Medical Image Analysis 73 (2021) 102189

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Deep-gated recurrent unit and diet network-based genome-wide association analysis for detecting the biomarkers of Alzheimer's disease

Meiyan Huang[a,b,c,*], Haoran Lai[a], Yuwei Yu[a], Xiumei Chen[a], Tao Wang[a], Qianjin Feng[a,b,c,*], The Alzheimer's Disease Neuroimaging Initiative

[a] School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[b] Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou 510515, China
[c] Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University, Guangzhou 510515, China

ARTICLE INFO

Article history:
Received 28 January 2021
Revised 30 May 2021
Accepted 16 July 2021
Available online 24 July 2021

Keywords:
Alzheimer's disease
Biomarker detection
Deep learning
Genome-wide association analysis

ABSTRACT

Genome-wide association analysis (GWAS) is a commonly used method to detect the potential biomarkers of Alzheimer's disease (AD). Most existing GWAS methods entail a high computational cost, disregard correlations among imaging data and correlations among genetic data, and ignore various associations between longitudinal imaging and genetic data. A novel GWAS method was proposed to identify potential AD biomarkers and address these problems. A network based on a gated recurrent unit was applied without imputing incomplete longitudinal imaging data to integrate the longitudinal data of variable lengths and extract an image representation. In this study, a modified diet network that can considerably reduce the number of parameters in the genetic network was proposed to perform GWAS between image representation and genetic data. Genetic representation can be extracted in this way. A link between genetic representation and AD was established to detect potential AD biomarkers. The proposed method was tested on a set of simulated data and a real AD dataset. Results of the simulated data showed that the proposed method can accurately detect relevant biomarkers. Moreover, the results of real AD dataset showed that the proposed method can detect some new risk-related genes of AD. Based on previous reports, no research has incorporated a deep-learning model into a GWAS framework to investigate the potential information on super-high-dimensional genetic data and longitudinal imaging data and create a link between imaging genetics and AD for detecting potential AD biomarkers. Therefore, the proposed method may provide new insights into the underlying pathological mechanism of AD.

1. Introduction

As a common neurodegenerative disease, Alzheimer's disease (AD), which is characterised by memory loss, cognitive impairment, and personality changes, often appears in people aged > 65 years (Assoc, 2018). Although no effective cure for AD has been developed, studies have demonstrated that the early detection of AD may improve its treatment and slow down its progression (Huang et al., 2019c). Therefore, AD-related potential biomarkers should be detected. The risk of AD is affected by genetic variants an individual carries, and it can be measured accurately from birth (Lambert et al., 2013). Hence, detecting AD-associated genetic biomarkers may uncover the possible biological mechanisms and lead to early prediction and diagnosis of AD.

Genome-wide association analysis (GWAS) can be applied to detect the potential biomarkers of diseases by combining multiple phenotypic variables and whole genomes. Therefore, disease status (i.e., diagnostic labels, such as AD and normal control [NC] in this study) can be used as phenotype to detect disease-related genetic biomarkers for biological interpretation of the disease. However, genetic data do not explicitly encode disease status, which indicates that directly associating genetic data with disease status might lead to inaccurate or incorrect results (Bi et al., 2017). Moreover, brain imaging data, such as MRI, can contribute to the understanding of AD-related neural changes (Huang et al., 2019a; Huang et al., 2019c; Ning et al., 2018; Zhuo et al., 2018). Hence, imaging data can be used as endophenotypes in GWAS to anal-

* Corresponding authors at: School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.
E-mail addresses: huangmeiyan16@163.com (M. Huang), haoranlai@163.com (H. Lai), 2523750358@qq.com (Y. Yu), chenxiumei97@163.com (X. Chen), wangtao_9802@163.com (T. Wang), fengqj99@smu.edu.cn (Q. Feng).

https://doi.org/10.1016/j.media.2021.102189
1361-8415/

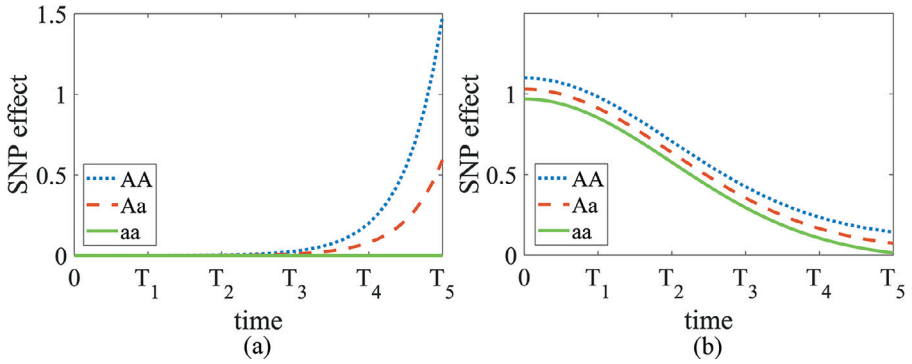

Fig. 1. Genome-wide association analysis has a greater detection power when longitudinal data are used. (a) and (b) are two examples of time-dependent patterns of the effect of SNPs on the phenotype, which are difficult to detect using the data from a single time point.

yse the associations between imaging and genetic data and bridge the gap between genetic data and disease. To date, GWAS has been used to detect the potential biomarkers of diseases by analysing the associations between imaging and genetic data (e.g., single nucleotide polymorphism [SNP]) (Huang et al., 2017; Huang et al., 2019a; Huang et al., 2015; Huang et al., 2019b). In contrast to methods involving candidate phenotypes or genotypes, GWAS does not require prior knowledge of disease pathology to select candidate phenotypes and genotypes; consequently, GWAS can reduce the probability of missing important genes and brain clusters (Huang et al., 2015). However, several problems in GWAS should be considered.

