

# Predictive models and representations for neuroimaging and genetic data

A Dissertation Presented

by

**Syed Fahad Sultan**

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

**Doctor of Philosophy**

in

**Computer Science**

Stony Brook University

**May 2022**

**Stony Brook University**

The Graduate School

**Syed Fahad Sultan**

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

**Dr. Steven Skiena – Dissertation Advisor**  
**Professor, Department of Computer Science**

**Dr. Andrew Schwartz – Chairperson of Defense**  
**Associate Professor, Department of Computer Science**

**Dr. Stanley Bak**  
**Assistant Professor, Department of Computer Science**

**Dr. Lilianne Mujica-Parodi**  
**Professor, Biomedical Engineering Department**  
**Stony Brook University**

This dissertation is accepted by the Graduate School.

**Eric Wertheimer**  
**Dean of the Graduate School**

Abstract of the Dissertation

**Predictive models and representations for  
neuroimaging and genetic data**

by

**Syed Fahad Sultan**

**Doctor of Philosophy**

in

**Computer Science**

Stony Brook University

**2022**

In this work, I present multiple computational models for extracting relevant features and learning predictive representations for neuroimaging and genetic data in UK-Biobank. These models are designed with the goal of providing improved diagnoses and a better understanding of medical conditions and other phenotypes. This includes a machine learning based predictive framework called Neuropredictome that identifies statistically significant linkages between 4928 phenotypes and neuroimaging features of 19,831 subjects. I also provide a novel quantitative method that uses deep learning based text embeddings to evaluate how well Neuropredictome's results align with 14,371 previously published peer-reviewed research articles. Next, I present a generalized framework based on state space systems that bridges the gap between network theory and control theory and extracts fMRI derived control circuits. This framework has the scalability required to mine mega-scale datasets, hitherto not possible using existing methods. In a purely data-driven manner, without priors, I demonstrate that the framework

identifies thalamus-linked prefrontal-limbic and ventral stream sub-circuits, selectively engaged during sensorimotor processing of affective and non-affective stimuli. I demonstrate that circuit-wide dysregulation, defined by degree of drift from healthy trajectories, tracks symptom severity for neuroticism, depression, and bipolar disorder. I also present methods for constructing low-dimensional vector representations (embeddings) of large-scale genotyping data, capable of reducing genotypes of hundreds of thousands of SNPs to 100-dimensional embeddings that retain substantial predictive power for inferring medical phenotypes. I also demonstrate how these genotype embeddings can be used for sharing sensitive medical data while preserving subject anonymity. Finally, using structural and functional neuroimaging in conjunction with cognitive tests, I show that type 2 diabetes mellitus accelerates brain aging and cognitive decline. Together, I believe such computational techniques can significantly advance modern medicine and treatment while enabling several scientific discoveries that revolutionize human health.

# Contents

<b>List of Figures</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Predictive models and representations . . . . .	3
1.3 Publications . . . . .	8
<b>2 NeuroPredictome: A Data-Driven Predictome Linking Neuroimaging to Phenotype</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.2 Methodology . . . . .	12
2.2.1 Data . . . . .	12
2.2.2 Preprocessing Phenotypes . . . . .	13
2.2.3 Data Representations . . . . .	14
2.2.4 Identifying Statistically Significant Predictions . . . . .	15
2.2.5 Evaluation against Neurosynth . . . . .	16
2.3 Results . . . . .	18
2.3.1 Choosing a representation and a classifier . . . . .	18
2.3.2 Neuropredictome identifies brain-phenotype linkages, both expected and new within neuroimaging. . . . .	21
2.3.3 Brain maps learned by our classifier align with results from 14,371 published papers. . . . .	23
2.4 Conclusion . . . . .	23
<b>3 Quantifying Dysregulation of fMRI-Derived Control Circuits for Computational Psychiatry</b>	<b>27</b>
3.1 Introduction . . . . .	27
3.2 Methods . . . . .	29
3.2.1 Image Acquisition . . . . .	31
3.2.2 Task . . . . .	32
3.2.3 Clinical Variables . . . . .	32

