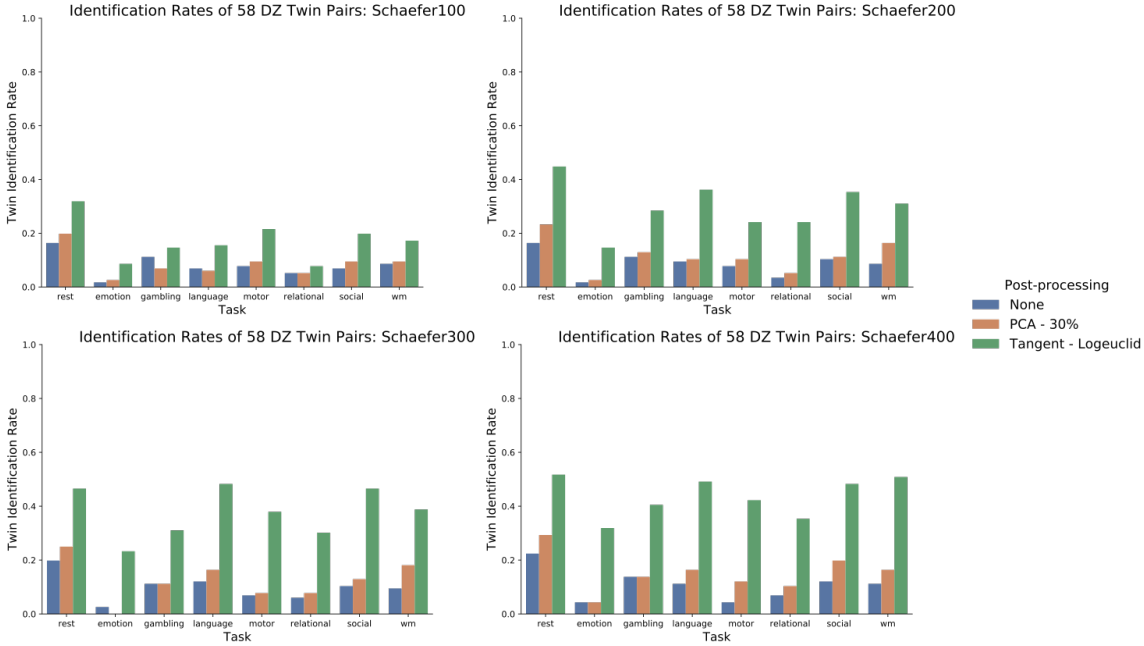


(b) Digyzotic twin pairs

Fig. 3.8. Twin identification rates of 106 monozygotic (MZ) and 58 dizygotic (DZ) twin pairs across 7 HCP tasks and resting state functional connectivity. Greater parcellation granularity results in higher performance across all 8 categories for MZ twins.

