Bioinformatics, 34, 2018, i866–i874
doi: 10.1093/bioinformatics/bty557
ECCB 2018

Quantitative trait loci identification for brain endophenotypes via new additive model with random networks

Xiaoqian Wang[1], Hong Chen[1], Jingwen Yan[2], Kwangsik Nho[2], Shannon L. Risacher[2], Andrew J. Saykin[2], Li Shen[3] and Heng Huang[1,*] for the ADNI

[1]Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA, [2]Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN 46202, USA and [3]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

*To whom correspondence should be addressed.

Abstract

Motivation: The identification of quantitative trait loci (QTL) is critical to the study of causal relationships between genetic variations and disease abnormalities. We focus on identifying the QTLs associated to the brain endophenotypes in imaging genomics study for Alzheimer's Disease (AD). Existing research works mainly depict the association between single nucleotide polymorphisms (SNPs) and the brain endophenotypes via the linear methods, which may introduce high bias due to the simplicity of the models. Since the influence of QTLs on brain endophenotypes is quite complex, it is desired to design the appropriate non-linear models to investigate the associations of genotypes and endophenotypes.

Results: In this paper, we propose a new additive model to learn the non-linear associations between SNPs and brain endophenotypes in Alzheimer's disease. Our model can be flexibly employed to explain the non-linear influence of QTLs, thus is more adaptive for the complex distribution of the high-throughput biological data. Meanwhile, as an important computational learning theory contribution, we provide the generalization error analysis for the proposed approach. Unlike most previous theoretical analysis under independent and identically distributed samples assumption, our error bound is based on *m*-dependent observations, which is more appropriate for the high-throughput and noisy biological data. Experiments on the data from Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort demonstrate the promising performance of our approach for identifying biological meaningful SNPs.

Availability and implementation: An executable is available at https://github.com/littleq1991/additive_FNNRW.

Contact: heng.huang@pitt.edu

1 Introduction

Alzheimer's Disease (AD) is the most common form of dementia, which triggers memory, thinking and behavior problems. The genetic causal relationship of AD is complex (Avramopoulos, 2009) and therefore presents difficulties in the prevention, diagnosis and treatment of this disease. Recent advances in multimodal neuroimaging and high throughput genotyping and sequencing techniques bring an emerging research field, imaging genomics, which provides exciting new opportunities to ultimately improve our understanding of brain disease, their genetic architecture and their influences on cognition and behavior.

The rapid progress in neuroimaging techniques has provided insights into early detection and tracking of neurological disorders (Weiner *et al.*, 2013). Later research interest in imaging neuroscience has focused on Genome Wide Association Studies (GWAS) to examine the association between genetic markers, called Single Nucleotide Polymorphisms (SNPs), and imaging phenotypes (Cooper-Knock *et al.*, 2014; Waring and Rosenberg, 2008), with the

goal of finding explanations for the variability observed in brain structures and functions. However, these research works typically study associations between individual SNPs and individual phenotypes and overlook interrelated structures among them. To better understand the genetic causal factors of brain imaging abnormalities, previous works have laid great emphasis on identifying relevant QTL (Ryan *et al.*, 2016; Vounou *et al.*, 2010), which related high-throughput SNPs to imaging data and enhanced the progress and prosperity of neuroscience research.

Several machine learning models were established to depict the relations between SNPs and brain endophenotypes (Huo *et al.*, 2018; Wang *et al.*, 2012a; Wang *et al.*, 2017; Yang *et al.*, 2015; Zhu *et al.*, 2016). In Wang *et al.* (2012a), Zhu *et al.* (2016), Wang *et al.* (2017) and Huo *et al.* (2018), the authors used the low-rank learning models or structured sparse learning models to select the imaging features that share common effects in the regression analysis. Yang *et al.* (2015) applied the LASSO regression model to discover the significant SNPs that are associated with brain imaging features. However, previous works use linear models to predict the relations between genetic biomarkers and brain endophenotypes, which may introduce high bias during the learning process. Since the influence of QTL is complex, it is crucial to design appropriate non-linear model to investigate the genetic biomarkers (due to the limited size of biological data, deep learning models don't work well for our problem). Besides, most previous computational models on genotype and phenotype studies did not provide theoretical analysis on the performance of the models, thus leaves uncertainty in the validity of the models.

To tackle with these challenging problems, in this paper, we propose a novel and efficient nonlinear model for the identification of QTL. We apply our model to the QTL identification of Alzheimer's disease (AD), the most common cause of dementia. By means of feedforward neural networks, our model can be flexibly employed to explain the non-linear associations between genetic biomarkers and brain endophenotypes, which is more adaptive for the complicated distribution of the high-throughput biological data. We would like to emphasize the following contributions of our work:

- We propose a novel additive model with generalization error analysis. In particular, different from conventional analysis with independent samples, our error bound is under m-dependent observations, which is a more general assumption and more appropriate for the high-throughput complex genotypes and phenotypes.
- Our model is efficient in computation. The time complexity of our model is linear to the number of samples and number of features in the data. Experimentally we showed that it only takes a few minutes to run our model on the ADNI data.
- Experimental results demonstrate that our model not only identifies several well-established AD-associated genetic variants, but also finds out new potential SNPs.

Notation: Throughout this paper, unless specified otherwise, upper case letters denote matrices, e.g. X, Y. Bold lower case letters denote vectors, e.g. $\mathbf{w}$, $\mathbf{b}$. Plain lower case letters denote scalars, e.g. a, γ. w_i denotes the i-th element of vector $\mathbf{w}$. $\mathbf{w}_i$ denotes the i-th row of matrix W. $\mathbf{w}^j$ or $\mathbf{w}^{(j)}$ denotes the j-th column of W. w_{ij} denotes the ij-th element of matrix W. $||\mathbf{w}||_2$ or $||\mathbf{w}||$ denotes the ℓ_2-norm of vector $\mathbf{w}$: $\sqrt{\sum_i w_i^2}$. $\|W\|_F$ denotes the Frobenius norm of matrix W: $\|W\|_F = \sqrt{\sum_i \sum_j w_{ij}^2} = \sqrt{\sum_i \|\mathbf{w}_i\|^2}$. $||W||_1$ denotes the ℓ_1 norm: $||W||_1 = \sum_i \sum_j |w_{ij}|$. $||W||_*$ denotes the trace norm (*a.k.a.* nuclear norm): $||W||_* = \sum_i \sigma_i$, where σ_i is the i-th singular value of W.

Specially, d denotes the dimension of the feature vector, i.e. number of SNPs. n denotes the number of patients. c represents the number of QTs. $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n \times d}$ denotes the input SNP matrix, where each row of X represents the genetic variants of each patient. $Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^{\mathrm{T}} \in \mathbb{R}^{n \times c}$ represents the input imaging feature matrix where each row of Y denotes the phenotype of one patient. I stands for the identity matrix, and $\mathbf{1}$ stands for a vector with all elements being 1.

2 Related work

In QTL identification of brain imaging abnormalities, the goal is to learn a prediction function which estimates the imaging feature matrix $Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^{\mathrm{T}} \in \mathbb{R}^{n \times c}$ given the genetic information $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n \times d}$. Meanwhile, we want to weigh the importance of each SNP in the prediction according to the learning model. The most straightforward method is least square regression, which learns a weight matrix $W \in \mathbb{R}^{d \times c}$ to study the relations between SNPs and brain endophenotypes. W is an intuitive reflect of the importance of each SNP for the prediction of each endophenotype.

Based on least square regression, several models were proposed for QTL identification. In Tibshirani (1996) and Yang *et al.* (2015), the authors employed sparse regression models for the discovery of predominant genetic features. In Fazel (2002) and Wang *et al.* (2012a), low-rank constraint was imposed to uncover the group structure among SNPs in the association study.