3.3	Results . . . . .	34
3.3.1	Recovering circuit motifs from dynamic outputs . . . . .	34
3.3.2	Data-Driven Circuit Discovery Using Human fMRI . . . . .	35
3.3.3	Prefrontal-Limbic and Ventral Stream Subcircuits . . . . .	37
3.3.4	Trajectory Drift as Dysregulation . . . . .	38
3.4	Discussion . . . . .	40
<b>4</b>	<b>Low-Dimensional Genotype Embeddings for Predictive Models</b>	<b>46</b>
4.1	Introduction . . . . .	46
4.2	Literature Review . . . . .	48
4.3	Data . . . . .	49
4.4	Embedding Methods . . . . .	49
4.4.1	Genome Sequence Processing . . . . .	50
4.4.2	Base selection . . . . .	51
4.4.3	Chromosome partitioning . . . . .	51
4.4.4	Matrices . . . . .	52
4.4.5	Matrix compression . . . . .	53
4.4.6	Evaluation . . . . .	54
4.5	Results . . . . .	56
4.5.1	Design choices in genotype embeddings . . . . .	56
4.5.2	Dimensionality of Embeddings . . . . .	58
4.5.3	Privacy Preserving Embeddings . . . . .	58
4.5.4	Phenotype prediction . . . . .	58
4.5.5	Interpreting the Embeddings: Population structure . . . . .	59
4.6	Conclusion . . . . .	60
<b>5</b>	<b>Type 2 diabetes mellitus accelerates brain aging and cognitive decline: complementary findings from UK Biobank and meta-analyses.</b>	<b>63</b>
5.1	Introduction . . . . .	63
5.2	Methods . . . . .	64
5.2.1	Analysis of UK Biobank Dataset (N=26,113) . . . . .	64
5.2.2	Meta-Analysis of Published Literature (N=24,185) . . . . .	66
5.3	Results . . . . .	69
5.3.1	Cognitive correlates with Age and T2DM . . . . .	69
5.3.2	Neurobiological correlates with Age and T2DM . . . . .	69
5.3.3	Neurocognitive Changes associated with T2DM and Age Overlap, Suggesting Common Pathways. . . . .	70
5.3.4	T2DM Chronicity Exacerbates Neurocognitive Symptoms. . . . .	71

5.3.5	T2DM Patients Treated with Metformin Do Not Demonstrate Improved Neurocognitive Symptoms. . . . .	71
5.4	Discussion . . . . .	71
5.5	Figures . . . . .	73
5.6	Supplementary Tables . . . . .	81
<b>6</b>	<b>Conclusion and Future Work</b>	<b>86</b>
	<b>Bibliography</b>	<b>88</b>

PREVIEW

# List of Figures

2.1	<b>Schematic overview of NeuroPredictome.</b> For each neuroimaging modality, brain features of 19,831 subjects, along with age and sex, are used to train 4,926 classifiers, one for each binarized-phenotype, using 20-fold cross validation. Prediction scores from each classifier is compared to that of a baseline classifier to measure statistical significance. Weights learned by each classifier is used to generate an activation map for each phenotype that is then compared to activation maps reported in Neurosynth papers. UK-Biobank phenotypes and Neurosynth papers are similarly compared using text features and we then observe how well do similarities using text features align with similarities using brain features. . . . .	10
2.2	<b>Neuropredictome identifies new brain-phenotype linkages, along with some expected ones.</b> Physical and lifestyle phenotypes, typically not looked at in light of neuroimaging, showed strong brain effects. Across neuroimaging modalities, resting state functional connectivity (rfMRI) broadly yields the best prediction scores and predicts a larger number of phenotypes compared to structural DTI, T1-weighted or task fMRI. . . . .	18
2.3	<b>Neuropredictome results strongly align with results reported in 14,371 Neurosynth papers.</b> When Neuropredictome maps a Neurosynth paper to a UK-Biobank phenotype based on similarity in brain activity, then the phenotype-paper pair are also similar in terms of textual descriptions. The more confident Neuropredictome is about its predictions, the stronger the alignment. Most significant phenotypes (**) $r = 0.25$ ; moderately significant $r = 0.18$ (*); for non-significant (N.S) phenotypes $r = 0.09$ . . . . .	22