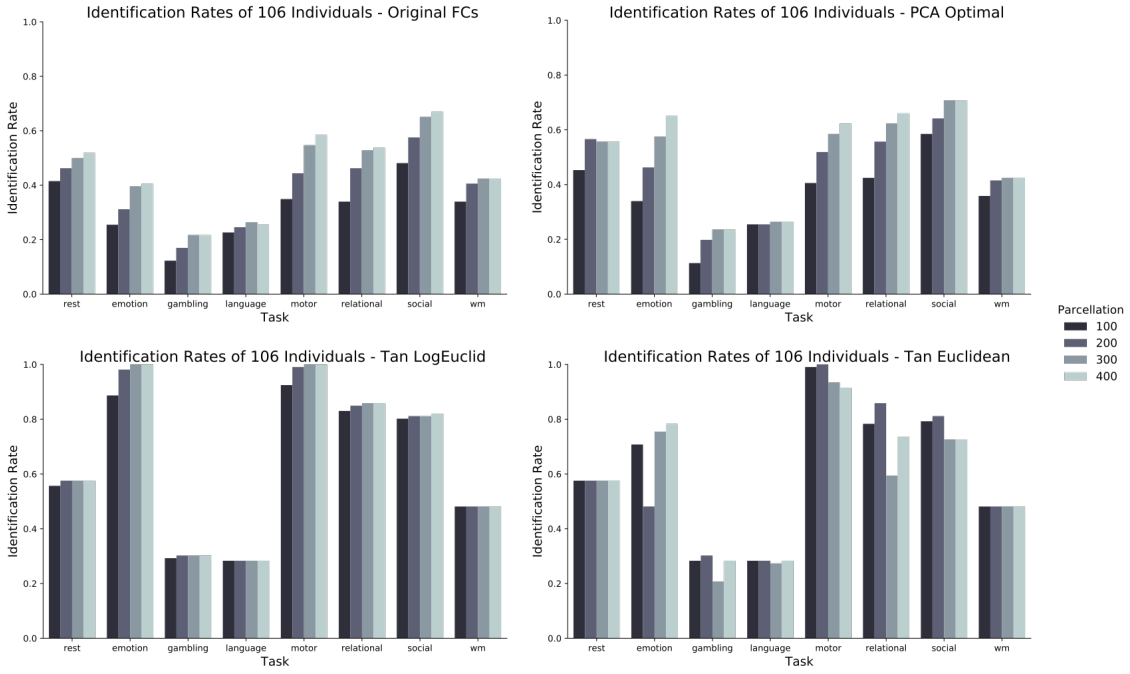


(a) Monozygotic twin pairs

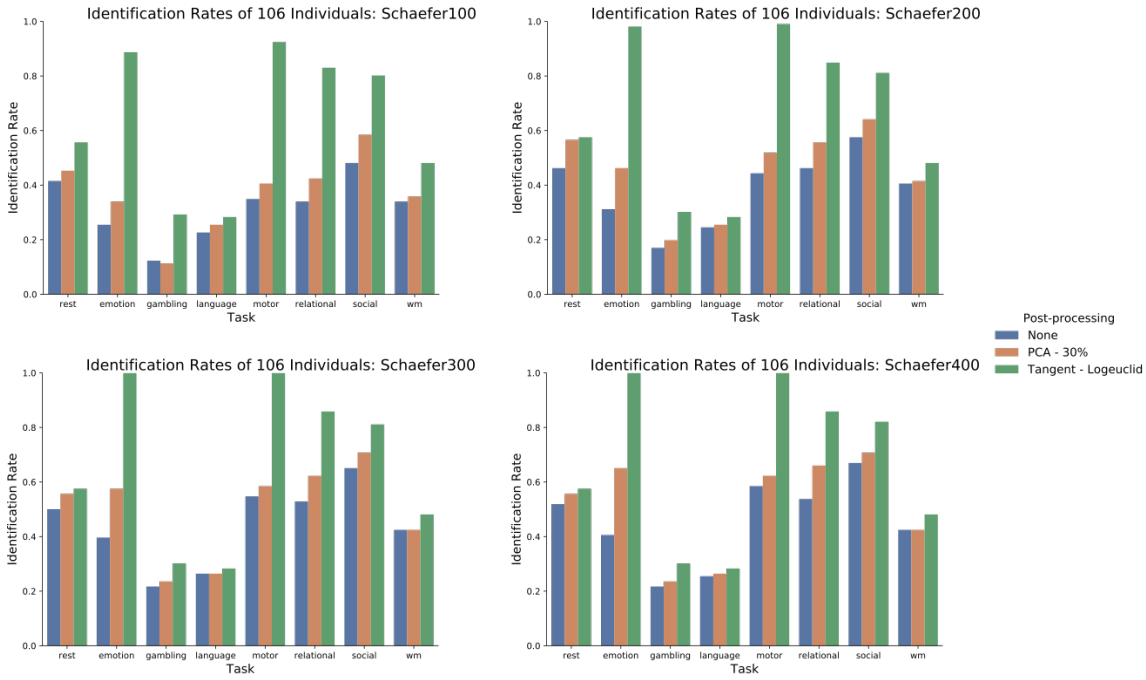


(b) Digyzotic twin pairs

Fig. 3.9. Twin identification rates of 106 monozygotic (MZ) and 58 dizygotic (DZ) twin pairs across 7 HCP tasks and resting state functional connectivity. Both post-processing methods of PCA optimal reconstruction and tangent projection with the logarithmic Euclidean reference improve twin identification rates.



(a) Parcellation effects on subject identification rates



(b) Post-processing method effects on subject identification rates

Fig. 3.10. Subject identification rates obtained with test and retest scans of 106 individuals across resting state and 7 HCP task functional connectivity. In (a), higher parcellation granularity improves subject identification rate. In (b), both optimal PCA reconstruction and tangent projection with logarithmic Euclidean references improves subject identification rates across all tasks.

4. DISCUSSION

Advancements in analytical methods and the increase in size of publicly available neuroimaging datasets have resulted in many studies applying predictive modeling to functional connectivity. For example, it has been shown that fully connected deep neural networks can be used successfully for connectome-based classification [18]. However, any type of analysis of fMRI data is hindered by the lack of a common post-processing and analysis pipeline within the neuroimaging community. Predictive models that use functional connectivity typically follow three main steps: parcellating the brain into regions, estimating the interactions between these defined regions, and feeding the data into a classifier for prediction [16]. In this study, we attempted to address all three steps by experimenting with various post-processing methods with two different classifiers with the goal to identify a best-practice, robust data pipeline functional connectivity analysis. In this section, we discuss the results and our overall recommendations and their rationale.

4.1 Impact of k on Identification Rate

In the context of task identification, we find that a value of $k = 10$ yielded the best overall performance. In K-Nearest Neighbors classifiers, the greater the value of k, the less of an effect noise has on the results. This is because a larger voting population smooths out the variance in the individual predictors. Because our best-performing value of $k = 10$ is relatively small compared to the number of samples, $6,784$, we infer that the nearest neighbors were usually representative of the actual underlying space. For values of k greater than 10, we observed practically identical performances but with the sacrifice of greater computational expense. For twin identification, for each label there are only two samples within the test subset: the retest scans for each

twin member. Therefore, a value of $k = 1$ was used to prevent ties in classification. A KNN classifier with $k = 1$ is identical to the identification rate framework proposed by Finn [4]. We also find that cosine similarity and correlation were the best predictors in the KNN classifier as shown in Figure 3.1, with marginal differences between these two similarity measures.

