Studying Temporal Lobe Epilepsy using Machine Learning

By

Gyujoon Hwang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Medical Physics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2020

Date of final oral examination: 03/02/2020

The dissertation is approved by the following members of the Final Oral Committee:

M. Elizabeth Meyerand, Professor, Medical Physics and Biomedical Engineering

Rasmus M. Birn, Associate Professor, Psychiatry

Diego Hernando, Assistant Professor, Medical Physics and Radiology

Vivek Prabhakaran, Associate Professor, Medical Physics and Radiology

Oliver Wieben, Professor, Medical Physics

Abstract

Machine learning is changing the field of medical imaging. Studying complex neurological diseases like epilepsy can substantially benefit from its use. It can offer valuable insight onto the disease characteristics and also train predictive models to be used in various applications. Using both imaging and neuropsychological data provided by the Epilepsy Connectome Project, this work explores using machine learning to study temporal lobe epilepsy population in three steps. First, it exploits the feature extraction ability of machine learning to find that the frequency range between 0.1 – 0.073Hz is best at capturing abnormal resting-state functional connectivity in temporal lobe epilepsy compared to healthy controls, and that the impaired processing speed is the most informative among other neuropsychological tests in separating between the two groups. Second, it builds machine learning classification and regression models that can make various predictions on temporal lobe epilepsy patients. One finding reveals that temporal lobe epilepsy patients exhibit functional brains that are predicted to be on average 8.3 years older compared to their chronological ages. Third, the relationship between the sample size and binary classification accuracy is systematically explored using neuroimaging data. A number of guidelines are proposed for future research, as well as an equation for the sample size relationship that can be used to predict future accuracies given limited samples. Finally, it ends with suggestions of future research directions. Overall, this work presents how machine learning can facilitate epilepsy research and suggests ways that the limited sample size problems can be addressed.

Acknowledgements

First and foremost, I would like to thank God Almighty for giving me the wisdom, perseverance and strength thus far.

I wish to express my sincere appreciation to my advisors, Dr. Elizabeth Meyerand and Dr. Vivek Prabhakaran, for inviting me to be a part of the lab and for becoming my true mentors. They have shown faith in me and given me the freedom to pursue my research, while ensuring that I stay on course. I would not be here without their amazing leadership and persistent help.

I am grateful for my thesis committee, Dr. Rasmus Birn, Dr. Diego Hernando, and Dr. Oliver Wieben, for their support and guidance that have helped progress this work to completion.

I would like to pay my special regards to Dr. Bruce Hermann who has inspired amazing research ideas and mentored me through numerous projects. His experience and knowledge in the field have been critical resources to my work.

A special gratitude goes out to Veena Nair for her amazing dedication to the lab and her genuine care for her colleagues. I cannot even count the number of times I knocked on her door with questions, and every time, she would willingly spare her precious time and attention.

A big thank you also goes to all my fellow lab members, both past and present, Cole Cook, Rosaleena Mohanty, Charlene Rivera-Bonet, Neelima Tellapragada, Gengyan Zhao who have been there through my everyday struggles in the lab. It was a wonderful opportunity to work with such a delightful team of talented researchers.

Additionally, I would like to acknowledge the funding support from the National Institutes of Health grants U01NS093650 (Epilepsy Connectome Project), 1UF1AG051216-01A1 (Alzheimer's Disease Connectome Project), and U01MH93765 (Human Connectome Project).

I am forever indebted to my parents Younghun Hwang, Jihye Lee and also Heejin Yang and Jeonghee Hwang, whose wholehearted love and care have carried me through these past years.

I would like to dedicate this work to the best wife in the world, Hyeri Yang whose devoted support has resulted in this achievement. We have shared many special moments in our lives during the past four years, including our sons Louis and Logan, and through all of our spectacular life events, she has been the best friend and a loving supporter. Her prayers have been the source of my strength and blessings, which I would not have been here without.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADCP	Alzheimer's Disease Connectome Project
AED	Anti-Epileptic Drugs
ALFF	Amplitude of Low Frequency Fluctuations
AUC	Area-Under-the-Curve
CI	Cognitive Impairment
DVARS	Derivative of Variance Root-mean-Squared
ECP	Epilepsy Connectome Project
EEG	Electroencephalogram
fALFF	Fractional Amplitude of Low Frequency Fluctuations
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
HCP	Human Connectome Project
Lasso	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LFO	Low Frequency Oscillation
LOOCV	Leave-One-Out Cross Validation
MAE	Mean Absolute Error
MRI	Magnetic Resonance Imaging
NB	Naïve Bayes
NIH	National Institutes of Health
PCA	Principal Component Analysis

PCPS	Pattern Completion Processing Speed
RAVLT	Rey Auditory Verbal Learning Test
RMS	Root-Mean-Squared
RMSE	Root-Mean-Squared Error
RSFC	Resting State Functional Connectivity
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
TE	Echo Time
TLE	Temporal Lobe Epilepsy
TR	Repetition Time
WASI-II	Wechsler Abbreviated Scale of Intelligence-2

Chapter 1

Introduction

The concept of machine learning has existed for a long time, but it is now beginning to change the field of medicine[1-4]. Over time, many machine learning algorithms have been proven effective to perform complex classification and regression problems in many industrial fields[5]. The promise of machine learning is that given datasets of sufficient size, it will solve these complex problems more efficiently and with greater accuracy than the traditional approaches[6]. However, its applications in medical imaging are being developed at a much slower pace compared to in industrial fields and are largely limited to a few classes of problems such as image segmentation, registration or 2-dimensional image classification[3, 7-10].

Machine learning can be divided into two large branches: supervised and unsupervised learning[1]. Supervised learning starts with the goal of predicting a known output. Its models are trained from a set of input-output (feature-label), or annotated (labeled) datasets. The performance is then evaluated using the accuracy of predicting the outputs of unseen data. In contrast, in

unsupervised learning, there are no outputs to predict. Instead, the goal is to find naturally occurring patterns or groupings within the data. A method known as the cluster analysis is one of the most common algorithms of unsupervised learning[11]. The assessment of the performance is inherently more challenging due to the lack of ground truth, and is generally done by indirectly evaluating within-group similarities and between-group differences using other attributes.

Supervised machine learning can solve two types of problems: classification and regression. Classification is at the basis of any medical diagnosis problems. Currently, in radiology, we mostly rely on radiologists to look through medical images to make diagnoses. With advancements in medical imaging and the growing complexity of imaging modalities, it is becoming more and more difficult to make crucial decisions in a timely manner[4, 12]. Regression is another group of problems machine learning can solve[13]. This is when the degree of a continuous outcome is to be modeled, such as age, tumor grade, success rate of a surgery or the rate of a disease progression[14]. Since the outcome is continuous, regression problems may require more training dataset to create high-performing and reliable models, compared to classification problems where the outcome variable is categorical.

One of the most limiting factors in applying machine learning to medical imaging is the lack of sample size[15-19]. Acquiring and handling large amounts of medical data are difficult due to issues with recruitment, cost, storage, patient data privacy and more[20, 21]. Higher dimensional datasets that require large storage and working memory, such as high-resolution magnetic resonance imaging (MRI) or functional MRI (fMRI), are especially challenging to handle and to train machine learning models with. With small sample sizes, the performance and the reliability of a trained model are expected to be deficient, because of issues with overfitting[22, 23]. This limits utilizing machine learning to study diseases whose symptoms or biomarkers can only be captured

with high dimensional imaging modalities. For example, epilepsy is a neurological disorder that causes unprovoked, recurrent seizures in affected patients[24, 25]. Currently, diagnosing and characterizing patients with epilepsy primarily rely on high-dimensional imaging modalities such as electroencephalogram (EEG), magnetoencephalography (MEG)[25, 26], or MRI[27]. Acquiring enough imaging data from epilepsy patients to build reliable machine learning models is clinically, practically, and computationally challenging.

Machine learning research has two broad goals: first, to reveal useful patterns in a dataset related to solving specific problems, and second, to make accurate predictions of unseen data. These two goals must accompany each other. For example, imagine that a machine learning model has been trained and it has developed a certain algorithm to utilize a certain pattern of data in order to make its predictions. Even if the mechanism seems reasonable, if the model's prediction accuracies on unseen data are poor, it is doubtful. On the other hand, if a model shows good prediction accuracies, then the subsequent analyses of its underlying algorithms as well as the quality of the training dataset must follow, in order to verify that the good performance was not resulted from faulty algorithms or biased datasets. An ideal machine learning model shows good prediction results, as well as reveals informative patterns that can be reliably used to deliver future predictions. And, especially in medicine, it is much more preferred when the extracted patterns are understandable and at least partially co-align with prior clinical knowledge, which then potentially reveal the underlying biomarkers for the disease.

It is difficult to achieve satisfactory results with machine learning when sample sizes are small[15-19]. However, a powerful advantage of using machine learning is simply the ability of the model to improve itself with more data. Therefore, in machine learning research with limited sample sizes, not only the model performance, but also the potential of the model is important:

whether the model has room for improvements if more data points are recruited, or else it has reached its maximum potential. To make this assessment, it is necessary to study the relationship between the model performance and the sample size.

1.1 Specific Aims

This work studies the most common form of epilepsy in adults called temporal lobe epilepsy (TLE) using machine learning[28]. High-resolution MRI images as well as neuropsychological test data of TLE patients were acquired from the Epilepsy Connectome Project (ECP)[29, 30], sponsored by the National Institutes of Health (NIH). More details of this study and of the dataset can be found in Chapter 2. The specific aims addressed for the completion of this work are as follows:

1. *Investigate imaging and neuropsychological biomarkers of TLE using machine learning,*
2. *Build machine learning models to make clinical predictions on TLE patients,*
3. *Investigate the relationship between machine learning performance and the sample size.*

1.2 Thesis Outline

In accordance with the aforementioned aims of this work, the remainder of the thesis will be structured as follows.

- **Chapter 2** introduces the main research project (ECP) where data from the TLE patients as well as most of the healthy control volunteers were taken from. The first section provides

a brief introduction of TLE. The following sections summarize the project aims and its acquired dataset.

- **Chapter 3** discusses using machine learning to search for biomarkers of TLE patients. The first section searches for biomarkers in resting-state fMRI images and the second section in neuropsychological test results. The following sections then discuss the potential discordance between diagnostic methods developed by humans and by machine learning.
- **Chapter 4** discusses using machine learning for building prediction models on TLE patients. The first section introduces the proper methods to assess the generalizability of a model performance. Then, in the following sections, some examples of building both classification and regression machine learning models are introduced using TLE patients' data.
- **Chapter 5** explores the relationship between machine learning binary classification performance and the sample size. First, the relationship is investigated with respect to a number of machine learning training hyperparameters, and then, an equation is fitted to study the trends. Based on the findings, a number of research directions are proposed for future machine learning research with limited sample sizes.
- **Chapter 6** provides a final discussion of the takeaway points from this work and discusses potential directions for future work.

Chapter 2

Epilepsy Connectome Project (ECP)

A recent NIH-sponsored project known as the Human Connectome Project (HCP)[31] which ended in 2018 laid out a thorough neuroimaging blueprint of young adults. It collected a comprehensive MRI and neuropsychological data from 1,200 healthy young adults between ages of 22 and 35. Then the focus moved towards finding abnormalities in patient populations compared to this normative dataset available. More than a dozen Connectome projects related to human disease were launched and applied HCP-style data collection protocols[32, 33]. One of these sister studies known as the Epilepsy Connectome Project (ECP) investigated temporal lobe epilepsy (TLE) population[29, 30]. Most of the studies included in this work involve data from the ECP and therefore, in this chapter, a brief introduction to the study is provided.

2.1 Temporal Lobe Epilepsy (TLE)

Epilepsy, a brain disorder characterized by recurring seizures, affects an estimated 1.2% of the United States population (3.4 million persons) and is associated with a high risk of cognitive and psychosocial dysfunction, and enormous healthcare costs[24, 34]. The number of affected people

worldwide is 50 million, which is expected to increase further due to the rising life expectancy and the increasing proportion of people surviving epilepsy-provoking insults, such as birth trauma, traumatic brain injury (TBI), brain infection and stroke[35]. Even with adequate diagnosis and treatment, 30 – 40% of epilepsy patients still experience recurring seizures that are uncontrolled by medication[34, 36], who are then considered to have refractory epilepsy. Powerful imaging tools are now available for quantitatively characterizing the structural and functional connections between brain regions that make up epileptic networks[37-39], providing a promising new approach for understanding, predicting, and treating refractory epilepsy.

TLE is the most common form of epilepsy in adults, and the largest group among those with medically refractory seizures[28]. It is characterized by seizure activities emanating from the temporal lobe, which is where the most damage to the structural brain occurs, although this damage can extend to thalamus, insula, and other cortical regions[40-42]. Resting-state fMRI analyses have also demonstrated both temporal and extra-temporal functional connectivity abnormalities[43, 44]. Chronic TLE is associated with abnormalities in cognition, brain structure and brain connectivity in midlife[45-48]. Finding reliable biomarkers is crucial in prevention therapy and drug development, but has so far only been modestly successful[49, 50].

2.2 Enrollment Criteria

The ECP (grant number U01NS093650) is a two-site, prospective research project based in the Medical College of Wisconsin and the University of Wisconsin-Madison. The enrollment period spanned from 2015 and ended in 2019. The Medical College of Wisconsin and Froedtert Hospital Institutional Review Board approved the use of human participants for this study. All participants provided written informed consent prior to their participation.

TLE patients were enrolled if they were between the ages of 18 and 60 (inclusive), had tested full-scale intelligence quotient (IQ) at or above 70, spoke English fluently, with no medical contraindications to MRI. The diagnosis of TLE was supported by two or more of the following: 1) described or observed clinical semiology consistent with seizures of temporal lobe origin, 2) EEG evidence of either temporal intermittent rhythmic delta activity (TIRDA) or temporal lobe epileptiform discharges, 3) temporal lobe onset of seizures captured on video EEG monitoring, or 4) MRI evidence of mesial temporal sclerosis or hippocampal atrophy. Patients with any of the following were excluded: 1) lesions other than mesial temporal sclerosis causative for seizures, and 2) an active infectious/autoimmune/inflammatory etiology of seizures. The TLE group was a combination of refractory and better-controlled patients (45% reported having at least one seizure during the past year).

The controls were healthy adults between the ages of 18 and 60. Exclusion criteria included: Edinburgh laterality (handedness) quotient less than +50; primary language other than English; history of any learning disability, brain injury or illness, substance abuse, or major psychiatric illness (major depression, bipolar disorder, or schizophrenia); current use of vasoactive medications; and any medical contraindications to MRI.

2.3 Data Types

2.3.1 Neuropsychological Assessment

All controls and TLE patients underwent neuropsychological evaluation targeting assessment of intelligence, language, visuoperceptual/constructional skills, learning and memory, executive functions, and cognitive/psychomotor speed. A total of 18 cognitive indices resulted which included assessment of intelligence (Wechsler Abbreviated Scale of Intelligence-2 [WASI-

II] Vocabulary and Block Design subtests)[51], verbal learning and memory (Rey Auditory Verbal Learning Test [RAVLT]) including total words learned across trials[52], object naming (Boston Naming Test)[53], letter fluency (Controlled Oral Word Association Test)[54, 55], semantic fluency (Animal Naming)[55, 56], spatial orientation (Judgement of Line Orientation)[57], face recognition (Facial Recognition Test)[57], speeded fine motor dexterity (Grooved Pegboard, dominant and non-dominant hands)[58], and selected subtests from the NIH Toolbox-Cognitive Battery including the Pattern Comparison Processing Speed (PCPS)[59, 60], Dimensional Change Card Sort, List Sorting Working Memory, Flanker Inhibitory Control and Attention, Picture Vocabulary, Oral Reading Recognition, and Picture Sequence Memory tests.

2.3.2 Neuroimaging

MRI was performed on 3T GE (General Electric) 750 scanners at both institutions. T1-weighted structural images were acquired using magnetization prepared gradient echo sequence (MPRAGE, repetition time [TR]/echo time [TE] = 604ms/2.516ms, inversion time = 1060.0ms, flip angle = 8°, field-of-view = 25.6cm, voxel size = 0.8mm isotropic). Cube T2-weighted structural images were also acquired (TR/TE = 2,500ms/94.641ms, flip angle = 90°, field-of-view = 25.6cm, 0.8mm isotropic).

Resting-state fMRI images were acquired using whole-brain simultaneous multi-slice imaging[61] (8 bands, 72 slices, TR/TE = 802ms/33.5ms, flip angle = 50°, matrix = 104 ×104, field-of-view = 20.8cm, voxel size = 2.0mm isotropic) and a Nova 32-channel receive coil. The participants were asked to fixate on a white cross at the center of a black screen during the scans for better reliability[62]. Time-series from four 5-minute resting-state fMRI scans acquired in a single session were concatenated for more reliable analysis[63].

2.4 Neuroimaging Data Processing

2.4.1 Preprocessing

Imaging data were pre-processed using the HCP minimal processing pipelines version 3.4.0[64] which is primarily based on FreeSurfer[65] and FSL (FMRIB Software Library)[66]. In brief, the function of this pipeline is to nonlinearly register T1- and T2-weighted images to the MNI (Montreal Neurological Institute) space, segment the volume into predefined structures, reconstruct white and pial cortical surfaces, and perform FreeSurfer's standard folding-based surface registration to a surface atlas (the "fsaverage" template). The functional portion of the pipelines removes nonlinear spatial distortions using spin echo unwarping maps, realigns volumes to compensate for subject motion, registers the fMRI data to the structural images, reduces the bias field, normalizes the 4D image to a global mean, masks the data with the final brain mask and maps the voxels within the cortical gray matter ribbon onto the native cortical surface space. More details on the HCP processing pipelines can be found in Glasser et al.[64]

2.4.2 Glasser Parcellation

The Glasser parcellation atlas[67] was used for studying resting-state fMRI images throughout this study. This parcellation is a recent development from the HCP consortium for surface-based morphometry. It consists of 180 cortical parcels per hemisphere. These parcels were delineated using a multi-modal approach and the authors reported that the parcellation is highly reproducible[67]. One limiting factor is that this excludes subcortical brain regions. Therefore, 19 subcortical regions from the FreeSurfer subcortical segmentation[68] were additionally analyzed: a total of 379 regions of interest per brain.

Chapter 3

Searching for Biomarkers of Temporal Lobe Epilepsy using Machine Learning

When machine learning results are reported, what catches the public eyes is the prediction accuracy. Whether or not a machine can perform better in a particular task compared to humans is typically the question that generates excitement and disappointment from the crowds. However, what is almost equally important is how the machine was able to achieve the superior accuracy. A machine learning research not only focuses on producing the best performance, but also investigates the underlying features that drive the performance. In medical imaging or medicine in general, this is especially important, because this may reveal important biomarkers of a disease that humans were not able to decipher. Then the underlying features highlighted by machine learning models can be compared with the prior clinical knowledge to better characterize the disease in question. This chapter provides a few examples of using machine learning for searching for biomarkers of TLE patients.

3.1 Resting-state Functional Connectivity[§]

The human brain is a complex dynamic system characterized by spontaneous oscillations in multiple frequency bands[69]. Traditionally, the analysis of resting-state fMRI data focused on the low-frequency oscillation range (LFO; 0.01 – 0.1Hz, although exact cutoffs vary slightly), because the signals in this range seemed to be less contaminated by low/high frequency noise and to capture relevant resting-state information[70]. Some investigators have tested narrower frequency bands within and around the LFO, labeled Slow-5 (0.01 – 0.027 Hz), Slow-4 (0.027 – 0.073 Hz), Slow-3 (0.073 – 0.198 Hz) and Slow-2 (0.198 – 0.50 Hz) by Buzsáki et al.[69]. Zuo et al. suggested that the Slow-5 and Slow-4 bands reflect signal changes from the gray matter, while Slow-3 and Slow-2 signal changes from the white matter[71] (**Figure 1**). A recent work by Gohel & Biswal revealed that functional integration between brain regions at rest occurs in multiple frequency bands[72].

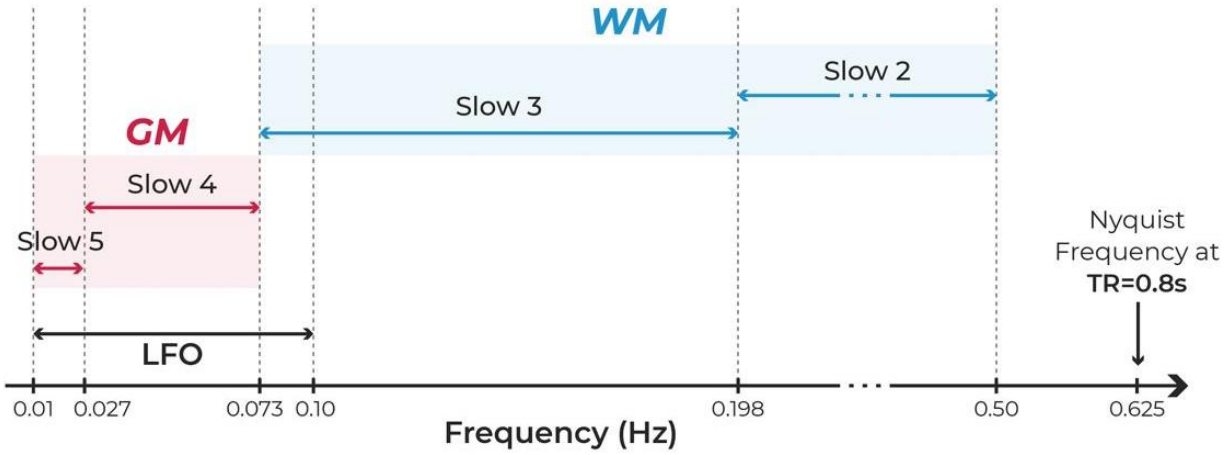

Figure 1. Functional MRI Frequency Bands. It has been suggested that the Slow-5 and Slow-4 bands reflect signal changes from the gray matter (GM), while Slow-3 and Slow-2 from the white matter (WM)[71]. 0.625Hz is the highest frequency that can be captured by a functional MRI scan with the repetition time (TR) of 0.8 seconds. †LFO = low frequency oscillations.

[§] Portions of this work have been published in: Hwang G, Nair VA, Mathis J, Cook CJ, Mohanty R, Zhao G, Tellapragada N, Ustine C, Nwoke OO, Rivera-Bonet C, Rozman M, Allen L, Forseth C, Almane DN, Kraegel P, Nencka A, Felton E, Struck AF, Birn R, Maganti R, Conant LL, Humphries CJ, Hermann B, Raghavan M, DeYoe EA, Binder JR, Meyerand E, Prabhakaran V. Using low-frequency oscillations to detect temporal lobe epilepsy with machine learning. Brain Connect. 2019;9(2):184-93

Based on these findings, our hypothesis was that seizure activity in TLE patients, which generally occurs at much higher frequencies than these slow bands, produces alterations in grey matter connectivity that can be detected at lower frequencies with fMRI. Since the raw voxel-based signal data are 4-dimensional, highly complicated, and very large in size, three summary measures were calculated: resting-state functional connectivity (RSFC)[70], amplitude of low frequency fluctuations (ALFF)[70, 73], and fractional ALFF (fALFF)[74]. RSFC measures correlations between blood-oxygen-level dependent (BOLD) time series of two brain regions, while ALFF and fALFF capture intensity-based measures of the signal changes at a single region of interest. The goal was to reveal which combinations of a resting state measure and a frequency band capture the most valuable information to discriminate between TLE patients and healthy controls.

Previous studies that investigated these measures in TLE patients reported abnormalities in different regions of the resting brain. These abnormalities include decreased RSFC within the epileptic temporal lobe, between hippocampi, and between the hippocampus and the orbito-frontal region[75], and increased RSFC in the lateral portions of the non-epileptic hemisphere[76]. Zhang et al.[77] reported that TLE patients with medial temporal sclerosis (a common structural abnormality in TLE) show increased ALFF in the medial temporal lobe and thalamus and decreased ALFF in the default-mode network. A difference in fALFF was noted between left and right TLE patients in the thalamus[78].

To create a reliable machine learning model, one needs to select an informative set of features for training, then narrow this set down to key components[79, 80]. Therefore, knowing what information is useful is essential, but typically difficult to determine *a priori*. In this section, 20 different combinations of resting fMRI measures and frequency bands were examined for the machine learning training. Note that it is not necessary to consider the "All" band with fALFF,

because fALFF is defined as ALFF of a specific frequency band over that of the "All" band. A feature selection method using a least absolute shrinkage and selection operator (Lasso)[81] was employed to remove uninformative features[82, 83].

3.1.1 Participants

Data from 60 TLE patients (mean age = 39.5 ± 12.0 years, 34 females, five left-handed, epilepsy duration = 18.7 ± 14.4 years, 38 drug-resistant TLE), and 59 healthy controls (mean age = 36.0 ± 14.4 years, 32 females, all right-handed) were analyzed. The two groups did not differ in the mean age (p = 0.16, two-tailed t-test), and gender ratio (p = 0.79, Chi-squared test). The patient group consisted of 29 individuals with left TLE, 15 with right TLE, and four who had bilateral onsets based on either interictal EEG, imaging (hippocampal sclerosis) or ictal monitoring. Twelve patients had uncertain lateralization. To closely match the mean age and gender ratio between the TLE and control samples, 12 of the healthy control data were taken from the Alzheimer's Disease Connectome Project (ADCP)[84], which used the same set of MRI scanners and the same imaging protocols for structural and resting-state fMRI scans as the ECP. ADCP is also an NIH-sponsored disease Connectome project that launched in 2016 with aims to study populations with Alzheimer's disease and mild cognitive impairment (MCI). The Medical College of Wisconsin Institutional Review Board has approved the use of human participants for ADCP and the sharing of de-identified datasets from this study.

3.1.2 Data Processing

In addition to the pre-processing described in Section 2.4.1, additional processing was performed using AFNI (Analysis of Functional NeuroImages)[85], which included motion regression using 12 motion parameters, and regression-based removal of signal changes in the white matter,

cerebrospinal fluid, and the global signal. Bandpass filtering was applied to select frequency bands of interest: Slow-2, Slow-3, Slow-4, Slow-5, Slow-4+5 (covering both Slow-4 and Slow-5), LFO, and All (no bandpass filtering; approximately 0.00 – 0.62Hz) (**Figure 1**). 379 time series signals from the combined parcellation scheme described in Section 2.4.2 have been extracted per subject.

Pairwise Pearson correlations were computed to generate RSFC matrices to be used as machine learning training features. For ALFF, the filtered resting-state fMRI signals in the time domain were Fourier transformed to the frequency domain, and the mean of the square root values within the frequency range of interest was calculated[73]. fALFF was calculated as the ALFF of the selected frequency range over the ALFF of the All range[74]. For ALFF and fALFF, the number of possible features in the training was 379 for each.

3.1.3 Motion Outliers

Resting-state fMRI images can be heavily affected by subject motion in the scanner[86]. However, it is also not desired to build a classifier model based on highly selected data, because an ideal model should be able to classify participants despite moderate levels of motion in the scanner. To achieve this, the machine learning model needs to be exposed to a sufficient number of data points contaminated by motion. One needs to be cautious, however, not to train a model that classifies based on the differences between high and low motion, instead of between TLE patients and healthy controls.

Therefore, instead of performing a rigorous motion scrubbing, we used three different motion metrics to determine if a MRI run was acceptable: relative mean root-mean-squared (RMS), absolute mean RMS, and derivative of variance RMS (DVARS)[86, 87]. These are common quality control measures for resting-state fMRI scans, where the RMS's measure pure subject motion, while DVARS measures the combination of motion and the scanner instabilities. These three

motion measures were calculated per each run of five minutes, and transformed into the standard scores. Subjects who had z > 3 on any of the three measures in any of the four runs being concatenated were defined as motion outliers (**Figure 2**).