First, running GWAS poses remarkable computational challenges because it usually runs genome-wide associations ($N_G \sim 10^6$known variants) with signals at tens (prespecified regions) or millions (voxels) of locations in the brain (Huang et al., 2017; Huang et al., 2019a; Huang et al., 2015; Huang et al., 2019c). Consequently, traditional GWAS methods prefer to model imaging data as a linear function of genetic data to reduce computational costs. However, the relationship between imaging and genetic data is so complex, in which formulating it via a simple linear function is difficult (Hu et al., 2019; Zhou et al., 2019a). Thus, using a nonlinear function to describe the relationship between imaging and genetic data may improve the detection accuracy of AD biomarkers in GWAS.

Second, traditional GWAS is performed using single-locus models (i.e., independent testing of the association between each SNP and a given phenotype) (Guo et al., 2019). However, complex traits are often collectively controlled by multiple SNPs (Dudbridge, 2016; Li et al., 2015). Therefore, apart from the association between imaging and genetic data, correlations among imaging data and correlations among genetic data should be considered. Multivariate sparse regression methods have been introduced to achieve this goal. These regression methods regress either a set of SNPs for a few candidate phenotypes or a set of phenotypes for a few candidate SNPs; thus, the relationship among phenotypes or SNPs is considered via regularization (Zhou et al., 2019a; Zhu et al., 2017; Zhu et al., 2018). However, these methods are usually performed on a small set of SNP data (~10^3) because of complex calculation.

Third, traditional GWAS methods involve the use of phenotypes (imaging data in this study) from a single time point (e.g., baseline visit). Therefore, various associations between genetic and imaging data may be missed, and the power of biomarker detection for GWAS decreases (Marchetti-Bowick et al., 2016a). Fig. 1 illustrates two hypothetical patterns of the effect of SNP on the phenotypes. An association between a phenotype and an SNP may exist if the three SNP genotypes (denoted as AA, Aa, and aa) have differential effects on the phenotype in traditional GWAS. In the first case (Fig. 1 -a-), the effects of the three SNP genotypes only differ in $t \in [T_3, T_5]$ time interval. If a single time point is arbitrarily selected, then the effective association may be missed, whereas this association may be detected using longitudinal data that consider data from varied time points. In the second case (Fig. 1 b), the difference between the effects of the three genotypes in an SNP is small but consistent over time; the effects may be too weak to be interpreted as a remarkable association by using data from a single time point. However, this weak signal may strengthen once the data from the entire time series are considered. In particular, AD is a progressive neurodegenerative disease. The brain structure, brain function, and disease diagnosis change over time. For example, a NC subject may progress to mild cognitive impairment (MCI) and to AD several years later. On the contrary, some patients with MCI can never progress to AD or revert to NC. Thus, longitudinal imaging data should be considered to discover the effect of genetic data on the trajectory of disease progression and enhance the detection power of potential biomarkers.

Fourth, existing GWAS methods directly involve the use of imaging data as endophenotypes to identify endophenotype-associated genetic biomarkers; this step further requires to confirm or refute a suggestive link based on imaging endophenotypes because the identified genetic biomarkers may or may not be associated with the disease (e.g., AD in this study) (Xu et al., 2017). Therefore, a direct link between genetic data and AD should be constructed to improve the accuracy of detecting AD potential biomarkers.

Deep-learning methods have been widely proposed to detect potentially useful information and achieve interesting results in AD-related research, such as the prediction and classification of AD (Ghazi et al., 2019; Lee et al., 2019; Lin et al., 2018; Lu et al., 2018; Ning et al., 2018). On the one hand, deep-learning methods bridge inputs and outputs with nonlinear mapping to approximate a highly accurate and complex relationship between inputs and outputs. On the other hand, correlations among inputs are examined through inner product and nonlinear mapping via deep-learning methods. Therefore, deep-learning methods can be incorporated into a GWAS framework to accurately model the following relationships: among imaging data, among genetic data, and between imaging and genetic data. Li et al. proposed a deep canonically correlated sparse autoencoder to analyse the associations between brain imaging and genetic data (Li et al., 2020). This method first combined deep canonical correlation analysis and sparse autoencoder for extraction of imaging and genetic features, which can maximize the canonical correlation between the two extracted features while minimizing the reconstruction errors of the sparse autoencoder. Then, a classifier, such as a support vector machine (SVM), was applied on the extracted features to classify patients with schizophrenia and healthy controls (Li et al., 2020). Although

good classification performance is achieved by using the method proposed in Li et al. (2020), the feature extraction and classification steps are separated, whose extracted genetic features are more related to imaging features than the disease. Therefore, the biological interpretation of this method related to the disease is unexploited (Li et al., 2020). Recently, Hu et al. proposed a Grad-CAM-guided convolutional collaborative learning to analyse the associations between brain imaging and genetic data and to perform automated diagnosis in a whole network (Hu et al., 2021). The results can be well interpreted by using this method because the imaging and genetic data are related to the disease status (Hu et al., 2021). However, this method is performed on a small set of preselected SNP data (only 750 SNPs were used), which may miss some disease-related SNPs. Moreover, the methods proposed in Hu et al. (2021) and Li et al. (2020) are performed on imaging data from a single time point; thus, the effect of genetic data on the trajectory of disease progression contained in the longitudinal imaging data may be lost.

Although existing deep-learning-based imaging genetic studies can be applied to model the complex associations between imaging and genetic data, the computational challenges imposed by super-high-dimensional genetic data and the information loss caused by using imaging data from a single time point remain unsolved. With regard to computational challenges, if a fully connected layer with 100 hidden units is used after the input layer, then approximately 10^8 (number of SNPs $\times$ number of hidden units in the fully connected layer: $\sim 10^6 \times 100 \approx 10^8$) network parameters exist, and the number of parameters increases considerably as the number of fully connected layers increases. In addition to the high dimensionality of SNP data, sequence–context information is unavailable in SNP (values with 0, 1, or 2), and convolutional networks are unsuitable for SNP data in reducing the number of parameters. The goals are as follows: to introduce a modified diet network (Romero et al., 2017), reduce the number of network parameters, accelerate GWAS calculation, and address this problem. Moreover, a gated recurrent unit (GRU) is applied to integrate longitudinal imaging data and discover various associations between genetic and imaging data.