4.2 PCA Reconstruction

We applied PCA reconstruction with varied levels of principal components (PCs) to the datasets. In the context of the differential identifiability framework with respect to within-task subject identifiability, it has been shown that PCA reconstruction with slightly less than 50% of PCs results in optimal performance [33]. The functional connectivity included in basic functions of the human brain are represented by the first few principal components. These components also explain the most variance and as the explained variance decreases, task and subject-level variability are introduced. Then, in the last 50% of PCs, scanner artifacts and motion noise are likely introduced due to the decline in differential identifiability. In this study, we see some similarities with the findings in Amico et al. 2018. In twin identification, PCA reconstruction was used for each task separately, similar to Amico's study. As shown in Figure 3.3, optimal reconstruction was observed at using approximately 50% and 30% of the total principal components of MZ and DZ twins pairs, respectively. For task identification, we reach optimal accuracy with task identifiability at only 80 of the 6,784 PCs included as shown in Figure 3.2. This is likely due to the much larger dimensionality of this dataset and therefore many more principal components. In fact, the first few principal components disproportionately explain the most variance with over 40.8% of the total variance explained in the first 1% of components. We observe that the explained variance in an equal proportion of components is greater when the number of components is high. The added principal components that result from an increase in dimensionality do not scale linearly in terms of explained

variance. Intuitively, it is apparent that most of the variance in the data attributed to base human brain function is contained within the first few components. Variance attributed to brain networks involved in cognitive tasks are likely included within the next few components, up to approximately 80 components. We hypothesize this because the task identification rate increases dramatically until the optimal reconstruction at 80 components. From there, since task identification rate decreases, we hypothesize that noise is slowly added back in. Sources of this noise may include scanner artifacts, motion noise, and variance due to subject-specific brain networks or 'fingerprinting' [4].

As shown in Figure 3.3, the KNN classifier on the original Schaefer 100-region parcellation FCs achieved a task identification rate of 0.778 with the correlation measure. With PCA reconstruction with the optimal 80 PCs, the task identification rate increased to 0.891, an increase of 0.113, with the correlation measure. In task identification with the CNN classifier, PCA reconstruction with 80 PCs increased the task identification rate of the original FCs from an average of 0.926 to an average of 0.945, an increase of 0.019. Further, the standard deviation of the classifier's performance also decreased from 0.0571 to 0.0280, a decrease of 0.0291. Similar improvements were observed in the twin identification results in Figure 3.2. The use of PCA reconstruction is promising as a data cleaning method for functional connectivity data as we show here. We extend the differential identifiability framework to successfully clean FC data for classification. However, a limitation of this approach is that there is no universal best proportion of principal components for reconstruction of the datasets - it depends on the dataset itself. We recommend experimenting with different proportions of PCs used in reconstruction to find an optimal configuration. With more expensive classifiers such as the convolutional neural network used in this study, this is often infeasible and, as such, a less expensive classifier can be used as a stand-in to evaluate the optimal configuration of PCs. In summary, we show that the performance of both the KNN and CNN classifiers achieve greater results on both task and twin identification problems with PCA reconstruction than with the original FCs.

Furthermore, PCA reconstruction decreases the variance in the deep learning task classifier as shown in Figure 3.1 and yields more consistent results.

4.3 Tangent Space Projections

Functional connectivity matrices are computed by correlating BOLD time series data and are positive definite. Therefore, they lie within a non-linear surface called the positive semidefinite cone. Because of this, their geometry is non-Euclidean and classic distance measures such as correlation distance or euclidean distance cannot be used if geometry is to be preserved. To preserve geometry and to use the classifiers such as those exemplified this study, the Pearson correlated FC matrices should be first projected into the tangent space [16]. Geodesic distance, a non-Euclidean distance metric that accounts for the manifold on which the data lies, improves participant identification compared to the Pearson correlation distance metric [41] by 2% to 20% accuracy. To project FCs into the tangent space, we use Equation 2.3 with various reference matrices C_g shown in Table 2.1. First, we look at task identification. We found that in the context of twin identification for both monozygotic (MZ) and dizygotic (DZ) twin pairs, the projected FCs with Riemann, Kullback, and logarithmic Euclidean references resulted in the highest twin identification rates. In the original FCs, resting state performed significantly better than the other 7 tasks. Interestingly however, when projected into the tangent space with the three aforementioned reference matrices, the across-task performance variability decreased. In fact, for MZ twin pairs, language FCs outperformed resting state FCs as shown in subfigure (a) in Figure 3.9. The Euclidean mean reference boosts the twin identification rate of resting state and working memory FCs significantly but fails to achieve high performance in the other task FCs. This suggests that the underlying networks that are unique to the other tasks may not be translated well with tangent projection with the Euclidean reference. Finally, the harmonic mean resulted in very poor performance compared to the other reference matrices. This may be the result of the

scale of the harmonic mean being very small in magnitude compared to the others as shown in Figure 2.4.3 and errors in calculation were introduced. For both DZ and MZ twin matching, twin identification rates skyrocketed with tangent projection compared both to the original FCs and PCA-reconstructed FCs as we see in the subfigures in Figure 3.7. The twin identification rates were improved across all 7 tasks and resting state. Again, these improvements were especially pronounced in language, motor, and relational FCs where the twin identification rates were originally very low. Typically, clinical fMRI studies use resting state fMRI for prediction. We suggest that the tangent projection of FCs especially enhances within-subject identifiability of task FCs. It may be feasible or even preferred to use task fMRI for the identification of clinical biomarkers in functional connectivity data.