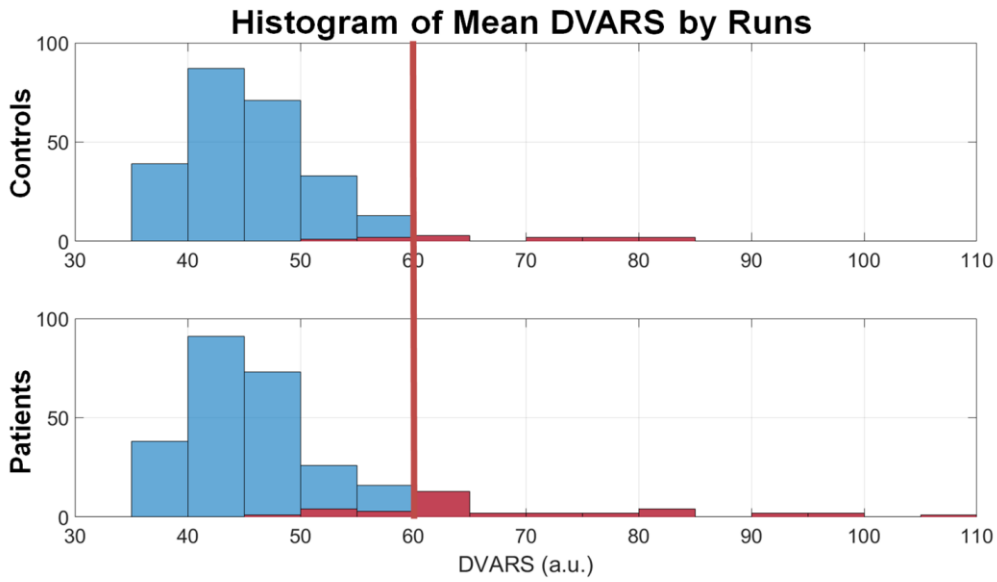

Figure 2. Histogram of Mean DVARS by Runs. Most subjects showed acceptable in-scanner motion, while a few outliers existed, with respect to a metric called the derivative of variance root-mean-squared (DVARS). Subjects with high in-scanner motion were excluded based on a criteria described in Section 3.1.3.

3.1.4 Machine Learning Models

All machine learning analyses in this section were done in MATLAB R2016a with the Statistics and Machine Learning Toolbox[88]. Three different binary classifiers were examined: support vector machine (SVM)[89], linear discriminant analysis (LDA)[90], and naïve Bayes (NB)[91] classifiers. These three traditional classifiers were trained instead of one to get a general sense of the expected machine learning classification performance.

Leave-one-out-cross-validation (LOOCV) was used to estimate model performance[92]. In each LOOCV loop, one participant was taken out and the machine learning model was trained with $N - 1$ participants. Then the left out participant was used as a testing sample for the trained

model. This procedure was repeated until every participant had been left out once. The classification performance was averaged to give the LOOCV accuracy. This method is known to give the most unbiased estimate of the test error and is a good method for small sample cases[92, 93], which will be discussed more in Section 5.2.5. Receiver operating characteristic area-under-the-curve (AUC) was also computed by adjusting the misclassification cost function during the training. A random classifier would give 50% LOOCV accuracy with AUC = 0.5.

3.1.5 Feature Selection

To reduce feature dimensionality, Lasso regression analysis was performed on the training set in each cross validation loop, with the regularization coefficient (lambda) at 0.1[81, 82]. Only features with non-zero Lasso coefficients were used in the training of the machine learning models. This technique was selected over other common feature selection techniques such as principle component analysis (PCA)[94] in order to preserve the original features in the training. Features that received non-zero coefficients in all 119 cross validation loops were marked for further analysis. Recursive feature elimination[95] was employed within each loop based on the Lasso coefficients to further reduce the dimensionality.

3.1.6 Results

The highest LOOCV classification accuracies using RSFC were in the low to mid-80%, with the AUC close to 90%. The highest cross validation accuracies were only in the mid-70% using ALFF and fALFF measures. These results are summarized in **Table I**.

Using RSFC, the Slow-4+5 band produced the best overall model performance in classifying the TLE patients and healthy controls, with around 83% LOOCV accuracy. The highest

cross validation accuracies from the three machine learning models were also consistent: 83.2 ± 1.4%. Using ALFF and fALFF, LOOCV accuracies were not as consistent as using RSFC.

19 Slow-4+5 RSFC features were selected by the Lasso feature selection every time in all 119 cross validation loops, and these are summarized in

Table II, as well as shown in **Figure 3**. Only 5 of these features also received significant *p*-values (Benjamini-Hochberg false discovery rate [FDR] corrected based on the standard alpha level of 0.05)[96] from the group *t*-test. Connection between right fusiform face complex (R_FFC) and right area posterior 9-46v (R_p9-46v, a part of Brodmann area 46) was the most significant feature based on both Lasso and *t*-test (corrected $p < 0.001$) analyses, and was stronger (less negative correlation) in TLE compared to the healthy group (**Figure 4**).

The 19 significant features did not include any exclusively temporal lobe connections. Repeating the analysis using only Slow-4+5 RSFC within the temporal lobe (24 regions, 276 connections), the maximum LOOCV accuracy was only 68.9%.

Table I. Classification Results of Separating TLE Patients and Healthy Controls. The three resting-state measures and seven frequency bands tested are organized in the two leftmost columns. The three traditional machine learning models trained are organized in the top row. The accuracies are the LOOCV accuracies. "Features" columns indicate the number of features selected from the recursive feature elimination feature selection. Best LOOCV accuracies were achieved with Slow-4+5 RSFC features. †LFO = low frequency oscillations. AUC = area under the curve.

Measure	Frequency Band	SVM			LDA			NB		
		Accuracy	AUC	Features	Accuracy	AUC	Features	Accuracy	AUC	Features
RSFC	Slow2	57.14	0.52	57	63.87	0.62	3	61.34	0.60	3
	Slow3	65.55	0.67	14	63.03	0.62	12	63.03	0.61	21
	Slow4	52.10	0.44	3	53.78	0.43	3	52.94	0.42	3
	Slow5	75.63	0.80	10	75.63	0.80	8	76.47	0.79	11
	Slow4+5	**84.87**	**0.86**	**31**	**81.51**	**0.86**	**36**	**83.19**	**0.88**	**29**
	LFO	72.27	0.72	5	69.75	0.71	5	73.95	0.79	60
	All	72.27	0.72	37	69.75	0.73	27	68.07	0.69	26
ALFF	Slow2	52.94	0.43	25	53.78	0.49	34	57.98	0.56	33
	Slow3	63.03	0.59	4	67.23	0.69	1	66.39	0.67	1
	Slow4	69.75	0.71	17	68.91	0.72	17	69.75	0.73	17
	Slow5	**78.99**	**0.81**	**11**	77.31	0.81	13	**73.11**	**0.76**	**12**
	Slow4+5	64.71	0.65	3	67.23	0.68	3	64.71	0.67	3
	LFO	73.95	0.72	14	**78.15**	**0.81**	**14**	69.75	0.72	15
	All	62.18	0.56	6	61.34	0.61	8	63.87	0.64	10
fALFF	Slow2	53.78	0.46	22	59.66	0.58	2	60.50	0.57	2
	Slow3	54.62	0.42	12	**73.11**	**0.78**	**6**	**72.27**	**0.75**	**6**
	Slow4	64.71	0.55	2	63.87	0.64	2	65.55	0.65	2
	Slow5	**70.59**	**0.70**	**6**	68.07	0.70	6	64.71	0.69	2
	Slow4+5	56.30	0.46	6	55.46	0.53	25	56.30	0.53	24
	LFO	58.82	0.50	16	55.46	0.54	3	63.03	0.57	20

Table II. Slow-4+5 Features Repeatedly Selected by Lasso Feature Selection. These 19 Slow-4+5 connections were selected by Lasso feature selection repeatedly in all 119 cross validation loops. Features with positive Lasso weights were stronger in TLE patients, and vice versa for those with negative weights. Only 5 out of 19 features showed significant group differences based on the *t*-test. Feature names and abbreviations follow the nomenclature in Glasser parcellation[67]. †FDR = false discovery rate.

	Lasso Features					*t*-test
No	**From**		**To**		**Lasso Weight**	***p*_FDR**
1	R_Fusiform Face Complex	R_FFC	R_Area posterior 9-46v	R_p9-46v	0.713	<0.001***
2	R_RetroSplenial Complex	R_RSC	R_Area 46	R_46	-0.622	0.003**
3	L_Entorhinal Cortex	L_EC	L_Area V3CD	L_V3CD	0.584	> 0.1
4	L_Area 11l	L_11l	L_Area 52	L_52	0.575	> 0.1
5	R_RetroSplenial Complex	R_RSC	R_Rostral Area 6	R_6r	-0.552	> 0.1
6	L_Area 2	L_2	R_Area PGs	R_PGs	0.451	0.087
7	R_Area posterior 9-46v	R_p9-46v	R_Area PH	R_PH	0.377	0.005**
8	L_Area 9 anterior	L_9a	L_Amygdala	L_Amygdala	-0.376	0.087
9	L_VentroMedial Visual Area 3	L_VMV3	R_Frontal Opercular Area 4	R_FOP4	0.336	0.063
10	L_Area 23d	L_23d	L_Hippocampus	L_H	-0.297	0.095
11	L_Area OP4/PV	L_OP4	L_Area PFt	L_PFt	-0.292	0.039*
12	L_Ventral Visual Complex	L_VVC	R_Ventral Visual Complex	R_VVC	-0.290	0.002**
13	R_Fusiform Face Complex	R_FFC	R_Medial Belt Complex	R_Mbelt	0.274	> 0.1
14	L_Area anterior 10p	L_a10p	R_Area IFJa	R_IFJa	-0.260	> 0.1
15	L_Premotor Eye Field	L_PEF	L_ParaBelt Complex	L_Pbelt	-0.231	> 0.1
16	L_Primary Auditory Cortex	L_A1	R_Area 8Av	R_8Av	-0.225	> 0.1
17	R_Area IntraParietal 2	R_IP2	R_Area PGs	R_PGs	-0.214	> 0.1
18	R_Medial Area 7P	R_7Pm	R_Area 5m	R_5m	0.201	> 0.1
19	R_Area 31p ventral	R_31pv	R_Area 46	R_46	-0.191	0.057

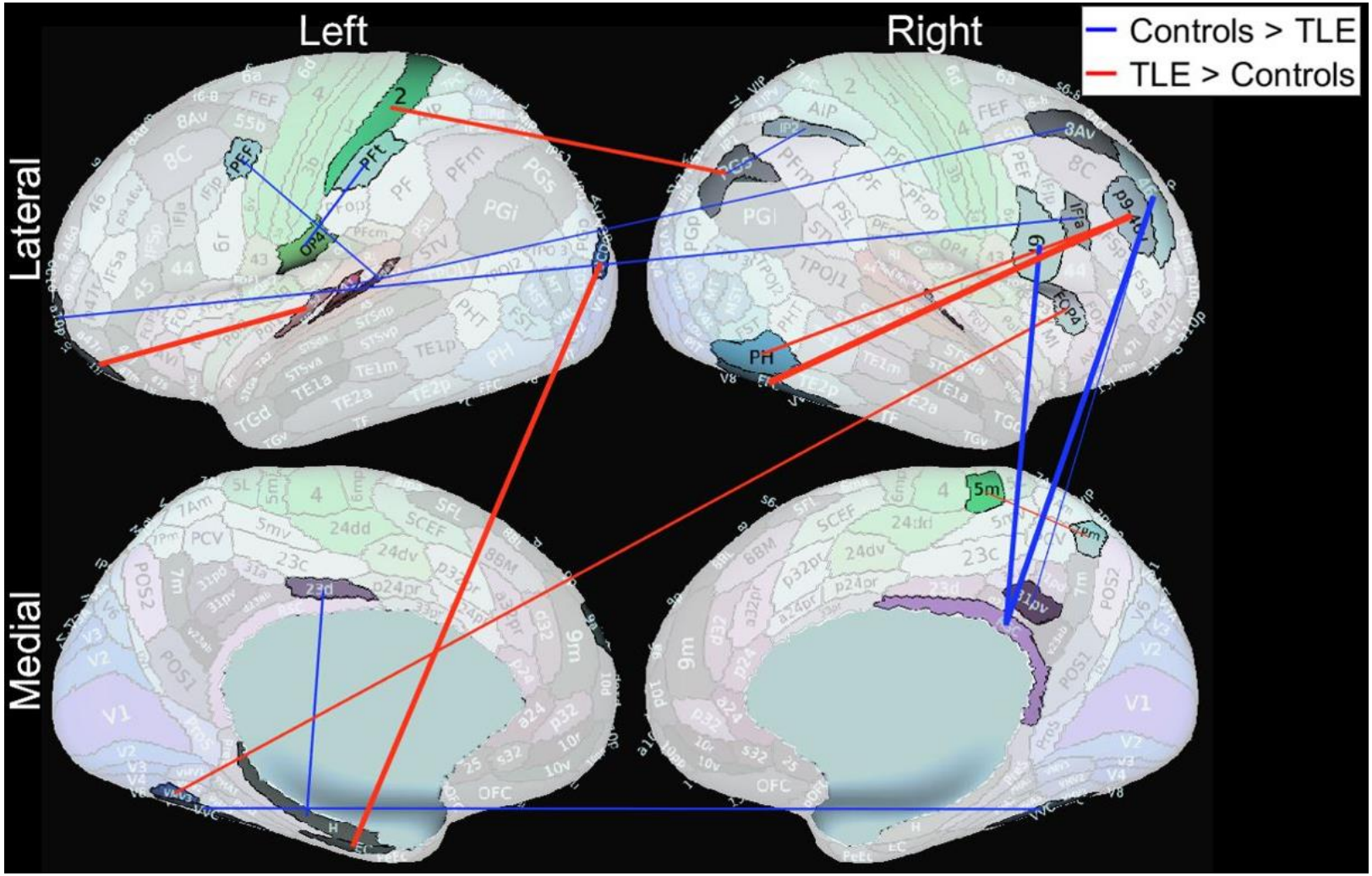

Figure 3. Slow-4+5 Features Repeatedly Selected by Lasso Feature Selection. 18 significant Slow-4+5 RSFC cortical features based on Lasso feature selection are shown. Subcortical connection between left area 9 anterior (L_9a) and left amygdala is not shown in the picture from

Table **II**. This picture suggests that the changes in the temporal lobe epilepsy (TLE) brains are throughout the whole brain, not only in the temporal lobes. The background brain images have been generated with the Connectome Workbench and show Glasser parcellation.[67]

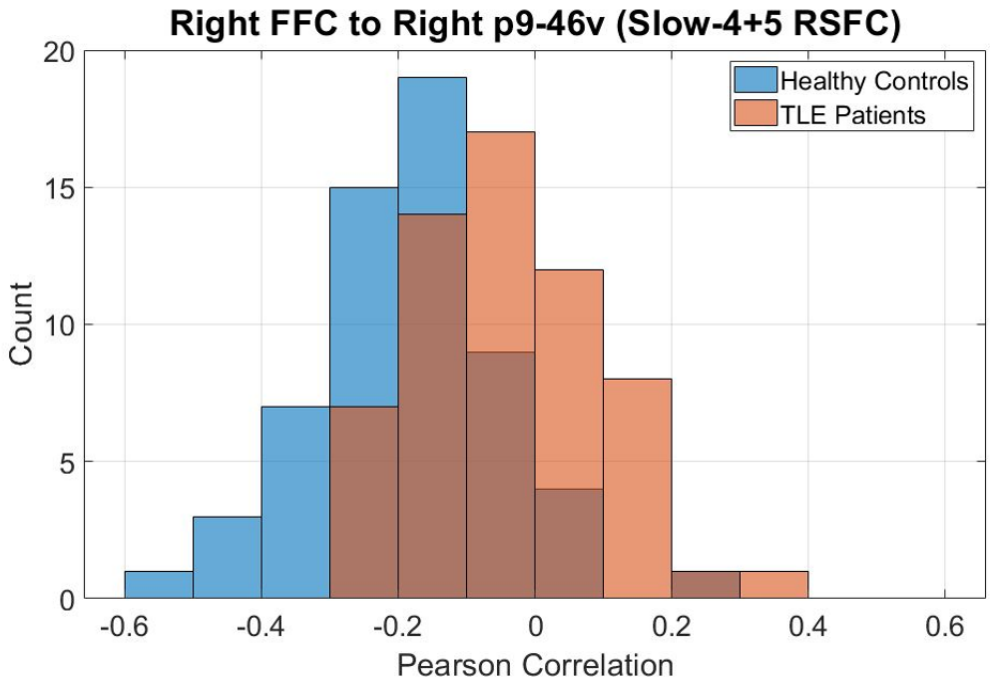

Figure 4. Most Contributing Connection to Classification Model. This histogram shows the distributions of pearson correlations between signals from right fusiform face complex (Right FFC) and right area posterior 9-46v (Right p9-46v). An increased correlation, or decreased negative correlation, was found in the temporal lobe epilepsy (TLE) group, and this was the most significant feature based on both Lasso and *t*-test analyses.

3.1.7 Discussion

Seven frequency ranges with three different measures of the resting functional brain signals were used to train three different traditional machine learning models. This extensive search for good training features was an attempt to cover all possible measures using the resting-state fMRI images. In brief, the results suggest that functional brain alterations in the TLE patients are indeed detectable and are captured best by RSFC using the Slow-4+5 range. The machine learning models were able to use this information to separate TLE patients from age- and gender-matched healthy controls in our samples with approximately 83% cross validation accuracy (more discussion on the difference between a cross validation accuracy and a test accuracy found in Section 4.1)[97]. Also notably, the features separating between the TLE patients and healthy controls were located

throughout the entire brain, and not just within the temporal lobe, which is consistent with previous findings[98, 99].

There have been many papers in the recent literature describing the development of reliable machine learning models to make more accurate decisions from complex clinical datasets. For example, there are reports on using machine learning to predict post-surgical outcome of TLE patients using non-imaging data[100], structural MRI data[101, 102], or intracranial EEG[103]. Machine learning has also been applied in determining the lateralization of TLE seizure focus, based on resting-state fMRI[78] or positron emission tomography (PET)[104], and also in separating TLE patients and healthy controls using structural imaging[105], diffusion imaging[106], or both[107]. It was also applied in separating epilepsy patients overall and healthy controls using RSFC[108, 109].

Without machine learning, or a similar automated method, humans are limited in their ability to comprehend high-dimensional data, especially when the patterns are complex. Also, the true nature of a clinical question may be significantly non-linear than one may assume at the outset. Instead of trying to extract multi-dimensional patterns from a complex set of features by hand, one can consult machine learning models.

One of the biggest limitations of training traditional machine learning models is the need to select input features. In most cases, we do not know *a priori* what combination of features would contain the most useful information for the models. In this study, a Lasso-based feature selection method along with recursive feature elimination was employed, in order to preserve original features for the feature analysis. There are other feature reduction methods that aim to maximize classification accuracies (more discussion on feature reduction in Section 5.2.2). At present, identifying the best set of features and the correct non-linearity of the model (or kernel) remains a trial-and-error process. It is advisable to think broadly, considering a wide range of potential

features available, while actively narrowing it down so that the models are not clogged with noisy information.

In this section, three traditional machine learning models were trained, in order to get a general sense of the expected machine learning classification performance using traditional techniques. The best overall cross validation accuracy was achieved with the Slow-4+5 RSFC features and it was very comparable between the three classifiers. These traditional models are more straight-forward and understandable compared to highly non-linear models such as deep learning. Therefore, they allow us to easily analyze the underlying features contributing the most to the models.

The set of Slow-4+5 RSFC features that contributed most to the models, as visualized in **Figure 3**, suggested widespread functional connectivity alterations in TLE patients. This list included no exclusively temporal lobe connections, perhaps due to the heterogeneity of our TLE patient group. It is notable that using the whole-brain connectivity yielded better classification results, compared to using temporal lobe connections alone. The decreased negative connection between right fusiform face complex and right area posterior 9-46v found in TLE patients is consistent with the findings of Riley et al., who reported altered functional connectivity of the cortical face processing networks in TLE[110]. Abnormal structure and function of the retrosplenial cortex have also been reported[111, 112]. These results are promising for future applications of machine learning in diagnosing and understanding the basic pathophysiology of TLE.

3.2 Neuropsychological Test[§]

The domains of memory, language and executive function are among the most studied cognitive complications of the epilepsies[45, 113, 114] with an increasing number of imaging investigations focused on the disrupted regions and networks associated with these cognitive anomalies[115-120]. Psychomotor slowing is also a common but arguably less investigated cognitive abnormality of the epilepsies. While known to be exacerbated by many anti-epileptic drugs (AED)[121-123], cognitive and/or psychomotor slowing is evident in new onset adult and pediatric patients prior to administration of AEDs[124, 125], and has been observed to persist following remission of epilepsy and cessation of medication treatment[126, 127]. Thus, cognitive and psychomotor slowing is an inherent neuropsychological morbidity of the epilepsies.

The relative salience of slowed processing speed relative to other potential cognitive abnormalities in epilepsy remains uncertain. Abnormalities in memory, language and executive function are of clear importance, but the set of abnormalities that most reliably discriminates persons with epilepsy compared to healthy controls, and the role of slowing of processing speed in this discrimination, remains to be determined. To address this issue we utilize machine learning to characterize the relative power of various cognitive abilities, including processing speed, to classify or discriminate patients with epilepsy compared to controls. As machine learning builds multidimensional models using multiple variables, it offers the ability to analyze neuropsychological measures together as a group, instead of individually. For example, a combination of several, individually non-significant features may classify two groups better than

[§] Portions of this work have been published in: Hwang G, Dabbs K, Conant L, Nair VA, Mathis J, Almane DN, Nencka A, Birn R, Humphries C, Raghavan M, DeYoe EA, Struck AF, Maganti R, Binder JR, Meyerand E, Prabhakaran V, Hermann B. Cognitive slowing and its underlying neurobiology in temporal lobe epilepsy. *Cortex*. 2019;117:41-52

the most significant feature itself. In this investigation, we apply SVM to standardized neuropsychological test scores to classify groups (epilepsy and controls) and identify the salient predictors.

3.2.1 Participants

Research participants included 55 TLE patients and 58 healthy controls from the ECP. The difference in the mean age ($p < 0.01$) between the TLE (range 19 – 60 years) and control groups (range 18 – 56 years) was addressed by using age-corrected cognitive scores. The two groups did not significantly differ with regard to gender ($p = 0.85$), with a modest trend in years of education ($p = 0.06$). In the TLE group, 14 subjects had right TLE, 26 had left TLE, and 2 had bilateral onsets based on either interictal EEG, imaging (hippocampal sclerosis) or ictal monitoring. Thirteen subjects had uncertain lateralization. TLE participants were taking 0 to 4 AEDs with a mean of 2.1, with chronic epilepsy (mean = 20 years) characterized by onset in late adolescence (mean = 19 years). A subset of the sample underwent Wada testing or fMRI language assessment and none showed reversed cerebral dominance.

3.2.2 Processing Speed

At its most basic level, processing speed can be defined as either the amount of time it takes to process a specific quantity of information, or the quantity of information that can be processed within a specific unit of time[128]. There has been little consistency in the metrics used to assess cognitive and psychomotor slowing in epilepsy, as speed-based performances have been assessed with a variety of measures including simple and complex reaction time, finger tapping, mental scanning, motor assembly tasks, and others[129]. One common approach, across diverse disorders, has been the use of digit symbol substitution tests, with applications to examine speeded

performance in schizophrenia[130-132], multiple sclerosis[133], normal aging[134, 135] as well as epilepsy[136]. Further investigation of the task has shown that it is driven in part by speed-dependent processes (graphomotor speed, perceptual speed), with contributions of visual scanning efficiency, learning/memory and executive function[137, 138].

An alternative measure of central processing speed is the Pattern Comparison Processing Speed Test (PCPS) of the NIH Toolbox Battery-Cognition Battery (NIHTB-CB) which is an efficient visually-based measure of choice reaction time adapted for computerized presentation. This test has applicability across the lifespan, sound test-retest reliability, appropriate age-related performance characteristics, and demonstrated construct validity[59]. Furthermore, there is less confounding of psychomotor issues with quantification of central information processing speed compared to digit symbol substitution tests.

PCPS requires the subject to identify whether two simultaneously presented visual patterns are the "same" or "not the same". Patterns are either identical or vary in: 1) color, 2) adding/taking something away, or 3) one versus many. The score reflects the number of correct items (out of a possible 130) completed in 90 seconds[59, 60]. The distribution of PCPS scores for the TLE and control groups is shown in **Figure 5**.

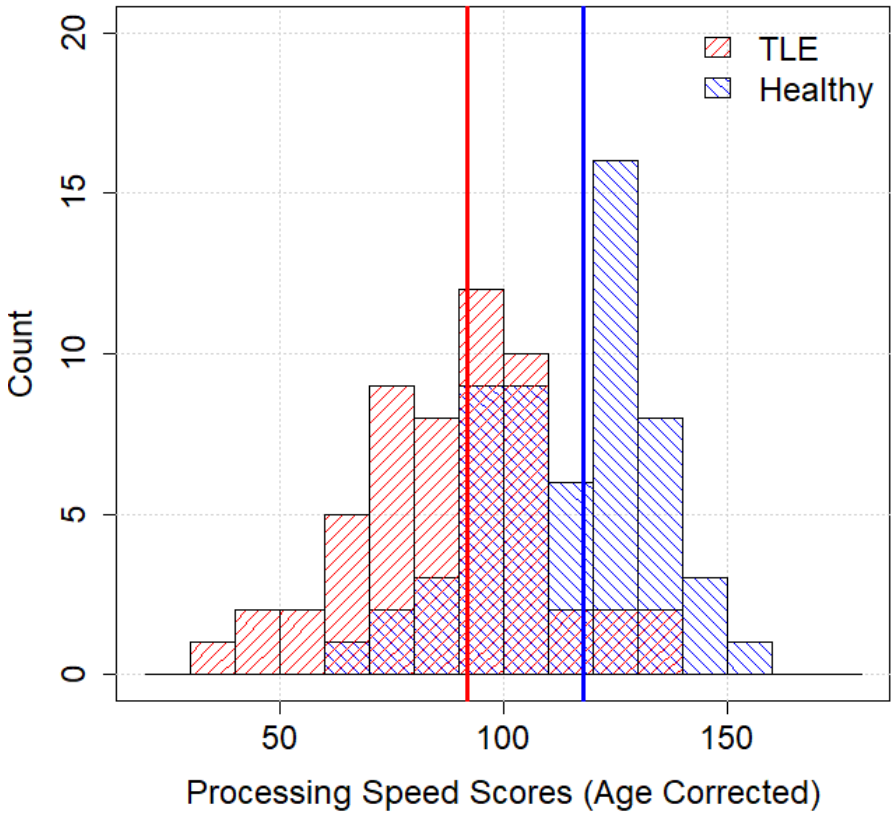

Figure 5. Pattern Completion Processing Speed Score Histogram. TLE patients (red) overall scored significantly lower than the age- and education-matched healthy controls (blue) on Pattern Completion Processing Speed (PCPS). The scores are the age-corrected standard scores. The vertical lines indicate the median scores for each group, which were 92 and 118.

3.2.3 Neuropsychological Test Results

Fourteen of the neuropsychological tests from Section 2.3.1 that were administered to both TLE patients and healthy controls have been selected to be the training features to machine learning. For all 14 measures the age-corrected standard scores were used. All test scores were normally distributed in both TLE and control groups (p's > 0.15, Kolmogorov-Smirnov test), except for the Judgement of Line Orientation test (p's < 0.05). Therefore, the Wilcoxon Rank-Sum Test was performed on this test and two-sample t-tests on the others.

TLE patients as a group performed significantly worse on 13 of the 14 administered

neuropsychological tests (**Table III**, Columns 5 and 6). PCPS had the largest effect size (1.27) followed by Grooved Pegboard (dominant hand 1.12, non-dominant hand 1.07), WASI-II Vocabulary (0.91), RAVLT (total words 0.89, delayed recall 0.86), and Dimensional Change Card Sort Test (0.86). Medium effect sizes were evident for WASI-II Block Design (0.78), Judgement of Line Orientation, and Boston Naming Test (0.71). Small effect sizes were observed for Flanker (0.49), Working Memory (0.48), Semantic Fluency (0.45) and Controlled Oral Word Association (0.24).