In this study, a novel GWAS method, namely, deep GRU and diet Network-based GWAS (dGDN-GWAS), is proposed to detect potential AD genetic biomarkers. Our contributions are as follows. First, a novel deep-learning framework is incorporated into the GWAS method to construct a nonlinear relationship between imaging and genetic data and consider the correlations among imaging data and correlations among genetic data. Second, a modified diet network is applied to reduce the number of parameters in the genetic network and effectively perform GWAS. Third, considering the variable lengths of the longitudinal imaging data, GRU is used to integrate these data and determine various associations between genetic and imaging data. Lastly, the proposed method provides a novel way to investigate the association between imaging and genetic data while creating a link between imaging genetics and AD. To validate our method, a set of experiments are performed on the simulated data and a real AD dataset. For the experiments of simulated data, the results show that the proposed method can accurately detect relevant biomarkers and achieve high classification accuracy. For the experiments of the real AD dataset, 708 subjects are included, each subject has 501,584 SNPs and a set of longitudinal imaging data. The AD/NC classification accuracy of the 501,584 SNPs using the proposed method is 0.709, which is comparable to that of some relevant methods with SNPs to classify subjects into AD and NC. Moreover, some previously reported AD-associated genes and some novel AD-related risk genes are detected by using the proposed method. Based on previous studies, no research has incorporated a deep-learning model into the GWAS framework to investigate the potential information in super-high-dimensional genetic and longitudinal imaging data and establish a link between imaging genetics and AD for the detection of potential AD biomarkers. Therefore, the proposed method may provide insights into the underlying pathological mechanism of AD.

2. Methods

2.1. Overview of the proposed method

A novel dGDN-GWAS method was proposed for potential AD biomarker detection. Fig. 2 presents an overview of the proposed dGDN-GWAS method that primarily consists of three steps. In the first step, the potentially integrated information of the longitudinal imaging data for image representation was extracted by using a GRU-based network (Fig. 2 a). Apparently, the original imaging measurements might not be enough to characterise the disease well, such as region of interests (ROIs) and image voxels, although the imaging measurements were more informative with regard to the disease than genetic data. Therefore, image representations, which contained rich information of longitudinal imaging data and disease status, were extracted using a GRU-based classification network. In the second step, the modified diet network was used to extract the potential genetic information and examine the association between the imaging and genetic data (Fig. 2 b). Using this step, the extracted genetic representations were closely related to the image representations, which indicated that the gap between genetic and imaging data was shortened. However, the gap between the genetic representations and AD was still observed, and a direct association analysis between them was required in bridging the gap and detecting AD-related biomarkers. Therefore, in the third step, a link between imaging genetics and AD for potential AD biomarker detection was established (Fig. 2 c).

2.2. Mathematical formulation

n independent subjects were considered, and each subject was associated with a set of imaging data and genetic markers. R is the whole-brain region that contains N_R image features; r is the image feature in R ($r \in R$); G is the set of genetic loci containing N_G SNPs, and g is the locus in G ($g \in G$). For each individual i ($i = 1, \ldots, n$), the following parameters were considered: an $N_G \times 1$ vector of genetic data denoted by $\boldsymbol{X}_i = \{x_i(g)\}_{i=1,\ldots,n}$ and an $N_R \times 1$ vector of image measurement denoted by $\boldsymbol{Y}_{it} = \{y_{it}(r)\}_{i=1,\ldots,n;t=1,\ldots,N_T^i}$, where N_T^i is the number of time points for the ith subject. The main notations used in this study are listed in Table 1.

2.3. Integration of longitudinal imaging data with GRU

A recurrent neural network (RNN) was composed of multiple recurring processing layers to obtain a representation of sequential data (LeCun et al., 2015), and variable lengths of time-series data can be processed using the RNN (Lee et al., 2019). Given the variable number of time points for different subjects in this study, this advantage was valuable in longitudinal data processing. GRU, a widely used type of RNN, was introduced to allow each recurrent unit to adaptively capture the dependencies of different time scales; it can also be used in gradient vanishing found in the RNN (Cho et al., 2014). Therefore, GRU was applied to integrate the longitudinal imaging data in this study.

Integrating the longitudinal imaging data for image representation was illustrated by using GRU. GRU had reset and update gates to modulate the flow of information inside the unit. The parameters of these two gates were studied to determine how $\boldsymbol{Y}_t$ was processed (Eqs. (1)–(4)). The update gate z_t was computed using Eq. (1), which determined the amount of information from the previous hidden state s_{t-1} that was carried over to the current hidden