When we employ the KNN classifier on tangent projected FCs in the context of task identification, we obtained similar results to those described above. Specifically, the task identification rates as shown in Figure 3.3 are very low for the harmonic reference. We also showed that the logarithmic Euclidean, Kullback, and Riemann mean references outperform the original FCs in task identification. However, the results of the CNN classifier tell a different story. In Figure 3.1, these three reference matrices resulted in task identification rates lower than that of the harmonic mean and Euclidean mean. In fact, the task identification rates resulting from the five tangent reference matrices are essentially flipped in the CNN classifier versus the KNN classifier. Our conclusion from this observation is that the performance of each tangent reference matrix is highly classifier-dependent. The relatively simple distance-based KNN classifier may have trouble identifying tasks due to the nature of the tangent projections. The CNN classifier, however, is a more sophisticated deep learning algorithm that can iteratively learn and modulate itself over hundreds of training epochs. As shown in Figures 3.4 and 1 in the Appendix, the training and validation loss and identification rates were relatively low in the first few epochs of training. It is not until multiple iterations of weight optimizations that the classifier produced excellent results. Therefore, while the tangent projected FCs with the har-

monic mean reference fared poorly with the KNN classifier, this may be a great data transformation for deep learning applications. We recommend that all five reference matrices proposed in Pervaiz's study [16] be experimented depending on the classifier chosen for predictive modeling. All in all, tangent projection of functional connectivity preserves geometry and increases predictive power of the data in both task and twin identification.

4.4 Impact of Brain Parcellation Granularity on Identification Rates

Recent studies have found that a 268-node atlas generated by Finn et al. 2015 resulted in higher identification rates than the FreeSurfer atlas, which has only 68 nodes. This suggests that higher resolution parcellations allows better detection of individual features [58]. In fact, we show consistent findings with the KNN classifier in the context of twin identification. Figure 3.3 shows that higher parcellation granularity increases performance of the classifier across all post-processing methods, suggesting that the algorithm successfully utilizes the increase in input data size. Due to computational limits of the CNN classifier, the same experiment was infeasible to test parcellation effects on task identification. We expect that a similar trend is also true for this application for both the KNN and CNN classifier, although additional research is necessary. We recommend studies that use predictive modeling to maximize parcellation granularity of their datasets given their practical constraints. The increase in information of larger parcellations has shown to lead to greater performance in the twin identification portion of this study. However, we must also proceed cautiously in increasing parcellation resolution, as it has been suggested that doing so may have some unintended consequences. Possibilities of these include creating insignificant individual differences due to misalignment in the scanner or motion-related artifacts [4]. Furthermore, we encountered memory issues while training the CNN on the 400-region parcellation where the KNN classifier had no issues. Schaefer parcel-

lations increase in increments of 100 up to 1000 brain regions, resulting in a matrix size of over 1 million parameters [10].

4.5 Areas of Improvement and Future Work

In respect to the CNN architecture and design, a more structured hyperparameter search would have been better suited to determine parameters such as the learning rate, optimizing function, and loss function. Methods such as a grid or random search on the logarithmic scale should be used here [59]. Future work in CNNs with parcellation granularity should consider using big data software such as Spark to handle the training of large datasets. It would be an interesting challenge to use all the Schaefer parcellations up to 1000 brain regions.

A more detailed analysis of why and when certain reference matrices perform well and fail to perform well is necessary before generalizing this post-processing method to other FC applications. It would be helpful to analyze the difference in the encoding of brain networks as a result of the tangent space projection with the different references. Until then, we recommend testing multiple reference matrices when using tangent projection of functional connectivity data. We have also shown PCA reconstruction to be a consistent improvement upon using the minimally processed raw FC data in both the simple distance classifier and the CNN classifier. The number of PCs included in PCA reconstruction should be customized for each dataset depending on finding the optimal performance metric. This is also true for the twin identification where ideally, a customized number of PCs should be used for each task instead of only using resting state FC as we did in this study. We did not attempt to combine both PCA reconstruction and tangent space projection because of the intricacies involved in the geometry, but this could be a potential avenue of research. Additionally, due to the black-box nature of deep learning, it is difficult to get meaningful insights of the underlying causes of the classifier's decisions. Future work should be dedicated to investigating the intermediate features of the convolutional neural network and

their practical significance. We are optimistic that these post-processing methods are robust enough to be used on clinical applications of fMRI, where previous predictive modeling pipelines fail to achieve satisfactory results and if understood better, the determining features may aid in diagnosis and tracking disease progression.

5. CONCLUSION

Our results show that the use of post-processing techniques, including optimal reconstruction via PCA as well as projection of the data to tangent spaces, may improve functional connectivity classification methods. We have shown that optimal PCA reconstruction consistently outperforms the original FCs and decreases variance in both the convolutional neural network and K-Nearest Neighbors classifiers' performances. In task identification, we demonstrated that reconstruction with a small proportion of principal components (PCs) (approximately 1%) achieves an optimal task identification rate. In twin identification, PCA reconstruction was used on each task independently with 50% and 30% of total PCs in monozygotic (MZ) and dizygotic (DZ) twin pairs, respectively. To properly use mathematical functions in the data's native geometry, we projected FC matrices into the tangent space with various reference matrices. We showed that reference matrices are likely classifier-specific in performance. Specifically, we observe that the harmonic reference matrix performs best in conjunction with the CNN classifier but performs poorly with the distance-based KNN classifier. Conversely, the logarithmic Euclidean, Kullback, and Riemann references perform well with the KNN classifier yet perform relatively worse with the CNN classifier. Future study in the underlying effects of these reference matrices is necessary. Additionally, we have demonstrated that increasing parcellation size increases twin identification rate. However, the large size of these parcellations can restrict the use of complex deep learning classifiers.