2.4	<b>Glove text embeddings [1] over individual phenotype/variable descriptions and pearson as the brain similarity measure, together, yielded the best alignment score.</b> We conducted an extensive search over the hyperparameter search space to optimize for the best alignment scores. Other fixed hyperparameters were as follows: the brain space was MNI, the text-to-text similarity measure was cosine and the measure for the final alignment score was pearson correlation coefficient. . . . .	23
3.1	<b>Trajectory Drift as a Measure of Feedback Control Error, and Thus Circuit Dysregulation.</b> (a) In a classic engineering control application, such as autopilot, a vehicle corrects for deviations from its desired trajectory through negative feedback. As such, the difference between the autopilot’s actual versus desired trajectories provides a measure of its <i>control error</i> , or dysregulation. (b) As per the autopilot example, we use trajectory drift as control error to calculate dysregulation across fMRI-derived control circuits, and demonstrate its application for three psychiatric use cases: <i>neuroticism</i> , <i>depression</i> , and <i>bipolar disorder</i> (c) Schematic of the pipeline for the discovery of circuit architecture and dynamics from human fMRI and simulated time series, using Time-Varying Autoregressive Model with Exogenous Inputs (TVARX) and other state space models. We use trajectory drift between predicted and actual trajectories to quantify circuit dysregulation across subjects with varying degrees of severity for psychiatric symptoms. . . . .	41

3.2 <b>Inferring Closed-Loop Circuits</b>	
(a) Standard connectivity (“network”) analyses depend upon linear regressions, which are only capable of modeling a very specific topology: parallel inputs. In contrast to parallel inputs $A$ , most neurobiological circuits of relevance to psychiatry also require serial $B$ and feedback $C$ components, structures that could lead to an explosion of error propagation using standard statistical methods.	
(b) Impulse response for two hemodynamic response functions (HRFs) with different relaxation times and transfer functions.	
(c) To illustrate how transfer function structure changes with different circuit topologies, we show three transfer functions, each of which corresponds to a different kind of “motif,” with series, parallel and feedback connections. By using pairs of inputs $u$ and outputs $y$ to obtain their transfer function, we systematically infer circuit topology.	
(d) Block diagram representation of state-space equations	
(e) Dynamic effective connectivity as a time-stamped temporal graph. . . . .	43

PREVIEW

**3.3 Using Synthetic Data, We Compare the Performance (Accuracy and Speed) of System Identification Algorithms in Recovering Control Circuit Architecture and Dynamics** (a) In this work, we fill the gap in literature between under-specified models that model networks, rather than circuits, and over-specified models that cannot scale computationally and therefore fail to extract circuits for large number of regions using shorter ( 5 minute) time series. (b) Accuracy scores for classification of simulated circuits and their components, including correct classification of no connection:  $\phi$ ) using different state space models discussed in Table 3.1. TVARX performed best at recovering the original circuit topology (c) AIC scores across models with respect to order of autoregressive component, which accounts for increasing complexity. TVARX performs best even when penalized for having the larger number of parameters. (d) Comparison of TVARX with (stochastic) Dynamic Causal Modeling (DCM) on human task-based fMRI. DCM fails to converge for shorter ( 5 minute) fMRI time series as well as for larger number of nodes. *sFCN=static Functional Connectivity Networks, X=eXogenous inputs, AR=AutoRegressive, dFCN=dynamic Functional Connectivity Networks, TVX=Time Varying with eXogenous inputs, ARX=AutoRegressive with eXogenous inputs, TVARX=Time Varying AutoRegressive with eXogenous inputs, DCM=Dynamic Causal Modeling. . . . .* 44

**3.4 Psychiatric Symptom Severity Tracks Degree of Circuit-Wide Dysregulation.** (a) Two subcircuits, interacting and linked via the thalamus, were overwhelmingly implicated across subjects for each matching task; the *prefrontal-limbic* subcircuit selectively engaged during processing of affective (faces) stimuli while the *ventral stream* subcircuit selectively engaged during processing of non-affective (shapes) stimuli. (b) Results show selective dominance of each subcircuit with changing stimuli (the y-axis is the relative dominance of one circuit versus the other, defined as absolute sum of relevant entries of the feed-forward matrix  $D$  in our state space equation (Eqn. 3.3). For prefrontal-limbic subcircuit  $M = \{thalamus, hippocampus, OFC, vmPFC\}$  and for ventral stream subcircuit  $M = \{thalamus, ITF, IFG\}$  (c) We use trajectory drift (measured as the mean squared error between actual and predicted trajectories) as a measure of feedback control error, or dysregulation, of each subcircuit. Dysregulation of the prefrontal-limbic subcircuit was measured as the error in the thalamus trajectory as predicted from negative feedback by the hippocampus (hippocampus  $\bar{\rightarrow}$  thalamus). Dysregulation of the ventral stream was measured as error in the thalamus trajectory as predicted from negative feedback by the inferior frontal gyrus (inferior frontal gyrus  $\bar{\rightarrow}$  thalamus). Trajectory drift of the ventral stream subcircuit tracks severity of neuroticism, while trajectory drift of the prefrontal-limbic circuit tracks severity of depression. For bipolar disorder, the thalamic trajectory could not be predicted from either hippocampal or IFG trajectories, and thus reflects dysregulation of the full circuit. This could be due to either more systemic problems with feedback across both circuits, or that the full circuit is receiving dysregulated inputs from another, different, circuit not identified by these tasks. Bonferroni corrected  $*P < 0.05$ ;  $**P < 0.01$  . . . . . 45