In the identification of QTL, previous works mainly use linear models for the prediction. However, according to previous studies, the biological impact of genetic variations is complex (Meyer-Lindenberg *et al.*, 2006) and the genetic influence on brain structure is complicated (Peper *et al.*, 2007). Thus, the relations between genetic biomarkers and brain-imaging features may not be necessarily linear and the prediction with linear models is likely to trigger large bias.

To depict the non-linear association between genetic variations and endophenotypes, neural networks introduce a convenient and popular framework. Schmidt *et al.* (1992) proposed feed forward neural networks with random weights (FNNRW), which can be formed as:

$$f(\mathbf{x}) = \sum_{t=1}^{h} a_t \phi(\langle \mathbf{v}_t, \mathbf{x} \rangle + b_t), \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_d] \in \mathbb{R}^d$ is the input data, h is the number of hidden nodes, $\mathbf{v}_t|_{t=1}^h = [v_{t1}, v_{t2}, \ldots, v_{td}] \in \mathbb{R}^d$ is the parameter in the hidden layer for t-th hidden node, $b_t \in \mathbb{R}$ is the corresponding bias term, $\langle \mathbf{v}_t, \mathbf{x} \rangle = \sum_{j=1}^d v_{tj} x_j$ represents Euclidean inner product, $\phi(.)$ is the activation function and $a_t \in \mathbb{R}$ is the weight for the t-th hidden node.

As is analyzed in Igelnik and Pao (1995) and Rahimi and Recht (2009), FNNRW enjoys an obvious advantage in computational efficiency over neural nets with back propagation. In Equation (1), $\mathbf{v}_t$ and b_t are randomly and independently chosen before hand, and the randomization in parameter largely relieves the computational burden. FNNRW is aimed at estimating only the weight parameter $a_t|_{t=1}^h$ thus is extremely efficient. Such property makes FNNRW more appropriate for analysis of the high-throughput data in Alzheimer's research.

Rahimi and Recht (2009) constructed a classifier using FNNRW where they conduct classification on the featurized data as shown in Equation (1). The classification model can be easily extended to the regression scenario with the objective function formulated as:

$$\min_{\mathbf{a}_t|_{t=1}^h} \left\| Y - \sum_{t=1}^{h} \phi\left(X\mathbf{v}_t^{\mathrm{T}} + b_t\mathbf{1}\right)\mathbf{a}_t \right\|_F^2 + \gamma \sum_{t=1}^{h} \|\mathbf{a}_t\|_2^2, \tag{2}$$

where γ is the hyper-parameter for the regularization term and $\mathbf{a}_t = [a_1,\ a_2, \ldots,\ a_c] \in \mathbb{R}^c$ is the weight parameter of the t-th hidden node for c different endophenotypes. As discussed above, Problem (2) can be adopted to efficiently estimate the nonlinear associations between genetic variations and brain endophenotypes. However, since the parameters of hidden layer is randomly assigned, traditional FNNRW model makes it hard to evaluate the importance of each feature.

To tackle with these problems, we propose a novel additive model in next section, which not only maintains the advantage of computational efficiency of FNNRW but also integrates the flexibility and interpretability of additive models.

3 New additive model for identifying quantitative trait loci of brain endophenotypes

We propose new Additive Model via Feedforward Neural networks with random weights (FNAM) as:

$$f_a(X) = \sum_{t=1}^{h} \sum_{j=1}^{d} \phi\left(v_{tj}\mathbf{x}_j + b_t\mathbf{1}\right)\mathbf{a}_t, \tag{3}$$

where we distinguish the contribution of each feature $\mathbf{x}_j$ and formulate the model in an additive style for the prediction. Similar to that of FNNRW, we propose to optimize the least square loss between the ground truth endophenotype matrix Y and the estimation $f_a(X)$ with ℓ_2-norm penalization, then we propose the following objective function:

$$\min_{\mathbf{a}_t|_{t=1}^h} \left\| Y - \sum_{t=1}^{h} \sum_{j=1}^{d} \phi\left(v_{tj}\mathbf{x}_j + b_t\mathbf{1}\right)\mathbf{a}_t \right\|_F^2 + \gamma \sum_{t=1}^{h} \|\mathbf{a}_t\|_2^2, \tag{4}$$

For simplicity, if we define $A = [\mathbf{a}_1,\ \mathbf{a}_2, \ldots,\ \mathbf{a}_h]^{\mathrm{T}} \in \mathbb{R}^{h\times c}$ as the weight parameter for hidden nodes, and $G \in \mathbb{R}^{n\times h}$ such that

$$G = \begin{bmatrix} \sum_{j=1}^{d}\phi\left(v_{1j}x_{1j} + b_1\right) & \ldots & \sum_{j=1}^{d}\phi\left(v_{hj}x_{1j} + b_h\right) \\ \vdots & \ldots & \vdots \\ \sum_{j=1}^{d}\phi\left(v_{1j}x_{nj} + b_1\right) & \ldots & \sum_{j=1}^{d}\phi\left(v_{hj}x_{nj} + b_h\right) \end{bmatrix}, \tag{5}$$

then we could rewrite our objective function Problem (4) as:

$$\min_A \|Y - GA\|_F^2 + \gamma\|A\|_F^2. \tag{6}$$

Take derivative *w.r.t.* A in Problem (6) and set it to 0, we get the closed form solution of A as below:

$$A = \left(G^{\mathrm{T}}G + \gamma I\right)^{-1}G^{\mathrm{T}}Y. \tag{7}$$

As discussed in the previous section, one obvious advantage of FNAM over FNNRW is that FNAM considers the role of each feature independently in the prediction, thus makes it possible to interpret the importance of each SNP in the identification QTL, which is a fundamental goal of jointly studying genetic and brain imaging features.

Here, we discuss how to estimate the role of each feature in FNAM. To separate the contribution of each feature, we rewrite Equation (3) as below:

$$f_a(X) = \sum_{j=1}^{d}\left(\sum_{t=1}^{h}\phi\left(v_{tj}\mathbf{x}_j + b_t\mathbf{1}\right)\mathbf{a}_t\right), \tag{8}$$

which indicates that the prediction function $f_a(X)$ can be regarded as the summation of d terms, where the j-th term $\sum_{t=1}^{h}\phi\left(v_{tj}\mathbf{x}_j + b_t\mathbf{1}\right)\mathbf{a}_t$ denotes the contribution of the j-th feature.

Naturally, if we normalize the magnitude of the j-th term with the ℓ_2-norm of $\mathbf{x}_j$, we could get a good estimation of the significance of the j-th feature. As a consequence, we could define a weight matrix $W \in \mathbb{R}^{d\times c}$ to show the importance of the d SNPs in the prediction of the c imaging features, respectively, such that:

$$w_{jl} = \frac{\left\|\sum_{t=1}^{h}\phi\left(v_{tj}\mathbf{x}_j + b_t\mathbf{1}\right)a_{tl}\right\|}{\|\mathbf{x}_j\|}, j = 1, \ldots d,\ \ l = 1, \ldots c, \tag{9}$$

Algorithm 1 Optimization Algorithm of FNAM for QTL Identification.

Input:

SNP matrix $X \in \mathbb{R}^{n\times d}$, endophenotype $Y \in \mathbb{R}^{n\times c}$, number of hidden nodes h, parameter γ.

Output:

Weight matrix $A \in \mathbb{R}^{h\times c}$ for the hidden nodes. Weight matrix $W \in \mathbb{R}^{d\times c}$ showing the relative importance of the d SNPs in the prediction.

Initialize the weight matrix $V \in \mathbb{R}^{h\times d}$ randomly according to uniform distribution $\mathcal{U}(0,\ 1)$.

Initialize the bias term $b \in \mathbb{R}^h$ randomly according to uniform distribution $\mathcal{U}(0,\ 1)$.