In the context of twin classification, state-of-the-art results have identification rates of 0.64 and 0.25 for MZ and DZ twins, respectively, and sample sizes of $n = 25$ twin pairs. We achieved an identification rate of 0.943 with language FCs for MZ twins and 0.517 with resting state FCs for DZ twins, both with Schaefer's 400-region parcellation with sample sizes of $n = 106$ and $n = 58$ twin pairs, respectively. Finally,

in the context of task classification, we achieve an identification rate of over 0.986 with tangent projection with the harmonic mean reference, surpassing state-of-the-art results in literature of 0.937 [18]. These post-processing methods are promising for future research in functional connectome predictive modeling and, if optimized further, can likely be extended into clinical applications.

REFERENCES

REFERENCES

[1] T. M. Press, "Networks of the Brain | The MIT Press," library Catalog: mitpress.mit.edu Publisher: The MIT Press. [Online]. Available: https://mitpress.mit.edu/books/networks-brain

[2] *Fundamentals of Brain Network Analysis*. Elsevier, 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/C2012006036X

[3] F. X. Castellanos, A. Di Martino, R. C. Craddock, A. D. Mehta, and M. P. Milham, "Clinical applications of the functional connectome," *NeuroImage*, vol. 80, pp. 527–540, Oct. 2013.

[4] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable, "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity," *Nature Neuroscience*, vol. 18, no. 11, pp. 1664–1671, Nov. 2015. [Online]. Available: http://www.nature.com/articles/nn.4135

[5] D. H. Schultz and M. W. Cole, "Higher Intelligence Is Associated with Less Task-Related Brain Network Reconfiguration," *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 36, no. 33, pp. 8551–8561, 2016.

[6] A. Fornito and E. T. Bullmore, "Connectomics: a new paradigm for understanding brain disease," *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, vol. 25, no. 5, pp. 733–748, May 2015.

[7] B. A. Seitzman, C. Gratton, T. O. Laumann, E. M. Gordon, B. Adeyemo, A. Dworetsky, B. T. Kraus, A. W. Gilmore, J. J. Berg, M. Ortega, A. Nguyen, D. J. Greene, K. B. McDermott, S. M. Nelson, C. N. Lessov-Schlaggar, B. L. Schlaggar, N. U. F. Dosenbach, and S. E. Petersen, "Trait-like variants in human functional brain networks," *Proceedings of the National Academy of Sciences*, vol. 116, no. 45, pp. 22 851–22 861, Nov. 2019. [Online]. Available: http://www.pnas.org/lookup/doi/10.1073/pnas.1902932116

[8] M. Rajapandian, E. Amico, K. Abbas, M. Ventresca, and J. Goñi, "Uncovering differential identifiability in network properties of human brain functional connectomes," *Network Neuroscience*, pp. 1–16, Apr. 2020. [Online]. Available: https://www.mitpressjournals.org/doi/abs/10.1162/netn$_{a0}$0140

[9] H. Niu, Z. Zhu, M. Wang, X. Li, Z. Yuan, Y. Sun, and Y. Han, "Abnormal dynamic functional connectivity and brain states in Alzheimer's diseases: functional near-infrared spectroscopy study," *Neurophotonics*, vol. 6, no. 2, p. 025010, Jun. 2019, publisher: International Society for Optics and Photonics. [Online].

Available: https://www.spiedigitallibrary.org/journals/Neurophotonics/volume-6/issue-2/025010/Abnormal-dynamic-functional-connectivity-and-brain-states-in-Alzheimers-diseases/10.1117/1.NPh.6.2.025010.short

[10] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo, "Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, Sep. 2018. [Online]. Available: https://academic.oup.com/cercor/article/28/9/3095/3978804

[11] D. O. Svaldi, J. Goñi, K. Abbas, E. Amico, D. G. Clark, C. Muralidharan, M. Dzemidzic, J. D. West, S. L. Risacher, A. J. Saykin, and L. G. Apostolova, "Optimizing Differential Identifiability Improves Connectome Predictive Modeling of Cognitive Deficits in Alzheimer's Disease," *arXiv:1908.06197 [q-bio]*, Dec. 2019, arXiv: 1908.06197. [Online]. Available: http://arxiv.org/abs/1908.06197

[12] C. J. Brown and G. Hamarneh, "Machine Learning on Human Connectome Data from MRI," p. 51.

[13] J. A. K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers," p. 8.