4.1 Schematic overview of the evaluation framework for classification 54  
 4.2 Effect of embedding dimensionality on prediction scores. Increasing the length of the embeddings improves the prediction of ethnicity across different categories, but the improvements taper off around  $d = 100$ . . . . . 59

4.3	Anonymizing genotype embeddings by adding Gaussian noise to each dimension. The embeddings prove more robust to added noise as dimensionality increases. The nearest neighbor to the noisy representation becomes increasingly unlikely to be the original subject (as the top subfigure shows), while even at the noise level of $1.5 \sigma$ , it still preserves much of the predictive power of the embedding (as shown in the bottom). . . . .	60
4.4	Prediction of UK Biobank phenotypes using our genomic embeddings ( $d = 100$ ). Each circle in the scatter below represents a phenotype, and its height represents the improvement over our baseline. . . . .	61
4.5	First two genetic principal components in UK Biobank, computed from 18,321 samples and 101,284 SNPs. . . . .	62
5.1	<b>Cognitive deficits are apparent with respect to both age and T2DM diagnosis.</b> A: Using the UK Biobank dataset, we performed a quantitative analysis of the effects related to age on cognitive performance across five cognitive domains. Associated changes were derived from estimated regression coefficients as percentages and are shown on the y axis. Age was associated with significant deficits in all five domains, with the strongest effects observed in executive function and processing speed. B: Using the same dataset, we also analyzed cognitive performance in T2DM, with negative values on the y-axis represent performance below that of age, sex, and education-matched HC. As per age effects, executive function and processing speed showed the highest magnitude changes. C: Cognitive deficits identified in UK Biobank data were confirmed by our meta-analysis, which included 11 domains from 34 studies. Average effect sizes (Cohen's $d$ ) corresponding to T2DM are shown on the y axis. Values below the cut-off line ( $y=0$ ) indicate cases in which subjects with T2DM performed less well than age, sex, and education-matched HC. Numbers next to labels identify domains common across panels. Marker sizes represent sample sizes scaled as indicated in the bottom left corner of each panel. On panel C, sample size indicates the number of individual studies. Error bars are 95% CI. . . . .	74

5.2 : Widespread gray matter atrophy can be observed with respect to both age and T2DM diagnosis status. Using the UK Biobank dataset, we measured gray matter atrophy across 45 anatomical regions. Associated changes were derived from estimated regression coefficients as percentages and are shown on the x axes. A: We observed significantly decreased gray matter volume in both cortical and subcortical brain regions with respect to age in HC. Age was associated with an average of 0.5% brain-wide decrease in gray matter volume per year, most prominently for the ventral striatum and Heschl's gyrus B: Gray matter atrophy was also seen in patients diagnosed with T2DM compared to age matched HC, most prominently for the ventral striatum, cerebellum, and putamen. The distribution of T2DM-related effects overlapped with those associated with age, with degeneration of the ventral striatum and preservation of the thalamus and caudate. Error bars are 95% CI. \*P ≤ 0.05; \*\*P ≤ 0.01; \*\*\*P ≤ 0.001, Bonferroni corrected. . . . . 75