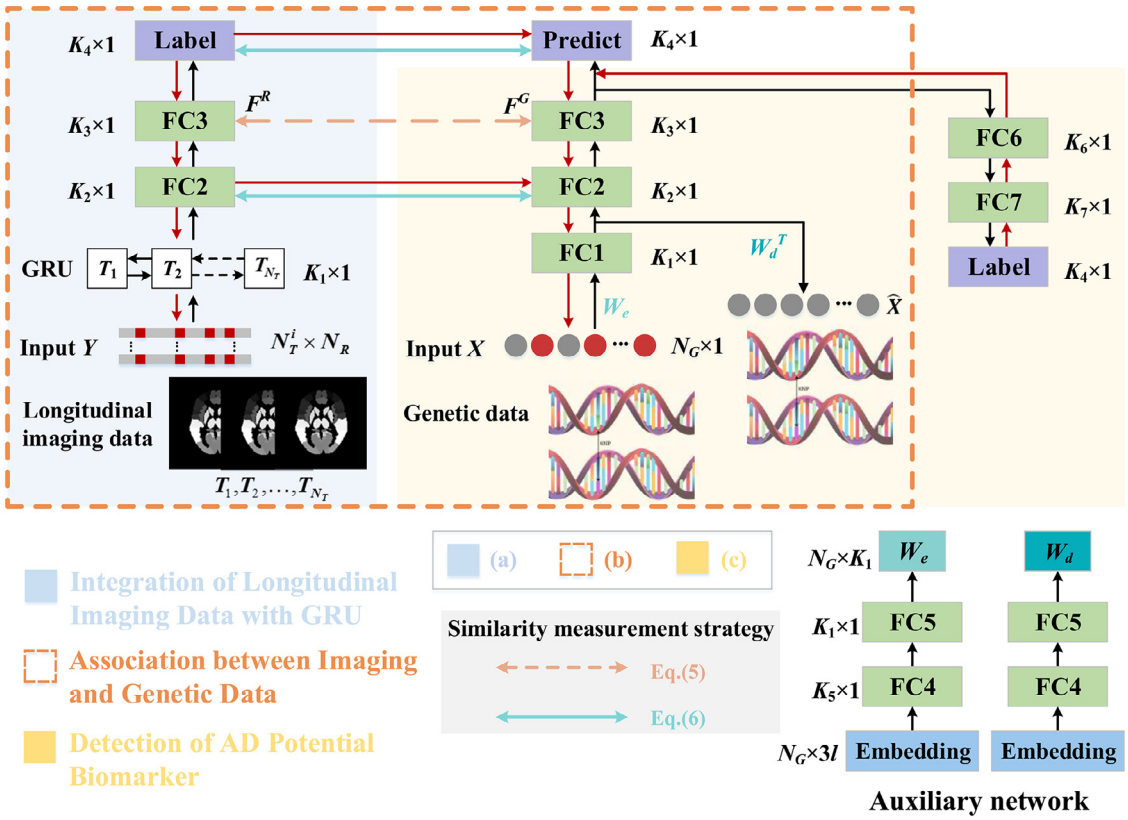

Fig. 2. Overview of the proposed dGDN-GWAS method, where the symbol "$p \times q$" next to the layer represents the output size of the layer, and "FC" denotes the fully connected layer. (a) The potentially integrated information of the longitudinal imaging data is extracted for image representation by using the GRU-based network, where the dashed black arrows within the GRU layer represents abbreviatory GRU cells. (b) The modified diet network is used to extract the potential genetic information and examine the association between imaging and genetic data, in which two strategies are used to analyse the association between imaging and genetic data. (c) A link between imaging genetics and AD is established for potential AD biomarker detection. The red arrows, blocks, and circles indicate the path of partial derivatives and the potential AD-related imaging and genetic biomarkers, respectively.

Table 1
Main notations used in this study.

Symbol	Size	Description	Symbol	Size	Description
$\boldsymbol{X}_i$	$N_G \times 1$	Genetic data with N_G loci for the ith subject	$\boldsymbol{x}(g)$	$n \times 1$	The gth locus for n subjects
$\boldsymbol{Y}_{it}$	$N_R \times 1$	Imaging data with N_R image features in the tth time point for the ith subject	$\boldsymbol{y}_t(r)$	$n \times 1$	The rth image feature at the tth time point for n subjects
$c(g)$/ $c(r)$	–	Contribution score of the gth locus/the rth image feature	$c(g, r)$	$N_G \times N_R$	Association score between the gth SNP and rth ROI
$\boldsymbol{H}_j^R$/ $\boldsymbol{H}_j^G$/ $\boldsymbol{H}_j^A$/ $\boldsymbol{H}_j^C$	$K_j \times 1$	Outputs of the jth layer for the GRU-based network/the main/auxiliary networks of the modified diet network/the network that links the imaging genetics and the diagnostic results of AD	$h_j(r)$/ $h_j(g)$	$K_j \times 1$	The derivatives of units in the jth layer of the GRU-based network/ the modified diet network with respect to the inputs
f_i^R	$K_3 \times 1$	Image representation obtained by using the GRU-based network for the ith subject	f_i^G	$K_3 \times 1$	Genetic representation obtained by using the modified diet network for the ith subject
W, U	–	Weights	$z(\cdot)$	–	Nonlinear mapping function
K_j	–	Unit number of the jth layer	$\lambda_1, \lambda_2, \lambda_3$	–	Tunable coefficients
$\boldsymbol{V}$/$\boldsymbol{M}$	$q \times t$/ $p \times 1$	Sparse matrix used to generate the longitudinal imaging/genetic data	n_q/n_p	–	Number of nonzero values for the simulated imaging/ genetic data

state s_t. For example, if z_t was computed as 1 by using Eq. (1), then only the previous information was reserved, whereas the newly computed hidden value $\tilde{s}_t$ was disregarded (Eq. (4)). Moreover, the reset gate r_t was computed using Eq. (2), which determined the amount of information from the previous hidden state s_{t-1} that should be forgotten. When r_t was close to 0, the reset gate effectively forced the unit to act as if it is reading the first symbol of an input, allowing it to forget the previously computed state (Eq. (3)).

$$z_t = \sigma\left(W_z Y_t + U_z s_{t-1}\right) \tag{1}$$

$$r_t = \sigma\left(W_r Y_t + U_r s_{t-1}\right) \tag{2}$$

$$\tilde{s}_t = \tanh\left(W Y_t + U\left(r_t \odot s_{t-1}\right)\right) \tag{3}$$

$$s_t = z_t s_{t-1} + (1 - z_t)\tilde{s}_t \tag{4}$$

where $\odot$ is element-wise multiplication.