There were few lateralized cognitive findings. 26 left TLE and 14 right TLE patients did not differ in age ($p > 0.10$), gender ratio ($p = 0.50$), education ($p = 0.40$), AED count ($p = 0.96$), or duration of epilepsy ($p = 0.81$). The right TLE group performed significantly worse than the left TLE group on the Dimensional Change Card Sort Test ($p = 0.027$, $t = 2.30$). There were no other significant lateralized cognitive findings. The majority of cognitive tests were significantly lower than controls in both the left TLE (11 of 14 tests, all p's < 0.02) and right TLE (13 of 14 tests, all p's < 0.03) groups. Thus, cognitive anomalies were generalized in nature in the context of lateralized epilepsy.

Spearman correlations examined the relationship between the number of AEDs and cognitive performance. AED effects were observed on measures of dominant and non-dominant hand speeded fine motor dexterity (ρ's = -0.287 and -0.271, p's = 0.034 and 0.046 respectively) and working memory ($\rho = -0.266$, $p = 0.049$). There were no other significant associations between AED number and cognition including PCPS ($\rho = 0.122$, $p = 0.374$).

3.2.4 Machine Learning Feature Selection

The ability of the 14 neuropsychological tests to classify TLE and healthy control participants was tested using machine learning. SVM binary classification models[89] were trained

using the *z*-transformed age-corrected standard scores as the features. **Figure 6** provides a diagram for the SVM training and testing procedures employed in this study. 10-fold cross validation was used, where 10% of the samples were kept as a testing set. Randomization seeds were used for repeatability. When feature selection was on, a cross validation loop was added within the procedure and the feature with the lowest average absolute weight (or smallest contribution to the classification model) was removed per loop (recursive feature elimination). The feature selection continued until the 10-fold classification loss reached the minimum. This survived "optimum set" of features was then used in the final testing. The 10-fold test accuracy was recorded. This entire procedure in **Figure 6** was repeated 10 times (10 iterations), both with and without the feature selection. The optimum sets of features were then analyzed based on their normalized weights, where the maximum absolute weight was one.

Table III. Neuropsychological Tests by Contribution to Machine Learning Classification. The 14 neuropsychological test scores are sorted by their average absolute weights (Column 8) from the support vector machine (SVM) analysis without feature selection. Pattern Completion Processing Speed (PCPS) is the biggest contributor to the classification model, both collectively (SVM weight) and individually (effect size). †SD = standard deviation.

No	Feature Name	TLE (Mean ± SD)	Control (Mean ± SD)	2 sample *t*-test *p*-value	Effect Size *d* (Cohen's)	SVM Weight
1	Pattern Completion Processing Speed	89.24 ± 16.16	102.78 ± 14.18	<0.001	1.27	0.66
2	Grooved Pegboard Dominant Hand	87.35 ± 16.69	100.72 ± 14.27	<0.001	1.12	0.47
3	Dimensional Change Card Sort	86.87 ± 15.30	97.88 ± 15.68	<0.001	0.86	0.46
4	Boston Naming Test	96.47 ± 13.28	108.00 ± 16.19	<0.001	0.71	0.41
5	RAVLT Delayed Recall	96.42 ± 10.59	107.19 ± 12.85	<0.001	0.86	0.36
6	Grooved Pegboard Non-Dominant Hand	90.22 ± 17.49	97.95 ± 17.13	<0.001	1.07	0.36
7	Flanker Inhibitory	99.49 ± 16.18	109.27 ± 9.24	0.010	0.49	0.28
8	WASI-II Vocabulary	85.73 ± 14.98	102.16 ± 14.33	<0.001	0.91	0.26
9	RAVLT Total Words	89.27 ± 14.32	104.14 ± 13.38	<0.001	0.89	0.24
10	Working Memory	90.09 ± 17.98	93.93 ± 14.38	0.013	0.48	0.23
11	WASI-II Block Design	94.13 ± 16.46	108.34 ± 16.71	<0.001	0.78	0.19
12	Controlled Oral Word Association	84.80 ± 12.73	91.33 ± 13.73	0.214	0.24	0.18
13	Semantic Fluency	94.91 ± 16.54	102.36 ± 14.66	0.019	0.45	0.16
14	Judgement of Line Orientation	89.11 ± 21.53	115.14 ± 19.24	<0.001	0.74	0.16

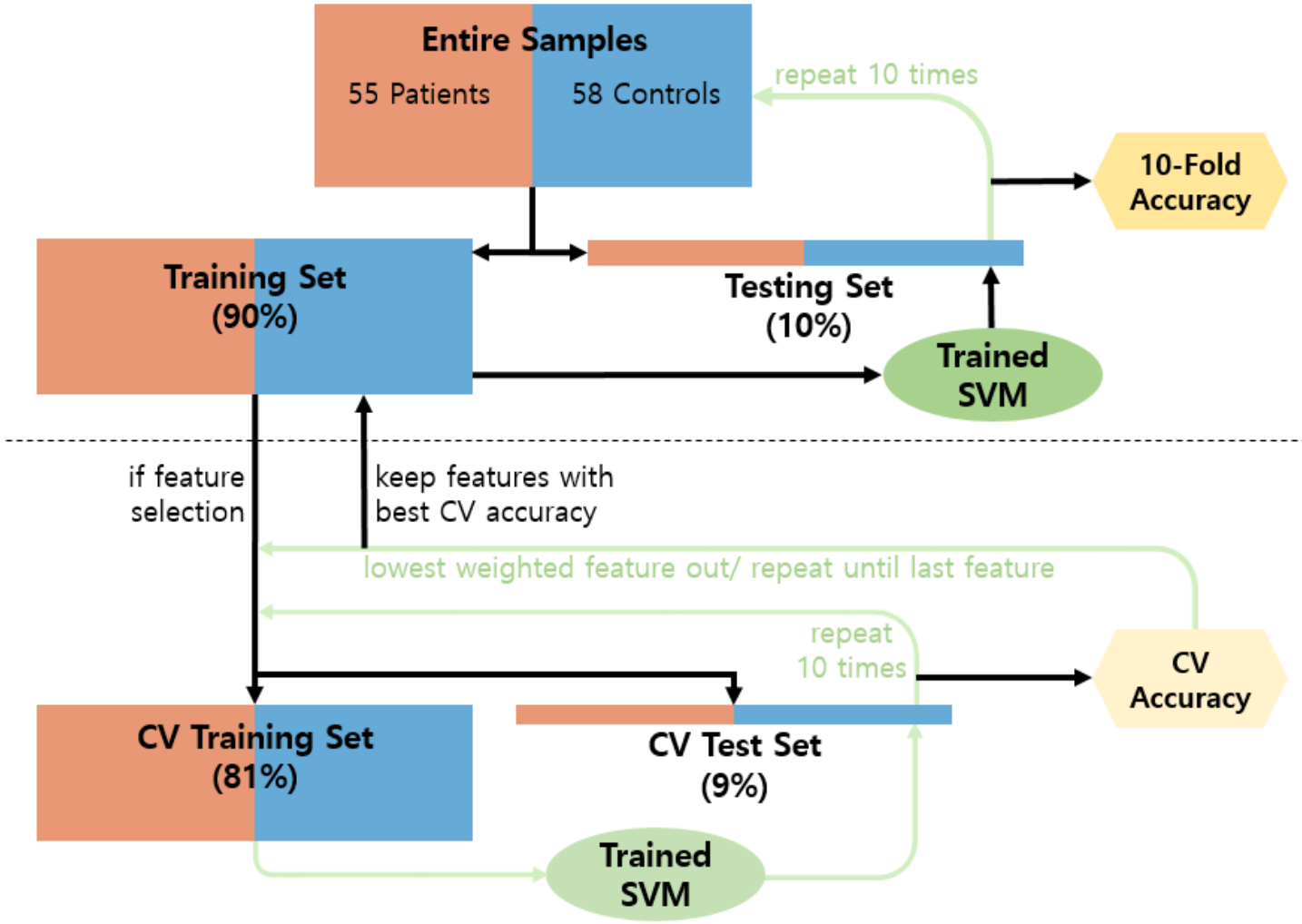

Figure 6. Diagram of 10-fold Machine Learning Training and Testing. Without feature selection (top half), a support vector machine (SVM) model gets trained on 90% of the entire samples and tested on the other 10%, which is repeated 10 times (exhaustive). With feature selection (bottom half), the training set is further split and the cross validation (CV) takes place. In the case of recursive feature elimination, CV is repeated, every time with the lowest weighted feature removed. The set of features that produce the best CV accuracy ("optimum set") is kept to train the entire training set for the final testing.

3.2.5 Results

The 14 neuropsychological test scores were able to train an SVM model that reliably classified TLE and control participants with 73.4 ± 2.7% test accuracy without feature selection. The PCPS score received the highest average absolute weight (w = 0.66) among all 14 scores in all 10 testing loops, followed by Grooved Pegboard Dominant Hand (w = 0.47) (**Table III**, Column 7). With feature selection, PCPS was most reliably and repeatedly present in every cross validation loop (9 out of 10 iterations) in the optimum set of features, followed by Grooved Pegboard-Dominant (5 out of 10) and Boston Naming Test (1 out of 10).

3.2.6 Discussion

For people with epilepsy the cognitive (and affective) comorbidities associated with the disorder create as much disability as the seizures themselves[139, 140]. The results of this investigation demonstrated that, in a non-surgical cohort of TLE participants, cognitive slowing is a powerful marker of TLE. While it has been recognized that processing speed is among the cognitive morbidities of chronic epilepsy, its relative standing among the other cognitive morbidities of epilepsy has not been fully appreciated. In fact, it was the most salient measure in separating the TLE and control groups (**Table III**, Column 7, SVM weight). Other measures of interest (Boston Naming Test, RAVLT) discriminated the TLE and control groups as expected, but not as powerfully as processing speed (**Table III**, Columns 5 and 7). Even though SVM does not assume feature independence, it is still possible that if two features are highly correlated, one of the two will receive less attention or weight, which explains the case with Grooved Pegboard Non-Dominant Hand score. Even with this in mind, we can conclude that PCPS is the best contributor to the classification model.

3.3 Data-driven Cognitive Phenotyping[§]

A longstanding pursuit in the neuropsychology of epilepsy has been an understanding of the signatures of cognitive abnormality associated with the disordered pathophysiology of specific epilepsy syndromes[117, 141]. This classic approach led to early appreciation of impaired memory in TLE, dysexecutive function in frontal lobe epilepsy, attentional disruption in absence epilepsy, language problems in Rolandic epilepsy, and dysexecutive behavior in juvenile myoclonic epilepsy[142, 143]. This general model, tracking cognition as a function of the taxonomy of the epilepsies and their associated clinical features, has served the field well[144, 145].

But incongruities in the classic model have accumulated over the years, in part due to studies involving broad-based neuropsychological assessment comprehensively overviewing human cognitive function as well as by head-to-head cognitive comparisons of epilepsy syndromes. Rather than finding the expected selective cognitive abnormalities linked to syndrome-specific pathophysiology, either a) more widespread and arguably unexpected cognitive anomalies have been reported when epilepsy syndromes are studied in depth (e.g., widespread cognitive abnormalities in focal epilepsies)[146-151] or, b) in head-to-head comparisons of two or more epilepsy syndromes, more shared than unique syndrome-specific cognitive abnormality is evident[146-156], or c) particular cognitive impairments (e.g., dysexecutive function) have been found to cut across multiple epilepsy syndromes[157-162].

A comprehensive neuropsychological test battery found in Section 2.3.1 from 111 TLE patients and 83 controls was reduced to core cognitive domains (language, memory, executive,

---

[§] Portions of this work are currently being reviewed: Hermann B, Conant L, Cook C, Hwang G, Garcia-Ramos C, Dabbs K, Nair V, Mathis J, Rivera-Bonet C, Allen L, Almane D, Arkush K, Birn R, DeYoe E, Felton E, Maganti R, Nencka A, Raghavan M, Shah U, Sosa V, Struck A, Ustine C, Reyes A, Kaestner E, McDonald C, Prabhakaran V, Binder J, Meyerand M. Network, Clinical and Familial Features of Cognitive Phenotypes in Temporal Lobe Epilepsy. *NeuroImage: Clinical.* Under Review

visuospatial, motor speed) which were then subjected to k-means clustering, a type of a cluster analysis (unsupervised learning). The resulting cognitive subgroups were compared in regard to sociodemographic and clinical epilepsy characteristics as well as variations in brain structure and functional connectivity.

Three cognitive subgroups were identified: *Generalized Cognitive Impairment* (Generalized-CI) (N = 20, 18% of TLE group) reflecting significant impairment affecting all domains, *Focal Cognitive Impairment* (Focal-CI) (N = 34, 31%) demonstrated by particularly abnormal language, memory and executive function/processing speed, and *No Cognitive Impairment* (No-CI) (N = 57, 51%) where performance was intact and comparable to controls across all domains (**Figure 7**). The Generalized-CI group was characterized by an earlier age at medication initiation ($p < 0.05$), fewer patient ($p < 0.001$) and parental years of education ($p < 0.05$), greater racial diversity ($p < 0.05$), and greater number of lifetime generalized seizures ($p < 0.001$) (**Table IV**). The three groups also differed in an orderly manner across total intracranial ($p < 0.001$) and bilateral cerebellar cortex volumes ($p < 0.01$), but not in regional measures of cortical thickness or volume. In contrast, large-scale patterns of cortical-subcortical covariance networks revealed significant differences across groups in global and local measures of community structure and distribution of hubs. Resting-state fMRI revealed stepwise anomalies as a function of cluster membership, with the most abnormal patterns of connectivity evident in the generalized impairment group and no significant differences from controls in the cognitively intact group (**Figure 8**).

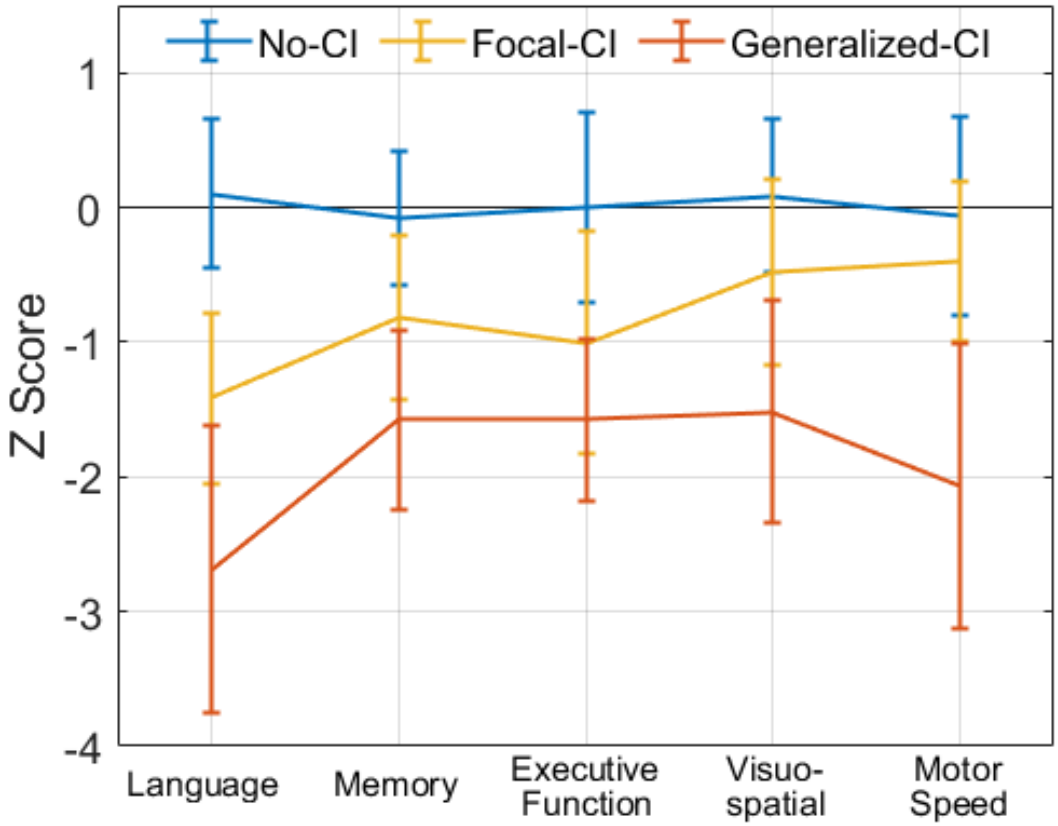

Figure 7. Cognitive Performance of Three Identified Subgroups of TLE. Three clusters within temporal lobe epilepsy group were identified, with Generalized Cognitive Impairment (Generalized-CI) (red, *N* = 20) being the most impaired overall, then Focal Cognitive Impairment (Focal-CI) (yellow, *N* = 34), and No Cognitive Impairment (No-CI) the most intact (blue, *N* = 57).

Table IV. Summary of Demographics of Three Identified Subgroups. Generalized cognitive impairment (Generalized-CI) group showed significantly fewer years of patient and parental education and larger proportion of non-Caucasian participants compared to focal (Focal-CI) and no cognitive impairment (No-CI) groups, which suggested the influence of socioeconomic risk factors in the cognitive impairment of temporal lobe epilepsy (TLE).

Groups	N	Age (years)	Gender (Male / Female)	Education (years)	Mother Education (year)	Father Education (years)	Duration of Seizures (years)	Race (Caucasian/ Non-Caucasian)
Controls	83	33.8 ± 10.6	36 / 47	15.8 ± 2.7	14.6 ± 2.7	14.8 ± 2.8	-	74 / 9
All TLE	111	39.6 ± 11.5	43 / 68	14.7 ± 2.7	13.5 ± 2.7	13.8 ± 2.9	16.8 ± 13.9	91 / 20
Generalized-CI	20	38.2 ± 13.5	8 / 12	12.3 ± 2.0	12.6 ± 2.7	11.9 ± 2.1	21.3 ± 16.7	10 / 10
Focal-CI	34	36.6 ± 11.1	15 / 19	13.6 ± 1.7	13.8 ± 2.3	13.2 ± 2.0	13.2 ± 12.9	26 / 8
No-CI	57	41.9 ± 10.4	20 / 37	16.2 ± 2.4	13.6 ± 2.9	14.7 ± 3.1	17.4 ± 13.2	55 / 2

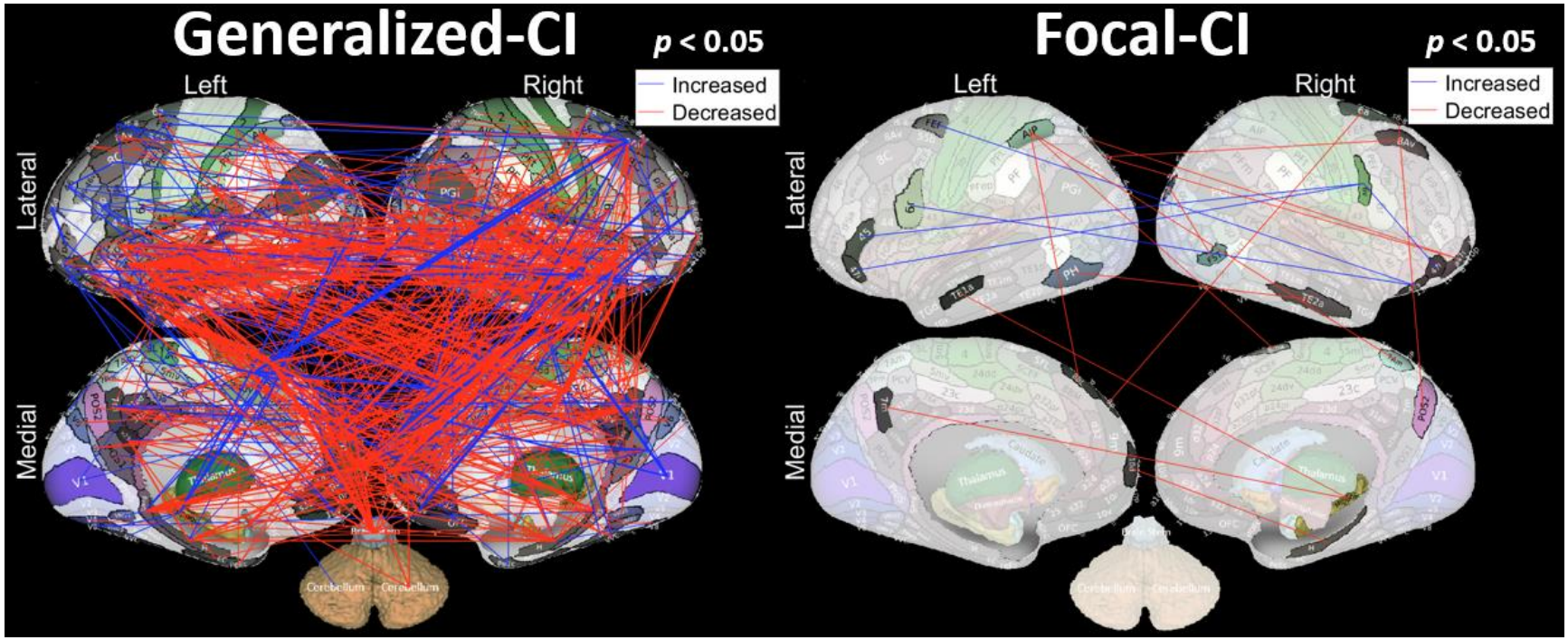

Figure 8. Resting-state Connectivity Changes in TLE Subgroups. Resting-state connectivity changes of temporal lobe epilepsy (TLE) patients in Generalized Cognitive Impairment (Generalized-CI, left) and Focal Cognitive Impairment (Focal-CI, right) groups, compared to healthy controls. Red lines indicate decreased connectivity (hypoconnectivity) in the patients, while blue lines indicate increased connectivity (hyperconnectivity). No-CI subjects did not show any significant changes in the connectivity.

Overall, patients with TLE are composed of distinct underlying cognitive phenotypes that harbor systematic relationships with clinical, familial, demographic and neuroimaging correlates. Cognitive phenotype variations in patient and familial education and ethnicity, with linked variations in total intracranial volume, suggest an early and persisting socioeconomic-status related neurodevelopmental impact with additional contributions of clinical epilepsy factors (e.g., lifetime generalized seizures). The neuroimaging features of cognitive phenotype membership are most notable for disrupted large scale cortical-subcortical networks and patterns of functional connectivity, and cerebellar atrophy.

There is a taxonomy of cognitive abnormality in TLE that only partially overlaps with the syndrome-specific pathophysiology of the disorder. This taxonomy is influenced by diverse epilepsy, non-epilepsy, and neuroimaging features reflecting the combined influence of socioeconomic, neurodevelopmental and neurobiological risk factors. The fact that there seems to be three distinct subgroups within TLE group should have influenced the results in Sections 3.1 and 3.2 negatively, since all TLE patients were originally assumed to be showing similar imaging and neuropsychological abnormalities compared to healthy controls. On the other hand, separating the TLE group into these three subgroups in order to reflect this finding substantially diminishes the sample sizes and is detrimental to machine learning research. Perhaps, there is a better method of diagnosing and sub-grouping epilepsy based on other phenotypes than seizure focus, which has been convenient for human clinicians.

3.4 Concluding Remarks

In the era of machine learning and big data, the methods that humans have developed to make medical diagnoses and to offer appropriate treatments will be tested and contested. Decisions

will need to be made when to trust criteria and standards proposed by machines over those of humans. There will also be cases where machine learning results will completely contradict diagnoses by human physicians. In this case, is the machine picking up something that the human physicians have not noticed, or is it an error? What if the underlying algorithm that the machine has used to reach the answer is too complex to be assessed, or perhaps inconsistent with established clinical knowledge?

Human decisions can be influenced by prior knowledge and prejudice. Confirmation bias drives humans to quickly accept scientific results that conform to the standard knowledge. It takes more effort to divert a scientific mistake than to establish one. With increased amounts of medical information as well as advanced computational capacity and techniques, leaning towards using more data-driven, objective approaches in medicine seem reasonable. However, this must come along with careful research on polishing the data-driven methods to eliminate all sources of technical errors. In other words, we need better understanding on the data-driven techniques such as machine learning. Careful use of these techniques will take the current medicine to the next level. It will give human clinicians new insights to complex datasets, as well as correct unrecognized mistakes. As discussed in Section 3.3, it may redefine our understanding of a disease and even allow for more personalized medicine in the future.

In this chapter, the ability of machine learning to extract important features and patterns from complex datasets was discussed. In Chapter 4, the focus will be on building reliable machine learning models that can make accurate predictions on unseen data.

Chapter 4

Building Predictive Models of Temporal Lobe Epilepsy using Machine Learning

The attractive strength of machine learning is in its predictive power. Whether it is classifying a group of data into categories (classification) or predicting a continuous variable (regression)[13], a machine learning has worth only when it reliably makes correct predictions. In Chapter 3, we have discussed the feature extraction ability of machine learning, but if the models did not achieve enough predictive power and accuracies, the extracted features would not have carried much weight. Therefore, achieving high performance is often the top priority in a machine learning research, although the analysis of the trained model and of the underlying features is still crucial. Section 4.2 will consider using machine learning for classification, and Section 4.3 will introduce one example of machine learning regression to make predictions on TLE patients.

4.1 Assessment of Machine Learning Models

Reliability and reproducibility are essential in machine learning; especially in medicine, because it deals with human lives. In order to strictly assess a machine learning model's performance, a clear distinction must be made between a cross validation accuracy and a test accuracy[97]. When assessing model performance for generalizability, the initial dataset needs to be split into two groups: training and testing sets. The entire training procedure must happen strictly within the training set, so that the data in the testing set become truly new observations for the trained model. Otherwise, the observed performance may be overestimated. This simple rule is easy to be violated in practice.

For example, the results reported in Section 3.1.6 and **Table I** describe cross validation accuracy, because in order to determine the optimum number of features, the model performance was checked multiple times on the testing set during the recursive feature elimination. This was acceptable because the goal of the study was to compare the general classification performance between multiple models on multiple cases, but this accuracy should not be confused with a test accuracy. In other words, if these models were tested on strictly independent testing sets, their test accuracies would have likely been lower, or close at best. On the other hand, the accuracy reported in Section 3.2.5 was a test accuracy, because the procedure depicted in **Figure 6** restricted the crosstalk between the training and testing sets.

4.2 Classifying between TLE Patients and Healthy Controls

Having access to the neuroimaging and neuropsychological testing data of TLE patients as well as healthy control volunteers from the ECP, the first and the most intuitive research direction

was to train machine learning models to separate the two groups. So a number of training features (features described in Sections 3.1 and 3.2, and also structural T1-weighted MRI features) have been used to train machine learning models even including a few "shallow" deep learning models. However, the test accuracies were not satisfactory (around 70 – 75% at best, see Section 5.2), which was largely due to small sample sizes.

The poor accuracy does not necessarily equal bad hypothesis, because accuracy can improve significantly with more sample sizes. In this case, the correct question is to ask the potential of the model, instead of the current accuracy. This led me to study the relationship between the sample size and the machine learning classification performance. More on this topic will be discussed in Chapter 5.

Another topic of discussion is whether building a model that classifies between TLE patients and healthy controls is beneficial. The TLE patients enrolled in the ECP were already aware that they had epilepsy from seeing themselves simply having recurring seizures. A machine learning model that can tell whether a patient has an epilepsy or not may not be clinically useful in terms of the diagnostic gain. Rather, a better question is, for example, to classify between TLE and frontal lobe epilepsy patients, or between TLE patients with left and right seizure foci. In other words, the ultimate goal in studying epilepsy patients using machine learning classification perhaps is to be able to predict patient subgroups instead. However, this question is more demanding to address using machine learning because of the limited samples. The progress in this field is expected to be slower, compared to other more common diseases such as Alzheimer's disease. This is another reason to emphasize the understanding of the sample size relationship in machine learning, which will be discussed in Chapter 5.