1. Compute G matrix according to the definition in Equation (5).
2. Update A according to the solution in Equation (7)
3. Compute W according to the definition in Equation (9).

3.1 Time complexity analysis

We summarize the optimization steps of FNAM in Algorithm 1. In Algorithm 1, the time complexity of Step 1 (computing G) is $O(ndh)$, the time complexity of Step 2 (computing A) is $O(h^2n + hnc)$, and the time complexity of Step 3 (computing W) is $O(ndhc)$, where n is the number of patients, d denotes the number of SNPs, and c represents the number of brain endophenotypes. Typically, we have $d > h$ and $d > c$ in the identification of QTL, thus the total time complexity of Algorithm 1 is $O(ndhc)$.

4 Generalization ability analysis

In this section, based on the real situation of biological data, we provide theoretical analysis on the approximation ability of our FNAM model and derive the upper bound of generalization error.

In most previous works, theoretical analysis is based on the hypothesis of independent and identically distributed (i.i.d.) samples. However, the i.i.d. sampling is a very restrictive concept that occurs only in the ideal case. As we know, the acquisition of

high-throughput biological data involves complicated equipments, reagents as well as precise operation of highly trained technicians, which usually introduce variations to the data during the measurement process (Leek *et al.*, 2010). Thus, the i.i.d. sampling assumption is not appropriate for the high-throughput biological data analysis. In this section, we provide a learning rate estimate of our model in a much general setting, i.e. m-dependent observations (Modha and Masry, 1996).

For simplicity, here we consider the prediction of only one brain endophenotype $\mathbf{y} = [y_1, y_2, \ldots, y_n]^{\mathrm{T}} \in \mathbb{R}^n$, which could be easily extended to the case with multiple endophenotypes. Besides, we incorporate the bias term $\mathbf{b}$ into the weight matrix V by adding one feature valued 1 for all samples to the data matrix X. For analysis feasibility, we reformulate the general FNAM model as below.

Let $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ is a compact metric space and $\mathcal{Y} \subset [-k, k]$ for some constant $k > 0$. For any given $\mathbf{z} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n \in \mathcal{Z}^n$ and each $j \in \{1, 2, \ldots, d\}$, we denote $\phi_i^{(j)} = [\phi(v_{1j}, x_{ij}), \ldots, \phi(v_{bj}, x_{ij})]^{\mathrm{T}} \in \mathbb{R}^b$ and $\mathbf{v}^{(j)} = [v_{1j}, v_{2j}, \ldots, v_{bj}]^{\mathrm{T}} \in \mathbb{R}^b$, where each v_{tj}, $1 \le t \le h$, is generated i.i.d. from a distribution μ on $[0, 1]$.

The FNN with random weights in FNAM can be formulated as the following optimization problem:

$$\mathbf{a}_{\mathbf{z}} = \underset{\mathbf{a} \in \mathbb{R}^{bd}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{d} \left(\mathbf{a}^{(j)} \right)^{\mathrm{T}} \phi_i^{(j)} - y_i \right)^2 + \gamma \sum_{j=1}^{d} \left\| \mathbf{a}^{(j)} \right\|_2^2 \right\},$$

where $\mathbf{a}^{(j)} = \left[a_1^{(j)}, a_2^{(j)}, \ldots, a_b^{(j)} \right]^{\mathrm{T}} \in \mathbb{R}^b$.

The predictor of FNAM is:

$$f_{\mathbf{z}} = \sum_{j=1}^{d} \sum_{t=1}^{b} a_{\mathbf{z},t}^{(j)} \phi(v_{tj}, \cdot).$$

To investigate the generalization error bound of FNAM, we rewrite it from a function approximation viewpoint.

Define the hypothesis function space of FNAM as:

$$\mathcal{M}_b = \left\{ f = \sum_{j=1}^{d} f^{(j)} : f^{(j)} = \sum_{t=1}^{b} a_{tj} \phi(v_{tj}, \cdot), a_{tj} \in \mathbb{R} \right\}$$

and for any $j \in \{1, 2, \ldots, d\}$

$$\left\| f^{(j)} \right\|_{\ell_2}^2 = \inf \left\{ \left\| \mathbf{a}^{(j)} \right\|_2^2 : f = \sum_{t=1}^{b} a_{tj} \phi(v_{tj}, \cdot) \right\}.$$

Then, FNAM can be rewritten as the following optimization problem:

$$f_{\mathbf{z}} = \sum_{j=1}^{d} f_{\mathbf{z}}^{(j)} = \underset{f \in \mathcal{M}_b}{\operatorname{argmin}} \left\{ \mathcal{E}_{\mathbf{z}}(f) + \gamma \sum_{j=1}^{d} \left\| f^{(j)} \right\|_{\ell_2}^2 \right\},$$

where $\mathcal{E}_{\mathbf{z}}(f)$ is the empirical risk defined by $\mathcal{E}_{\mathbf{z}}(f) = \frac{1}{n} \sum_{i=1}^{n} (f(\mathbf{x}_i) - y_i)^2$.

For the regression problem, the goal of learning is to find a prediction function $f : \mathbf{x} \to \mathbb{R}$ such that the expected risk

$$\mathcal{E}(f) = \int_{\mathcal{Z}} \left(y - f(\mathbf{x}) \right)^2 d\rho(\mathbf{x}, y)$$

is as small as possible. It is well known that the Bayes function

$$f_\rho(\mathbf{x}) = \int_{\mathcal{Y}} y d\rho(y|\mathbf{x})$$

is the minimizer of $\mathcal{E}(f)$ over all measurable functions. Therefore, the excess expected risk $\mathcal{E}(f) - \mathcal{E}(f_\rho)$ is used as the measure to evaluate the performance of learning algorithm.

Since $\mathcal{Y} \subset [-k, k]$ and $\|f_\rho\|_\infty \le k$, we introduce the clipping operation

$$\pi(f) = \max(-k, \min(f(\mathbf{x}), k))$$

to get tight estimate on the excess risk of FNAM. Recall that FNAM in (4) depends on the additive structure and random weighted networks. Indeed, theoretical analysis of standard random weighted networks has been provided in Igelnik and Pao (1995) and Rahimi and Recht (2009) to characterize its generalization error bound. However, the previous works are restricted to the setting of i.i.d. samples, and do not cover the additive models. Hence, it is necessary to establish the upper bound of $\mathcal{E}(\pi(f_{\mathbf{z}})) - \mathcal{E}(f_\rho)$ with much general setting, e.g. m-dependent observations (Modha and Masry, 1996; Vidyasagar, 2013).

Now, we introduce some necessary definitions and notations for theoretical analysis.

Let $\{Z_i = (X_i, Y_i)\}_{i=1}^{\infty}$ be a stationary random process on a probability space $(\Omega, \mathcal{A}, P)$. Denote $\mathcal{A}_1^i$ and $\mathcal{A}_{i+m}^{\infty}$ as the σ-algebras of events generated by $(Z_1, Z_2, \ldots, Z_i)$ and $(Z_{i+m}, Z_{i+m+1}, \ldots)$, respectively.

Definition: For $m \ge 0$, if $\mathcal{A}_1^i$ and $\mathcal{A}_{i+m}^{\infty}$ are independent, we call $\{\mathcal{Z}_i\}_{i=1}^{\infty}$ m-dependent.

It is clear that $m = 0$ for i.i.d. observations.

It is a position to present the main result on the excess risk $\mathcal{E}(\pi(f_{\mathbf{z}})) - \mathcal{E}(f_\rho)$.