A GRU-based network was constructed to integrate the longitudinal imaging data. This network consisted of four layers, where the first layer was a single-layer GRU with K_1 units; the following two hidden layers were fully connected layers with K_2 and K_3 units, followed by a rectified linear unit (ReLU) activation, and the final layer was a sigmoid layer to calculate class-membership probabilities. Therefore, the longitudinal imaging data were fed into the network to integrate the longitudinal information and were nonlinearly mapped several times to produce the diagnostic probabilities of AD. After the training of the network, the weights of the network were fixed, and the output vectors of the second-last layer of the network were extracted to obtain the potentially integrated information of the longitudinal imaging data for image representation $\boldsymbol{F}^R = \{f_i^R\} \in \mathbb{R}^{n \times K_3}$. AD and NC subjects were used rather than MCI subjects to train this network because MCI can be regarded as an intermediate stage between AD and NC, and it possessed the characteristics of AD and NC (Liu et al., 2015). Therefore, the network trained by using AD and NC subjects can build a more robust association with the diagnostic probabilities of AD and provide a better guidance to the following link between imaging genetics and AD.

2.4. Association between imaging and genetic data

Using the extracted image representations, the associations between these image representations and genetic data were considered. However, constructing a fully connected layer to associate genetic data with image representations was difficult because of the limited super-high-dimensional genetic data. A modification of the diet network (Romero et al., 2017) that can considerably reduce the number of network parameters was introduced to extract the potential genetic information in the super-high-dimensional genetic data, investigate the association between the imaging and genetic data, and address this problem. In this section, the original diet network (Romero et al., 2017) was briefly introduced, and then the proposed modified diet network was presented in details.

2.4.1. Diet network

A multi-task architecture was used in the diet network (Fig. 3), which consisted of a classification task and a parameter prediction task, to considerably reduce the parameter number during network training (Romero et al., 2017). The diet network indicated that a distributed representation for each SNP can be learned. Then, the mapping pattern between the learned distributed representation and the parameters in the classification network can be studied, and the large parameters can be obtained without training in the classification network. In fulfilling this idea, an auxiliary network was introduced to predict the input layer parameters (Fig. 3 b) of the classification network. In addition, another auxiliary network was used to predict the reconstruction layer parameters of the classification network (Fig. 3 c). In contrast to the classification network, which used raw SNP data as inputs, feature embeddings were used as inputs for these two auxiliary networks. Moreover, feature embeddings were designed as the per-class histogram of genetic data. For a given SNP, the value of each locus may be 0, 1, and 2, and the per-class proportion of the three values (0/1/2) of all the subjects can be calculated to replace the three values. With l classes, the feature embeddings were a matrix with a size of $N_G \times 3l$ by using this method. In this case, the sample number of the parameter prediction network (two auxiliary networks) was N_G (~10^6), and more samples were allowed to train the parameter prediction network. If the unit number of the first fully connected layer in the classification network was $K_c = 100$, then a fully connected layer was used in each of the auxiliary network, whose unit number was $K_a = 100$. The parameter number between the input layer and the first fully connected layer in the classification network was considered, and it greatly decreased from $N_G \times K_1 (\sim 10^6 \times 10^2 = 10^8)$ to $2 \times 3l \times 10^2 \approx 10^3$ because of the auxiliary networks. Therefore, the super-high-dimensional SNP data may be fed into the fully connected network with minimal memory consumption because of the interaction networks.

2.4.2. Modified diet network

The modified diet network consisted of one main network and two auxiliary networks (Fig. 2 b). The structure of the main network was similar to that of the GRU-based network; however, the first layer of the main network was a fully connected layer with K_1 units. Of the two auxiliary networks, one was used to predict the input layer parameters of the main network (W_e in Fig. 2 b), and the other one was applied to predict the reconstruction layer parameters of the main network (W_d in Fig. 2 b). The architecture of these two auxiliary networks was the same, and each of these networks included two fully connected layers, where the unit number of the first and second layers was K_5 and K_1, respectively. Moreover, each of the two fully connected layers was followed by a ReLU activation. Similar to the original diet network (Romero et al., 2017), feature embeddings, which were represented as the per-class histogram of genetic data, were used as inputs for these two auxiliary networks.

The similarity between image and genetic representations was considered to discover the potential information of the genetic data and investigate the association between the imaging and genetic data. The SNP data can be mapped directly to the space of imaging representation to achieve similarity, and the objective function of the modified diet network can be defined as follows:

$$\frac{1}{2n}\left(\lambda_1 \sum\nolimits_{i=1,\ldots,n} \left\| f_i^R - f_i^G \right\|^2 + (1-\lambda_1) \sum\nolimits_{i=1,\ldots,n} \left\| X_i - \hat{X}_i \right\|^2\right) \tag{5}$$

where $\hat{X}_i$ is the evaluated values of X_i, and $\boldsymbol{F}^G = \{f_i^G\} \in \mathbb{R}^{n \times K_3}$ is the output of the third layer of the main network. The first term of the objective function is used to measure the association of the potential imaging and genetic data; the second term of the objective function is used to measure the reconstruction error of the genetic data, and λ_1 is a tunable coefficient that controls the tradeoff between association and reconstruction errors. However, the distance between imaging and genetic data was large, and simple similarity measurement between imaging and genetic representations may be difficult for the extraction of potential genetic representations. Two correlated views $(\boldsymbol{F}^R, \boldsymbol{F}^G)$ of a target can be nonlinearly mapped into a new space $[z(\boldsymbol{F}^R), z(\boldsymbol{F}^G)]$ on the basis of some previous works (Chen et al., 2020; Hinton et al., 2015), and the distance between $z(\boldsymbol{F}^R)$ and $z(\boldsymbol{F}^G)$ can be shortened to increase similarity between $\boldsymbol{F}^R$ and $\boldsymbol{F}^G$. Therefore, the objective function can