4.3 Predicting Brain Ages of TLE with Machine Learning Regression[§]

Chronic TLE is associated with abnormalities in cognition, brain structure and brain connectivity in midlife[45-48], findings that have raised concern regarding the future course of cognitive and brain aging and the risk of cognitive disorders of aging including dementia[163]. While different models of cognitive aging in epilepsy have been proposed (progressive decline, accelerated aging [two hit model], stable non-progressive abnormality)[164], consensus has yet to be achieved. Importantly, all models predict significantly more impaired cognition in aging individuals with chronic epilepsy compared to controls[165-167]. Similarly, cross-sectional modeling of structural brain aging has suggested greater abnormality in chronic epilepsy compared to controls with advancing age[40, 168].

In a novel approach, Pardoe et al.[169] trained a machine learning regression model using T1-weighted structural MRI scans of 2,001 healthy controls to predict their chronological ages. They then used the model to predict the ages of 94 medically refractory focal epilepsy patients and showed that these patients had structural brains that were on average 4.5 years older than the healthy controls. Sone et al.[170] recently reported findings from a similar study examining different types of epilepsy including TLE using T1-weighted images, and found the same trend of accelerated aging (10.9 years older for TLE patients with inter-ictal psychosis, and 5.3 years without).

There are many paths of exploration from these studies that can be considered. First, will the functional brains of epilepsy patients similarly show accelerated brain aging (or premature

---

[§] Portions of this work have been published in: Hwang G, Hermann B, Nair V, Conant L, Dabbs K, Mathis J, Cook C, Rivera-Bonet CN, Mohanty R, Zhao G, Almane D, Nencka A, Felton E, Struck AF, Birn R, Maganti R, Humphries CJ, Raghavan M, DeYoe EA, Bendlin BB, Prabhakaran V, Binder JR, Meyerand ME. Brain Aging in Temporal Lobe Epilepsy: Chronological, Structural, and Functional. *Neuroimage: Clinical*. 2020;

brain aging in Pardoe et al., 2017)? Accelerated brain aging in epilepsy has been investigated mainly in the structural brain. While many studies have reported changes in the functional connectivity of epilepsy patients[171, 172], whether the changes resemble accelerated aging is unknown.

Second, what factors are associated with age accelerated structural and functional brains? Possibilities include clinical seizure characteristics (e.g., age of onset, seizure frequency), treatment factors (e.g., number or type of AED use), and of course demographic characteristics. Previous studies have reported that brain volume reductions in epilepsy may be independent of or only weakly related to seizure activity[173] and potentially more related to AED use[174]. Pardoe et al.[169] and Sone et al.[170] in their secondary analyses briefly reported that increased brain age difference (or brain-PAD: predicted age – chronological age in Sone et al., 2019) in epilepsy was associated with earlier age of onset, but not with epilepsy duration nor AED use. More systematic search of potential correlates of accelerated brain aging is desired.

Third, is accelerated brain aging in epilepsy directly related to cognitive status and cognitive decline over time? Are brain ages better predictors of cognitive performance than the patients' chronological ages? Cognitive aging and its core dimensions (crystallized and fluid cognitive abilities) in epilepsy have yet to be examined in relation to potential age-accelerated alterations in functional connectivity patterns and brain structure. Whether they have explanatory power beyond chronological age remains to be determined.

4.3.1 Participants

Participants included 104 TLE patients (mean age = 40.4 ± 11.8 years, range = 19 – 60 years, 64 females) and 151 healthy controls (mean age = 53.7 ± 19.4 years, range = 18 – 89 years, 88 females). All TLE patients were from the ECP. 57 controls were from the ECP, and additionally

94 controls who matched the ECP criteria were drawn from ADCP[84]. The use of healthy controls from the two projects allowed investigation of participants with a wider age range than provided by either project alone, without compromising scanner, site or protocol variabilities (See Section 3.1.1 for more on ADCP). 42 TLE patients and 51 controls were scanned at the Medical College of Wisconsin, while 62 TLE patients and 100 controls were scanned at the University of Wisconsin-Madison.

4.3.2 Data Processing

HCP minimal pre-processing was performed as described in Section 2.4.1. 254 structural features generated by FreeSurfer's standard reconstruction (recon-all) were extracted from the T1-weighted images, including cortical thicknesses, surface areas, volumes and also subcortical and global volumes. Surface areas and volumes were divided by the total surface area and total gray matter volume respectively to normalize for brain size. Then the structural features were normalized through *z*-score transform.

Additional pre-processing was performed on the resting-state fMRI images using AFNI (Analysis of Functional Neuro-Images)[85]. This included motion regression using 12 motion parameters, regression-based removal of signal changes in the white matter, cerebrospinal fluid (CSF), global signal, and band-pass filtering (0.01 – 0.1Hz). There are trade-offs of regressing out the global signal from the raw signals, such as potential false negative correlations[175]. Therefore, another machine learning model was trained without the global signal regression to confirm whether similar results were obtained.

Using the Connectome Workbench (version 1.1.1) (Marcus et al., 2011), time-series data from four 5-minute resting-state fMRI scans acquired in a single session were concatenated. 360 time-series from Glasser Parcellation[67] plus 19 FreeSurfer subcortical regions[68] were extracted per

subject (see Section 2.4.2). Pairwise Pearson correlations between 379 timeseries were computed and Fisher-*z* transformed for generating connectivity matrices.

A subset of connectivity features were found to be affected by the subject motion in the scanner (absolute and relative mean RMS motion)[176]. Therefore, absolute mean RMS motion was linearly regressed out from features that showed significant correlation (raw $p < 0.05$), first separately for healthy controls and then for TLE patients, by combining the two groups (in order to restrict crosstalk between the two groups). Without this regression, the accelerated functional brain ages were significantly correlated with motion ($p < 0.01$), while regressing it out from the entire matrices resulted in the opposite correlation ($p < 0.05$).

4.3.3 Support Vector Regression (SVR)

Two age-prediction support vector regression (SVR) models[177, 178] were built in Python using the scikit-learn library[179]: with structural and functional (resting-state correlation matrices) features from the healthy controls. A linear kernel was used with no feature selection. First, the SVR models were trained and tested on the healthy controls using 10-fold cross validation. A linear correction that was suggested by Le et al.[180] was applied to remove known systematic bias caused by regression dilution and regression towards the mean (old subjects predicted young, and vice versa)[181]. The accuracy of the models were quantified using the correlation between chronological age and predicted, the amount of variance in age explained by the model (R^2), the mean absolute error (MAE) and the root mean squared error (RMSE).

The final models were trained with the entire healthy control dataset and applied on the TLE patients. The predicted ages (brain ages) were compared to the chronological ages. Accelerated ages (brain age − chronological age) were calculated. The entire training and testing process is summarized in **Figure 9**.

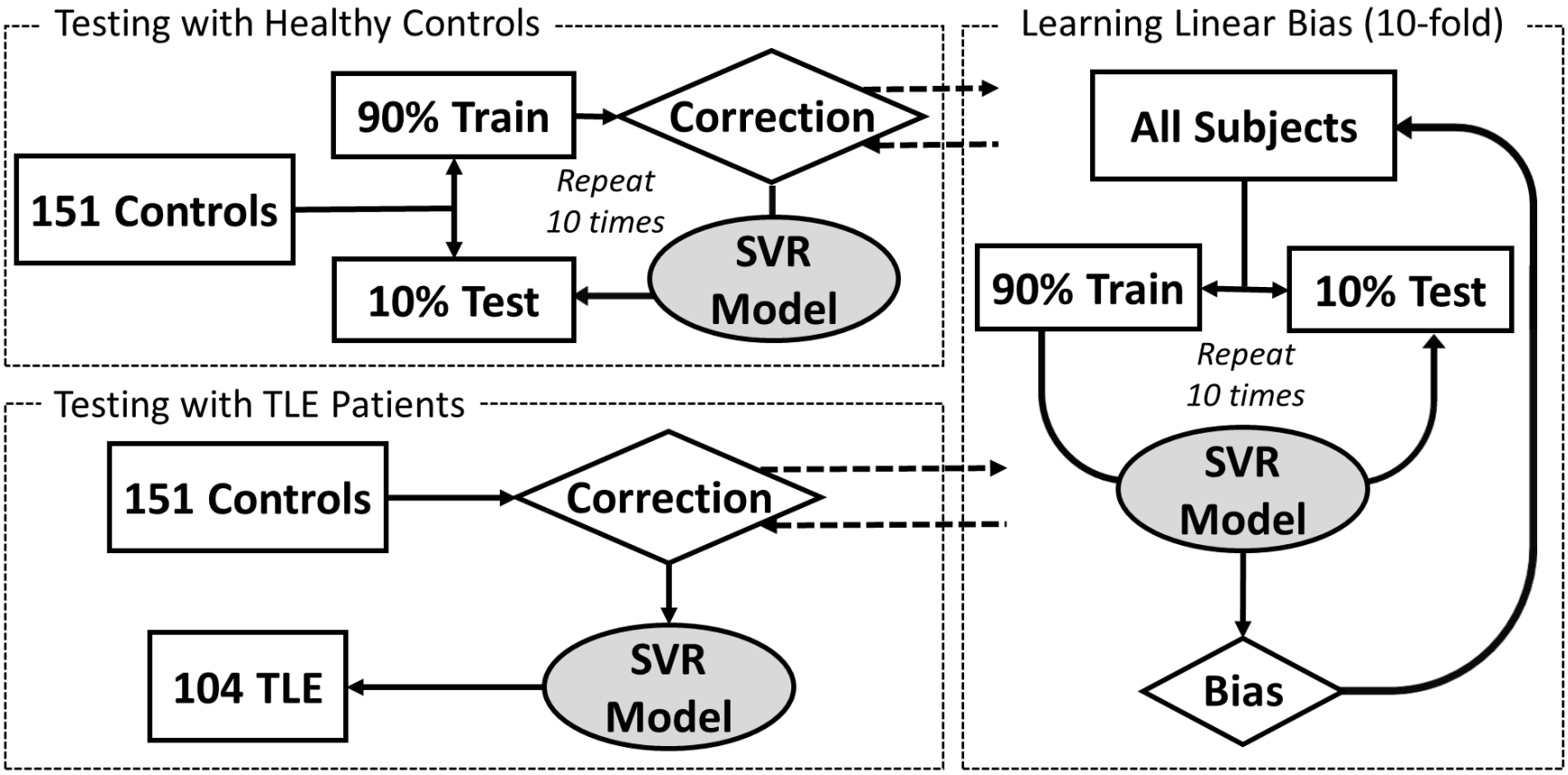

Figure 9. Diagram of 10-fold Brain Age Model Training with Bias Correction. This diagram summarizes the process of support vector regression (SVR) model training and testing procedure. 10-fold cross validation on the healthy controls were first performed (left top), and then separately the testing on the temporal lobe epilepsy (TLE) patients (left bottom). Linear correction suggested by Le et al.[180] was preformed to remove systematic bias caused by regression dilution and regression towards the mean.

4.3.4 Brain Age Prediction Results

The cross validation results of healthy controls are visualized in **Figure 10** in blue dots (r = 0.82, R^2 = 0.67, MAE = 10.7, RMSE = 13.65 for structural, r = 0.91, R^2 = 0.83, MAE = 6.94, RMSE = 8.86 for functional model). The variance was significantly larger ($p < 0.001$, two-sample F-test for equal variances) in the accelerated structural ages compared to the functional ages.

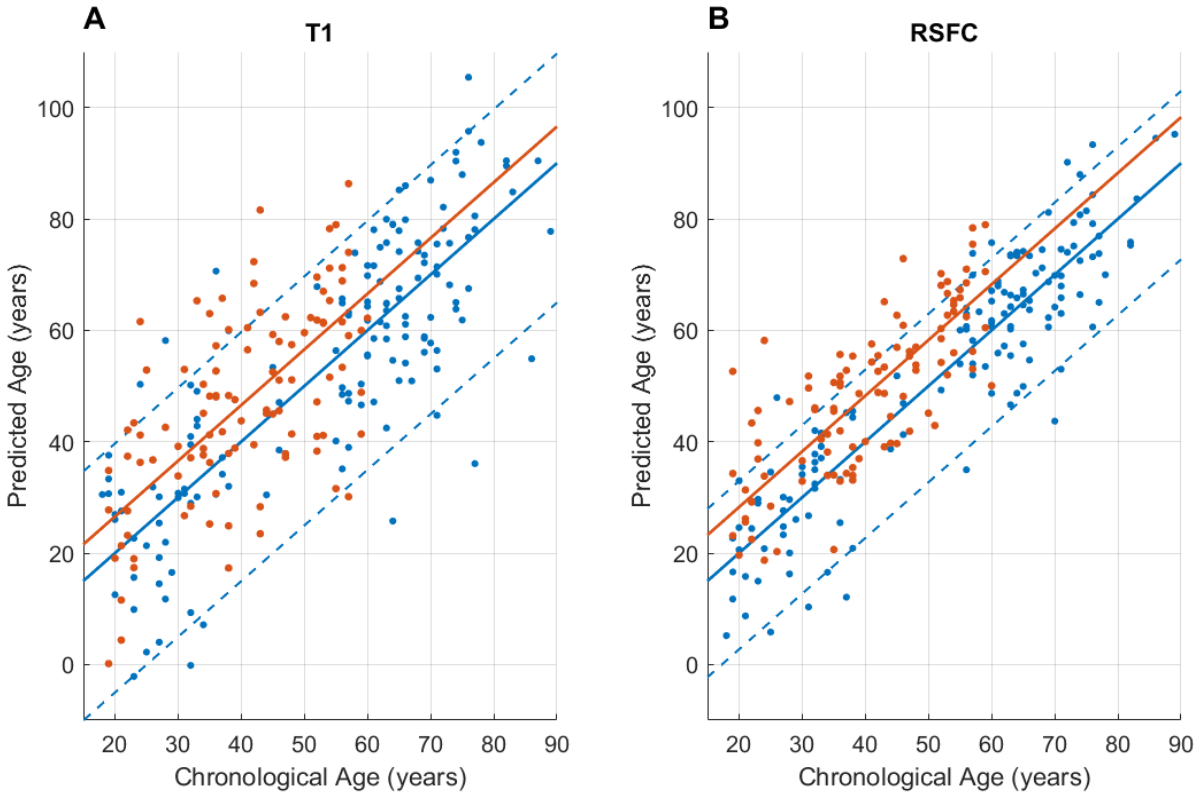

Figure 10. Scatter Plots of Predicted versus Actual Ages. These scatter plots show the support vector regression (SVR) age prediction results of both healthy controls (blue) and temporal lobe epilepsy (TLE) patients (orange): **(A)** with structural, and **(B)** functional features. The dotted lines indicate the 5th and the 95th percentiles of the cross validation results. †RSFC = resting-state functional connectivity.

The orange dots in **Figure 10** represent TLE patients. The brain aging effect in TLE was found in all age groups. The 5th and 95th percentiles of the cross-validation (healthy control) results were marked. 17 TLE patients (16%) showed structural brain ages greater than the 95th percentile (>19.7 years of acceleration), and 34 patients (33%) showed functional brain ages greater than the 95th percentile (>12.9 years of acceleration), with seven patients who overlapped. There was no significant correlation between the two accelerated ages ($r < 0.01$, $p = 0.94$ in healthy controls, $r = 0.14$, $p = 0.15$ in TLE patients) (**Figure 11**).

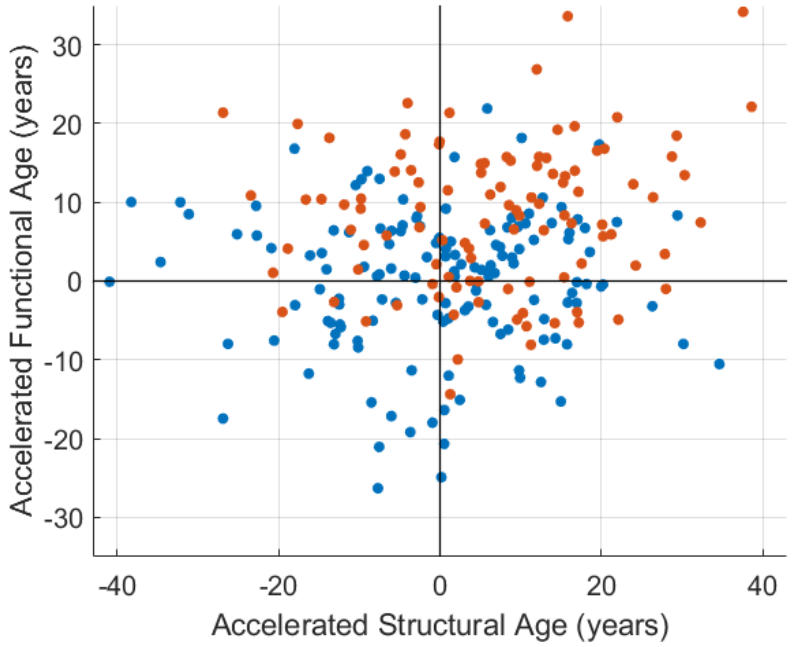

Figure 11. Scatter Plot of Two Accelerated Brain Ages. This scatter plot shows the relationship between two accelerated ages. No statistically significant relationship was found in neither healthy controls (blue, $r < 0.01$, $p = 0.94$, Pearson correlation), nor TLE patients (orange, $r = 0.14$, $p = 0.15$).

In healthy controls, structural ($r = 0.82$) and functional ($r = 0.91$) brain ages were highly correlated with chronological age, and also were inter-correlated ($r = 0.74$). The three ages were still correlated in TLE ($r = 0.60$, $r = 0.77$, $r = 0.53$ correspondingly), but to a significantly lesser degree compared to healthy controls (p's < 0.01, $z = 3.54$, $z = 3.87$, $z = 2.75$).

Figure 12 shows the histograms of the accelerated brain ages of the TLE patients. The final SVR model with the linear correction predicted their structural brain ages to be on average 6.6 years older than their chronological ages ($p < 0.001$, paired *t*-test). Their structural brain ages were significantly older than those of the healthy controls ($p < 0.001$, unpaired *t*-test). The accelerated structural ages (structural brain age – chronological age) ranged from -27 (brain age younger than chronological age) to +39 years (brain age older than chronological age), with the standard deviation of 13.7 years, which was the same as in healthy controls. There was no specific structural feature whose value was significantly associated with the accelerated structural ages (Spearman correlation).

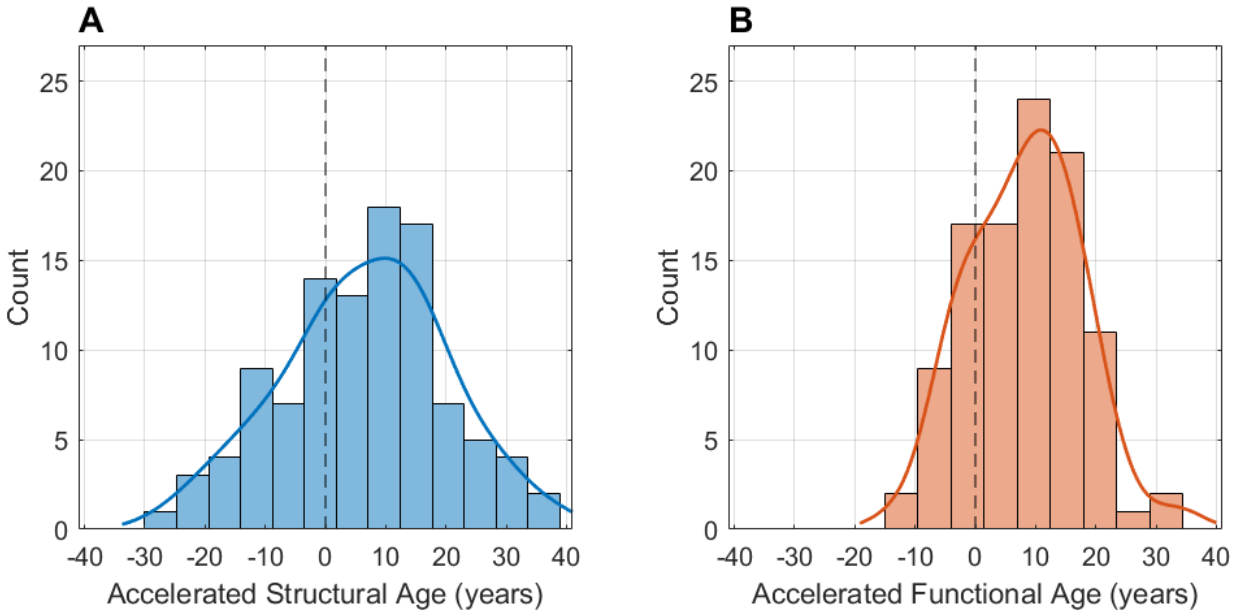

Figure 12. Accelerated Brain Ages of Temporal Lobe Epilepsy. These histograms show the accelerated brain ages of 104 temporal lobe epilepsy (TLE) patients: **(A)** with structural, and **(B)** functional features. Accelerated aging in TLE was observed both in the structural (6.6 ± 13.7 years) and functional brains (8.3 ± 9.2 years).

The final SVR model predicted the functional brain ages of the TLE patients to be on average 8.3 years older than their chronological ages ($p < 0.001$, paired *t*-test). Without the global signal regression of the raw signals, a similar results were found with the TLE patients' functional

brain ages predicted to be on average 5.1 years older than their chronological ages ($p < 0.001$, paired *t*-test). Their functional brain ages were significantly older than those of the healthy controls ($p < 0.001$, unpaired *t*-test). The accelerated functional ages ranged from -14 to +34 years with the standard deviation of 9.2 years, which was similar to 8.9 years in healthy controls. They were not significantly associated with the absolute/relative mean RMS motion (p's > 0.6, r's < 0.05). The variance was significantly larger ($p < 0.001$, two-sample *F*-test for equal variances) in the accelerated structural ages compared to the functional ages.

8,341 out of 71,631 connectivity features were significantly associated (corrected *p*-values < 0.05, Spearman correlation) with the accelerated functional ages, with the top 48 features (ρ's < -0.53) all showing negative correlation (weaker connection associated with more accelerated functional age). Most of these 48 connections were bilateral temporal or frontal lobe connections (**Figure 13**).

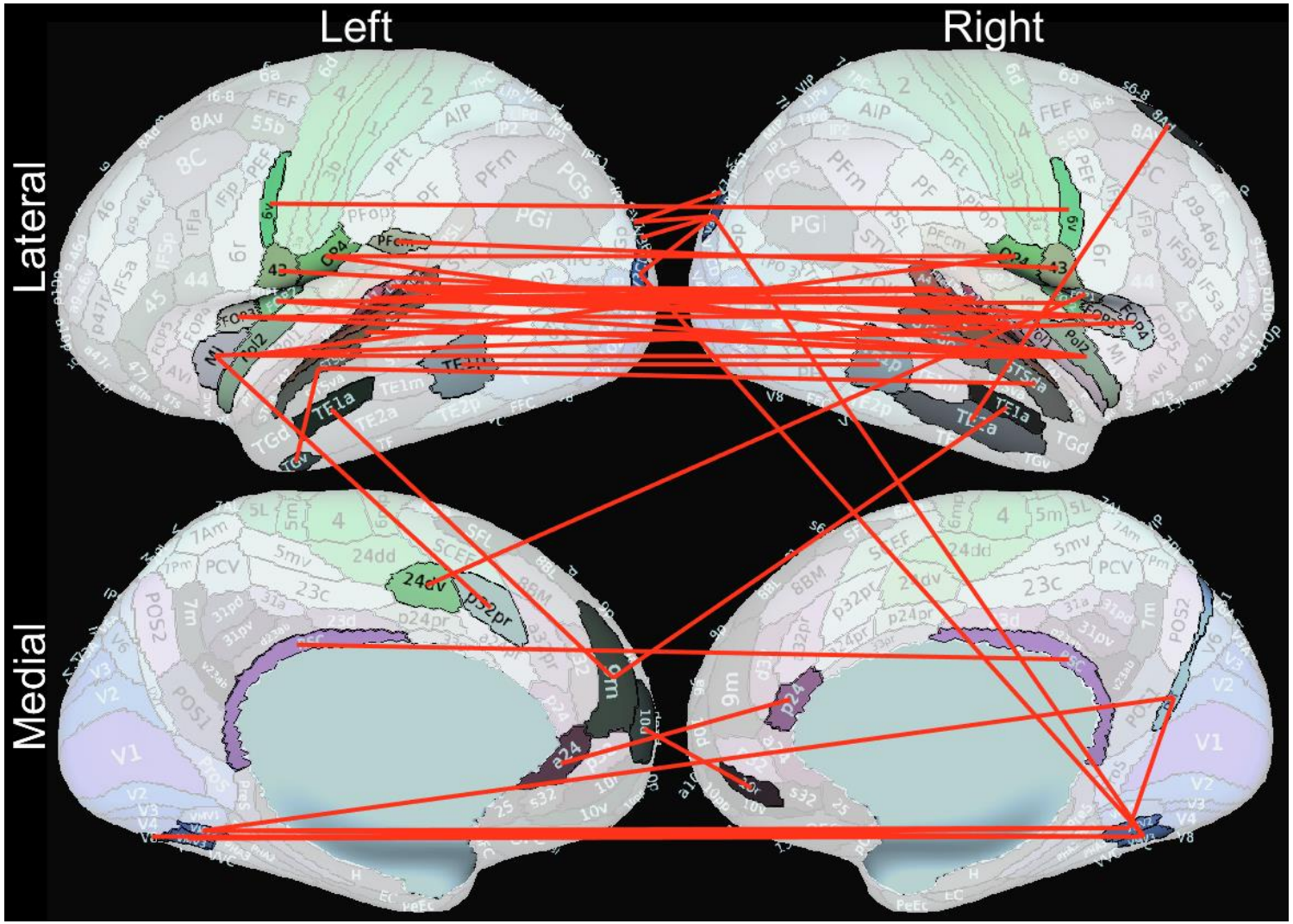

Figure 13. Resting-state Connectivity Associated with Functional Brain Aging in TLE. These 48 resting-state functional connections are most significantly associated with accelerated functional brain aging (corrected *P*-values < 0.0001, ρ's < -0.53). Weaker correlations in these connections are associated with more accelerated functional brain aging.

4.3.5 Clinical and Cognitive Correlates

Out of 104 TLE patients that were examined, 74 reported having had complex partial seizures (49 currently) and 62 reported secondary generalized seizures (22 currently). After correcting the *p*-values for multiple comparisons with Benjamini-Hochberg FDR correction[96], only trend-to-significant relationships were found between the functional accelerated ages of the TLE patients and their complex partial seizure frequency ($p = 0.07$) and AED count ($p = 0.07$). Patients who reported having at least one seizure during the past year were taking a greater number of AEDs ($p < 0.01$) compared to those who were seizure-free the past year, although there were no significant differences in the accelerated brain ages between the two groups.

Table V shows the correlation results between the three ages of TLE patients and their cognitive test scores. The FDR multiple comparison correction was performed on the *p*-values within each age measure and cognition type.

Chronological age was significantly associated (corrected $p < 0.05$) with four of seven tests, with trends (corrected $p < 0.1$) seen for two others. Structural age was not significantly associated with any test. Functional age was significantly associated with four of seven tests, with trends seen for one other: all fluid cognitive tests. Brain ages were not significantly associated with the crystallized subtests. Chronological age was significantly more associated with Picture Vocabulary than the brain ages ($Z = 2.31$, $p = 0.02$ for structural, $Z = 2.13$, $p = 0.03$ for functional age, Steiger's *Z*-test).

Three of seven tests (Dimensional Change Card Sort, Picture Sequence Memory, and Pattern Comparison Processing Speed) were significantly associated with both chronological and functional age measures. Subsequent mediation analyses addressed the question of whether structural or functional brain age mediated the association between chronological age and these

cognitive scores. Structural age was never a significant mediator while functional age partially mediated the relationship between chronological age and performance on three tests: Picture Sequence Memory ($p < 0.001$), Dimensional Change Card Sort ($p = 0.004$) and Flanker Inhibitory Control and Attention ($p = 0.03$) (**Figure 14**).