[14] "Underfitting and Overfitting in Machine Learning," Nov. 2017, library Catalog: www.geeksforgeeks.org Section: Advanced Computer Subject. [Online]. Available: https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/

[15] Q. Xu, M. Zhang, Z. Gu, and G. Pan, "Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs," *Neurocomputing*, vol. 328, pp. 69–74, Feb. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231218309524

[16] U. Pervaiz, D. Vidaurre, M. W. Woolrich, and S. M. Smith, "Optimising network modelling methods for fMRI," *NeuroImage*, vol. 211, p. 116604, May 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811920300914

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, number: 7553 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nature14539

[18] X. Wang, X. Liang, Z. Jiang, B. A. Nguchu, Y. Zhou, Y. Wang, H. Wang, Y. Li, Y. Zhu, F. Wu, J.-H. Gao, and B. Qiu, "Decoding and mapping task states of the human brain via deep learning," *Human Brain Mapping*, vol. 41, no. 6, pp. 1505–1519, 2020, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.24891. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24891

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[20] R. J. Meszlényi, K. Buza, and Z. Vidnyánszky, "Resting State fMRI Functional Connectivity-Based Classification Using a Convolutional Neural Network Architecture," *Frontiers in Neuroinformatics*, vol. 11, 2017, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fninf.2017.00061/full

[21] "TensorFlow Image Classification: Three Quick Tutorials," library Catalog: missinglink.ai. [Online]. Available: https://missinglink.ai/guides/tensorflow/tensorflow-image-classification/

[22] "International Neuroimaging Data-sharing Initiative." [Online]. Available: http://fcon$_1$000.*projects.nitrc.org*/

[23] "ADNI | Access Data," library Catalog: adni.loni.usc.edu. [Online]. Available: http://adni.loni.usc.edu/data-samples/access-data/

[24] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, and M. Jenkinson, "The minimal preprocessing pipelines for the Human Connectome Project," *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811913005053

[25] D. S. Marcus, J. Harwell, T. Olsen, M. Hodge, M. F. Glasser, F. Prior, M. Jenkinson, T. Laumann, S. W. Curtiss, and D. C. Van Essen, "Informatics and Data Mining Tools and Strategies for the Human Connectome Project," *Frontiers in Neuroinformatics*, vol. 5, Jun. 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3127103/

[26] D. C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. E. J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss, S. Della Penna, D. Feinberg, M. F. Glasser, N. Harel, A. C. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S. E. Petersen, F. Prior, B. L. Schlaggar, S. M. Smith, A. Z. Snyder, J. Xu, and E. Yacoub, "The Human Connectome Project: A data acquisition perspective," *NeuroImage*, vol. 62, no. 4, pp. 2222–2231, Oct. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811912001954

[27] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U. R. Acharya, R. Khosrowabadi, and V. Salari, "Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network," *Frontiers in Neuroscience*, vol. 13, 2020, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2019.01325/full

[28] S. Sarraf and G. Tofighi, "Classification of Alzheimer's Disease using fMRI Data and Deep Learning Convolutional Neural Networks," *arXiv:1603.08631 [cs]*, Mar. 2016, arXiv: 1603.08631. [Online]. Available: http://arxiv.org/abs/1603.08631

[29] Y. Liang, Y. Chen, H. Li, T. Zhao, X. Sun, N. Shu, and D. P. a. Z. Zhang, "Disrupted Functional Connectivity Related to Differential Degeneration of the Cingulum Bundle in Mild Cognitive Impairment Patients," Feb. 2015, issue: 3 Library Catalog: www.eurekaselect.com Pages: 255-265 Volume: 12. [Online]. Available: http://www.eurekaselect.com/128979/article

[30] P. Patel, P. Aggarwal, and A. Gupta, "Classification of Schizophrenia versus normal subjects using deep learning," in *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing - ICVGIP '16*. Guwahati, Assam, India: ACM Press, 2016, pp. 1–6. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3009977.3010050

[31] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature reviews. Neuroscience*, vol. 10, pp. 186–98, Mar. 2009.

[32] M. D. Fox and M. Greicius, "Clinical applications of resting state functional connectivity," *Frontiers in Systems Neuroscience*, vol. 4, 2010, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2010.00019/full

[33] E. Amico and J. Goñi, "The quest for identifiability in human functional connectomes," *Scientific Reports*, vol. 8, no. 1, p. 8254, Dec. 2018. [Online]. Available: http://www.nature.com/articles/s41598-018-25089-1

[34] S. Maknojia, N. W. Churchill, T. A. Schweizer, and S. J. Graham, "Resting State fMRI: Going Through the Motions," *Frontiers in Neuroscience*, vol. 13, 2019, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2019.00825/full

[35] S. Noble, M. N. Spann, F. Tokoglu, X. Shen, R. T. Constable, and D. Scheinost, "Influences on the Test-Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility," *Cerebral Cortex (New York, N.Y.: 1991)*, vol. 27, no. 11, pp. 5415–5429, 2017.