Theorem 1: Let $f_{\mathbf{z}}$ be defined in (4) associated with m-dependent observations $\mathbf{z} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$. There holds

$$E_{\rho^n} E_{\mu^b} \|\pi(f_{\mathbf{z}}) - f_\rho\|_{L^2_{\rho_{\mathcal{X}}}}^2 \le c \sqrt{\frac{\log n^{(m)} - \frac{1}{2}\log\gamma}{n^{(m)}}} + \inf_{f \in \mathcal{M}_b} \left\{ \left\| f - f_\rho \right\|_{L_{\rho_{\mathcal{X}}}}^2 + \gamma \sum_{j=1}^{d} \left\| f^{(j)} \right\|_{\ell_2}^2 \right\},$$

where $n^{(m)} = \left\lfloor \frac{n}{m+1} \right\rfloor$, $\|\cdot\|_{L^2_{\rho_{\mathcal{X}}}}$ is norm of square integral function space $L^2_{\rho_{\mathcal{X}}}$, and c is a positive constant independent of $n^{(m)}, \gamma$.

Theorem 1 demonstrates that FNAM can achieve the learning rate $O\left(\sqrt{\frac{\log n^{(m)}}{n^{(m)}}}\right)$ as the hypothesis space satisfies

$$\inf_{f \in \mathcal{M}_b} \left\{ \left\| f - f_\rho \right\|_{L^2_{\rho_{\mathcal{X}}}}^2 + \gamma \sum_{j=1}^{d} \left\| f^{(j)} \right\|_{\ell_2}^2 \right\} = O\left(\sqrt{\frac{\log n^{(m)}}{n^{(m)}}} \right).$$

When $f_\rho \in \mathcal{M}_b$, we have

$$\lim_{n \to \infty} E_{\rho^n} E_{\mu^b} \|\pi(f_{\mathbf{z}}) - f_\rho\|_{L^2_{\rho_{\mathcal{X}}}}^2 = 0,$$

which means the proposed algorithm is consistency. The current result extends the previous theoretical analysis with i.i.d samples (Igelnik and Pao, 1995; Rahimi and Recht, 2009) to the m-dependent observations. Indeed, we can also obtain the error bound for strong mixing samples by the current analysis framework.

The following Bernstein inequality for m-dependent observations [Theorem 4.2 in Modha and Masry (1996)] is used for our theoretical analysis.

Lemma 2: Let $\{Z_i\}_{i=1}^{\infty}$ be a stationary m-dependent process on probability space $(\Omega, \mathcal{A}, P)$. *Let* $\psi : \mathbb{R} \to \mathbb{R}$ be some measurable

function and $U_i = \psi(Z_i), 1 \le i \le \infty$. Assume that $|U_1| \le d_1$ and $EU_1 = 0$. Then, for all $n \ge m+1$ and $\epsilon > 0$,

$$P\left\{\frac{1}{n}\sum_{i=1}^{n} U_i \ge \epsilon\right\} \le \exp\left\{-\frac{n^{(m)}\epsilon^2}{2\left(E|U_1|^2 + \frac{\epsilon d_1}{3}\right)}\right\},$$

where $n^{(m)} = \left\lfloor \frac{n}{m+1} \right\rfloor$ is the number of 'effective observations'.

The covering number is introduced to measure the capacity of hypothesis space, which has been studied extensively in (Cucker and Smale, 2001; Cucker and Zhou, 2007; Zou *et al.*, 2009).

Definition: The covering number $\mathcal{N}(\mathcal{F}, \epsilon)$ of a function set $\mathcal{F}$ is the minimal integer l such that there exists l disks with radius ϵ covering $\mathcal{F}$.

Considering the hypothesis space $\mathcal{M}_h$ in Section 4, we define its subset

$$\mathcal{B}_R = \left\{f \in \mathcal{M}_h : \sum_{j=1}^{d} \|f^{(j)}\|_{\ell_2}^2 := \sum_{j=1}^{d}\sum_{t=1}^{h} |a_{tj}|^2 \le R^2\right\}.$$

Now we present the uniform concentration estimate for $f \in \mathcal{B}_R$

Lemma 3: Let $\mathbf{z} = \{z_i\}_{i=1}^n := \{(\mathbf{x}_i, y_i)\}_{i=1}^n \in \mathcal{Z}^n$ be m-dependent observations. Then

$$P\left\{\sup_{f\in\mathcal{B}_R} \left(\mathcal{E}(\pi(f)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right)\right) \ge \epsilon\right\}$$
$$\le \mathcal{N}\left(\mathcal{B}_R, \frac{\epsilon}{16k}\right) \cdot \exp\left\{-\frac{n^{(m)}\epsilon^2}{512k^2 + 22k\epsilon}\right\}.$$

Proof: Set $U_i = \psi_f(z_i) = \mathcal{E}(\pi(f)) - \mathcal{E}(f_\rho) - \left((y_i - \pi(f)(\mathbf{x}_i))^2 - (y_i - f_\rho(\mathbf{x}_i))^2\right)$. It is easy to verify that $|U_i| \le 8k^2$ and $EU_i = 0$. From Lemma 1 we obtain, for any given m-dependent samples $z = \{(\mathbf{x}_i, y_i)\}_{i=1}^n \in \mathcal{Z}^n$ and measurable function f,

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_f(z_i) \ge \epsilon\right\}$$
$$= P\left\{\mathcal{E}(\pi(f)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right) \ge \epsilon\right\}$$
$$\le \exp\left\{-\frac{n^{(m)}\epsilon^2}{128k^2 + 16k\epsilon/3}\right\}.$$

Let $J = \mathcal{N}\left(\mathcal{B}_R, \frac{\epsilon}{16k}\right)$ and $\{f_j\}_{j=1}^J$ be the centers of disks D_j such that $\mathcal{B}_R \subset \cup_{j=1}^J D_j$. Observe that, for all $f \in D_j$ and $z \in \mathcal{Z}^n$,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\psi_f(z_i) - \psi_{f_j}(z_i)\right)$$
$$= \left|\mathcal{E}(\pi(f)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right)\right.$$
$$\left. - \left[\mathcal{E}(\pi(f_j)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f_j)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right)\right]\right|$$
$$= \left|\mathcal{E}(\pi(f)) - \mathcal{E}(f_j) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f)) - \mathcal{E}_{\mathbf{z}}(f_j)\right)\right|$$
$$\le 8k\|f - f_j\|_\infty \le \frac{\epsilon}{2}.$$

It means that

$$\sup_{f\in D_j} \mathcal{E}(\pi(f)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right) \ge \epsilon$$
$$\Rightarrow \mathcal{E}(\pi(f_j)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f_j)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right) \ge \frac{\epsilon}{2}.$$

Then

$$P\left\{\sup_{f\in\mathcal{B}_R} \left(\mathcal{E}(\pi(f)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right)\right) \ge \epsilon\right\}$$
$$\le \sum_{j=1}^{J} P\left\{\sup_{f\in D_j} \left(\mathcal{E}(\pi(f_j)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f_j)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right)\right)\right\}$$
$$\le \mathcal{N}\left(\mathcal{B}_R, \frac{\epsilon}{16k}\right)\exp\left\{-\frac{n^{(m)}\epsilon^2}{4(128k^2 + 16k\epsilon/3)}\right\}.$$

This completes the proof. □

Proof of Theorem 1: According to the definition of $\mathcal{E}(f)$ and f_ρ, we deduce that

$$\mathcal{E}(\pi(f_{\mathbf{z}})) - \mathcal{E}(f_\rho) = \|\pi(f_{\mathbf{z}}) - f_\rho\|_{L_{\rho,\mathcal{X}}}^2 = E_1 + E_2, \tag{10}$$

where $E_1 = \mathcal{E}(\pi(f_{\mathbf{z}})) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f_{\mathbf{z}})) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right)$ and $E_2 = \mathcal{E}_{\mathbf{z}}(\pi(f_{\mathbf{z}})) - \mathcal{E}_{\mathbf{z}}(f_\rho) + \gamma\sum_{j=1}^{d}\|f_{\mathbf{z}}^{(j)}\|_{\ell_2}^2$.