Table V. Cognitive Correlates of TLE Brain Ages. This table summarizes the correlation results between the three brain ages of the temporal lobe epilepsy (TLE) patients and their cognitive test scores. False discovery rate (FDR) correction was made on the *p*-values within each age measure and cognition type. Overall, fluid cognition was well associated with both chronological and functional ages. Chronological age was the best predictor among the three age measures of Picture Vocabulary ($Z > 2.1$, $p < 0.05$). *corrected $p < 0.05$.

Cognition Type	NIH Toolbox Cognition Battery (NIHTB-CB)	Subdomain	Chronological Age Correlation		Structural Age Correlation		Functional Age Correlation	
			r	*p*	*r*	*p*	*r*	*p*
Fluid	Flanker Inhibitory Control and Attention	Executive Function	-0.174	0.091	-0.094	0.605	-0.172	0.088
	Dimensional Change Card Sort		**-0.239**	**0.028***	-0.070	0.608	**-0.214**	**0.041***
	List Sorting Working Memory		-0.171	0.091	-0.033	0.748	**-0.221**	**0.041***
	Picture Sequence Memory	Episodic Memory	**-0.290**	**0.008***	-0.254	0.055	**-0.335**	**0.005***
	Pattern Comparison Processing Speed	Processing Speed	**-0.313**	**0.008***	-0.092	0.605	**-0.231**	**0.041***
Crystallized	Picture Vocabulary	Language	**0.293**	**0.006***	0.094	0.508	0.154	0.180
	Oral Reading Recognition		0.143	0.156	0.067	0.508	0.135	0.180

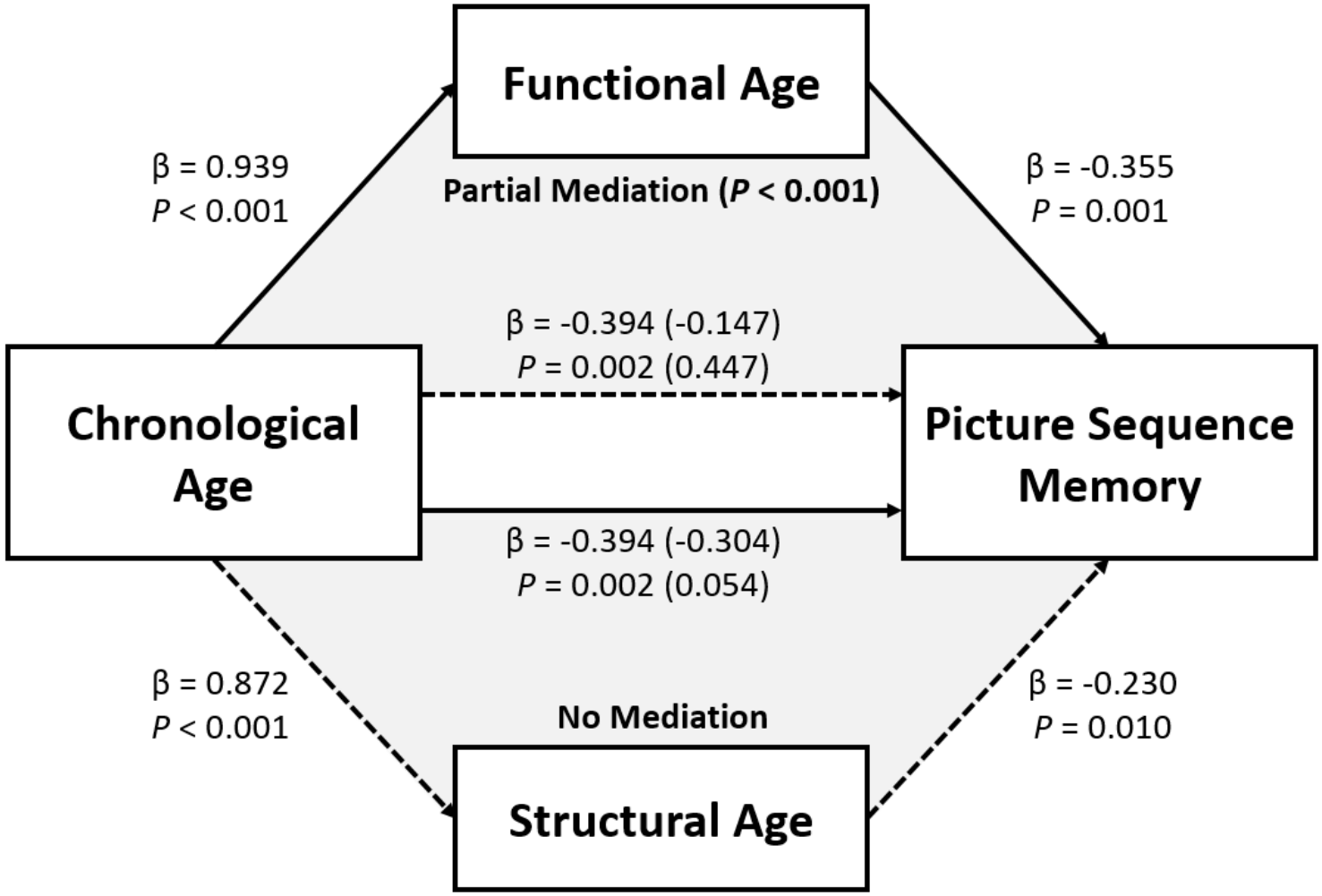

Figure 14. Mediation Analysis of Brain Age Correlation. This diagram shows the results from the mediation analysis for Picture Sequence Memory test. The independent variable was the chronological age of the TLE patients. The mediator was either their functional or structural brain age. Functional age partially mediated ($P < 0.001$) the association between chronological age and the test score (top triangle), whereas structural age did not (bottom triangle). Numbers in parentheses are results after the mediator was introduced.

4.3.6 Discussion

Accelerated aging is evident not only in the structural brains of patients with TLE, but also in their functional brains. This confirms and expands prior findings, here in a TLE group. Pardoe et al.[169] and Sone et al.[170] trained their regression models with a larger number of healthy control data (N = 2,001 and 1,196 respectively). Although the present study comparably lacks power in the trained age regression models (N = 151), the parameters and qualities of MRI images here are more controlled and the comparisons between the structural and functional brain ages provide novel insights into different dimensions of the brain aging effect in TLE.

While the overall structural and functional brain ages are indeed accelerated compared to chronological age (**Figure 12**), inspection of the age discrepancy plots (**Figure 10**) shows that this accelerated aging effect is evident across the chronological age range of the TLE participants examined here. We did not observe increased accelerated brain aging in the older compared to younger TLE participants, nor in participants with longer history of seizures compared to shorter.

It is worth noting that the correlations among the chronological and the two brain ages were significantly weaker in TLE patients compared to healthy controls ($p's < 0.01$), suggesting a detectable dissociation of brain ages from chronological age. Weintraub et al.[182] reported correlations between chronological age and NIH Toolbox Battery-Cognition Battery (NIHTB-CB) test scores in healthy controls ($N > 230$), and observed significantly stronger negative correlations in fluid cognitive abilities ($p's < 0.001$, $-0.46 > r's > -0.65$) compared to those seen in the TLE patients in our study ($p's < 0.03$, $Z > 2.2$). Together with the finding that the functional age mediated the relationship between chronological age and cognition in TLE patients, this leads us to conclude that judgment of cognitive abilities in the TLE patients based on their chronological ages may be less predictable compared to healthy controls.

There were significantly smaller variances ($p < 0.001$) in the predicted accelerated functional brain ages compared to the structural ages, both from the healthy control and TLE groups (**Figure 12** and **Figure 10**), although the opposite was expected given the increased complexity of the model (71,631 dimensions in functional, compared to 254 in structural). This suggests that the functional brain age calculated from resting-state functional connectivity is a more stable measure of brain age.

It was hypothesized that accelerated brain aging in TLE was related to either or both the clinical features of the epilepsy and AED use. Accelerated functional brain age was correlated with both complex partial seizure frequency (corrected $p = 0.07$) and the number of AEDs (corrected $p = 0.07$), suggesting that the accelerated functional brain aging in TLE patients may be related to both seizure burden and related polytherapy. Results from this study confirm those from Pardoe et al.[169] and Sone et al.[170] which reported that there was no significant relationship between epilepsy duration and the accelerated brain age. However, the relationship between age of seizure onset and the accelerated brain age in our TLE population was found insignificant. Current data in this study were not sufficient to reveal definitive clinical correlates of accelerated brain aging.

Table V depicts the dynamic nature of the relationships between chronological and brain ages with crystallized and fluid cognitive abilities. In regard to crystalized abilities, only chronological age predicted improvement on one of the two measures, predictably showing improving naming ability with age. In contrast, the interplay of chronological age with brain ages was more dynamic for fluid abilities. Moreover, functional brain age partially mediated the relationship for three measures including memory (Picture Sequence Memory) and selected measures of executive function (Dimensional Change Card Sort, Flanker Inhibitory Control and Attention) (**Figure 14**). Importantly, these brain age relationships are detected in a predominantly

young to middle age sample (mean age = 40.3) who have yet to enter the epoch where age exerts stronger and more diverse effects. It will be important to continue to monitor these relationships prospectively to confirm their change over time and linkages to changing cognition.

4.3.7 Limitations

One limitation of this investigation is the relatively small sample sizes. In order to control for the scanner variability, scan protocols and procedures, only data from the two Disease Connectome Studies (ECP and ADCP) were used. This resulted in a smaller training sample size compared to previous studies[169, 170], while allowing us to expand the study to investigate the functional brain aging and other clinical and cognitive traits in TLE.

The age range of our TLE population (19 – 60 years) was towards the younger spectrum of that of our control population (18 – 89 years). The results using this dataset should remain valid, since 1) the age range of the training set covered that of the testing set, and 2) the testing results on the healthy controls confirmed the performance of the linear correction. Before the linear correction, the bias in the regression model over-estimated the ages of young test subjects, making the prediction of TLE brain ages unreliable. The correction mitigated, if not completely removed, this bias effect[180]. Use of larger training sample sizes in conjunction with accurate non-linear regression models will create more robust age-predicting models. Future work is also desired to confirm the findings from the current study in older TLE population.

4.4 Concluding Remarks

Only a modest success has been achieved at building reliable machine learning models using the ECP data, where the performance was most likely limited by the small sample sizes.

Although the state-of-the-art quality MRI images and a thorough neuropsychological battery were used in the training, building clinically useful machine learning models seemed to require much larger samples.

Having enough sample sizes is a relative matter: it largely depends on the difficulty of the problem (classification or regression) and the given set of clues (or training features). For example, we may expect to build a reasonable classification model that separates between males and females with their height and weight as features, but not as well with eyesight and number of fingers. On the other hand, a very small sample size is required to build such a model if one of the training features happens to be the sex chromosome. Also, we intuitively expect separating between males and females to be easier than, for example, between a married person and a single. Likewise, in medical imaging, each problem is unique and has its own target sample sizes for training effective machine learning models.

We can hypothesize that the general relationship between the sample size and the classification accuracy, however, follows a predictable trend. One model may reach a target accuracy faster due to the simplicity of the problem compared to others, but the general shape of the relationship curve may look similar. Understanding this relationship can aid the assessment of machine learning models that have been trained in the field so far, because it is likely that they have not reached their maximum potential with the limited sample sizes available at the time of training. In Chapter 5, therefore, we will systematically explore this relationship in binary classification.

Chapter 5

Sample Size Limitations of Applying Machine Learning in Medical Imaging[§]

With advancements in medical imaging, the amount of data to evaluate exponentially grows and so does the complexity of clinical problems[12]. It is becoming more impossible for human radiologists to analyze every detail in the high quality, high dimensional images that the state-of-the-art imaging devices offer. The need for developing automated systems to help processing these images is clear. Then, when do we start trusting machines to the point where we confidently give them the same responsibilities as the human radiologists? Understandably, this question is loaded with not only technical, but also moral and logistical issues. However, in order to start the discussion, a thorough inspection of the current status in developing such machines must precede. And as discussed earlier, when the sample sizes are limited, we must assess the problem with the sample size in mind.

[§] Portions of this work are currently being reviewed: Hwang G, Nair VA, Bendlin BB, Prabhakaran V, Meyerand ME. Support Vector Machine Binary Classification for Diagnosis in Neuroimaging: the Sample Size Limitations. *American Journal of Neuroradiology*. Under Review

5.1 Challenges of Whole-brain MRI Classification

Machine learning has a potential to make significant impact in medical imaging, but it is still at its infancy[183]. Areas with the most number of successful applications are image segmentation and registration[3, 7, 8]. Applications in image segmentation include delineating tissue interfaces or detecting abnormal cells, such as tumor. They are substantially reducing the amount of human work required to solve these problems and increasing the output quality. One of the key reasons why we already see high-performing models in this area is that accumulating large training dataset is relatively easy with data multiplication techniques such as data augmentation[184]. Data augmentation allows a single training data point to multiply to be many. For example, when the problem is to segment a tumor region out of a 2-dimensional image, one labeled image can be translated, reflected, rotated, stretched or down-sized, so that the model is exposed to many different examples from a single image. So, if a thousand images are required to train a reliable machine learning model, this number can be reduced to perhaps a hundred, with a proper use of data augmentation. The same technique can be applied to image registration problems, which effectively tackles sample size limitations.

Data augmentation is not always straight-forward or even possible. For example, if the problem is to classify people into two groups based on their region-of-interested-based structural brain MRI features, such as cortical thicknesses or volumes, then the features cannot be rotated or stretched. The MRI images themselves can be, but they would be re-aligned before the features get extracted. In this case, without the help of the traditional data augmentation, the problem is much more difficult to solve with machine learning, compared to the tumor segmentation problem above. The problem requires much larger sample sizes, while the complexity of the problem may

be worse. In medical imaging, unfortunately, acquiring enough samples for machine learning is often impractical without data augmentation.

Various feature extraction techniques have been developed, in order to reliably reduce the number of training features while conserving the most useful information for the model[185]. They are generally perceived as tools to reduce the high-dimensionality of the problem. However, from a different perspective, they can also be viewed as means to reduce the required sample sizes for solving a problem, because the lower dimensionality reduces the chance of overfitting and we can expect to build a reliable model with smaller samples. There are numerous proposed feature extraction techniques, in which some conserve the original values and only remove unwanted features[81, 95, 186], whereas in others they use combinations of features to create more meaningful features using techniques such as PCA[94], or singular value decomposition (SVD)[187, 188]. Some studies have compared the effectiveness of using these techniques in a given application by comparing the accuracies[189], but none of them approached it from the perspective of reducing the required sample sizes.

5.2 Sample Size and Machine Learning Classification

Various machine learning classification models have been tried in medical imaging and SVM[30, 89] is among the most popular techniques. It has been shown effective in high-dimensional classification problems[190]. There have been a number of studies on the sample size requirements when using traditional machine learning classification models including SVM[191]. Mukherjee et al. and Figueroa et al. fitted SVM classification error curves to inverse power law models and introduced methods to predict the true accuracy given small sample sizes[16, 192]. Dobbin & Simon, and Guo et al. trained different types of omics data to test the sample size requirements[193].

5.2.1 Participants

MRI data from the Connectome studies were examined: HCP for healthy young adults[64], ECP for TLE patients[29] and ADCP for Alzheimer's disease and mild cognitive impairment (MCI) patients[84]. The processed images from the HCP were publically available on the ConnectomeDB web database[194]. All images were acquired with 3T Siemens or GE (General Electric) scanners using simultaneous multi-slice imaging (8 bands, TR < 802ms, voxel size = 2mm isotropic). Although the exact scanner parameters vary slightly, this was not the main focus of the current study. The images were processed using the Human Connectome Project pipelines as described in Section 2.4.1.

Four binary classification problems were defined (**Table VI**): classifying between 1) healthy males and females (HCP–Sex, N = 440 per group), 2) healthy twenties and thirties (HCP–Age, N = 445 per group), 3) healthy controls and temporal lobe epilepsy patients (ECP, N = 94 per group), 4) healthy controls and the combination of Alzheimer's disease and mild cognitive impairment (MCI) patients (ADCP, N = 63 per group). HCP–Sex problem represents a relatively easy problem, whereas HCP–Age problem a relatively challenging problem.

In each classification problem, two types of training feature set were investigated in order to consider two cases with widely different feature dimensionalities (or number of training features): structural (254 dimensions) and functional (71,631 dimensions) brain feature sets. 254 structural features were extracted from the T1-weighted images using FreeSurfer[65], including cortical thicknesses, surface areas, volumes and also subcortical and global volumes. These structural features were transformed to z-scores. From the resting state images, 360 timeseries from the Glasser parcellation[67] plus 19 FreeSurfer subcortical regions[68] were extracted (see Section 2.4.2), and the Pearson correlation was used to generate connectivity matrices and then normalized

with Fisher-*z* transformation. Taking the upper triangles of the matrices resulted in 71,631 features for the training. Therefore, in total, there were eight distinct classification problems to be investigated.

Dataset	Group	*N*	Age (years)	Sex (Male / Female)
HCP-Sex	Male	445	28.5 ± 3.2	445 / 0
	Female	445	28.2 ± 3.5	0 / 445
HCP-Age	20s	440	26.3 ± 2.2	180 / 260
	30s	440	32.3 ± 1.7	161 / 279
ECP	TLE	94	40.7 ± 12.4	41 / 53
	Control	94	42.7 ± 16.2	43 / 51
ADCP	AD + MCI	63	72.1 ± 8.9	34 / 29
	Control	63	70.8 ± 6.9	31 / 32

Table VI. Summary of Four Binary Classification Problems. Two problems involved only healthy controls (HCP-Sex, HCP-Age) and the other two patient populations (ECP, ADCP). This table summarizes the demographics of each group. †AD = Alzheimer's Disease. MCI = Mild Cognitive Impairment.

5.2.2 Hyperparameters Tested

SVM binary classification model training and testing were implemented in MATLAB R2018a. For the baseline, a linear kernel and leave-two-out cross validation (one subject from each group left out for the testing) were used with no feature reduction. Note that the samples in this study are only split into training and testing sets, and the term "cross validation" is used to denote that the testing is repeated exhaustively. This is not to be confused with a validation set[97].

The goal was to study the relationship between the sample size and the classification performance. For simplicity, the number of samples was kept equal between the two groups in comparison throughout the study, which minimized the concern of unbalanced sensitivity and

specificity[195] (more discussion in Section 5.2.3). In the rest of Chapter 5, N will represent the total number of samples in each group, instead of the combined.

In each problem, subsets of the entire dataset, with varying size, were randomly selected, and used to train and test SVM models. The sample size of a subset in each group is denoted as N_{sample} ($N \geq N_{sample}$). For each problem, 15 logarithmically spaced sample sizes or N_{sample}'s between 5 and N were tested. Training and testing for each N_{sample} were repeated until 95% confidence level was reached that the mean accuracy was within ±1.0%. The mean and the standard deviation of the classification accuracy over all randomly selected subsets were calculated per N_{sample}.

Two other generic kernels were tested while keeping other parameters match the baseline setup: 3rd order polynomial and radial basis function (RBF, or Gaussian) kernels. Three other K-fold values were tested while keeping other parameters match the baseline setup: 2, 5, and 10-fold testing. $N_{sample}{}'s < 10$ were skipped for 10-fold testing.

Three feature reduction methods were tested while keeping other parameters match the baseline setup: PCA[94], SVD[187, 188], and Lasso methods[30, 81]. For PCA, the number of features was reduced to the number of training sample minus one ($N_{train} - 1$), or kept equal for cases where $N_{train} > number\ of\ original\ features$ ($N_{train} + N_{test} = N_{sample}$). Then, the features were normalized to z-scores. For SVD, the number of features was reduced to the number of training samples (N_{train}), or kept equal where $N_{train} \geq number\ of\ original\ features$. For Lasso method, lambda of 0.1 was used and features with zero regression coefficients were eliminated. The reduction was applied first on the training set, and the information such as the coefficients were kept for later applying the same reduction independently on the testing set. Note that this

procedure does not allow the K-fold holdout testing set to influence the training, and therefore, the testing is unbiased.

5.2.3 Accuracy and Precision

Only equal-sized groups were considered in this work, in order to minimize the concern of unbalanced sensitivity and specificity[195]. One measure of this unbalance is the precision of the model.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

Precision of 50% would indicate perfectly balanced sensitivity and specificity. The precision of our results ranged from 42% to 56%. This range was considered acceptable for the purposes of our work, because with the balanced group sizes, the total accuracy should not vary significantly within this range of precision.

5.2.4 Effects of Kernels

In all eight classification problems, the error rate (1 – accuracy) steadily decreased following a power law, while still showing room for improvements at $N_{sample} = N$ (**Figure 15**, A and B). As expected, the HCP–Sex classification model yielded superior performance compared to the HCP–Age model, whose problem was selected to represent a relatively difficult task (separating between brains of 20s and 30s). Using the structural features, the trend in performance of the ADCP model resembled that of the HCP–Sex model, while the ECP model showed the worst trend. Using the functional features, the performance of the two disease models were in between the two healthy control (HCP) models.

The change of kernels showed mixed effects on the classification accuracy (**Figure 15**, A and B): in some cases, the change of kernel significantly improved the accuracy, and in others, it decreased it. The effects of changing kernels were unpredictable.

Leave-two-out cross validation was used for the baseline analyses, and the classification accuracy as N_{sample} approached N seemed unstable in many examples, especially with the disease models with smaller N (**Figure 15**, C and D). Standard deviation of the accuracy was as high as $\pm 18.5\%$ in T1 problems and $\pm 19.7\%$ in RSFC problems at $N_{sample} = 5$, and gradually decreased to around $\pm 5\%$ as N_{sample} increased to 50 (per group).

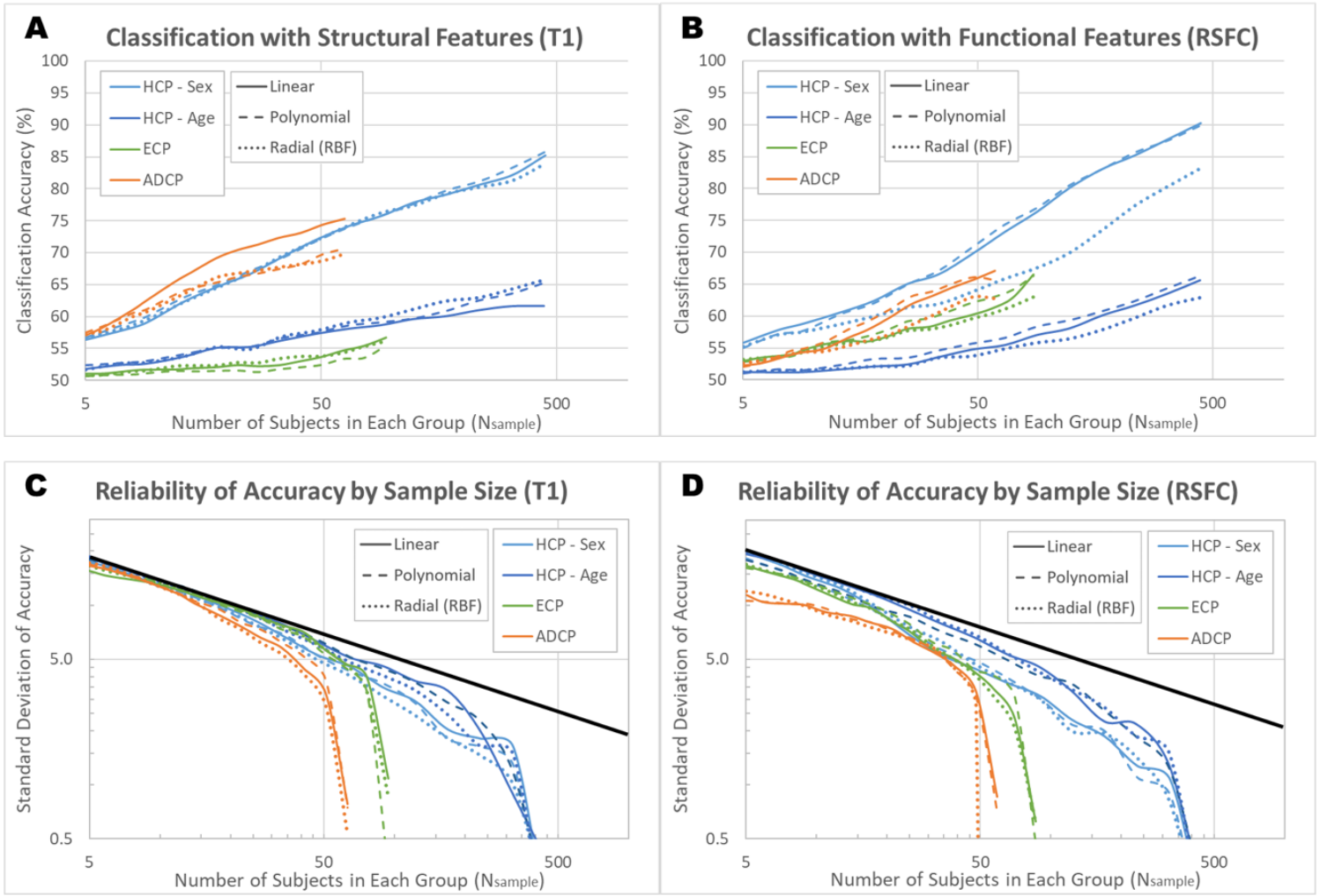

Figure 15. Accuracy of Eight Classification Models with Varying Kernels. Support vector machine (SVM) classification results using leave-two-out cross validation and no feature selection. Three generic kernels were tested: linear, polynomial (3rd order) and radial-based function (RBF) kernels. (A) and (B) summarize the sample size relationship with binary classification accuracy, and (C) and (D) with the standard deviation of the accuracy. N_{sample} represents the number of subjects in each group, instead of the combined. †RSFC = resting-state functional connectivity.