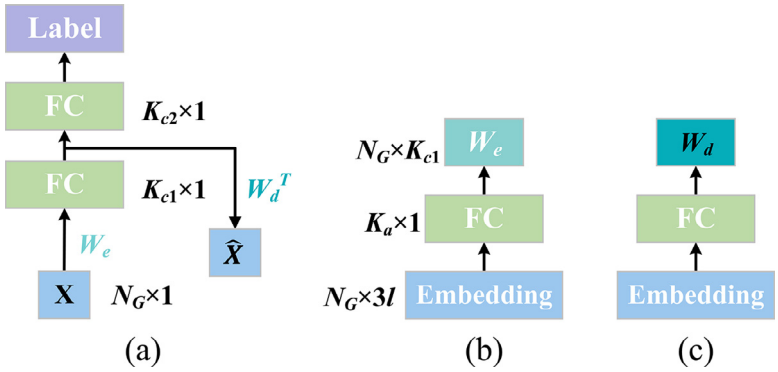

Fig. 3. Network structure of the original diet network. (a) A classification network with a reconstruction term of the input. (b) and (c) are auxiliary networks, which are applied to predict the input layer parameters (W_e) and reconstruction layer parameters (W_d) of the classification network, respectively.

be formulated as follows:

$$\frac{1}{2n}\left(\lambda_2 \sum_{i=1,\ldots,n} \left\|X_i - \hat{X}_i\right\|^2 + \sum_j \lambda_3^j \sum_{i=1,\ldots,n} \left\|z\left(f_i^R\right) - z\left(f_i^G\right)\right\|^2\right) \tag{6}$$

where λ_2 and λ_3^K are the tunable coefficients that control the tradeoff between reconstruction error and similarity loss. After the training of the network, the weights of the network were fixed, and the potential genetic representation $\boldsymbol{F}^G$ can be obtained. AD and NC subjects were used to generate feature embeddings, whereas NC, MCI, and AD subjects are utilized to train the modified diet network. The AD/NC dataset contained more distinct knowledge to discover AD-related biomarkers in the subsequent analyses. Moreover, the whole dataset can alleviate the overfitting caused by high-dimensional SNP data.

2.5. Detection of AD potential biomarker

A link between the imaging genetics and diagnostic results of AD will be created to guide the detection of potential genetic biomarkers associated with AD. Herein, a fully connected network with three layers was constructed to link the imaging genetics and diagnostic results of AD, where the input was the potential genetic representation $\boldsymbol{F}^G$ obtained from the modified diet network; the first and second fully connected layers had K_6 and K_7 units, respectively, followed by a ReLU activation; and the final layer was a sigmoid layer, which could be used to calculate class-membership probabilities. The AD and NC subjects were utilized in this network. Given that this study aimed to detect potential AD biomarkers, the most special characteristics associated with AD were primarily explored. Moreover, the split pattern of the AD and NC subjects in the experiments should be the same in all three networks to avoid overfitting. In particular, the AD and NC subjects in the training, validation, and testing sets in the GRU-based network and modified diet network should be in the network described in this section.

For the whole well-trained network, a strategy was introduced to identify the potential AD biomarkers by using contribution scores and to provide the importance of input features to network outputs. Given that the networks in the second and third steps were multi-layer fully connected networks, the importance of an input feature can be calculated by using partial derivatives (Ning et al., 2018). For the gth locus $\boldsymbol{x}(g) = \{x_i(g)\}_{i=1,\ldots,n}$, the derivative of the predicted probability $\tilde{L}_i$ of a subject i with AD was obtained with regard to $\boldsymbol{x}(g)$, and the absolute value of the derivative among all subjects was averaged. Finally, the contribution score of $\boldsymbol{x}(g)$ can be calculated as $\boldsymbol{c}(g) = E[A(\frac{\partial \tilde{L}_i}{\partial \boldsymbol{x}(g)})]$, where $E[\boldsymbol{.}]$ and $A(\boldsymbol{.})$ represent the functions of average and absolute operations, respectively. $\boldsymbol{x}(g)$ with a high contribution score indicates a close association with AD. Thus, these loci were potential AD-related biomarkers.

Moreover, the contribution of each image feature, such as ROI, in the MRI scans at each time point to the network outputs can be estimated by using the abovementioned strategy. For the rth image feature at the tth time point $\boldsymbol{y}_t(r) = \{y_{it}(r)\}_{i=1,\ldots,n}$, the derivative of the predicted probability $\tilde{L}_i$ of a subject i with AD with regard to $\boldsymbol{y}_t(r)$ was obtained from the GRU-based network, and the absolute value of the derivative among all subjects was averaged. The contribution score of $\boldsymbol{y}_t(r)$ can be calculated as $\boldsymbol{c}_t(r) = E[A(\frac{\partial \tilde{L}_i}{\partial \boldsymbol{y}_t(r)})]$. Finally, the contribution score of each image feature for all time points was denoted as $\boldsymbol{c}(r) = E[\boldsymbol{c}_t(r)]$. Therefore, the high contribution score of an image feature indicated its important role in the network and close association with AD.

2.6. Summary of the proposed method

We have provided a pseudo-code in Algorithm 1 to elucidate the proposed method.

3. Experimental results

The proposed method was evaluated on a set of simulated data and a real AD dataset obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI). Moreover, the proposed method was implemented using Pytorch. Python environment was used for computing on a computer with Ubuntu 16.04, Nvidia GTX GPU, and 12G memory. The details of the experimental setup, such as data partition (training/validation/testing sets) and hyperparameter turning, are provided in Section 3.2.2.