Now we turn to bound E_1 in terms of Lemma 2. According to the definition of $f_{\mathbf{z}}$, we get

$$\gamma\sum_{j=1}^{d}\|f_{\mathbf{z}}^{(j)}\|_{\ell_2}^2 \le \mathcal{E}_{\mathbf{z}}(0) \le k^2.$$

It means that $f_{\mathbf{z}} \in \mathcal{B}_R$ with $R = \frac{k}{\sqrt{\gamma}}$. By proposition 5 in Cucker and Smale (2001), we know that:

$$\log\mathcal{N}(\mathcal{B}_R, \epsilon) \le hd\log\left(\frac{4R}{\epsilon}\right).$$

Integrating these facts into Lemma 2, we obtain:

$$P\{E_1 \ge \epsilon\}$$
$$\le P\left\{\sup_{f\in\mathcal{B}_R} \left(\mathcal{E}(\pi(f)) - \mathcal{E}(f_\rho) - \left(\mathcal{E}_{\mathbf{z}}(\pi(f)) - \mathcal{E}_{\mathbf{z}}(f_\rho)\right)\right) \ge \epsilon\right\}$$
$$\le \exp\left\{hd\log\left(\frac{64kR}{\epsilon}\right) - \frac{n^{(m)}\epsilon^2}{512k^2 + 22k\epsilon}\right\}.$$

Then, for any $\eta \ge \frac{64k^2}{n^{(m)}}$,

$$E_{\rho^n}(E_1) = \int_0^\infty P\{E_1 \ge \epsilon\}d\epsilon$$
$$\le \eta + \int_\eta^\infty \exp\left\{hd\log\left(\frac{64k^2}{\epsilon\sqrt{\gamma}}\right) - \frac{n^{(m)}\epsilon^2}{512k^2 + 22k\epsilon}\right\}d\epsilon$$
$$\le \eta + \gamma^{-\frac{hd}{2}}\exp\left\{\frac{n^{(m)}\epsilon^2}{512k^2 + 22k\epsilon}\right\}\cdot\int_\eta^\infty\left(\frac{64k^2}{\epsilon}\right)^{hd}d\epsilon$$
$$\le \eta + \gamma^{-\frac{hd}{2}}\exp\left\{\frac{n^{(m)}\epsilon^2}{512k^2 + 22k\epsilon}\right\}\cdot\left(\frac{64k^2}{\epsilon}\right)^{hd}\eta\cdot\frac{1}{hd-1}$$
$$\le \eta + \gamma^{-\frac{hd}{2}}\exp\left\{\frac{n^{(m)}\epsilon^2}{512k^2 + 22k\epsilon}\right\}\cdot\left(n^{(m)}\right)^{hd}\cdot\frac{\eta}{hd-1}.$$

Setting $\eta = \gamma^{-\frac{hd}{2}}\exp\left\{-\frac{n^{(m)}\eta^2}{512k^2+22k\eta}\right\}\left(n^{(m)}\right)^{hd}\frac{\eta}{hd-1}$, we get:

$$\left(\frac{\sqrt{\gamma}}{n^{(m)}}\right)^{hd}(hd-1) = \exp\left\{-\frac{n^{(m)}\eta^2}{512k^2 + 22k\eta}\right\}.$$

From this equation, we can deduce that

$$\eta \le \frac{khd\left[\log\left(n^{(m)}\right) - \log\sqrt{\gamma}\right]}{n^{(m)}} + 50k^2\sqrt{\frac{hd\left(\log n^{(m)} - \log\sqrt{\gamma}\right)}{n^{(m)}}}.$$

Hence,

$$E_{\rho^n}(E_1) \leq 2\eta \leq \frac{2khd(\log(n^{(m)}) - \log\sqrt{\gamma})}{n^{(m)}} + 100k^2\sqrt{\frac{hd(\log(n^{(m)}) - \log\sqrt{\gamma})}{n^{(m)}}}.$$

On the other hand, the definition $f_{\mathbf{z}}$ tells us that

$$\begin{aligned} & E_{\rho^n}(E_2) \\ &= E_{\rho^n}\left(\inf_{f\in\mathcal{M}_h}\left\{\mathcal{E}_{\mathbf{z}}(f) - \mathcal{E}_{\mathbf{z}}(f_\rho) + \gamma\sum_{j=1}^{d}\left\|f^{(j)}\right\|_{\ell_2}^2\right\}\right) \\ &\leq \inf_{f\in\mathcal{M}_h}\left\{E_{\rho^n}(\mathcal{E}_{\mathbf{z}}(f) - \mathcal{E}_{\mathbf{z}}(f_\rho)) + \gamma\sum_{j=1}^{d}\left\|f^{(j)}\right\|_{\ell_2}^2\right\} \\ &\leq \inf_{f\in\mathcal{M}_h}\left\{\int_{\mathcal{X}}(f(x) - f_\rho(x))^2 d\rho_{\mathcal{X}}(x) + \gamma\sum_{j=1}^{d}\left\|f^{(j)}\right\|_{\ell_2}^2\right\}. \end{aligned} \tag{11}$$

Combining Equations (10) and (11), we get the desired result in Theorem 1. □

5 Experimental results

In this section, we conduct experiments on the ADNI cohort. The goal of QTL identification is to predict brain imaging features given the SNP data. Meanwhile, we expect the model to show the importance of different SNPs, which is fundamental to understanding the role of each genetic variant in Alzheimer's disease.

5.1 Data description

The data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). One of the goals of ADNI is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For the latest information, see www.adni-info.org. The genotype data (Saykin *et al.*, 2010) for all non-Hispanic Caucasian participants from the ADNI Phase 1 cohort were used here. They were genotyped using the Human 610-Quad BeadChip. Among all the SNPs, only SNPs within the boundary of $\pm 20K$ base pairs of the 153 AD candidate genes listed on the AlzGene database (www.alzgene.org) as of April 18, 2011 (Bertram *et al.*, 2007), were selected after the standard quality control (QC) and imputation steps. The QC criteria for the SNP data include (i) call rate check per subject and per SNP marker, (ii) gender check, (iii) sibling pair identification, (iv) the Hardy–Weinberg equilibrium test, (v) marker removal by the minor allele frequency and (vi) population stratification. As the second pre-processing step, the QC'ed SNPs were imputed using the MaCH software (Li *et al.*, 2010) to estimate the missing genotypes. As a result, our analyses included 3123 SNPs extracted from 153 genes (boundary: $\pm 20KB$) using the ANNOVAR annotation (http://www.openbioinformatics.org/annovar/).

As described previously, two widely employed automated MRI analysis techniques were used to process and extract imaging phenotypes from scans of ADNI participants (Shen *et al.*, 2010). First, Voxel-Based Morphometry (VBM) (Ashburner and Friston, 2000) was performed to define global gray matter (GM) density maps and extract local GM density values for 90 target regions. Second, automated parcellation via FreeSurfer V4 (Fischl *et al.*, 2002) was conducted to define volumetric and cortical thickness values for 90 regions of interest (ROIs) and to extract total intracranial volume (ICV). Further details are available in (Shen *et al.*, 2010). All these measures were adjusted for the baseline ICV using the regression weights derived from the healthy control (HC) participants. All 749 participants with no missing MRI measurements were included in this study, including 330 AD samples, and 210 MCI samples and 209 health control (HC) samples. In this study, we focus on a subset of these 90 imaging features which are reported to be related with AD. We extract these QTs from roughly matching regions of interest (ROIs) with VBM and FreeSurfer. Please see (Wang *et al.*, 2012b) for details. We select 26 measures for FreeSurfer, 36 measures for VBM and summarize these measures in Tables 1 and 2.