[36] R. M. Birn, E. K. Molloy, R. Patriat, T. Parker, T. B. Meier, G. R. Kirk, V. A. Nair, M. E. Meyerand, and V. Prabhakaran, "The effect of scan length on the reliability of resting-state fMRI connectivity estimates," *NeuroImage*, vol. 83, pp. 550–558, Dec. 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104183/

[37] S. Gao, A. S. Greene, R. T. Constable, and D. Scheinost, "Combining multiple connectomes improves predictive modeling of phenotypic measures," *NeuroImage*, vol. 201, p. 116038, Nov. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811919306196

[38] D. O. Svaldi, J. Goñi, A. B. Sanjay, E. Amico, S. L. Risacher, J. D. West, M. Dzemidzic, A. Saykin, and L. Apostolova, "Towards Subject and Diagnostic Identifiability in the Alzheimer's Disease Spectrum based on Functional Connectomes," *arXiv:1809.09757 [q-bio]*, vol. 11044, pp. 74–82, 2018, arXiv: 1809.09757. [Online]. Available: http://arxiv.org/abs/1809.09757

[39] S. Bari, E. Amico, T. Talavage, and J. Goñi, "Inter-Scanner Identifiability Based On Resting-State Functional Connectomes," Oct. 2018.

[40] K. Abbas, M. Liu, M. Venkatesh, E. Amico, J. Harezlak, A. D. Kaplan, M. Ventresca, L. Pessoa, and J. Goñi, "Regularization of functional connectomes and its impact on geodesic distance and fingerprinting," p. 16.

[41] M. Venkatesh, J. Jaja, and L. Pessoa, "Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification," *NeuroImage*, vol. 207, p. 116398, Feb. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811919309899

[42] G. A. M. Blokland, M. A. Mosing, K. J. H. Verweij, and S. E. Medland, "Twin Studies and Behavior Genetics," p. 21.

[43] M. Cole, D. Bassett, J. Power, T. Braver, and S. Petersen, "Intrinsic and Task-Evoked Network Architectures of the Human Brain," *Neuron*, vol. 83, no. 1, pp. 238–251, Jul. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0896627314004000

[44] NIH, "Human Connectome Project." [Online]. Available: http://www.humanconnectomeproject.org/data/

[45] E. Amico and J. Goñi, "Mapping hybrid functional-structural connectivity traits in the human connectome," *Network Neuroscience*, vol. 2, no. 3, pp. 306–322, Sep. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6145853/

[46] E. Amico, A. Arenas, and J. Goñi, "Centralized and distributed cognitive task processing in the human connectome," *Network Neuroscience*, vol. 3, no. 2, pp. 455–474, Jan. 2019. [Online]. Available: https://www.mitpressjournals.org/doi/abs/10.1162/netn$_{a0}$0072

[47] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. Laumann, K. L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Xu, E. Yacoub, K. Uğurbil, D. C. Van Essen, and M. F. Glasser, "Resting-state fMRI in the Human Connectome Project," *NeuroImage*, vol. 80, pp. 144–168, Oct. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811913005338

[48] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, Aug. 2012.

[49] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage*, vol. 90, pp. 449–468, Apr. 2014.

[50] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fMRI," *NeuroImage*, vol. 84, Jan. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3849338/

[51] "Components of the Human Connectome Project - Task fMRI - Connectome." [Online]. Available: https://www.humanconnectome.org/study/hcp-young-adult/project-protocol/task-fmri

[52] B. Sen and K. K. Parhi, "Predicting Tasks from Task-fMRI Using Blind Source Separation," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, Nov. 2019, pp. 2201–2205, iSSN: 2576-2303.

[53] A. W. Chung, E. Pesce, R. P. Monti, and G. Montana, "Classifying HCP Task-fMRI Networks Using Heat Kernels," *arXiv:1604.08912 [q-bio]*, Apr. 2016, arXiv: 1604.08912. [Online]. Available: http://arxiv.org/abs/1604.08912

[54] A. Gritsenko, M. Lindquist, and M. K. Chung, "Twin Classification in Resting-State Brain Connectivity," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2020, pp. 1391–1394, iSSN: 1945-8452.

[55] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017, publisher: Taylor & Francis. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01570120

[56] "Cosine Similarity - an overview | ScienceDirect Topics." [Online]. Available: https://www.sciencedirect.com/topics/computer-science/cosine-similarity

[57] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980

[58] T. Vanderwal, J. Eilbott, E. S. Finn, R. C. Craddock, A. Turnbull, and F. X. Castellanos, "Individual differences in functional connectivity during naturalistic viewing conditions," *NeuroImage*, vol. 157, pp. 521–530, 2017.

[59] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, vol. 13, no. 10, pp. 281–305, 2012. [Online]. Available: http://jmlr.org/papers/v13/bergstra12a.html