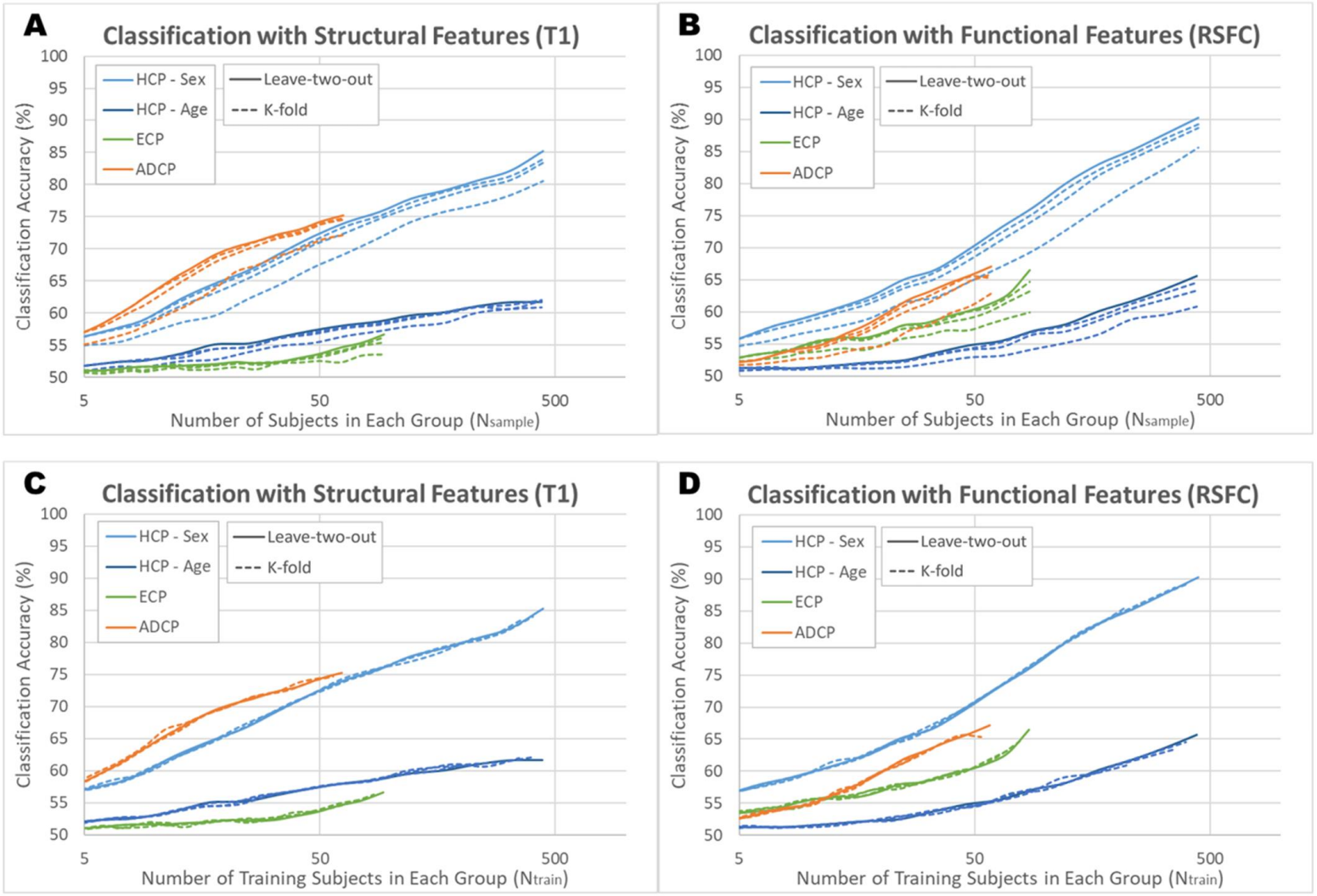

Figure 16. Accuracy of Eight Classification Models with Varying K-Fold. Support vector machine (SVM) classification results using linear kernel and no feature selection. Four K-fold settings were tested: 2, 5, 10-fold, as well as leave-two-out (N_{sample}-fold) testing. (A) and (B) show the results with the x-axis being the number of subjects in the subset per group (N_{sample}), whereas (C) and (D) the number of training subjects in the subset per group (N_{train}). †RSFC = resting-state functional connectivity.

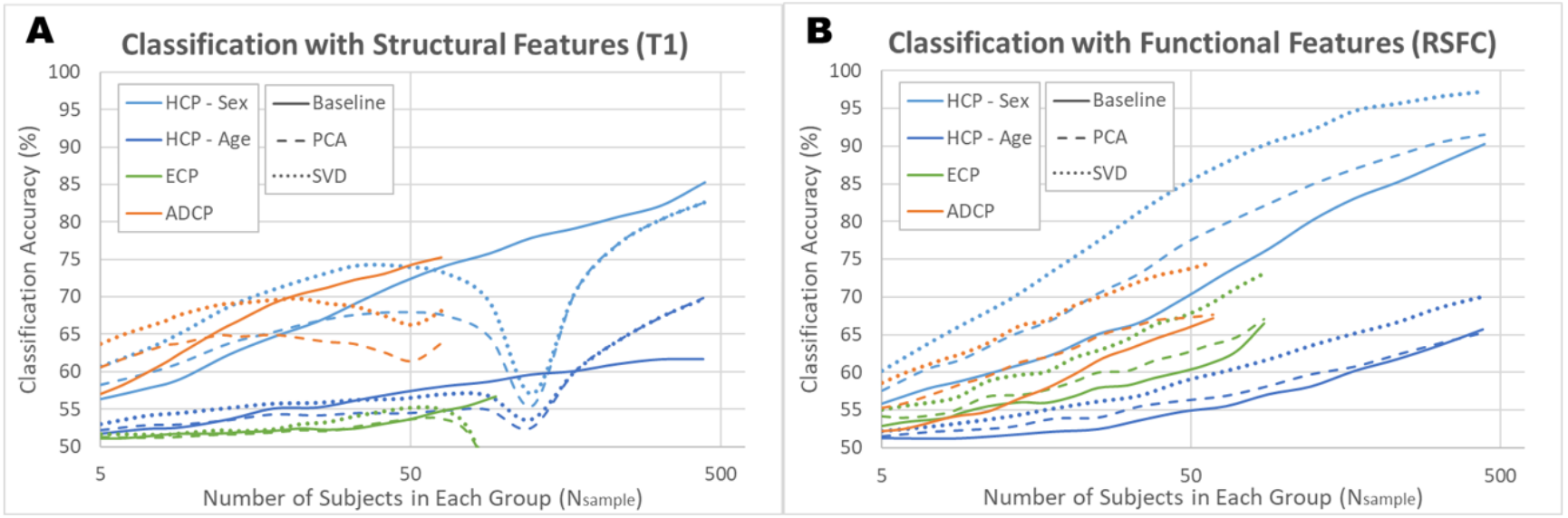

Figure 17. Accuracy of Eight Classification Models with Feature Reduction. Support vector machine (SVM) classification results using linear kernel and leave-two-out cross validation. Feature reduction using singular value decomposition (SVD, dotted) improved the accuracies significantly more than using principal component analysis (PCA, dashed) in all examples, especially when $N_{sample} \ll number\ of\ original\ features$. †RSFC = resting-state functional connectivity.

5.2.5 Effects of K-fold

Using smaller number of K-folds, or holding out bigger testing sets consistently decreased the accuracies (**Figure 16**, A and B). Plotting the results with the number of training subjects per group ($N_{train} = N_{sample} - N_{test} = N_{sample} \times \frac{K-1}{K}$) as the x-axis (**Figure 16**, C and D) revealed that the classification accuracy is closely related to N_{train}, or the number of subjects per group that the model was trained on.

5.2.6 Effects of Feature Reduction

Feature reduction using SVD methods most significantly improved the classification accuracy, whereas Lasso methods were the least effective (**Figure 17**). In fact, Lasso feature reduction significantly decreased the final accuracy ($N_{sample} = N$) for seven out of eight examples.

With the functional models, both PCA and SVD methods significantly improved the accuracies ($p < 0.001$), while SVD performing much better than PCA ($p < 0.001$). SVD feature reduction achieved boost in the final accuracy ($N_{sample} = N$) by 4 – 7%, while PCA by 0 – 1%. With the structural models using PCA or SVD, the classification accuracies were improved with relatively low N_{sample}, then there was a sudden decrease in accuracy as N_{sample} approched the number of original features. When $N_{sample} > number\ of\ original\ features$ (HCP problems), the outcome was unpredictable, with HCP-Age problem showing improvement, while not with HCP-Sex (**Figure 17**). This sudden decrease in accuracy did not appear in the functional models, because $N_{sample} \ll number\ of\ original\ features = 71{,}631$.

5.3 Sample Size Prediction Model

5.3.1 Sample Size and Classification Accuracy

It was previously suggested that the relationship between the sample size and the error rate (1 - classification accuracy) follow the inverse power law[16, 192], which would have a general form:

$$Error\ Rate = 1 - Accuracy = \frac{1}{x_1 \times N^{x_2}} \tag{2}$$

N is the total sample size, and x_1 and x_2 are the model parameters. This equation needs modification as not all model will reach a 100% accuracy even with infinitely large N. From empirical search, this slightly modified equation below gave the best nonlinear fit to the sample size relationship both visually and from the residual statistics:

$$\begin{aligned} Binary\ Classification\ Accuracy &= f(N) \\ &= 50 + (\mathrm{x}_1 - 50) \times \left(1 - \frac{1}{1 + x_2 \times N^{x_3}}\right)(\%) \end{aligned} \tag{3}$$

Notice that equation (3) was slightly modified from (2) so that the model parameter x_1 gives the maximum possible accuracy, and $f(N = 0) = 50\%$. Parameters x_2 and x_3 change how fast the accuracy improves as the sample size increases.

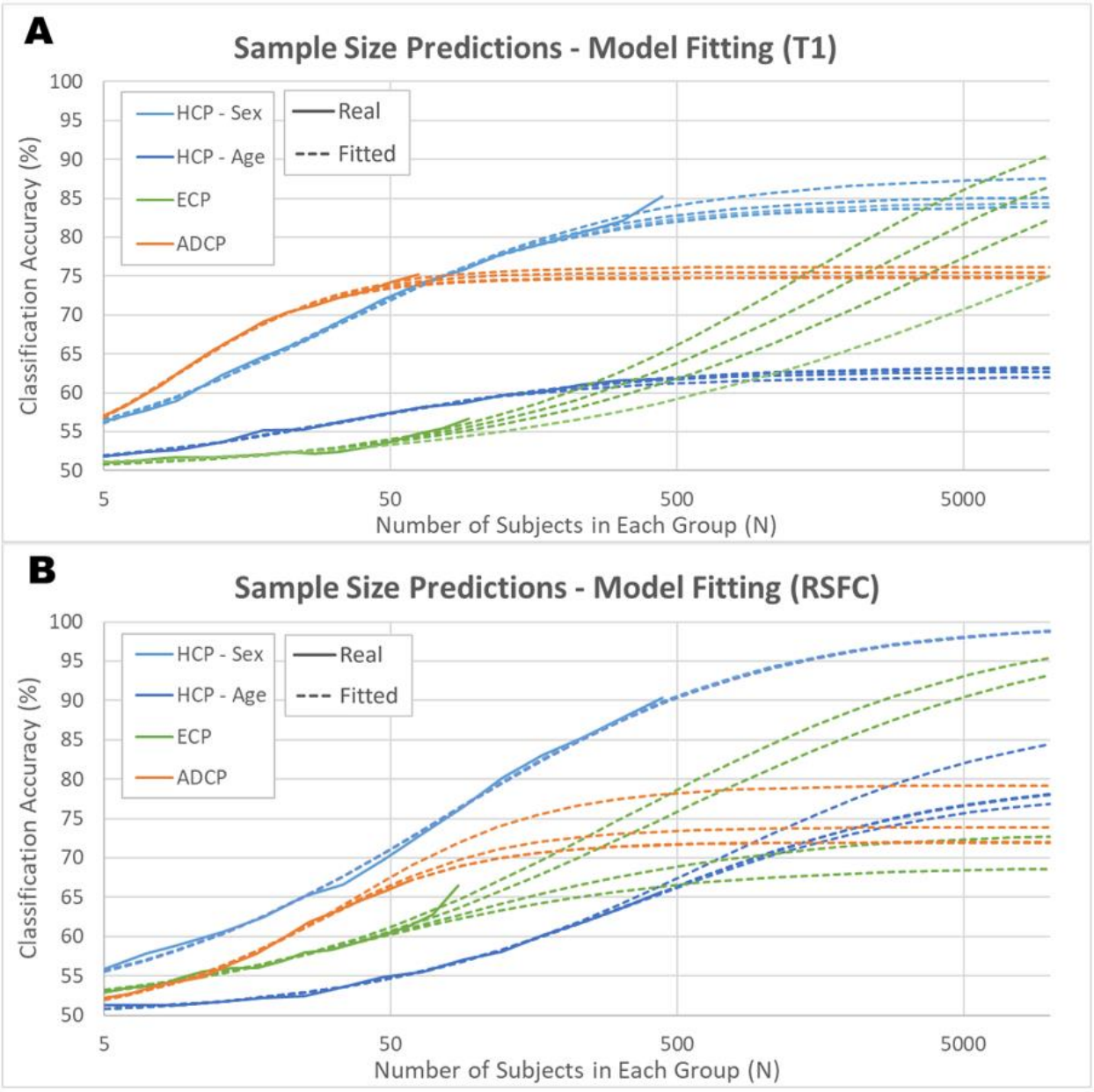

Figure 18. Predicting Future Accuracy of the Eight Models. Support vector machine (SVM) baseline classification results (solid line) and the fitting results (dotted lines). Same model (**Equation 3**) was fitted four times: first with all data points, and then with one, two, or three last data points out (15 logarithmically-spaced data points in total per problem). †RSFC = resting-state functional connectivity.

5.3.2 Sample Size Model Fitting

Equation 3 was fitted four times per problem: first with all data points, and then with one, two, or three last sample data points (N_{sample}) out (15 logarithmically-spaced data points in total per problem) (**Figure 18**). The fitted lines matched the available data points well both visually and based on the residual statistics. For problems with smaller total samples sizes (N), especially for ones that demonstrated higher degree of overfitting (ECP problems), the predictions were more unstable. For problems with larger N (HCP problems), the prediction seemed more reasonable and stable.

Fitted lines predicted that the HCP-Sex structural model would reach 84 – 88%, and its functional model nearly 100%. The HCP-Age structural model was predicted to reach 62 – 63%, and its functional model 79 – 88%. The ECP structural model was predicted to reach nearly 100%, and its functional model 69 – 100%. The ADCP structural model was predicted to reach 75 – 76%, and its functional model 72 – 79%.

5.4 Discussion

5.4.1 Reliability of Classification Accuracy

Machine learning classification accuracy is unreliable with small sample sizes, due to the large variance and overfitting[23]. In medical imaging, it is often difficult to accrue large amounts of training dataset for machine learning research, and many publications vaguely state the low sample size as the biggest limitation of their work. Therefore, it is important to study the relationship between sample size and the model performance. Here, we tested the SVM binary classification algorithm with a number of different kernels, testing K-folds, and feature reduction methods.

As discussed earlier, most machine learning publications do not consider the sample size relationships of their model performance. However, the true reliability of a model cannot be correctly assessed only with the given dataset as evident from **Figure 15**, (C) and (D). The standard deviation of the model accuracy starts from around $\pm 15\%$ with $N_{sample} = 5$, and steadily decreases until it drops sharply near $N_{sample} \approx N$, because this is analogous to a sample standard deviation, instead of a population standard deviation. The same phenomenon is also found with smaller K-fold settings, albeit to lesser degrees. In order to correctly assess the true reliability at a given sample size, a larger sample needs to be tested. For example, the standard deviation of ADCP T1 problem was found to be $\pm 0.78\%$ with $N_{sample} = 63$, but that of HCP-Sex T1 problem was $\pm 4.55\%$ with $N_{sample} = 65$, which should be closer to the true population standard deviation. The results suggest that, for similar problems, if $\pm 5\%$ standard deviation is desired, at least $N = 50$, or more conservatively $N = 100$ is required. If $\pm 1\%$ standard deviation is desired, around $N = 5{,}000$ may be required for similar problems.

5.4.2 Sample Size and Classification Accuracy

Using a smaller number of K-folds (or holding out a larger testing set) consistently decreased the classification accuracy. **Figure 16** effectively shows that the reason for this is due to the smaller training sample sizes. The number of training subjects in each group (N_{train}) is closely related to the classification accuracy. The common practice of preferring K-fold over leave-two-out or leave-one-out testing to minimize overfitting (evident from curved ends in **Figure 15**) is justified; however, it does so at the cost of the training sample size which is more difficult to afford in medical imaging.

Previous findings suggested that the relationship between the error rate (1 – accuracy) in binary classification and the sample size followed a power law function (**Equation 2**)[16, 192], and a slightly modified version (**Equation 3**) showed the best fitting results. Although this fitted model may conveniently provide predictions of future accuracies, overfitting can be problematic as seen from the ECP problems in **Figure 18**. In addition, the accuracies in the problems that we investigated had not reached their maximum accuracies yet, which still left the validity and universality of this fitted model for further investigation and confirmation.

Once the relationship is established, it can be used to make predictions and to examine whether a model has room for potential improvements with added training samples in the future. A classification model that gives only a sub-optimal accuracy can be promoted to a clinical tool, if it shows enough potential. Also, a classification model that shows little potential can be deserted quickly to save research time and effort. Moreover, the knowledge of this relationship can be informative when designing new clinical studies, as it suggests reasonable sample sizes to recruit per clinical group.

5.4.3 Feature Reduction

Three common feature reduction methods were explored and the best results in terms of improving classification accuracy were achieved using SVD in all eight problems, especially when $N_{sample} \ll number\ of\ original\ features$. PCA also consistently improved the accuracies, but to a lesser degree compared to SVD. Lasso methods did not improve the final accuracies ($N_{sample} = N$), most likely due to information loss.

Both SVD and PCA attempt to summarize high-dimensional information using smaller number of dimensions, and they are closely related. In fact, PCA can be performed using SVD,

because it can be considered as a special case of SVD[188, 196]. However, performing SVD is computationally less demanding compared to PCA, and therefore, has an inherent advantage when used in machine learning[197].

The present results are in favor of using SVD over the other two feature reduction methods when the number of features is much greater than the available sample size. However, the results may differ with other types of applications, and therefore, if enough time is available, testing a variety of feature reduction methods is the most ideal. In addition, only linear feature reduction methods were considered in this study. If enough samples are available, exploring nonlinear methods is a possibility, although, considering the complexity of this search, building and optimizing small neural networks may turn out to be a more efficient and effective option.

5.4.4 Machine Learning Research in Medical Imaging

Given the observations and discussions above, a number of approaches are proposed for future machine learning research in medical imaging, or wherever dealing with limited sample sizes. First, it is advisable not to jump to conclusions from results acquired with low sample sizes. Population standard deviation of 10% or more is expected for $N < 20$, and 5% or more for $N < 100$. Second, instead of using a K-fold testing, maximizing the training sample, while plotting the relationship between the model performance and sample size using subsampling methods introduced here, may provide better insight on the dataset. This will also allow assessing the trend of the performance to determine whether the model has room for improvements. Third, hastily generalizing the findings of increased or decreased performances from changing training parameters can be risky. They may be specific to the dataset and the best combination of parameters may not carry over to new datasets, possibly even to the same dataset with added

samples, especially when dealing with small sample sizes. The sample size plots help discriminate between good and bad parameter combinations per given dataset.

A machine learning model continues to learn after it is first trained, as it is exposed to more training samples, just as a human physician continues to learn throughout their residency and even while practicing. Therefore, machine learning research in medical imaging should focus less on the current accuracy, but more on its maximum potential accuracy that is reasonably achievable. Since the models can be trained much faster in clinical settings, compared to restricted research settings, more sub-optimal models showing enough potential should be promoted to clinical tools. They can first serve as a second eye to trained physicians and then, after a period of extra training with immense clinical data and reaching their target accuracy and reliability, can become stand-alone tools.

5.4.5 Limitations

In this work, only eight selected research problems which have similar data types and qualities have been tested. Although their accuracies showed very similar sample size relationships, and the fitting results in **Figure 18** were successful, these may not represent all problems in medical imaging. Future work is needed to confirm the findings here with other feature types and feature dimensionality.

Only equal-sized groups were considered in this work for simplicity, because unbalanced groups cause unbalanced sensitivity and specificity using default training parameters. However, many medical imaging classification problems involve unbalanced groups, especially when the disease to be diagnosed is rare. Future work is needed to expand the findings here to cases with significantly unbalanced sample sizes.

There is no one-size-fits-all solution to training machine learning models, and each model tackling unique problems must be fine-tuned in order to handle the specific dataset well. Significant improvements in performance can be gained by optimizing the hyperparameters. However, in a systematic search considering multiple models at multiple sample sizes, sufficient time for this fine-tuning cannot be allotted to each model, and only a few default settings can be reasonably tested. Also, models that require relatively longer training periods such as deep learning are difficult to consider. Therefore, the predicted performance from these research efforts should be regarded as the lower bounds of the actual performance. A carefully fine-tuned deep learning model is expected to out-perform a linear SVM model.

5.5 Concluding Remarks

Applying machine learning in medical imaging problems is mostly limited by the small sample sizes. When sample sizes are small, large uncertainty in performance measurements is expected, along with overfitting. Here a number of research approaches have been proposed, including plotting the performance over the sample sizes, maximizing the training sample, using SVD feature reduction, making predictions of future accuracy, and identifying sub-optimal models with potentials. These guidelines will help the accurate assessment of medical imaging classification models, and ultimately allow the field to reach its goal faster, which is to make timely, accurate, and reliable medical diagnoses.

Chapter 6

Conclusion and Future Works

Machine learning has been drawing a massive attention in the past decade. The amount of research publications and also commercial products using machine learning have exploded and do not yet show signs of slowing. It is changing the ways humans view things and perform tasks. The versatility of it is being exploited in literally every field of study. The predictions that it offers are intriguing, but systematic patterns within complex datasets that it detects along the way also offer tremendous insight.

In the midst of all the hypes and success stories in the media, however, there are many areas that are still in the early developing stages, mainly due to the limited sample sizes. Then, the question becomes when we can expect to see useful products from these areas, if that is feasible. One of these areas is the whole-brain MRI classification, and the problem becomes more challenging with rare diseases.

In this work, a neurological disorder known as temporal lobe epilepsy (TLE) was investigated using machine learning. The ECP provided comprehensive and high-quality imaging and neuropsychological testing data to train the models with. They would have been sufficient amount of information for many other traditional statistical analyses, but for machine learning, they seemed to fall short. Despite the limitations, we were able to make a number of noteworthy discoveries as well as train promising machine learning models. And these results suggest interesting research topics for the future.

6.1 Epilepsy Research

There is still a lot to be discovered about epilepsy. Because of the temporal nature of this neurological disorder (seizure activity), the direct search for the underlying cause, biomarkers, and the cure for epilepsy has only been effective with the developments of imaging devices capable of capturing temporal (functional) information (EEG or fMRI). With the advancements of imaging devices will come new discoveries on epilepsy. One example of this was introduced in Section 3.3, which showed that the classic model of associating cognitive abnormality in epilepsy with the disordered pathophysiology of specific epilepsy syndrome seemed to be challenged. Imaging results revealed that a large proportion of TLE patients instead suffered whole brain abnormality. The accelerated brain aging of TLE patients introduced in Section 4.3 also provided an interesting viewpoint of epilepsy.

6.1.1 Machine Learning Classification of Epilepsy Subgroups

As discussed in Section 4.2, the correct goal in applying machine learning classification on epilepsy data should not be to separate between epilepsy patients and healthy controls, but to

separate among subgroups in epilepsy. A clinically useful machine learning model should provide information to physicians which would otherwise not have been easily acquired, but which would help in the assessment of the disease or the treatment planning. A few examples of this would be predicting seizure origin (by lobes, or hemispheres), course of epilepsy, severity of structural or functional brain damage, developing symptoms, best AED to be treated (or else refractory), or success rate of lobectomy. According to the results in Sections 5.2 and 5.3, any of these problems would likely to require data from at least hundreds of epilepsy patients just to start assessing whether such model can reasonably be built.

On the other hand, as shown in Chapter 3, there are still lots of insight to be gained from machine learning research, even if the trained model itself may not be clinically useful. For example, studying a machine learning model that effectively separates between TLE patients and healthy controls can reveal patterns in the training dataset that have not been discovered previously. This means that the possibilities are endless in the use of machine learning in epilepsy research. If only a valid research question is posed and if a sensible machine learning model can be trained, the underlying features can be analyzed, similar to the approach in Section 4.3.

6.1.2 Data-driven Clustering of Epilepsy Subgroups

The class label is crucial in machine learning classification. For example, if a TLE patient is mislabeled as a healthy control, this is devastating for the training. Therefore, carefully screening for mistakes in the class labels is a critical step. However, it can be an issue even when there are no mistakes: if the classes are not well separable using the given feature set.

This can occur in two cases: an ill-posed problem, or ill-defined classes that are not supported by the data. First, the HCP–Age problem (separating between brain images of healthy 20s and 30s) in Section 5.2 is an example of an ill-posed problem. Intuitively, we expect that the

training of this model would be more challenging compared to the HCP–Sex problem (separating between males and females). Second, Section 3.3 suggests that simply designating one class to include all TLE patients may be ill-defined. Fixing this requires re-thinking the definitions that we are accustomed to. Was separating epilepsy patients based on their seizure focus the best strategy? Cluster analysis is a data-driven method which aims to organize data points into subgroups and can address this question. If the efforts in Section 6.1.1 turn out to be unsuccessful, these two cases can be checked for alternative solutions.

6.2 Sample Size Limitations

As in MRI research, where there are both researchers analyzing the images and ones developing hardware, in machine learning research, if there are the users, there must also be some that troubleshoot. A large amount of effort goes into building the best models, which grab more attention. Problems that are not showing high enough performance metrics get deserted quickly. Comparably few are devoted to troubleshooting problems that are currently not exciting due to either limited sample sizes or limited understanding of the problem. Chapter 5 explored the relationship between the sample size and binary classification accuracy, and proposed a few methods to identify machine learning models with enough potential. This can benefit many overlooked areas of study where previously limited by the lack of sample sizes. More of such systematic research effort is needed to achieve comprehensive understanding of the sample size issue, which can guide future machine learning research.

As discussed in Section 5.4.5, future work is needed to explore other types of classification problems in medical imaging: less controlled images (such as clinical), unbalanced sample sizes, significantly non-normal features (with skewed or bimodal distributions), other imaging

modalities and training features, etc. In each case, performance metrics other than the simple accuracy may be considered to be more appropriate, such as AUC, sensitivity/specificity, precision/recall, etc. Similar approach can also be used to study machine learning regression problems.

In order to combat the problem of low sample sizes directly instead of simply waiting for more data to be available, efforts are needed to develop effective feature extraction or feature reduction methods. Traditional data augmentation techniques are not straightforwardly applicable in many medical imaging diagnosis problems, for reasons discussed in Section 5.1. It would be beneficial not only to know the amount of sample sizes required to build reliable machine learning models, but also to have methods to reduce the sample size barrier because of the difficulty in accumulating large medical imaging data.

After gaining enough understanding on the relationship between sample size and machine learning performance, it would be intriguing to perform a meta-analysis to review published machine learning models for their true potential as most publications only assess their current performances. A thorough review may discover among a large pile of proposed models ones that actually show enough potential for pursuing further. This is the correct way of exploiting the important advantage of using machine learning, which is its ability to improve itself given more data. It is an exciting time to study medical imaging and epilepsy with the powerful tool of machine learning in hand.

Bibliography

1. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30. doi: 10.1161/Circulationaha.115.001593. PubMed PMID: WOS:000364632100010.

2. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sanchez CI. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88. Epub 2017/08/05. doi: 10.1016/j.media.2017.07.005. PubMed PMID: 28778026.

3. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. 2012;16(5):933-51. Epub 2012/04/03. doi: 10.1016/j.media.2012.02.005. PubMed PMID: 22465077; PMCID: PMC3372692.

4. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med. 2001;23(1):89-109. Epub 2001/07/27. doi: 10.1016/s0933-3657(01)00077-x. PubMed PMID: 11470218.

5. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015;349(6245):255-60. Epub 2015/07/18. doi: 10.1126/science.aaa8415. PubMed PMID: 26185243.

6. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016;375(13):1216-9. Epub 2016/09/30. doi: 10.1056/NEJMp1606181. PubMed PMID: 27682033; PMCID: PMC5070532.

7. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, Kim N. Deep Learning in Medical Imaging: General Overview. Korean J Radiol. 2017;18(4):570-84. Epub 2017/07/04. doi: 10.3348/kjr.2017.18.4.570. PubMed PMID: 28670152; PMCID: PMC5447633.

8. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 2017;19:221-48. Epub 2017/03/17. doi: 10.1146/annurev-bioeng-071516-044442. PubMed PMID: 28301734; PMCID: PMC5479722.

9. Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. Sci Rep. 2019;9(1):6381. Epub 2019/04/24. doi: 10.1038/s41598-019-42294-8. PubMed PMID: 31011155; PMCID: PMC6476887.

10. Davatzikos C. Machine learning in neuroimaging: Progress and challenges. Neuroimage. 2019;197:652-6. Epub 2018/10/09. doi: 10.1016/j.neuroimage.2018.10.003. PubMed PMID: 30296563; PMCID: PMC6499712.

11. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recogn Lett. 2010;31(8):651-66. doi: 10.1016/j.patrec.2009.09.011. PubMed PMID: WOS:000277552600002.

12. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. Radiographics. 2017;37(2):505-15. Epub 2017/02/18. doi: 10.1148/rg.2017160130. PubMed PMID: 28212054; PMCID: PMC5375621.