3.1. Simulation study

3.1.1. Generation of simulated data

We generated simulated data according to the following procedure. First, we randomly simulated two sparse matrices: $\boldsymbol{M} \in \mathbb{R}^{p\times 1}$ and $\boldsymbol{V} \in \mathbb{R}^{q\times t}$ (i.e., disease-related SNPs or imaging data were set to nonzero values, otherwise set to zero), where p and q are the feature dimensions for SNP and imaging data with values of 10,000 and 100, respectively. The number of nonzero values for SNP and imaging data was fixed to $n_p = 1000$ and $n_q = 10$, respectively. The number of time point t was set to 2. Then, we generated a latent vector $\boldsymbol{Z}_l \in \mathbb{R}^{n_l\times 1}$ for the lth class, where l was set to 0 or 1; n_l is the subject number, which was set to 300. Based on $\boldsymbol{Z}_l$, we can generate a set of longitudinal imaging data $\boldsymbol{Y} = \{\boldsymbol{Y}_t^l\}_{l=0,1;t=1,2}$, where $\boldsymbol{Y}_t^l \sim N(\boldsymbol{Z}_l\boldsymbol{V}_t, \mathbf{I}_{q\times q})$; $\boldsymbol{V}_t$ is the sparse vector for the tth time point, and $\mathbf{I}_{q\times q}$ is an identity matrix. We generated SNP data $\boldsymbol{X}$ by

Step 1. Integration of Longitudinal Imaging Data with GRU.
Inputs: Longitudinal imaging data $\mathbf{Y} \in \mathbb{R}^{n\times N_T\times N_R}$, labels $\boldsymbol{L} \in \mathbb{R}^{n\times 1}$, unit number of the GRU-based network $\{K_j\}_{j=1,\ldots,4}$
Process: Initialize the parameters of the GRU-based network.
repeat:
1. Sample minibatch of $\{\boldsymbol{Y}_i, L_i\}_{i=1}^{n}$ from $\boldsymbol{Y}$ and $\boldsymbol{L}$, respectively.
2. Calculate $\boldsymbol{H}_1^R$ based on the Eqs. (1)–(4)
3. for $k = 2$: 3
Follow with the fully connected layer with K_j units, output $\boldsymbol{H}_k^R \leftarrow \sigma(\boldsymbol{W}_k^T\boldsymbol{H}_{k-1}^R + \boldsymbol{b}_k)$.
end
4. Calculate classification probabilities $\tilde{\boldsymbol{L}} \leftarrow \sigma(\boldsymbol{W}_4^T\boldsymbol{H}_3^R + \boldsymbol{b}_4)$.
5. Calculate cross-entropy-loss$(\boldsymbol{L}, \tilde{\boldsymbol{L}})$, update parameters.
until Convergence
Outputs: $\boldsymbol{H}_2^R$, $\boldsymbol{H}_4^R$, and image representations $\boldsymbol{F}^R \leftarrow \boldsymbol{H}_3^R$

Step 2. Association between Imaging and Genetic Data.
Inputs: Genetic data $\boldsymbol{X} \in \mathbb{R}^{n\times N_G}$, $\boldsymbol{H}_2^R \in \mathbb{R}^{n\times K_2}$, $\boldsymbol{H}_4^R \in \mathbb{R}^{n\times K_4}$, feature embeddings of two auxiliary networks $\boldsymbol{E} \in \mathbb{R}^{N_G\times 3l}$, unit number of auxiliary network $\{K_j\}_{j=1,5}$, and unit number of main network $\{K_j\}_{j=1,\ldots,4}$
Process: Initialize the parameters of the modified diet network.
repeat:
1. Sample minibatch of $\{\mathbf{H}_{2i}^R, \mathbf{H}_{4i}^R, \mathbf{X}_i\}_{i=1}^{n}$ from $\boldsymbol{H}_2^R$, $\boldsymbol{H}_4^R$, and $\boldsymbol{X}$, respectively.
2. Input $\boldsymbol{E}$ to two auxiliary networks, output $W_e \leftarrow \boldsymbol{H}_2^{A_1}$ and $W_d \leftarrow \boldsymbol{H}_2^{A_2}$, respectively.
3. Calculate $\boldsymbol{H}_1^G \leftarrow \sigma(\boldsymbol{X}_i W_e)$ as outputs of the first layer in the main network.
4. for$k = 2$: 4
Follow with the fully connected layer with K_j units, output $\boldsymbol{H}_k^G \leftarrow \sigma(\boldsymbol{W}_k^T\boldsymbol{H}_{k-1}^G + \boldsymbol{b}_k)$
end.
5. Calculate the reconstruction error in the main network $\hat{\boldsymbol{X}}_i \leftarrow \sigma(\boldsymbol{H}_1^G W_d^T)$.
6. Calculate total loss by using Eq. (6), update parameters.
Until Convergence
Outputs: Potential genetic representations $\boldsymbol{F}^G \leftarrow \boldsymbol{H}_3^G$

Step 3(a). Establishing a link between imaging genetics and AD.
Inputs: Genetic representations $\boldsymbol{F}^G \in \mathbb{R}^{n\times K_3}$, labels $\boldsymbol{L} \in \mathbb{R}^{n\times 1}$, and unit number of fully connected layers $\{K_j\}_{j=6,7,4}$.
Process: Initialize the parameters of the classification network
repeat:
1. Sample minibatch of $\{\boldsymbol{f}_i^G, \boldsymbol{L}_i\}_{i=1}^{n}$ from $\boldsymbol{F}^G$ and $\boldsymbol{L}$, respectively.
2. for $k = 2$: 3
Follow with the fully connected layers with K_j units, output $\boldsymbol{H}_k^C \leftarrow \sigma(\boldsymbol{W}_k^T\boldsymbol{H}_{k-1}^C + \boldsymbol{b}_k)$ #$\boldsymbol{H}_1^C = \boldsymbol{f}_i^G$
end
3. Calculate classification probabilities $\tilde{\boldsymbol{L}} \leftarrow \sigma(\boldsymbol{W}_4^T\boldsymbol{H}_3^C + \boldsymbol{b}_4)$.
4. Calculate cross-entropy-loss$(\boldsymbol{L}, \tilde{\boldsymbol{L}})$, update parameters.
Until Convergence
Outputs: Classification probabilities $\tilde{\boldsymbol{L}} \in \mathbb{R}^{n\times 1}$