APPENDIX

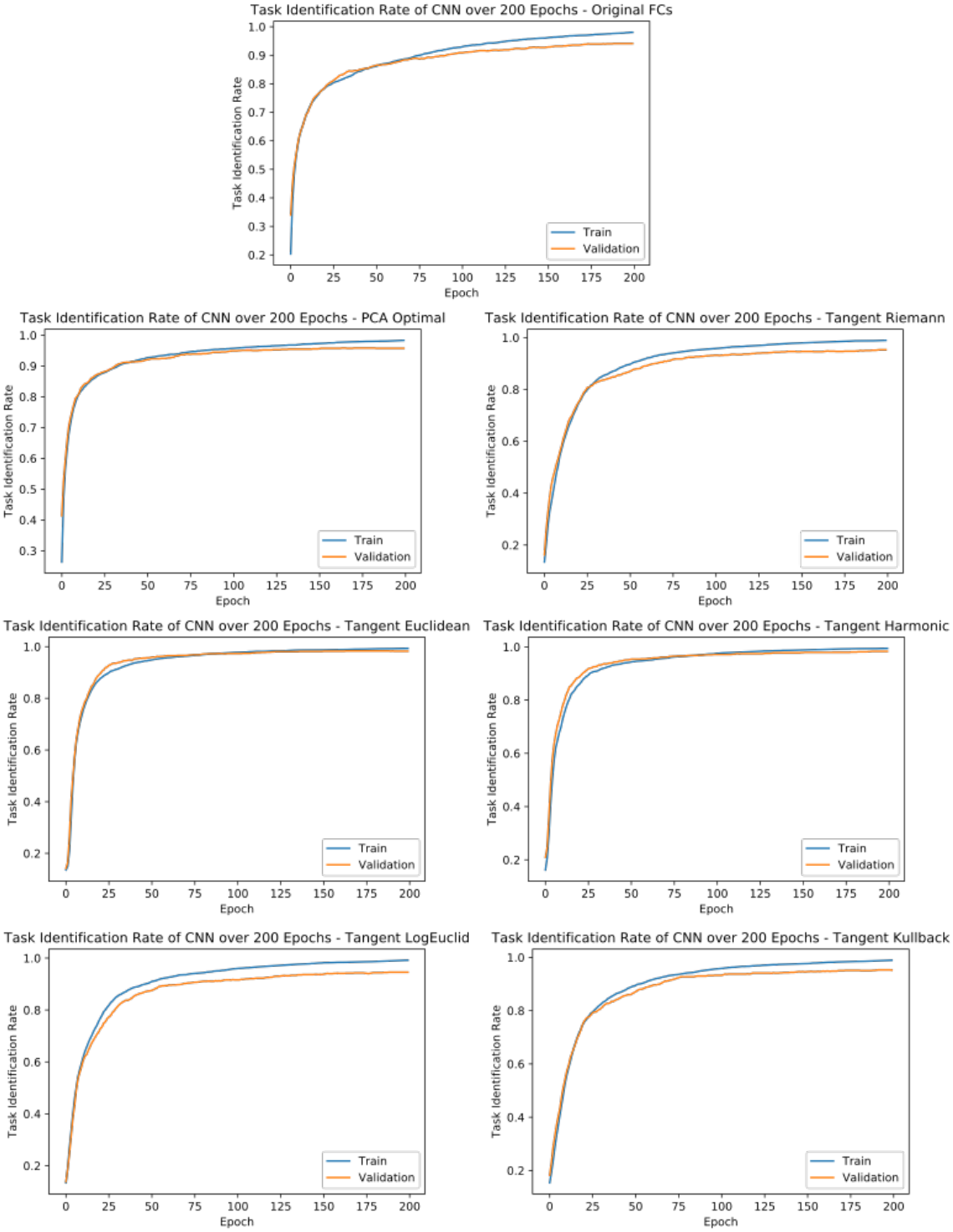


Fig. 1. Task identification rate progression during training of single instances of the CNN classifier on original function connectomes (FCs) and post-processed FCs over 200 epochs.

A confusion matrix of classification results of the same PCA reconstructed dataset with 80 PCs is is shown in Figure 2. The resting state task was predicted correctly most often, with only 1 false negative in motor task. However, other tasks such as Emotion and Relational saw many misclassifications. The most mixed up categories were Gambling scans predicted as Emotion and Working Memory scans predicted as Relational. Other pipelines saw very similar results in the labels that were mostly classified correctly and those that saw more misclassifications.

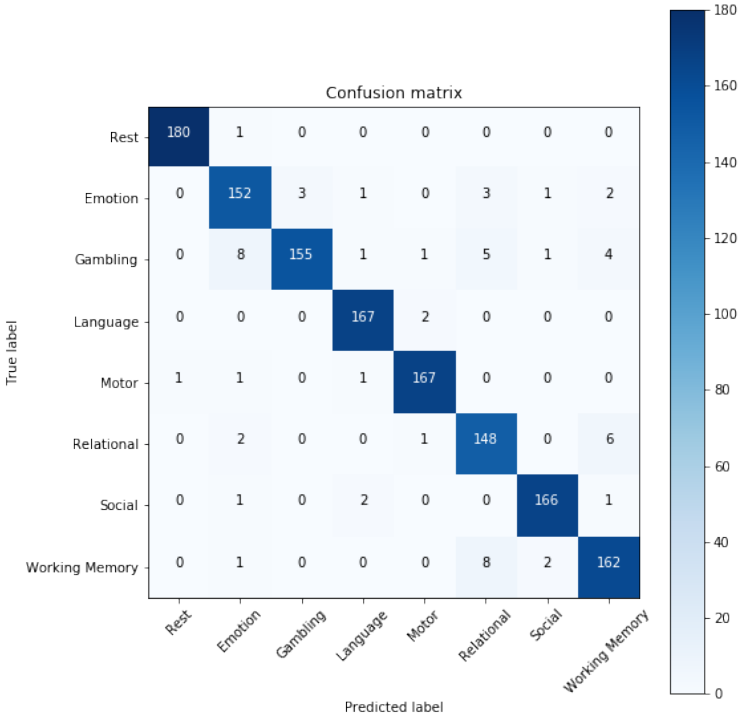


Fig. 2. Confusion matrix of CNN classification on Schaefer 100-region parcellated data after PCA reconstruction with 80 principal components