13. Wang Y, Fan Y, Bhatt P, Davatzikos C. High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. Neuroimage. 2010;50(4):1519-35. Epub 2010/01/09. doi: 10.1016/j.neuroimage.2009.12.092. PubMed PMID: 20056158; PMCID: PMC2839056.

14. Cole JH, Poudel RPK, Tsagkrasoulis D, Caan MWA, Steves C, Spector TD, Montana G. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. Neuroimage. 2017;163:115-24. Epub 2017/08/03. doi: 10.1016/j.neuroimage.2017.07.059. PubMed PMID: 28765056.

15. Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine Learning in Medical Imaging. IEEE Signal Process Mag. 2010;27(4):25-38. Epub 2010/07/01. doi: 10.1109/MSP.2010.936730. PubMed PMID: 25382956; PMCID: PMC4220564.

16. Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH. Predicting sample size required for classification performance. BMC Med Inform Decis Mak. 2012;12:8. Epub 2012/02/18. doi: 10.1186/1472-6947-12-8. PubMed PMID: 22336388; PMCID: PMC3307431.

17. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, Garcia-Pedrero A, Ramirez SC, Kong D, Moody AR, Tyrrell PN. Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review. Can Assoc Radiol J. 2019. Epub 2019/09/17. doi: 10.1016/j.carj.2019.06.002. PubMed PMID: 31522841.

18. Kohli MD, Summers RM, Geis JR. Medical Image Data and Datasets in the Era of Machine Learning-Whitepaper from the 2016 C-MIMI Meeting Dataset Session. J Digit Imaging. 2017;30(4):392-9. Epub 2017/05/19. doi: 10.1007/s10278-017-9976-3. PubMed PMID: 28516233; PMCID: PMC5537092.

19. Beleites C, Neugebauer U, Bocklitz T, Krafft C, Popp J. Sample size planning for classification models. Anal Chim Acta. 2013;760:25-33. doi: 10.1016/j.aca.2012.11.007. PubMed PMID: WOS:000314488300003.

20. Shaikhina T, Khovanova NA. Handling limited datasets with neural networks in medical applications: A small-data approach. Artif Intell Med. 2017;75:51-63. Epub 2017/04/02. doi: 10.1016/j.artmed.2016.12.003. PubMed PMID: 28363456.

21. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? arXiv [csLG]. 2015.

22. Domingos P. A few useful things to know about machine learning. Commun ACM. 2012;55(10):78-87. doi: 10.1145/2347736.2347755.

23. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. Neuroimage. 2018;180(Pt A):68-77. Epub 2017/06/29. doi: 10.1016/j.neuroimage.2017.06.061. PubMed PMID: 28655633.

24. Zack MM, Kobau R. National and State Estimates of the Numbers of Adults and Children with Active Epilepsy - United States, 2015. MMWR Morb Mortal Wkly Rep. 2017;66(31):821-5. Epub 2017/08/11. doi: 10.15585/mmwr.mm6631a1. PubMed PMID: 28796763; PMCID: PMC5687788.

25. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, Engel J, Jr., Forsgren L, French JA, Glynn M, Hesdorffer DC, Lee BI, Mathern GW, Moshe SL, Perucca E, Scheffer IE, Tomson T, Watanabe M, Wiebe S. ILAE official report: a practical clinical definition of epilepsy. Epilepsia. 2014;55(4):475-82. Epub 2014/04/16. doi: 10.1111/epi.12550. PubMed PMID: 24730690.

26. Barkley GL, Baumgartner C. MEG and EEG in epilepsy. J Clin Neurophysiol. 2003;20(3):163-78. Epub 2003/07/26. doi: 10.1097/00004691-200305000-00002. PubMed PMID: 12881663.

27. Ruber T, David B, Elger CE. MRI in epilepsy: clinical standard and evolution. Curr Opin Neurol. 2018;31(2):223-31. Epub 2018/02/02. doi: 10.1097/WCO.0000000000000539. PubMed PMID: 29389747.

28. Tellez-Zenteno JF, Hernandez-Ronquillo L. A review of the epidemiology of temporal lobe epilepsy. Epilepsy Res Treat. 2012;2012:630853. Epub 2012/09/08. doi: 10.1155/2012/630853. PubMed PMID: 22957234; PMCID: PMC3420432.

29. Cook CJ, Hwang G, Mathis J, Nair VA, Conant LL, Allen L, Almane DN, Birn R, DeYoe EA, Felton E, Forseth C, Humphries CJ, Kraegel P, Nencka A, Nwoke O, Raghavan M, Rivera-Bonet C, Rozman M, Tellapragada N, Ustine C, Ward BD, Struck A, Maganti R, Hermann B, Prabhakaran V, Binder JR, Meyerand ME. Effective Connectivity Within the Default Mode Network in Left Temporal Lobe Epilepsy: Findings from the Epilepsy Connectome Project. Brain Connect. 2019;9(2):174-83. Epub 2018/11/07. doi: 10.1089/brain.2018.0600. PubMed PMID: 30398367; PMCID: PMC6444922.

30. Hwang G, Nair VA, Mathis J, Cook CJ, Mohanty R, Zhao G, Tellapragada N, Ustine C, Nwoke OO, Rivera-Bonet C, Rozman M, Allen L, Forseth C, Almane DN, Kraegel P, Nencka A, Felton E, Struck AF, Birn R, Maganti R, Conant LL, Humphries CJ, Hermann B, Raghavan M, DeYoe EA, Binder JR, Meyerand E, Prabhakaran V. Using Low-Frequency Oscillations to Detect Temporal Lobe Epilepsy with Machine Learning. Brain Connect. 2019;9(2):184-93. Epub 2019/02/26. doi: 10.1089/brain.2018.0601. PubMed PMID: 30803273.

31. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, Consortium WU-MH. The WU-Minn Human Connectome Project: an overview. Neuroimage. 2013;80:62-79. Epub 2013/05/21. doi: 10.1016/j.neuroimage.2013.05.041. PubMed PMID: 23684880; PMCID: PMC3724347.

32. Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TE, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW, Della Penna S, Feinberg D, Glasser MF, Harel N, Heath AC, Larson-Prior L, Marcus D, Michalareas G, Moeller S, Oostenveld R, Petersen SE, Prior F, Schlaggar BL, Smith SM, Snyder AZ, Xu J, Yacoub E, Consortium WU-MH. The Human Connectome Project: a data acquisition perspective. Neuroimage. 2012;62(4):2222-31. Epub 2012/03/01. doi: 10.1016/j.neuroimage.2012.02.018. PubMed PMID: 22366334; PMCID: PMC3606888.

33. Glasser MF, Smith SM, Marcus DS, Andersson JL, Auerbach EJ, Behrens TE, Coalson TS, Harms MP, Jenkinson M, Moeller S, Robinson EC, Sotiropoulos SN, Xu J, Yacoub E, Ugurbil K, Van Essen DC. The Human Connectome Project's neuroimaging approach. Nat Neurosci. 2016;19(9):1175-87. Epub 2016/08/30. doi: 10.1038/nn.4361. PubMed PMID: 27571196.

34. Engel J, Jr. Approaches to refractory epilepsy. Ann Indian Acad Neurol. 2014;17(Suppl 1):S12-7. Epub 2014/05/03. doi: 10.4103/0972-2327.128644. PubMed PMID: 24791078; PMCID: PMC4001229.

35. Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon CS, Dykeman J, Pringsheim T, Lorenzetti DL, Jette N. Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. Neurology. 2017;88(3):296-303. Epub 2016/12/18. doi: 10.1212/WNL.0000000000003509. PubMed PMID: 27986877; PMCID: PMC5272794.

36. French JA. Refractory epilepsy: one size does not fit all. Epilepsy Curr. 2006;6(6):177-80. Epub 2007/01/30. doi: 10.1111/j.1535-7511.2006.00137.x. PubMed PMID: 17260051; PMCID: PMC1783491.

37. Holmes MD, Tucker DM. Identifying the epileptic network. Front Neurol. 2013;4:84. Epub 2013/07/13. doi: 10.3389/fneur.2013.00084. PubMed PMID: 23847586; PMCID: PMC3696895.

38. van Mierlo P, Holler Y, Focke NK, Vulliemoz S. Network Perspectives on Epilepsy Using EEG/MEG Source Connectivity. Front Neurol. 2019;10:721. Epub 2019/08/06. doi: 10.3389/fneur.2019.00721. PubMed PMID: 31379703; PMCID: PMC6651209.

39. Parker CS, Clayden JD, Cardoso MJ, Rodionov R, Duncan JS, Scott C, Diehl B, Ourselin S. Structural and effective connectivity in focal epilepsy. Neuroimage Clin. 2018;17:943-52. Epub 2018/03/13. doi: 10.1016/j.nicl.2017.12.020. PubMed PMID: 29527498; PMCID: PMC5842760.

40. Caciagli L, Bernasconi A, Wiebe S, Koepp MJ, Bernasconi N, Bernhardt BC. A meta-analysis on progressive atrophy in intractable temporal lobe epilepsy: Time is brain? Neurology. 2017;89(5):506-16. Epub 2017/07/09. doi: 10.1212/WNL.0000000000004176. PubMed PMID: 28687722; PMCID: PMC5539734.

41. Doucet GE, He X, Sperling M, Sharan A, Tracy JI. Gray Matter Abnormalities in Temporal Lobe Epilepsy: Relationships with Resting-State Functional Connectivity and Episodic Memory Performance. Plos One. 2016;11(5):e0154660. Epub 2016/05/14. doi: 10.1371/journal.pone.0154660. PubMed PMID: 27171178; PMCID: PMC4865085.

42. Keller SS, Roberts N. Voxel-based morphometry of temporal lobe epilepsy: an introduction and review of the literature. Epilepsia. 2008;49(5):741-57. Epub 2008/01/08. doi: 10.1111/j.1528-1167.2007.01485.x. PubMed PMID: 18177358.

43. Doucet G, Osipowicz K, Sharan A, Sperling MR, Tracy JI. Extratemporal functional connectivity impairments at rest are related to memory performance in mesial temporal epilepsy. Hum Brain Mapp. 2013;34(9):2202-16. Epub 2012/04/17. doi: 10.1002/hbm.22059. PubMed PMID: 22505284; PMCID: PMC3864618.

44. Cataldi M, Avoli M, de Villers-Sidani E. Resting state networks in temporal lobe epilepsy. Epilepsia. 2013;54(12):2048-59. Epub 2013/10/15. doi: 10.1111/epi.12400. PubMed PMID: 24117098; PMCID: PMC4880458.

45. Helmstaedter C, Witt JA. Clinical neuropsychology in epilepsy: theoretical and practical issues. Handb Clin Neurol. 2012;107:437-59. Epub 2012/09/04. doi: 10.1016/B978-0-444-52898-8.00036-7. PubMed PMID: 22938988.

46. Hermann BP, Loring DW, Wilson S. Paradigm Shifts in the Neuropsychology of Epilepsy. J Int Neuropsychol Soc. 2017;23(9-10):791-805. Epub 2017/12/05. doi: 10.1017/S1355617717000650. PubMed PMID: 29198272; PMCID: PMC5846680.

47. Baxendale S, Thompson P. The new approach to epilepsy classification: Cognition and behavior in adult epilepsy syndromes. Epilepsy Behav. 2016;64(Pt A):253-6. Epub 2016/10/25. doi: 10.1016/j.yebeh.2016.09.003. PubMed PMID: 27776297.

48. Tavakol S, Royer J, Lowe AJ, Bonilha L, Tracy JI, Jackson GD, Duncan JS, Bernasconi A, Bernasconi N, Bernhardt BC. Neuroimaging and connectomics of drug-resistant epilepsy at multiple scales: From focal lesions to macroscale networks. Epilepsia. 2019. Epub 2019/03/20. doi: 10.1111/epi.14688. PubMed PMID: 30889276.

49. Engel J, Jr., Pitkanen A, Loeb JA, Dudek FE, Bertram EH, 3rd, Cole AJ, Moshe SL, Wiebe S, Jensen FE, Mody I, Nehlig A, Vezzani A. Epilepsy biomarkers. Epilepsia. 2013;54 Suppl 4:61-9. Epub 2013/08/09. doi: 10.1111/epi.12299. PubMed PMID: 23909854; PMCID: PMC4131763.

50. Pitkanen A, Ekolle Ndode-Ekane X, Lapinlampi N, Puhakka N. Epilepsy biomarkers - Toward etiology and pathology specificity. Neurobiol Dis. 2019;123:42-58. Epub 2018/05/22. doi: 10.1016/j.nbd.2018.05.007. PubMed PMID: 29782966; PMCID: PMC6240498.

51. Wechsler D. Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II). San Antonio, TX: NCS Pearson; 2011.

52. Rey A. L'Examen clinique en psychologie , par André Rey,... 2e édition. Paris: Presses universitaires de France (Vendôme Impr. des P.U.F.); 1964. In-16 (8 cm), 224 , fig., couv. ill. 7 F. [D. L. 17930-64] p.

53. Kaplan EF, Goodglass H, Weintraub S. The Boston Naming Test (2nd ed.). Philadelphia, PA: Lea & Febiger; 1983.

54. Spreen O, Benton AL. Neurosensory center comprehensive examination for aphasia: Manual of directions. revised edition. Victoria, BC, Canada: Neuropsychology Laboratory, University of Victoria; 1977.

55. Heaton RK, Miller SW, Taylor MJ, Grant I. Revised Comprehensive Norms for an Expanded Halstead Reitan Battery: Demographically Adjusted Neuropsychological Norms for African American and Caucasian Adults. Lutz: Psychological Assessment Resources, Inc. 2004.

56. Strauss E, Sherman E, Spreen O. A Compendium of Neuropsychological Tests (3rd Edition). New York: Oxford University Press; 2006.

57. Benton AL, Hamsher KD, Varney NR, Spreen O. Contributions to Neuropsychological Assessment: A Clinical Manual. New York, NY: Oxford University Press; 1983.

58. Klove H. Clinical Neuropsychology. Med Clin North Am. 1963;47:1647-58. Epub 1963/11/01. PubMed PMID: 14078168.

59. Carlozzi NE, Tulsky DS, Chiaravalloti ND, Beaumont JL, Weintraub S, Conway K, Gershon RC. NIH Toolbox Cognitive Battery (NIHTB-CB): the NIHTB Pattern Comparison Processing Speed Test. J Int Neuropsychol Soc. 2014;20(6):630-41. Epub 2014/06/25. doi: 10.1017/S1355617714000319. PubMed PMID: 24960594; PMCID: PMC4424947.

60. Carlozzi NE, Beaumont JL, Tulsky DS, Gershon RC. The NIH Toolbox Pattern Comparison Processing Speed Test: Normative Data. Arch Clin Neuropsychol. 2015;30(5):359-68. Epub 2015/05/31. doi: 10.1093/arclin/acv031. PubMed PMID: 26025230; PMCID: PMC4542749.

61. Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Ugurbil K. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn Reson Med. 2010;63(5):1144-53. Epub 2010/05/01. doi: 10.1002/mrm.22361. PubMed PMID: 20432285; PMCID: PMC2906244.

62. Patriat R, Molloy EK, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V, Birn RM. The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. Neuroimage. 2013;78:463-73. Epub 2013/04/20. doi: 10.1016/j.neuroimage.2013.04.013. PubMed PMID: 23597935; PMCID: PMC4003890.

63. Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage. 2013;83:550-8. Epub 2013/06/12. doi: 10.1016/j.neuroimage.2013.05.099. PubMed PMID: 23747458; PMCID: PMC4104183.

64. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M, Consortium WU-MH. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage. 2013;80:105-24. Epub 2013/05/15. doi: 10.1016/j.neuroimage.2013.04.127. PubMed PMID: 23668970; PMCID: PMC3720813.

65. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage. 1999;9(2):179-94. Epub 1999/02/05. doi: 10.1006/nimg.1998.0395. PubMed PMID: 9931268.

66. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. Neuroimage. 2012;62(2):782-90. Epub 2011/10/08. doi: 10.1016/j.neuroimage.2011.09.015. PubMed PMID: 21979382.

67. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC. A multi-modal parcellation of human cerebral cortex. Nature. 2016;536(7615):171-8. Epub 2016/07/21. doi: 10.1038/nature18933. PubMed PMID: 27437579; PMCID: PMC4990127.

68. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002;33(3):341-55. Epub 2002/02/08. PubMed PMID: 11832223.

69. Buzsaki G, Draguhn A. Neuronal oscillations in cortical networks. Science. 2004;304(5679):1926-9. Epub 2004/06/26. doi: 10.1126/science.1099745. PubMed PMID: 15218136.

70. Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med. 1995;34(4):537-41. Epub 1995/10/01. PubMed PMID: 8524021.

71. Zuo XN, Di Martino A, Kelly C, Shehzad ZE, Gee DG, Klein DF, Castellanos FX, Biswal BB, Milham MP. The oscillating brain: complex and reliable. Neuroimage. 2010;49(2):1432-45. Epub 2009/09/29. doi: 10.1016/j.neuroimage.2009.09.037. PubMed PMID: 19782143; PMCID: PMC2856476.

72. Gohel SR, Biswal BB. Functional integration between brain regions at rest occurs in multiple-frequency bands. Brain Connect. 2015;5(1):23-34. Epub 2014/04/08. doi: 10.1089/brain.2013.0210. PubMed PMID: 24702246; PMCID: PMC4313418.

73. Zang YF, He Y, Zhu CZ, Cao QJ, Sui MQ, Liang M, Tian LX, Jiang TZ, Wang YF. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev. 2007;29(2):83-91. Epub 2006/08/22. doi: 10.1016/j.braindev.2006.07.002. PubMed PMID: 16919409.

74. Zou QH, Zhu CZ, Yang Y, Zuo XN, Long XY, Cao QJ, Wang YF, Zang YF. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. J Neurosci Methods. 2008;172(1):137-41. Epub 2008/05/27. doi: 10.1016/j.jneumeth.2008.04.012. PubMed PMID: 18501969; PMCID: PMC3902859.

75. Centeno M, Carmichael DW. Network Connectivity in Epilepsy: Resting State fMRI and EEG-fMRI Contributions. Front Neurol. 2014;5:93. Epub 2014/07/30. doi: 10.3389/fneur.2014.00093. PubMed PMID: 25071695; PMCID: PMC4081640.

76. Kucukboyaci NE, Kemmotsu N, Cheng CE, Girard HM, Tecoma ES, Iragui VJ, McDonald CR. Functional connectivity of the hippocampus in temporal lobe epilepsy: feasibility of a task-regressed seed-based approach. Brain Connect. 2013;3(5):464-74. Epub 2013/07/23. doi: 10.1089/brain.2013.0150. PubMed PMID: 23869604; PMCID: PMC3796326.

77. Zhang Z, Lu G, Zhong Y, Tan Q, Chen H, Liao W, Tian L, Li Z, Shi J, Liu Y. fMRI study of mesial temporal lobe epilepsy using amplitude of low-frequency fluctuation analysis. Hum Brain Mapp. 2010;31(12):1851-61. Epub 2010/03/13. doi: 10.1002/hbm.20982. PubMed PMID: 20225278.

78. Yang Z, Choupan J, Reutens D, Hocking J. Lateralization of Temporal Lobe Epilepsy Based on Resting-State Functional Magnetic Resonance Imaging and Machine Learning. Front Neurol. 2015;6:184. Epub 2015/09/18. doi: 10.3389/fneur.2015.00184. PubMed PMID: 26379618; PMCID: PMC4553409.

79. Hua JP, Xiong ZX, Lowey J, Suh E, Dougherty ER. Optimal number of features as a function of sample size for various classification rules. Bioinformatics. 2005;21(8):1509-15. doi: 10.1093/bioinformatics/bti171. PubMed PMID: WOS:000228401800034.

80. Vergun S, Gaggl W, Nair VA, Suhonen JI, Birn RM, Ahmed AS, Meyerand ME, Reuss J, DeYoe EA, Prabhakaran V. Classification and Extraction of Resting State Networks Using Healthy and Epilepsy fMRI Data. Front Neurosci. 2016;10:440. Epub 2016/10/13. doi: 10.3389/fnins.2016.00440. PubMed PMID: 27729846; PMCID: PMC5037187.

81. Tibshirani R. Regression shrinkage and selection via the Lasso. J Roy Stat Soc B Met. 1996;58(1):267-88. PubMed PMID: WOS:A1996TU31400017.

82. Tang J, Alelyani S, Liu H. Feature Selection for Classification: A Review. Data Classification: Algorithms and Applications. 2013;CRC Press.

83. Meier TB, Desphande AS, Vergun S, Nair VA, Song J, Biswal BB, Meyerand ME, Birn RM, Prabhakaran V. Support vector machine classification and characterization of age-related reorganization of functional brain networks. Neuroimage. 2012;60(1):601-13. Epub 2012/01/10. doi: 10.1016/j.neuroimage.2011.12.052. PubMed PMID: 22227886; PMCID: PMC3288439.

84. Hwang G, Cook CJ, Nair VA, Alexander A, Antuono PG, Asthana S, Birn R, Carlsson CM, Chen G, Edwards DF, Franczak M, Goveas JS, Johnson SC, Kecskemeti S, Kulkarni AP, Mohanty R, Nencka A, Okonkwo OC, Pasquesi M, Rivera-Bonet C, Taylor IK, Tellapragada N, Williams LM, Li S, Bendlin BB, Prabhakaran V. Characterizing Structural Brain Alterations in Alzheimer's Disease Patients with Machine Learning. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2018;14(7):P135-P6.

85. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 1996;29(3):162-73. Epub 1996/06/01. PubMed PMID: 8812068.

86. Power JD, Schlaggar BL, Petersen SE. Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage. 2015;105:536-51. Epub 2014/12/03. doi: 10.1016/j.neuroimage.2014.10.044. PubMed PMID: 25462692; PMCID: PMC4262543.

87. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage. 2014;84:320-41. Epub 2013/09/03. doi: 10.1016/j.neuroimage.2013.08.048. PubMed PMID: 23994314; PMCID: PMC3849338.

88. MathWorks. Statistics and Machine Learning Toolbox Release Notes. Natick, MA2017. Available from: https://www.mathworks.com/help/pdf_doc/stats/rn.pdf.

89. Cortes C, Vapnik V. Support-Vector Networks. Mach Learn. 1995;20(3):273-97. doi: Doi 10.1007/Bf00994018. PubMed PMID: WOS:A1995RX35400003.

90. Izenman AJ. Linear Discriminant Analysis. Springer Texts Stat. 2008:237-80. doi: 10.1007/978-0-387-78189-1_8. PubMed PMID: WOS:000266979300008.

91. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Mach Learn. 1997;29(2-3):131-63. doi: Doi 10.1023/A:1007465528199. PubMed PMID: WOS:000071159300003.

92. Evgeniou T, Pontil M. Leave one out error, stability, and generalization of voting combinations of classifiers. Mach Learn. 2004;55(1):71-97. doi: Doi 10.1023/B:Mach.0000019805.88351.60. PubMed PMID: WOS:000220203500004.

93. Vergun S, Nair V, Jensen M, Chacon M, Sattin J, Prabhakaran V. Support Vector Machine Classification of Stroke Using Resting State Functional Connectivity. Neurology. 2013;80. PubMed PMID: WOS:000332068602161.

94. Malhi A, Gao RX. PCA-based feature selection scheme for machine defect classification. Ieee T Instrum Meas. 2004;53(6):1517-25. doi: 10.1109/Tim.2004.834070. PubMed PMID: WOS:000225225500011.

95. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn. 2002;46(1-3):389-422. doi: Doi 10.1023/A:1012487302797. PubMed PMID: WOS:000171501800018.

96. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57(1):289-300. PubMed PMID: WOS:A1995QE45300017.

97. Ripley BD. Pattern recognition and neural networks. Cambirdge Univ. Press. pp. Glossary2009.

98. Liao W, Zhang Z, Pan Z, Mantini D, Ding J, Duan X, Luo C, Lu G, Chen H. Altered functional connectivity and small-world in mesial temporal lobe epilepsy. Plos One. 2010;5(1):e8525. Epub 2010/01/15. doi: 10.1371/journal.pone.0008525. PubMed PMID: 20072616; PMCID: PMC2799523.

99. Morgan VL, Abou-Khalil B, Rogers BP. Evolution of functional connectivity of brain networks and their dynamic interaction in temporal lobe epilepsy. Brain Connect. 2015;5(1):35-44. Epub 2014/06/06. doi: 10.1089/brain.2014.0251. PubMed PMID: 24901036; PMCID: PMC4313394.

100. Armananzas R, Alonso-Nanclares L, DeFelipe-Oroquieta J, Kastanauskaite A, de Sola RG, DeFelipe J, Bielza C, Larranaga P. Machine Learning Approach for the Outcome Prediction of Temporal Lobe Epilepsy Surgery. Plos One. 2013;8(4). doi: 10.1371/journal.pone.0062819. PubMed PMID: WOS:000319077300101.

101. Feis DL, Schoene-Bake JC, Elger C, Wagner J, Tittgemeyer M, Weber B. Prediction of post-surgical seizure outcome in left mesial temporal lobe epilepsy. Neuroimage-Clin. 2013;2:903-11. doi: 10.1016/j.nicl.2013.06.010. PubMed PMID: WOS:000209276800099.

102. Munsell BC, Wee CY, Keller SS, Weber B, Elger C, da Silva LAT, Nesland T, Styner M, Shen DG, Bonilha L. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. Neuroimage. 2015;118:219-30. doi: 10.1016/j.neuroimage.2015.06.008. PubMed PMID: WOS:000360630200022.

103. Memarian N, Kim S, Dewar S, Engel J, Staba RJ. Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. Comput Biol Med. 2015;64:67-78. doi: 10.1016/j.compbiomed.2015.06.008. PubMed PMID: WOS:000361412500007.

104. Kerr WT, Nguyen ST, Cho AY, Lau EP, Silverman DH, Douglas PK, Reddy NM, Anderson A, Bramen J, Salamon N, Stern JM, Cohen MS. Computer-Aided Diagnosis and Localization of Lateralized Temporal Lobe Epilepsy Using Interictal FDG-PET. Front Neurol. 2013;4:31. Epub 2013/04/09. doi: 10.3389/fneur.2013.00031. PubMed PMID: 23565107; PMCID: PMC3615243.

105. Bernhardt BC, Hong SJ, Bernasconi A, Bernasconi N. Magnetic resonance imaging pattern learning in temporal lobe epilepsy: classification and prognostics. Ann Neurol. 2015;77(3):436-46. Epub 2014/12/30. doi: 10.1002/ana.24341. PubMed PMID: 25546153.

106. Del Gaizo J, Mofrad N, Jensen JH, Clark D, Glenn R, Helpern J, Bonilha L. Using machine learning to classify temporal lobe epilepsy based on diffusion MRI. Brain Behav. 2017;7(10). doi: 10.1002/brb3.801. PubMed PMID: WOS:000413532400008.

107. Focke NK, Yogarajah M, Symms MR, Gruber O, Paulus W, Duncan JS. Automated MR image classification in temporal lobe epilepsy. Neuroimage. 2012;59(1):356-62. Epub 2011/08/13. doi: 10.1016/j.neuroimage.2011.07.068. PubMed PMID: 21835245.

108. Zhang J, Cheng W, Wang ZG, Zhang ZQ, Lu WL, Lu GM, Feng JF. Pattern Classification of Large-Scale Functional Brain Networks: Identification of Informative Neuroimaging Markers for Epilepsy. Plos One. 2012;7(5). doi: 10.1371/journal.pone.0036733. PubMed PMID: WOS:000305341200016.

109. Rajpoot K, Riaz A, Majeed W, Rajpoot N. Functional Connectivity Alterations in Epilepsy from Resting-State Functional MRI. Plos One. 2015;10(8). doi: 10.1371/journal.pone.0134944. PubMed PMID: WOS:000359121100087.

110. Riley JD, Fling BW, Cramer SC, Lin JJ. Altered organization of face-processing networks in temporal lobe epilepsy. Epilepsia. 2015;56(5):762-71. Epub 2015/04/01. doi: 10.1111/epi.12976. PubMed PMID: 25823855; PMCID: PMC4437862.