PREVIEW

- 5.3 Overlap between age and T2DM with respect to reorganization of brain activity. A: For functional MRI data obtained from the UK Biobank dataset, we used the amplitude of low-frequency fluctuation (ALFF) to quantify brain activity. Effects linked to age are shown in the form of an unthresholded z-map represented by the pink-green color gradient, with pink indicating increased activation and green showing decreased. T2DM related effects were thresholded (minimum cluster size 100mm<sup>3</sup>, FDR  $p < 0.05$ ) to result in significant clusters. The outlines of these significant clusters are overlaid on the age-related z-map to demonstrate overlapping effects. The largest significant clusters with respect to T2DM were in the subgenual area (increased), the caudate (decreased), and frontal eye fields (decreased). All highlighted regions were similarly impacted across age, indicating substantial overlap between the two contrasts. B: Using multimodal neuroimaging data, we performed a meta-analysis for the same contrast using NeuroQuery. We extracted contrast maps for age and T2DM with NeuroQuery and overlaid the outlines of thresholded (minimum cluster size 100mm<sup>3</sup>, FDR  $p < 0.05$ ) z-maps from T2DM on unthresholded z-maps belonging to age. The overlapping effects were evident in several regions, most importantly in the cingulate gyrus, thalamus and premotor cortex. These results support the hypothesis that neurodegeneration in both T2DM and aging may be associated with common mechanistic pathways. . . . . 76
- 5.4 Progression of T2DM disease is significantly associated with gray matter atrophy, accelerating neurodegenerative effects seen in brain aging. For a quantitative evaluation of the impact of T2DM progression on gray matter volume, we considered time since T2DM diagnosis as the main factor of interest from the UK Biobank dataset. The T2DM+ cohort was divided into two groups based on disease duration (separated at 10 years) with a HC cohort also included for visualization purposes. We matched age, sex, education across these three groups and performed linear regression within T2DM+ subjects focusing on disease duration. Evaluation of our sample suggested that time since diagnosis was a significant factor, with each year after diagnosis of T2DM associated with an additional 0.24 years of brain aging beyond that of age-matched T2DM-. Error bars are standard error of the mean. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ . . . . 77

- 5.5 Supplementary Figure 1: Aggregated whole brain measures represent the extent of accelerated brain aging with T2DM diagnosis. We used the UK Biobank dataset to address the extent by which aging is accelerated in individuals with T2DM. Subjects with T2DM were age, sex, and education matched with HC A: We quantified a gross cognitive metric from the combination of multiple z-scored performance scores from five cognitive domains. This metric yielded an effective representation of the general decline across age, the gap between HC versus subjects diagnosed T2DM, and the relative extent of these two phenomena. We observed significantly decreased cognitive performance in subjects with T2DM: an increase of 3.8-years in age-related cognitive decline. B: An equivalent analysis was performed using whole brain gray matter volume. This metric yielded even stronger results compared to cognition. T2DM diagnosis was associated with significant atrophy: an increase of 4.2-years in age-related neurodegeneration. Error bars are standard error of the mean. The seemingly constant gap between the two cohorts is explained by uniformly distributed disease duration across the lifespan. \*P  $\leq$  0.05; \*\*P  $\leq$  0.01; \*\*\*P  $\leq$  0.001. . . 78
- 5.6 Supplementary Figure 2: Effects of age and T2DM exhibited strong correlations within datasets and modalities, but no significant correlations were observed across modalities. We considered six sets of previously characterized changes in association with the following: 1. Age contrast, gray matter volume in UK Biobank; 2. T2DM contrast, gray matter volume in UK Biobank; 3. Age contrast, brain activation in UK Biobank; 4. T2DM contrast, brain activation in UK Biobank; 5. Age contrast, brain structure/activation (aggregate) from NeuroQuery; 6. T2DM contrast, brain structure/activation (aggregate) from NeuroQuery. Corresponding effects from region/domain specific analyses were considered as inputs for correlation measures, which were then determined for all combinations of the six sets of effects. Age and T2DM were significantly correlated (Pearson's r) within all modalities, suggesting common trajectories between age and T2DM related effects. No significant correlations were observed, however, across datasets or modalities. \*P  $\leq$  0.05; \*\*P  $\leq$  0.01; \*\*\*P  $\leq$  0.001, Bonferroni corrected. . . . 79

5.7 Supplementary Figure 3: Treatment of T2DM patients with metformin had no impact on cognitive deficits or gray matter atrophy. We evaluated the UK Biobank dataset to determine whether treatment with metformin would prevent gray matter atrophy or the development of cognitive deficits associated with T2DM. Among T2DM diagnosed subjects only, we compared those subjects who reported using metformin but no other medications to those who reported not taking any medications to treat T2DM. We matched subjects for age, sex, education and T2DM disease duration, and controlled for BMI. The direction of theoretical improvement by metformin is indicated on both panels by an arrow. A: No statistically significant ( $\alpha=0.05$ ) differences in cognitive performance were detected when comparing subjects on metformin to unmedicated subjects B: Neither our analysis of gray matter atrophy detected any significant ( $\alpha=0.05$ ) improvements associated with metformin treatment. Error bars are 95% CI. . . . . 80