Table 1. Twenty-six volumetric/thickness measures (FreeSurfer) selected from 'matching' ROIs

Volume/Thickness ID	ROI
LHippVol, RHippVol	Volume of hippocampus
LEntCtx, REntCtx	Thickness of entorhinal cortex
LParahipp, RParahipp	Thickness of parahippocampal gyrus
LPrecuneus, RPrecuneus	Thickness of precuneus
LCaudMidFrontal, RCaudMidFrontal	Mean thickness of caudal midfrontal
LRostMidFrontal, RRostMidFrontal	Mean thickness of rostral midfrontal
LSupFrontal, RSupFrontal	Mean thickness of superior frontal
LLatOrbFrontal, RLatOrbFrontal	Mean thickness of lateral orbitofrontal
LMedOrbFrontal, RMedOrbFrontal	Mean thickness of medial orbitofrontal gyri
LFrontalPole, RFrontalPole	Mean thickness of frontal pole
LInfTemporal, RInfTemporal	Mean thickness of inferior temporal
LMidTemporal, RMidTemporal	Mean thickness of middle temporal
LSupTemporal, RSupTemporal	Mean thickness of superior temporal gyri

5.2 Experimental setting

To evaluate the performance of our FNAM model, we compare with the following related methods: **LSR** (Least square regression), **RR** (Ridge regression), **Lasso** (LSR with ℓ_1-norm regularization), **Trace** (LSR with trace norm regularization) and **FNNRW** (Feedforward neural network with random weights), where we consider the Frobenius norm loss in the R_{emp} term of Rahimi and Recht (2009) for regression problem. We add a comparing method, **FNNRW-Linear** (FNNRW using linear activation function), which use linear activation function $\phi(\mathbf{x}) = \mathbf{x}$ to illustrate the contribution of the nonlinearity of activation function.

As for evaluation metric, we calculate root mean square error (RMSE) and correlation coefficient (CorCoe) between the predicted value and ground truth in out-of-sample prediction. We normalize the RMSE value via Frobenius norm of the ground truth matrix. In comparison, we adopt 5-fold cross validation and report the average performance on these five trials for each method.

We tune the hyper-parameter of all models in the range of $\{10^{-4}, 10^{-3.5}, \ldots, 10^{4}\}$ via nested 5-fold cross validation on the training data, and report the best parameter *w.r.t.* RMSE of each method. For methods involving feedforward neural networks, i.e. FNNRW, FNNRW-Linear and FNAM, we set $h = 50$. For FNNRW and FNAM, we set $\phi(.)$ as the tanh function which maps the input to $[-1, 1]$.

5.3 Performance comparison on ADNI cohort

We summarize the RMSE and CorCoe comparison results in Table 3. From the results we notice that FNAM outperforms all the counterparts in both FreeSurfer and VBM. Besides, from the comparison between Lasso, Trace and FNAM, we find that the assumptions imposed by Lasso (assumption of sparse structure) and Trace (low-rank assumption) may not be appropriate when the distribution of the real data does not conform to such assumptions. In contrast, FNAM is more flexible and adaptive since FNAM does not make such structure assumption on the data distribution. Moreover, from the comparison between FNNRW, FNNRW-Linear and FNAM, we find that both FNNRW and FNAM outperform FNNRW-Linear, which demonstrates the importance of the nonlinearity introduced by the activation function. FNNRW-Linear only involves linear functions, thus is not able to show the non-linear influence of QTL. As for FNNRW, we deem that the reason for FNAM to perform better than FNNRW lies in the additive mechanism of FNAM. Since FNNRW incorporates all features in each computation, it seems too complex for the prediction thus brings about high variance.

Table 2. Thirty-six GM density measures (VBM) selected from 'matching' ROIs

GM Density ID	ROI
LHippocampus, RHippocampus	Hippocampus
LParahipp, RParahipp	Parahippocampal gyrus
LPrecuneus, RPrecuneus	Precuneus
LInfFrontal_Oper, RInfFrontal_Oper	Inferior frontal operculum
LInfOrbFrontal, RInfOrbFrontal	Inferior orbital frontal gyrus
LInfFrontal_Triang, RInfFrontal_Triang	Inferior frontal triangularis
LMedOrbFrontal, RMedOrbFrontal	Medial orbital frontal gyrus
LMidFrontal, RMidFrontal	Middle frontal gyrus
LMidOrbFrontal, RMidOrbFrontal	Middle orbital frontal gyrus
LSupFrontal, RSupFrontal	Superior frontal gyrus
LMedSupFrontal, RMedSupFrontal	Medial superior frontal gyrus
LSupOrbFrontal, RSupOrbFrontal	Superior orbital frontal gyrus
LRectus, RRectus	Rectus gyrus
LRolandic_Oper, RRolandic_Oper	Rolandic operculum
LSuppMotorArea, RSuppMotorArea	Supplementary motor area
LInfTemporal, RInfTemporal	Inferior temporal gyrus
LMidTemporal, RMidTemporal	Middle temporal gyrus
LSupTemporal, RSupTemporal	Superior temporal gyrus

5.4 Important SNP discovery

Here, we look into the significant SNPs in the prediction. According to the definition in Equation (9), we calculate the importance of each SNP and select the top 10 SNPs that weigh the most in VBM analysis. We plot the weight map and brain map of the top 10 SNPs in Figure 1.

Table 3. Biomarker 'FreeSurfer' and 'VBM' prediction comparison via RMSE and CorCoe measurement

		FreeSurfer	VBM
RMSE	LSR	0.2579 ± 0.0123	0.1754 ± 0.0045
	RR	0.1837 ± 0.0132	0.1285 ± 0.0036
	Lasso	0.2525 ± 0.0124	0.1276 ± 0.0040
	Trace	0.1971 ± 0.0151	0.1388 ± 0.0048
	FNNRW-Linear	0.2439 ± 0.0191	0.1991 ± 0.0173
	FNNRW	0.2272 ± 0.0220	0.1683 ± 0.0269
	FNAM	**0.1815 ± 0.0133**	**0.1246 ± 0.0028**
CorCoe	LSR	0.9645 ± 0.0032	0.5849 ± 0.0192
	RR	0.9818 ± 0.0028	0.7435 ± 0.0138
	Lasso	0.9659 ± 0.0032	0.7482 ± 0.0144
	Trace	0.9789 ± 0.0032	0.7099 ± 0.0188
	FNNRW-Linear	0.9677 ± 0.0044	0.5122 ± 0.0589
	FNNRW	0.9718 ± 0.0058	0.6037 ± 0.0823
	FNAM	**0.9821 ± 0.0026**	**0.7592 ± 0.0114**

Note: The table shows the average RMSE/CorCoe value and the standard deviation in 5-fold cross validation. Better performance corresponds to lower RMSE value and higher CorCoe value. The best results are marked in bold.