111. Addis DR, Moscovitch M, McAndrews MP. Consequences of hippocampal damage across the autobiographical memory network in left temporal lobe epilepsy. Brain. 2007;130(Pt 9):2327-42. Epub 2007/08/08. doi: 10.1093/brain/awm166. PubMed PMID: 17681983.

112. Mueller SG, Laxer KD, Barakos J, Cheong I, Garcia P, Weiner MW. Widespread neocortical abnormalities in temporal lobe epilepsy with and without mesial sclerosis. Neuroimage. 2009;46(2):353-9. Epub 2009/03/03. doi: 10.1016/j.neuroimage.2009.02.020. PubMed PMID: 19249372; PMCID: PMC2799165.

113. Saling MM. Verbal memory in mesial temporal lobe epilepsy: beyond material specificity. Brain. 2009;132(Pt 3):570-82. Epub 2009/03/03. doi: 10.1093/brain/awp012. PubMed PMID: 19251757.

114. Hermann B, Loring DW, Wilson S. Paradigm Shifts in the Neuropsychology of Epilepsy. J Int Neuropsychol Soc. 2017;23(9-10):791-805. Epub 2017/12/05. doi: 10.1017/S1355617717000650. PubMed PMID: 29198272; PMCID: PMC5846680.

115. Diehl B, Busch RM, Duncan JS, Piao Z, Tkach J, Luders HO. Abnormalities in diffusion tensor imaging of the uncinate fasciculus relate to reduced memory in temporal lobe epilepsy. Epilepsia. 2008;49(8):1409-18. Epub 2008/04/10. doi: 10.1111/j.1528-1167.2008.01596.x. PubMed PMID: 18397294.

116. Vlooswijk MC, Vaessen MJ, Jansen JF, de Krom MC, Majoie HJ, Hofman PA, Aldenkamp AP, Backes WH. Loss of network efficiency associated with cognitive decline in chronic epilepsy. Neurology. 2011;77(10):938-44. Epub 2011/08/13. doi: 10.1212/WNL.0b013e31822cfc2f. PubMed PMID: 21832213.

117. Lin JJ, Mula M, Hermann BP. Uncovering the neurobehavioural comorbidities of epilepsy over the lifespan. Lancet. 2012;380(9848):1180-92. Epub 2012/10/02. doi: 10.1016/S0140-6736(12)61455-X. PubMed PMID: 23021287; PMCID: PMC3838617.

118. McDonald CR, Leyden KM, Hagler DJ, Kucukboyaci NE, Kemmotsu N, Tecoma ES, Iragui VJ. White matter microstructure complements morphometry for predicting verbal memory in epilepsy. Cortex. 2014;58:139-50. Epub 2014/07/13. doi: 10.1016/j.cortex.2014.05.014. PubMed PMID: 25016097; PMCID: PMC4188700.

119. He X, Bassett DS, Chaitanya G, Sperling MR, Kozlowski L, Tracy JI. Disrupted dynamic network reconfiguration of the language system in temporal lobe epilepsy. Brain. 2018;141(5):1375-89. Epub 2018/03/20. doi: 10.1093/brain/awy042. PubMed PMID: 29554279.

120. Reyes A, Uttarwar VS, Chang YA, Balachandra AR, Pung CJ, Hagler DJ, Jr., Paul BM, McDonald CR. Decreased neurite density within frontostriatal networks is associated with executive dysfunction in temporal lobe epilepsy. Epilepsy Behav. 2018;78:187-93. Epub 2017/11/12. doi: 10.1016/j.yebeh.2017.09.012. PubMed PMID: 29126704; PMCID: PMC5756677.

121. Kwan P, Brodie MJ. Neuropsychological effects of epilepsy and antiepileptic drugs. Lancet. 2001;357(9251):216-22. Epub 2001/02/24. doi: 10.1016/S0140-6736(00)03600-X. PubMed PMID: 11213111.

122. Loring DW, Marino S, Meador KJ. Neuropsychological and behavioral effects of antiepilepsy drugs. Neuropsychol Rev. 2007;17(4):413-25. Epub 2007/10/19. doi: 10.1007/s11065-007-9043-9. PubMed PMID: 17943448.

123. Witt JA, Elger CE, Helmstaedter C. Adverse cognitive effects of antiepileptic pharmacotherapy: Each additional drug matters. Eur Neuropsychopharmacol. 2015;25(11):1954-9. Epub 2015/08/25. doi: 10.1016/j.euroneuro.2015.07.027. PubMed PMID: 26296280.

124. Oostrom KJ, Smeets-Schouten A, Kruitwagen CL, Peters AC, Jennekens-Schinkel A, Dutch Study Group of Epilepsy in C. Not only a matter of epilepsy: early problems of cognition and behavior in children with "epilepsy only"--a prospective, longitudinal, controlled study starting at diagnosis. Pediatrics. 2003;112(6 Pt 1):1338-44. Epub 2003/12/05. PubMed PMID: 14654607.

125. Baker GA, Taylor J, Aldenkamp AP, group S. Newly diagnosed epilepsy: cognitive outcome after 12 months. Epilepsia. 2011;52(6):1084-91. Epub 2011/04/02. doi: 10.1111/j.1528-1167.2011.03043.x. PubMed PMID: 21453356.

126. Berg AT, Langfitt JT, Testa FM, Levy SR, DiMario F, Westerveld M, Kulas J. Residual cognitive effects of uncomplicated idiopathic and cryptogenic epilepsy. Epilepsy Behav. 2008;13(4):614-9. Epub 2008/08/05. doi: 10.1016/j.yebeh.2008.07.007. PubMed PMID: 18675938.

127. Aldenkamp AP, Alpherts WC, Blennow G, Elmqvist D, Heijbel J, Nilsson HL, Sandstedt P, Tonnby B, Wahlander L, Wosse E. Withdrawal of antiepileptic medication in children--effects on cognitive function: The Multicenter Holmfrid Study. Neurology. 1993;43(1):41-50. Epub 1993/01/01. PubMed PMID: 8423909.

128. Kalmar JH, Chiaravalloti ND. Information processing speed in multiple sclerosis: A primary deficit? In: John DeLuca PD, Jessica PD, Kalmar JH, editors. Information processing speed in clinical populations. New York: Taylor and Francis; 2008.

129. Grevers E, Breuer LE, Ijff DM, Aldenkamp AP. Mental slowing in relation to epilepsy and antiepileptic medication. Acta Neurol Scand. 2016;134(2):116-22. Epub 2016/02/27. doi: 10.1111/ane.12517. PubMed PMID: 26918421.

130. Dickinson D, Ramsey ME, Gold JM. Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. Arch Gen Psychiatry. 2007;64(5):532-42. Epub 2007/05/09. doi: 10.1001/archpsyc.64.5.532. PubMed PMID: 17485605.

131. Knowles EE, David AS, Reichenberg A. Processing speed deficits in schizophrenia: reexamining the evidence. Am J Psychiatry. 2010;167(7):828-35. Epub 2010/05/05. doi: 10.1176/appi.ajp.2010.09070937. PubMed PMID: 20439390.

132. Morrens M, Hulstijn W, Sabbe B. Psychomotor slowing in schizophrenia. Schizophr Bull. 2007;33(4):1038-53. Epub 2006/11/10. doi: 10.1093/schbul/sbl051. PubMed PMID: 17093141; PMCID: PMC2632327.

133. Benedict RH, Morrow SA, Weinstock Guttman B, Cookfair D, Schretlen DJ. Cognitive reserve moderates decline in information processing speed in multiple sclerosis patients. J Int Neuropsychol Soc. 2010;16(5):829-35. Epub 2010/07/09. doi: 10.1017/S1355617710000688. PubMed PMID: 20609273.

134. Salthouse TA, Toth J, Daniels K, Parks C, Pak R, Wolbrette M, Hocking KJ. Effects of aging on efficiency of task switching in a variant of the trail making test. Neuropsychology. 2000;14(1):102-11. Epub 2000/02/16. PubMed PMID: 10674802.

135. Tucker-Drob EM, Salthouse TA. Adult age trends in the relations among cognitive abilities. Psychol Aging. 2008;23(2):453-60. Epub 2008/06/25. doi: 10.1037/0882-7974.23.2.453. PubMed PMID: 18573019; PMCID: PMC2762546.

136. Garcia-Ramos C, Dabbs K, Meyerand ME, Prabhakaran V, Hsu D, Jones J, Seidenberg M, Hermann B. Psychomotor slowing is associated with anomalies in baseline and prospective large scale neural networks in youth with epilepsy. Neuroimage Clinical. 2018;19:222-31.

137. Joy S, Fein D, Kaplan E. Decoding digit symbol: speed, memory, and visual scanning. Assessment. 2003;10(1):56-65. Epub 2003/04/05. doi: 10.1177/0095399702250335. PubMed PMID: 12675384.

138. Ashendorf L, Reynolds E. Process analysis of the Digit Symbol task. In: Ashendorf L, Swenson R, Libon D, editors. The Boston process approach to neuropsychological assessment: A practitioner's guide. New York, NY: Oxford University Press; 2013. p. 77-87.

139. Perrine K, Hermann BP, Meador KJ, Vickrey BG, Cramer JA, Hays RD, Devinsky O. The relationship of neuropsychological functioning to quality of life in epilepsy. Arch Neurol. 1995;52(10):997-1003. Epub 1995/10/01. PubMed PMID: 7575228.

140. Fisher RS, Vickrey BG, Gibson P, Hermann B, Penovich P, Scherer A, Walker S. The impact of epilepsy from the patient's perspective I. Descriptions and subjective perceptions. Epilepsy Res. 2000;41(1):39-51. Epub 2000/08/05. PubMed PMID: 10924867.

141. Helmstaedter C, Witt JA. Multifactorial etiology of interictal behavior in frontal and temporal lobe epilepsy. Epilepsia. 2012;53:1765-73. doi: 10.1111/j.1528-1167.2012.03602.x. PubMed PMID: 22881602.

142. Elger CE, Helmstaedter C, Kurthen M. Chronic epilepsy and cognition. Lancet Neurol. 2004;3(11):663-72. Epub 2004/10/19. doi: 10.1016/S1474-4422(04)00906-8. PubMed PMID: 15488459.

143. MacAllister WS, Schaffer SG. Neuropsychological deficits in childhood epilepsy syndromes. Neuropsychol Rev. 2007;17(4):427-44. Epub 2007/10/27. doi: 10.1007/s11065-007-9048-4. PubMed PMID: 17963043.

144. Loring DW. History of neuropsychology through epilepsy eyes. Arch Clin Neuropsychol. 2010;25(4):259-73. Epub 2010/04/17. doi: 10.1093/arclin/acq024. PubMed PMID: 20395259; PMCID: PMC2872650.

145. Novelly RA. The debt of neuropsychology to the epilepsies. Am Psychol. 1992;47(9):1126-9. Epub 1992/09/01. doi: 10.1037//0003-066x.47.9.1126. PubMed PMID: 1416384.

146. Oyegbile TO, Dow C, Jones J, Bell B, Rutecki P, Sheth R, Seidenberg M, Hermann BP. The nature and course of neuropsychological morbidity in chronic temporal lobe epilepsy. Neurology. 2004;62(10):1736-42. Epub 2004/05/26. doi: 10.1212/01.wnl.0000125186.04867.34. PubMed PMID: 15159470.

147. Guimaraes CA, Li LM, Rzezak P, Fuentes D, Franzon RC, Augusta Montenegro M, Cendes F, Thome-Souza S, Valente K, Guerreiro MM. Temporal lobe epilepsy in childhood: comprehensive neuropsychological assessment. J Child Neurol. 2007;22(7):836-40. Epub 2007/08/24. doi: 10.1177/0883073807304701. PubMed PMID: 17715275.

148. Marques CM, Caboclo LO, da Silva TI, Noffs MH, Carrete H, Jr., Lin K, Lin J, Sakamoto AC, Yacubian EM. Cognitive decline in temporal lobe epilepsy due to unilateral hippocampal sclerosis. Epilepsy Behav. 2007;10(3):477-85. Epub 2007/03/21. doi: 10.1016/j.yebeh.2007.02.002. PubMed PMID: 17368105.

149. Rzezak P, Fuentes D, Guimaraes CA, Thome-Souza S, Kuczynski E, Li LM, Franzon RC, Leite CC, Guerreiro M, Valente KD. Frontal lobe dysfunction in children with temporal lobe epilepsy. Pediatr Neurol. 2007;37(3):176-85. Epub 2007/09/04. doi: 10.1016/j.pediatrneurol.2007.05.009. PubMed PMID: 17765805.

150. Braakman HM, Vaessen MJ, Jansen JF, Debeij-van Hall MH, de Louw A, Hofman PA, Vles JS, Aldenkamp AP, Backes WH. Aetiology of cognitive impairment in children with frontal lobe epilepsy. Acta Neurol Scand. 2015;131(1):17-29. Epub 2014/09/12. doi: 10.1111/ane.12283. PubMed PMID: 25208759.

151. Hwang G, Dabbs K, Conant L, Nair VA, Mathis J, Almane DN, Nencka A, Birn R, Humphries C, Raghavan M, DeYoe EA, Struck AF, Maganti R, Binder JR, Meyerand E, Prabhakaran V, Hermann B. Cognitive slowing and its underlying neurobiology in temporal lobe epilepsy. Cortex. 2019;117:41-52. Epub 2019/03/31. doi: 10.1016/j.cortex.2019.02.022. PubMed PMID: 30927560.

152. Wang WH, Liou HH, Chen CC, Chiu MJ, Chen TF, Cheng TW, Hua MS. Neuropsychological performance and seizure-related risk factors in patients with temporal lobe epilepsy: a retrospective cross-sectional study. Epilepsy Behav. 2011;22(4):728-34. Epub 2011/10/25. doi: 10.1016/j.yebeh.2011.08.038. PubMed PMID: 22019015.

153. Baxendale S, Thompson P. Beyond localization: the role of traditional neuropsychological tests in an age of imaging. Epilepsia. 2010;51(11):2225-30. Epub 2010/12/24. doi: 10.1111/j.1528-1167.2010.02710.x. PubMed PMID: 21175602.

154. Jackson DC, Dabbs K, Walker NM, Jones JE, Hsu DA, Stafstrom CE, Seidenberg M, Hermann BP. The neuropsychological and academic substrate of new/recent-onset epilepsies. J Pediatr. 2013;162(5):1047-53 e1. Epub 2012/12/12. doi: 10.1016/j.jpeds.2012.10.046. PubMed PMID: 23219245; PMCID: PMC3615134.

155. Smith ML. Rethinking cognition and behavior in the new classification for childhood epilepsy: Examples from frontal lobe and temporal lobe epilepsies. Epilepsy Behav. 2016;64(Pt B):313-7. Epub 2016/06/28. doi: 10.1016/j.yebeh.2016.04.050. PubMed PMID: 27346387.

156. Bremm FJ, Hendriks MPH, Bien CG, Grewe P. Pre- and postoperative verbal memory and executive functioning in frontal versus temporal lobe epilepsy. Epilepsy Behav. 2019;101(Pt A):106538. Epub 2019/11/05. doi: 10.1016/j.yebeh.2019.106538. PubMed PMID: 31678807.

157. Verche E, San Luis C, Hernandez S. Neuropsychology of frontal lobe epilepsy in children and adults: Systematic review and meta-analysis. Epilepsy Behav. 2018;88:15-20. Epub 2018/09/14. doi: 10.1016/j.yebeh.2018.08.008. PubMed PMID: 30212723.

158. Stretton J, Thompson PJ. Frontal lobe function in temporal lobe epilepsy. Epilepsy Res. 2012;98(1):1-13. Epub 2011/11/22. doi: 10.1016/j.eplepsyres.2011.10.009. PubMed PMID: 22100147; PMCID: PMC3398387.

159. Wandschneider B, Thompson PJ, Vollmar C, Koepp MJ. Frontal lobe function and structure in juvenile myoclonic epilepsy: a comprehensive review of neuropsychological and imaging data. Epilepsia. 2012;53(12):2091-8. Epub 2012/10/31. doi: 10.1111/epi.12003. PubMed PMID: 23106095.

160. Verrotti A, Matricardi S, Rinaldi VE, Prezioso G, Coppola G. Neuropsychological impairment in childhood absence epilepsy: Review of the literature. J Neurol Sci. 2015;359(1-2):59-66. Epub 2015/12/17. doi: 10.1016/j.jns.2015.10.035. PubMed PMID: 26671087.

161. Neri ML, Guimaraes CA, Oliveira EP, Duran MH, Medeiros LL, Montenegro MA, Boscariol M, Guerreiro MM. Neuropsychological assessment of children with rolandic epilepsy: executive functions. Epilepsy Behav. 2012;24(4):403-7. Epub 2012/06/12. doi: 10.1016/j.yebeh.2012.04.131. PubMed PMID: 22683244.

162. Conant LL, Wilfong A, Inglese C, Schwarte A. Dysfunction of executive and related processes in childhood absence epilepsy. Epilepsy Behav. 2010;18(4):414-23. Epub 2010/07/27. doi: 10.1016/j.yebeh.2010.05.010. PubMed PMID: 20656561.

163. Breteler MM, van Duijn CM, Chandra V, Fratiglioni L, Graves AB, Heyman A, Jorm AF, Kokmen E, Kondo K, Mortimer JA, et al. Medical history and the risk of Alzheimer's disease: a collaborative re-analysis of case-control studies. EURODEM Risk Factors Research Group. Int J Epidemiol. 1991;20 Suppl 2:S36-42. Epub 1991/01/01. PubMed PMID: 1833352.

164. Sen A, Capelli V, Husain M. Cognition and dementia in older patients with epilepsy. Brain. 2018;141(6):1592-608. Epub 2018/03/06. doi: 10.1093/brain/awy022. PubMed PMID: 29506031; PMCID: PMC5972564.

165. Breuer LE, Boon P, Bergmans JW, Mess WH, Besseling RM, de Louw A, Tijhuis AG, Zinger S, Bernas A, Klooster DC, Aldenkamp AP. Cognitive deterioration in adult epilepsy: Does accelerated cognitive ageing exist? Neurosci Biobehav Rev. 2016;64:1-11. Epub 2016/02/24. doi: 10.1016/j.neubiorev.2016.02.004. PubMed PMID: 26900650.

166. Helmstaedter C, Elger CE. Chronic temporal lobe epilepsy: a neurodevelopmental or progressively dementing disease? Brain. 2009;132(Pt 10):2822-30. Epub 2009/07/29. doi: 10.1093/brain/awp182. PubMed PMID: 19635728.

167. Baxendale S, Heaney D, Thompson PJ, Duncan JS. Cognitive consequences of childhood-onset temporal lobe epilepsy across the adult lifespan. Neurology. 2010;75(8):705-11. Epub 2010/08/25. doi: 10.1212/WNL.0b013e3181eee3f0. PubMed PMID: 20733146.

168. Dabbs K, Becker T, Jones J, Rutecki P, Seidenberg M, Hermann B. Brain structure and aging in chronic temporal lobe epilepsy. Epilepsia. 2012;53(6):1033-43. Epub 2012/04/05. doi: 10.1111/j.1528-1167.2012.03447.x. PubMed PMID: 22471353; PMCID: PMC3710695.

169. Pardoe HR, Cole JH, Blackmon K, Thesen T, Kuzniecky R, Human Epilepsy Project I. Structural brain changes in medically refractory focal epilepsy resemble premature brain aging. Epilepsy Res. 2017;133:28-32. Epub 2017/04/15. doi: 10.1016/j.eplepsyres.2017.03.007. PubMed PMID: 28410487.

170. Sone D, Beheshti I, Maikusa N, Ota M, Kimura Y, Sato N, Koepp M, Matsuda H. Neuroimaging-based brain-age prediction in diverse forms of epilepsy: a signature of psychosis and beyond. Mol Psychiatry. 2019. Epub 2019/06/05. doi: 10.1038/s41380-019-0446-9. PubMed PMID: 31160692.

171. Tracy JI, Doucet GE. Resting-state functional connectivity in epilepsy: growing relevance for clinical decision making. Curr Opin Neurol. 2015;28(2):158-65. Epub 2015/03/04. doi: 10.1097/WCO.0000000000000178. PubMed PMID: 25734954.

172. Constable RT, Scheinost D, Finn ES, Shen X, Hampson M, Winstanley FS, Spencer DD, Papademetris X. Potential use and challenges of functional connectivity mapping in intractable epilepsy. Front Neurol. 2013;4:39. Epub 2013/06/05. doi: 10.3389/fneur.2013.00039. PubMed PMID: 23734143; PMCID: PMC3660665.

173. Alvim MK, Coan AC, Campos BM, Yasuda CL, Oliveira MC, Morita ME, Cendes F. Progression of gray matter atrophy in seizure-free patients with temporal lobe epilepsy. Epilepsia. 2016;57(4):621-9. Epub 2016/02/13. doi: 10.1111/epi.13334. PubMed PMID: 26865066.

174. Pardoe HR, Berg AT, Jackson GD. Sodium valproate use is associated with reduced parietal lobe thickness and brain volume. Neurology. 2013;80(20):1895-900. Epub 2013/04/26. doi: 10.1212/WNL.0b013e318292a2e5. PubMed PMID: 23616155; PMCID: PMC3908352.

175. Murphy K, Fox MD. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. Neuroimage. 2017;154:169-73. Epub 2016/11/27. doi: 10.1016/j.neuroimage.2016.11.052. PubMed PMID: 27888059; PMCID: PMC5489207.

176. Marcus DS, Harms MP, Snyder AZ, Jenkinson M, Wilson JA, Glasser MF, Barch DM, Archie KA, Burgess GC, Ramaratnam M, Hodge M, Horton W, Herrick R, Olsen T, McKay M, House M, Hileman M, Reid E, Harwell J, Coalson T, Schindler J, Elam JS, Curtiss SW, Van Essen DC, Consortium WU-MH. Human Connectome Project informatics: quality control, database services, and data visualization. Neuroimage. 2013;80:202-19. Epub 2013/05/28. doi: 10.1016/j.neuroimage.2013.05.077. PubMed PMID: 23707591; PMCID: PMC3845379.

177. Smola AJ, Scholkopf B. A tutorial on support vector regression. Stat Comput. 2004;14(3):199-222. doi: Doi 10.1023/B:Stco.0000035301.49549.88. PubMed PMID: WOS:000222770200003.

178. Amoroso N, La Rocca M, Bellantuono L, Diacono D, Fanizzi A, Lella E, Lombardi A, Maggipinto T, Monaco A, Tangaro S, Bellotti R. Deep Learning and Multiplex Networks for Accurate Modeling of Brain Age. Frontiers in Aging Neuroscience. 2019;11(115). doi: 10.3389/fnagi.2019.00115.

179. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-30. PubMed PMID: WOS:000298103200003.

180. Le TT, Kuplicki RT, McKinney BA, Yeh HW, Thompson WK, Paulus MP, Tulsa I. A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. Front Aging Neurosci. 2018;10:317. Epub 2018/11/09. doi: 10.3389/fnagi.2018.00317. PubMed PMID: 30405393; PMCID: PMC6208001.

181. Liang H, Zhang F, Niu X. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. Hum Brain Mapp. 2019;40(11):3143-52. Epub 2019/03/30. doi: 10.1002/hbm.24588. PubMed PMID: 30924225.

182. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, Carlozzi NE, Slotkin J, Blitz D, Wallner-Allen K, Fox NA, Beaumont JL, Mungas D, Nowinski CJ, Richler J, Deocampo JA, Anderson JE, Manly JJ, Borosh B, Havlik R, Conway K, Edwards E, Freund L, King JW, Moy C, Witt E, Gershon RC. Cognition assessment using the NIH Toolbox. Neurology. 2013;80(11 Suppl 3):S54-64. Epub 2013/04/23. doi: 10.1212/WNL.0b013e3182872ded. PubMed PMID: 23479546; PMCID: PMC3662346.

183. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM, Xie W, Rosen GL, Lengerich BJ, Israeli J, Lanchantin J, Woloszynek S, Carpenter AE, Shrikumar A, Xu J, Cofer EM, Lavender CA, Turaga SC, Alexandari AM, Lu Z, Harris DJ, DeCaprio D, Qi Y, Kundaje A, Peng Y, Wiley LK, Segler MHS, Boca SM, Swamidass SJ, Huang A, Gitter A, Greene CS. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15(141). Epub 2018/04/06. doi: 10.1098/rsif.2017.0387. PubMed PMID: 29618526; PMCID: PMC5938574.

184. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1; Lake Tahoe, Nevada: Curran Associates Inc.; 2012. p. 1097-105.

185. Ghaddar B, Naoum-Sawaya J. High dimensional data classification and feature selection using support vector machines. Eur J Oper Res. 2018;265(3):993-1004. doi: 10.1016/j.ejor.2017.08.040. PubMed PMID: WOS:000417657200016.

186. Bron EE, Smits M, Niessen WJ, Klein S. Feature Selection Based on the SVM Weight Vector for Classification of Dementia. IEEE J Biomed Health Inform. 2015;19(5):1617-26. Epub 2015/05/15. doi: 10.1109/JBHI.2015.2432832. PubMed PMID: 25974958.

187. Swiniarski RW, Skowron A. Rough set methods in feature selection and recognition. Pattern Recogn Lett. 2003;24(6):833-49. doi: Doi 10.1016/S0167-8655(02)00196-4. PubMed PMID: WOS:000180351100002.

188. Wang YC, Zhu LG. Research and Implementation of SVD in Machine Learning. 2017 16th Ieee/Acis International Conference on Computer and Information Science (Icis 2017). 2017:471-5. PubMed PMID: WOS:000414478200079.

189. Cao LJ, Chua KS, Chong WK, Lee HP, Gu QM. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. Neurocomputing. 2003;55(1-2):321-36. doi: 10.1016/S0925-2312(03)00433-8. PubMed PMID: WOS:000186355100017.

190. Purnami SW, Andari S, Pertiwi YD. High-Dimensional Data Classification Based on Smooth Support Vector Machines. Procedia Comput Sci. 2015;72:477-84. doi: 10.1016/j.procs.2015.12.129. PubMed PMID: WOS:000373775700057.

191. Guo Y, Graber A, McBurney RN, Balasubramanian R. Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. BMC Bioinformatics. 2010;11:447. Epub 2010/09/08. doi: 10.1186/1471-2105-11-447. PubMed PMID: 20815881; PMCID: PMC2942858.

192. Mukherjee S, Tamayo P, Rogers S, Rifkin R, Engle A, Campbell C, Golub TR, Mesirov JP. Estimating dataset size requirements for classifying DNA microarray data. Journal of Computational Biology. 2003;10(2):119-42. doi: Doi 10.1089/106652703321825928. PubMed PMID: WOS:000182985200002.

193. Dobbin KK, Simon RM. Sample size planning for developing classifiers using high-dimensional DNA microarray data. Biostatistics. 2007;8(1):101-17. Epub 2006/04/15. doi: 10.1093/biostatistics/kxj036. PubMed PMID: 16613833.

194. Hodge MR, Horton W, Brown T, Herrick R, Olsen T, Hileman ME, McKay M, Archie KA, Cler E, Harms MP, Burgess GC, Glasser MF, Elam JS, Curtiss SW, Barch DM, Oostenveld R, Larson-Prior LJ, Ugurbil K, Van Essen DC, Marcus DS. ConnectomeDB--Sharing human brain connectivity data. Neuroimage. 2016;124(Pt B):1102-7. Epub 2015/05/03. doi: 10.1016/j.neuroimage.2015.04.046. PubMed PMID: 25934470; PMCID: PMC4626437.

195. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. Proc of Int'l Conf of Machine Learning. 2006:233-40.

196. Madsen RE, Hansen LK, Winther O. Singular value decomposition and principal component analysis2004. Available from: http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/4000/pdf/imm4000.pdf.

197. Kosinski M, Wang Y, Lakkaraju H, Leskovec J. Mining big data to extract patterns and predict real-life outcomes. Psychol Methods. 2016;21(4):493-506. Epub 2016/12/06. doi: 10.1037/met0000105. PubMed PMID: 27918179.