PREVIEW

# Chapter 1

## Introduction

### 1.1 Background

How can three pounds of flesh still outperform the most powerful computers to date, and that too while consuming less energy than an average lightbulb? The inner working of the human brain undoubtedly remains one of the biggest scientific mysteries to date. The set of technologies that hold the greatest promise to ever solving this great mystery are known as neuroimaging. These technologies allow us to peek inside the human brain from the outside, and capture images that reveal not only what it is made up of but also what it is doing at any point in time.

One neuroimaging technology in particular stands out as having revolutionized our ability to image the human brain: Magnetic Resonance Imaging (MRI). MRI provides us the ability to safely watch the human brain in action, which in turn allows us to understand how the brain performs many of its functions. MRI scans can be grouped into two broad categories: structural and functional MRI, each measuring a distinct but equally important aspect of the brain. Structural MRI captures the makeup of brain's anatomy, such as how much water or fat is present in the tissue. Because different parts of the brain contain different amounts of water and fat, they show up on the MRI image as either brighter or darker. Structural MRIs of the brain are often used for detecting diseases and for understanding differences in size and shape of different brain regions between people, but they fail to capture what the brain is doing. For that, we need functional MRI (fMRI). fMRI picks up on the after effects of brain activity by sensing the changes in amount of oxygen in the blood, for a particular area, over a window of time.

In many cases fMRI can even allow us to decode what people are experiencing or thinking about by looking at their brain activity, when they are

performing a task or simply resting. MRI has also shown us how experiences change the brain, and how individual human brains change over time from childhood to old age. It has also shown us that all human brains follow the same basic principle, while also exhibiting key differences across people, thus giving insights into brain dysfunctions that relate to various mental illnesses.

fMRI relies upon a set of critical chemical and biological dominoes that all had to be perfectly in line for the technology to work. The first biological fact that makes fMRI possible is that the firing of neurons is relatively localized in the brain. For example, all humans, in fact most animals, possess a specific part of the brain, often in the back of the head, that processes visual information, generally called the visual cortex. This part of the brain responds to information of different parts of the visual world coming from the eyes. Similarly, different parts of the brain are responsible for doing different things. In other words, there is some degree of modularity in our brains. This modularity and localization of function ultimately allows us to *decode* from the brain scans, what the person in the scanner might have been doing or thinking of at the time of the scan.

Another aspect of the brain that is critical to fMRI being extremely effective as a tool for studying human brains is that brains are organized in a relatively consistent manner across individuals. All humans, for instance, in addition to the visual cortex, have a region of the brain, at the rear of the frontal lobe, that controls the movement of hands and feet. The alignment of different brain regions across people is not always perfect, in fact, far from it, but it's good enough to allow group level analyses.

The third crucial factor is that the localized firing of neurons result in increase of blood flow at that location. Without such localization, we wouldn't be able to narrow down the precise region of interest relevant to the neurons that caused the activity.

The final important characteristic is that blood flow response to an active region of the brain active neurons brings with it glucose and oxygen for the neurons. The amount of glucose supplied is just the right amount to account for the energy needed by the neurons to fire, but the amount of oxygen sent is in surplus, relative to the small amount actually needed by the neurons. It is precisely this overflow of oxygenated blood that lets the fMRI scanner detect the activity of neurons.

fMRI has ultimately raised some extremely fundamental questions about how we view ourselves as humans. If all cognition is just neuronal activations, that we can visualize with MRI, does that mean telepathy is only a short step away? Would we be able to read each other's minds? Have we solved the mystery of human consciousness? In what sense are we truly responsible

for our choices? Is addiction the same as any other physiological disease or a failure of self control or both?

It is these kinds of questions have resulted in recent years witnessing an explosion of interest in neuroimaging research. Each year, thousands of brain-imaging studies explore the links between brain and behavior. Last year, over 35,000 fMRI articles were published [2], including more than 6,000 publications containing the term "neuroimaging" [3]. These include study of a truly wide array of behaviors and characteristics, ranging from mental illnesses, cognitive acuity and adolescent development to left-handedness, gambling and political leanings.