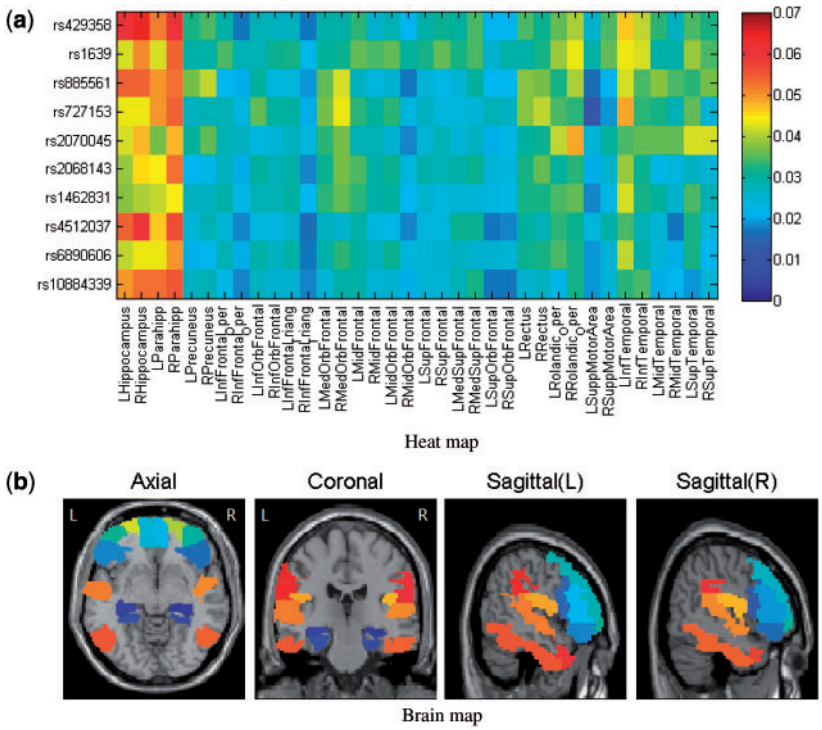

Fig. 1. Heat map and brain map of the top 10 SNPs in VBM analysis. The weight matrix is calculated on the whole VBM data so as to avoid the randomness introduced by fold split. (**a**) Heat map showing the weights calculated via Equation (9) of the top 10 SNPs in the prediction. (**b**) Weight matrix mapped on the brain for the VBM analysis. Different colors are employed to denote different ROIs

From the results, we notice that ApoE-rs429358 ranks the first in our prediction. As the major known genetic risk factor of AD, ApoE has been reported to be related with lowered parietal (Small *et al.*, 2000), temporal (van der Flier *et al.*, 2011) and posterior cingulate cerebral glucose metabolism (Liang *et al.*, 2008) of AD patients. Moreover, we present the LocusZoom plot (Pruim *et al.*, 2010) for the SNPs close to LIPA gene (10M boundary) in Chromosome 10 to show the AD-associated region around LIPA-rs885561 in Figure 2. Similar to ApoE, LIPA gene is also known to be involved in cholesterol metabolism (Papassotiropoulos *et al.*, 2005), where elevated cholesterol levels lead to higher risk of developing AD. In addition, we detect other SNPs that are established AD risk factors, e.g. rs1639-PON2 (Shi *et al.*, 2004) and rs2070045-SORL1 (Rogaeva *et al.*, 2007). Replication of these results demonstrate the validity of our model.

We also pick out SNPs with potential risks whose influence on AD has not been clearly revealed in literature. For example, rs727153-LRAT is known to be related with several visual diseases, including early-onset severe retinal dystrophy and Leber congenital amaurosis 14 (Perrault *et al.*, 2004). LRAT catalyzes the esterification of all-trans-retinol into all-trans-retinyl ester, which is essential for vitamin A metabolism in the visual system (Gollapalli and Rando, 2003). Clinically, vitamin A have been demonstrated to slow the progression of dementia and there are reports showing an trend of lower vitamin A level in AD patients (Ono *et al.*, 2004). Thus, it would interesting to look into the molecular role of LRAT in the progression of AD in future study. Such findings may provide insights into the discovery of new AD-associated genetic variations as well as the prevention and therapy of this disease.

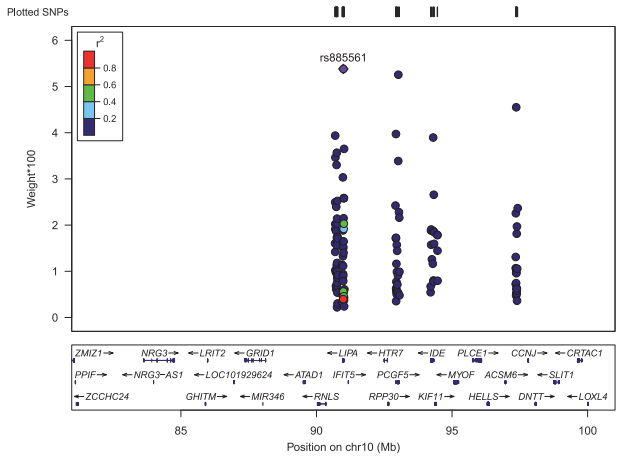

Fig. 2. LocusZoom plot showing AD associated region around rs885561-LIPA (10M boundary) in Chromosome 10. A total of 132 SNPs are plotted. The horizontal axis displays the chromosomal position and gene name of the SNPs. Positions of exons are presented, with an arrow indicating the transcribed strand. The vertical axis shows the weights of these 132 SNPs in the prediction of left hippocampus for VBM analysis. Node color indicates the local linkage disqeuilibrium (LD), i.e. r^2 value between a certain SNP and rs885561 in LIPA gene

5.5 Performance with varying hidden node number

In Algorithm 1, we need to predefine the number of hidden nodes h, thus it is crucial to test if the performance of FNAM is stable with different h. In this section, we analyze the stability of FNAM model *w.r.t.* the choice of hidden node number. Figure 3 display the RMSE and CorCoe comparison results of FNAM when h is set in the range of $\{10, 20, \ldots, 100\}$. From these results, we can find that our FNAM model performs quite stable *w.r.t.* the choice of hidden node number. As a consequence, we do not need to make much effort on tuning the number of hidden nodes. This is important to an efficient implementation in practice.

5.6 Running time analysis

Here, we present experimental results to analyze the runtime (in seconds) of FNAM with different number of hidden nodes. Our experiments are conducted on a 24-core Intel(R) Xeon(R) E5-2620 v3 CPU @ 2.40 GHz server with 65GB memory. The operating system is Ubuntu 16.04.1 and the software we use is Matlab R2016a (64-bit) 9.0.0. Seen from Figure 4, it only takes a few minutes to run our model on the ADNI data. The running time is roughly linear to the number of hidden nodes, which is consistent with our theoretical analysis that the time complexity of FNAM is $O(ndhc)$.

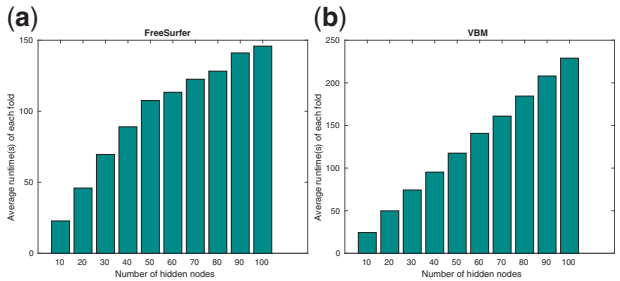

Fig. 4. Runtime (in seconds) comparison of FNAM using different number of hidden nodes. Y-axis shows the average runtime of 1-fold in the cross validation, including the time for tuning hyperparameter γ as well as the time for obtaining the prediction results

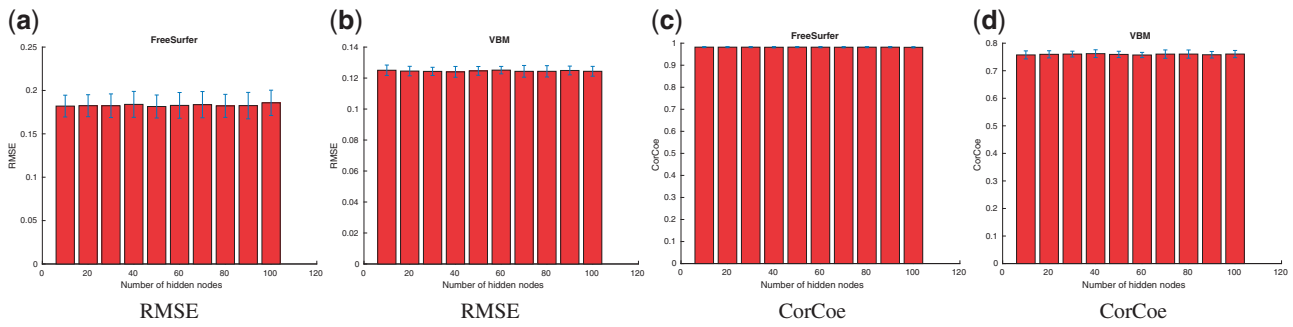

Fig. 3. RMSE and CorCoe comparison of FNAM when the number of hidden nodes is set as $\{10, 20, \ldots, 100\}$, respectively. We add an error bar (in blue color) to show the standard deviation in 5-fold cross validation

6 Conclusion

A novel additive model, called FNAM, was proposed for QTL identification, which can be easily adapted to depict the non-linear associations between SNPs and brain endophenotypes. The experimental results on the ADNI cohort indicated the promising performance of FNAM. In particular, we not only identified some SNPs validated in the previous literature, but also found new SNPs with potential risk for Alzheimer's. These empirical studies validate the effectiveness of our approach, and provide insights into the genetic causal relationships as well as early detection of neurological disorders. We also derived the generalization error bound of FNAM under a general assumption, i.e. *m*-dependent observations, thus is suitable to many other biological applications.