However most neuroimaging research involves probing single variables to make narrow predefined discoveries, typically on a small data set. Individual studies therefore, seldom have statistical power to establish fully trustworthy results [4, 5]. Most neuroimaging studies suffer from the problem of small sample sizes. Median sample size of fMRI studies in 2015 was 28.5 subjects and 75th percentile of sample size in cognitive neuroscience journals published between 2011 to 2014 was 28 subjects. Working with such small sample sizes raises concerns regarding representativeness of the sample and inflated false discovery rate, particularly in its usage for clinical purposes, as well as regarding reproducibility. Low statistical power reduces the probability of detecting statistically significant results. As a consequence, in neuroimaging, effect sizes are often overestimated and reproducibility of results is fairly low. Finding consistent aggregate trends in the knowledge acquired across these studies is crucial but daunting.

Technological advances in recent decades however, now allow generating, collecting and analyzing massive data sets. This new scale in availability of data necessitates moving away from traditional methods that are used to infer statistically relevant effects in carefully chosen variables towards more machine learning based pattern-recognition algorithms that can identify relevant biological signatures in an unbiased way and produce results that generalize better to new unseen data.

## 1.2 Predictive models and representations

In this document I use an unprecedentedly large-scale neuroimaging and genetic data set UKBiobank [6] (N=19,831) and detail multiple computational models and frameworks for effectively predicting disease and other phenotypes, while also shedding critical light on the underlying mechanisms at play.

In Chapter 2, I present a novel prediction based framework called *Neuropredictome* that allows identification of statistically significant linkages be-

tween phenotypes and neuroimaging features from four different modalities. I evaluate phenotype linkage to brain activity of 19,831 subjects with 4926 variables, pertaining to the health, physiology, psychology, social and economic state. I corroborate my identified regions of the brain with previous work by providing a novel quantitative evaluation of how well my results align with existing meta-analyses of 14,371 published neuroimaging research articles. My results show that neuroimaging reveals as many neurological links to physical and lifestyle factors as to cognitive factors, supporting a more integrative approach to medicine that considers disease interactions between multiple organs and systems. My analysis is presented as a public resource at <https://neuropredictome.com> providing an interpretable view of human brain organization and decoding, to assist in hypothesis generation and evaluating future studies.

Mental health disorders are increasingly believed to be biological disorders involving some form of breakdown in brain circuitry. However, most existing work in computational neuroscience focuses on networks rather than circuits. Techniques that *do* allow modeling circuits in the brain at the macro scale unfortunately do not scale well enough to allow studying circuits at the whole-brain level. This not only forgoes recent advances in high-resolution non-invasive human neuroimaging technologies and availability of large data sets, but also severely limits the clinical applicability of such work. In Chapter 3, I bridge the gap between network theory and control theory and present a scalable framework for extracting fMRI-derived (generative) control circuits, then use circuit trajectories to estimate their control error. Using synthetic circuits, I first demonstrate that my framework accurately identifies each circuit's architecture and models its dynamics by estimation of transfer functions at the individual node level. As a use case, I then apply the framework to human task-based functional MRI data from UK Biobank (N=19,831). In a purely data-driven manner, without priors, my framework identified thalamus-linked *prefrontal-limbic* and *ventral stream* subcircuits, selectively engaged during sensorimotor processing of affective and non-affective stimuli. Finally, I demonstrate that circuit-wide dysregulation, defined by degree of drift from healthy trajectories, tracks symptom severity for neuroticism (*ventral subcircuit*), depression (*prefrontal-limbic subcircuit*), and bipolar disorder (*full circuit*).

In Chapter 4, I develop methods for constructing low-dimensional vector representations (embeddings) of large-scale genotyping data, capable of reducing genotypes of hundreds of thousands of SNPs to 100-dimensional embeddings that retain substantial predictive power for inferring medical phenotypes. I demonstrate that embedding-based models yield an average F-score of 0.605

on a test of ten phenotypes (including BMI prediction, genetic relatedness, and depression) versus 0.339 for baseline models. Genotype embeddings also hold promise for creating sharing data while preserving subject anonymity: I show that they retain substantial predictive power even after anonymization by adding Gaussian noise to each dimension.

Type 2 diabetes mellitus (T2DM) is known to be associated with neurobiological and cognitive deficits; however, their extent, overlap with aging effects, and the effectiveness of existing treatments in the context of the brain are currently unknown. In Chapter 5, I characterize neurocognitive effects independently associated with T2DM and age in a large cohort of human subjects from the UK Biobank with cross-sectional neuroimaging and cognitive data. I then proceeded to evaluate the extent of overlap between the effects related to T2DM and age by applying correlation measures to the independently characterized neurocognitive changes. My findings were complemented by meta-analyses of published reports with cognitive or neuroimaging measures for T2DM and healthy controls (HC). I also evaluated in a cohort of T2DM diagnosed individuals using UK Biobank how disease chronicity and metformin treatment may influence the characterized deficiencies. All analyses were Bonferroni corrected. Duration of T2DM ranged from 0–45 years (mean  $9.7 \pm 7.9$  years); 559 were treated with metformin alone, while 473 were unmedicated. My meta-analysis evaluated 34 cognitive studies ( $N=22,231$ ) and 60 neuroimaging studies: 30 of T2DM ( $N=866$ ) and 30 of aging ( $N=1088$ ). As compared to age, sex, and education-matched HC, T2DM was associated with marked cognitive deficits, particularly in executive functioning and processing speed. Likewise, I found that the diagnosis of T2DM was significantly associated with gray matter atrophy and reorganization of brain activity. The structural and functional changes associated with T2DM show marked overlap with the effects correlating with age but appear earlier, with disease duration linked to more severe neurodegeneration. The neurocognitive impact of T2DM suggests marked acceleration of normal brain aging, by approximately 24%, made worse with chronicity. As such, neuroimaging-based biomarkers may provide a valuable adjunctive measure of T2DM progression and treatment efficacy based on neurological outcomes. Conventional T2DM treatments must be reevaluated with respect to restraining neurodegeneration.

Variables declared important by traditional null hypothesis significance testing can be incongruent with the variables that maximize predictive performance in new individuals or settings. This combined with small sample sizes is a major factor in the low reproducibility crisis [7] plaguing the scientific world today. The advent of “big data” in neuroscience and biomedicine in conjunction with recent advances in machine learning and deep learning

presents an tremendous opportunity to address these problems. The data-rich neuroscientist and geneticist can ask many new questions that could probably never be addressed quantitatively before.

In this document, I present models and analytical techniques that leverage the availability of large-scale data sets made possible by new modes of data dissemination and open science and also use state of the art machine learning techniques better suited for their analyses. These not only optimize for predictions and reproducibility of results for new previously unseen subjects but also evaluate them against decades of past and existing research. Most importantly, the results presented here argue that prediction does not have to come at the cost of interpretability. Mechanistic insight does not need be compromised by optimizing predictive accuracy and the two goals can be achieved together and synergistically.

PREVIEW

The remaining of the document is structured as follows:

1. **Chapter 2** - Building a machine learning based predictive framework called *Neuropredictome* that links 4928 phenotypes to brain features from fMRI scans of 19,831 subjects in UK-Biobank. The framework is then evaluated against 14,371 published peer-reviewed papers using deep learning based text embeddings.
2. **Chapter 3** - Modeling the brain as an amalgam of control circuits and modeling disease as the control circuit's failure to maintain homeostatic or allostatic control. Using a system of differential equations to model the complex causal relationships and feedback loops that constitute brain circuits, I show how the prefrontal-limbic circuit [8] and the cortico-thalamic circuit [9] can be discovered from task-fMRI scans in UK Biobank. Modeling prediction error as the measure of dysregulation, I also show how the two circuits are dysregulated across the following psychiatric disorders: neuroticism, depression and bipolar disorder.
3. **Chapter 4** - I develop methods for constructing low-dimensional vector representations (embeddings) of large-scale genotyping data, capable of reducing genotypes of hundreds of thousands of SNPs to 100-dimensional embeddings that retain substantial predictive power for inferring medical phenotypes. I also show that these embeddings retain substantial predictive power even after anonymization by adding Gaussian noise to each dimension.
4. **Chapter 5** - I characterize neurocognitive effects independently associated with Type 2 Diabetes (T2DM) and Aging in UK Biobank with cross-sectional neuroimaging and cognitive data. My results show a significant overlap in the effects related to T2DM and Aging, complemented by meta-analyses of published reports with cognitive or neuroimaging measures. I also evaluate how disease chronicity and metformin treatment may influence the characterized deficiencies. My results show T2DM accelerates normal brain aging, by approximately 24%, made worse with chronicity.