Acknowledgements

This work was supported by the National Science Foundation [IIS 1302675, IIS 1344152, DBI 1356628, IIS 1619308, IIS 1633753 to H.H.] at PITT and [IIS-1117335 to L.S.] at UPENN; and by the National Institutes of Health [R01 LM011360 to L.S. and A.S., U01 AG024904 to Michael Weiner and A.S., RC2 AG036535 to Michael Weiner and A.S., R01 AG19771 to A.S., P30 AG10133 to A.S.] and [R01 AG049371 to H.H.] at PITT. H.C. was partially supported by the National Natural Science Foundation of China 11671161. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). Detailed ADNI Acknowledgements information is available in http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Manuscript_Citations.pdf.

Conflict of Interest: none declared.

References

Ashburner,J. and Friston,K.J. (2000) Voxel-based morphometry—the methods. *Neuroimage*, **11**, 805–821.

Avramopoulos,D. (2009) Genetics of alzheimer's disease: recent advances. *Genome Med.*, **1**, 34.

Bertram,L. *et al*. (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat. Genet.*, **39**, 17–23.

Cooper-Knock,J. *et al*. (2014) The widening spectrum of c9orf72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype. *Acta Neuropathol.*, **127**, 333–345.

Cucker,F. and Smale,S. (2001) On the mathematical foundations of learning. *Bull. Am. Soc.*, **39**, 1–49.

Cucker,F. and Zhou,D.X. (2007) *Learning Theory: An Approximation Theory Viewpoint*, **Vol. 24**. Cambridge University Press.

Fazel,M. (2002) *Matrix rank minimization with applications*. Ph.D. Thesis, Electrical Engineering Department, Stanford University.

Fischl,B. *et al*. (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, **33**, 341–355.

Gollapalli,D.R. and Rando,R.R. (2003) All-trans-retinyl esters are the substrates for isomerization in the vertebrate visual cycle. *Biochemistry*, **42**, 5809–5818.

Huo,Z. *et al*. (2018) Genotype-phenotype association study via new multi-task learning model. In *Twenty-Third Pacific Symposium on Biocomputing (PSB 2018)*, pp. 353–364. World Scientific Publishing Company.

Igelnik,B. and Pao,Y.-H. (1995) Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE/ACM Trans. Netw.*, **6**, 1320–1329.

Leek,J.T. *et al*. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.*, **11**, 733–739.

Li,Y. *et al*. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.

Liang,W.S. *et al*. (2008) Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc. Natl. Acad. Sci. USA*, **105**, 4441–4446.

Meyer-Lindenberg,A. *et al*. (2006) Impact of complex genetic variation in comt on human brain function. *Mol. Psychiatry*, **11**, 867.

Modha,D.S. and Masry,E. (1996) Minimum complexity regression estimation with weakly dependent observations. *IEEE Trans. Inf. Theory*, **42**, 2133–2145.

Ono,K. *et al*. (2004) Vitamin a exhibits potent antiamyloidogenic and fibril-destabilizing effects in vitro. *Exp. Neurol.*, **189**, 380–392.

Papassotiropoulos,A. *et al*. (2005) A cluster of cholesterol-related genes confers susceptibility for alzheimer's disease. *J. Clin. Psychiatry*, **66**, 940–947.

Peper,J.S. *et al*. (2007) Genetic influences on human brain structure: a review of brain imaging studies in twins. *Hum. Brain Mapp.*, **28**, 464–473.

Perrault,I. *et al*. (2004) Retinal dehydrogenase 12 (rdh12) mutations in leber congenital amaurosis. *Am. J. Hum. Genet.*, **75**, 639–646.

Pruim,R.J. *et al*. (2010) Locuszoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.

Rahimi,A. and Recht,B. (2009) Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. In *Neural Information Processing Systems (NIPS)*, pp. 1313–1320.

Rogaeva,E. *et al*. (2007) The neuronal sortilin-related receptor sorl1 is genetically associated with alzheimer disease. *Nat. Genet.*, **39**, 168–177.

Ryan,N.S. *et al*. (2016) Clinical phenotype and genetic associations in autosomal dominant familial alzheimer's disease: a case series. *Lancet Neurol.*, **15**, 1326–1335.

Saykin,A.J. *et al*. (2010) Alzheimer's disease neuroimaging initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimers Dement.*, **6**, 265–273.

Schmidt,W.F. *et al*. (1992) Feedforward neural networks with random weights. In *Proc. 11th IAPR Int. Conf. Pattern Recognition Methodology Systems*. pp. 1–4. IEEE.

Shen,L. *et al*. (2010) Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. *Neuroimage*, **53**, 1051–1063.

Shi,J. *et al*. (2004) Possible association between cys311ser polymorphism of paraoxonase 2 gene and late-onset alzheimer's disease in chinese. *Brain Res. Mol. Brain Res.*, **120**, 201–204.

Small,G.W. *et al*. (2000) Cerebral metabolic and cognitive decline in persons at genetic risk for Alzheimer's disease. *Proc. Natl. Acad. Sci. USA*, **97**, 6037–6042.

Tibshirani,R. (1996) Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, **58**, 267–288.

van der Flier,W.M. *et al*. (2011) Early-onset versus late-onset alzheimer's disease: the case of the missing apoe ϵ4 allele. *Lancet. Neurol.*, **10**, 280–288.

Vidyasagar,M. (2013) *Learning and Generalisation: With Applications to Neural Networks*. Springer-Verlag, New York.

Vounou,M. *et al*. (2010) Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *Neuroimage*, **53**, 1147–1159.

Wang,H. *et al*. (2012) From phenotype to genotype: an association study of longitudinal phenotypic markers to alzheimer's disease relevant snps. *Bioinformatics*, **28**, i619–i625.

Wang,H. *et al*. (2012) Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adni cohort. *Bioinformatics*, **28**, 229–237.

Wang,X. *et al*. (2017) Longitudinal genotype-phenotype association study via temporal structure auto-learning predictive model. In *The 21st Annual International Conference on Research in Computational Molecular Biology (RECOMB 2017)*, pp. 287–302. Springer International Publishing.

Waring,S.C. and Rosenberg,R.N. (2008) Genome-wide association studies in alzheimer disease. *Arch. Neurol.*, **65**, 329–334.

Weiner,M.W. *et al*. (2013) The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dement.*, **9**, e111–e194.

Yang,T. *et al*. (2015) Detecting genetic risk factors for Alzheimer's disease in whole genome sequence data via lasso screening. In *Proc. IEEE Int. Symp. Biomed. Imaging*. IEEE, pp. 985–989.

Zhu,X. *et al*. (2016) Structured spare low-rank regression model for brain-wide and genome-wide associations. In *18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2016)*, pp. 344–352. Springer International Publishing

Zou,B. *et al*. (2009) The generalization performance of erm algorithm with strongly mixing observations. *Mach. Learn.*, **75**, 275–295.