Step 3(b). Detection of AD Potential Genetic Biomarker.
Inputs: Genetic data $\{\mathbf{X}_i\}_{i=1}^{n}$.
Process:
1. Combine our trained networks in Step 2 and Step 3(a).
2. Input $\{\mathbf{X}_i\}_{i=1}^{n}$ to our fully connected network to obtain predicted probabilities $\tilde{\boldsymbol{L}}$.
3. Calculate contribution scores $\boldsymbol{c}(g) = E[A(\frac{\partial \tilde{L}_i}{\partial \boldsymbol{x}(g)})]$
Outputs: Contribution scores $\boldsymbol{c}(g) \in \mathbb{R}^{N_G\times 1}$

Table 2
Causal biomarker rate (CBR) and classification accuracy (ACC) of SNPs and imaging data with different values in sparse matrices for linear and nonlinear associations, where $[\boldsymbol{V}_1, \boldsymbol{V}_2]/\boldsymbol{M}$ represents nonzero values set in $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, and $\boldsymbol{M}$.

	$[\boldsymbol{V}_1, \boldsymbol{V}_2]/\boldsymbol{M}$	Imaging data	SNP	
			Linear association	Nonlinear association
CBR	[1, 2]/1	1.000 ± 0.000	0.775 ± 0.008	0.755 ± 0.003
	[3, 4]/3	1.000 ± 0.000	0.957 ± 0.093	0.925 ± 0.075
	[5, 6]/5	1.000 ± 0.000	0.997 ± 0.002	1.000 ± 0.000
ACC	[1, 2]/1	0.933 ± 0.018	0.747 ± 0.046	0.750 ± 0.030
	[3, 4]/3	0.963 ± 0.016	0.927 ± 0.027	0.920 ± 0.040
	[5, 6]/5	1.000 ± 0.000	0.977 ± 0.008	0.970 ± 0.012

using a linear and a nonlinear association with $\boldsymbol{Y}$ to fully evaluate the performance of the proposed method.

① Linear association: $\boldsymbol{X}^l \sim N(\boldsymbol{Z}_l\boldsymbol{M}, \mathbf{I}_{p\times p})$,
② Nonlinear association: $\boldsymbol{X}^l \sim N(Q(\boldsymbol{Z}_l)\boldsymbol{M}, \mathbf{I}_{p\times p})$, where $Q(\boldsymbol{Z}) = \boldsymbol{Z}^{1/5} + \boldsymbol{Z}^{1/4} + \boldsymbol{Z}^{1/2}$.

Then, we applied a binomial distribution $\mathrm{B}(2, \mathrm{logit}^{-1}(\boldsymbol{X} + \mathrm{logit}(\omega)))$ to convert $\boldsymbol{X}$ into categorical variables at three levels (i.e., 0, 1, and 2) as introduced in Kim et al. (2020), where $\mathrm{logit}(\omega) = \log(\omega/(1-\omega))$ was the logit function and ω was drawn from a uniform distribution $\mathrm{U}(0.1, 0.7)$.

3.1.2. Results of simulated data

We used the causal biomarker rate to assess the biomarker detection performance of the proposed method on the simulated data, which was defined as the ratio of the number of detected SNP (or imaging) biomarkers in top n_p (or n_q) over the n_p (or n_q). Moreover, we applied classification accuracy to evaluate the classification performance of the proposed method on the simulated data. Table 2 lists the causal biomarker rate and classification accuracy of SNPs and imaging data for linear and nonlinear associations. Consequently, larger values in $\boldsymbol{M}$ or $\boldsymbol{V}$ led to higher classification accuracy and causal biomarker rate of SNPs and imaging data for either linear or nonlinear associations. All pre-set disease-related imaging data were detected by using the proposed method (causal biomarker rate = 1). High causal biomarker rate (larger than 0.9) was achieved when large and median values were set in $\boldsymbol{V}$ (nonzero values in $\boldsymbol{V}$ were equal to 5 or 3). Although the classification accuracy and causal biomarker rate decreased with a small value in $\boldsymbol{V}$ (nonzero values in $\boldsymbol{V}$ were equal to 1), more than half of the disease-related SNPs in 1000 disease-related SNPs can be detected, indicating the effectiveness of the proposed method on the simulated data. Moreover, the performance of the proposed method on the linear and nonlinear associations between SNPs and image data was similar, which indicated that the complex associations between SNPs and image data can be handled properly by using the proposed method.

We fixed nonzero values in $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, and $\boldsymbol{M}$ to 3, 4, and 3, respectively, to investigate the effect of Step 3 on the performance of disease-related biomarker detection and generated SNP data $\boldsymbol{X}$ by using the nonlinear association with $\boldsymbol{Y}$. In this experiment, if Step 3 was removed, then the causal biomarker rate of SNP was only 0.773, whereas the causal biomarker rate of SNP was improved (0.925) when Step 3 was included. The distance between the prediction layer of the GRU-based network and nonlinear map $z(\boldsymbol{F}^G)$ in the prediction space was measured in Step 2 to increase similarity between $\boldsymbol{F}^R$ and $\boldsymbol{F}^G$ for association analysis between imaging and genetic data. The SNP biomarkers detected in this step were closely related to imaging data, which may or may not be associ-

Table 3
Subject number at different time points.

Time point	Baseline visit	6 months	12 months	18 months	24 months	36 months
Subject number	708	640	601	251	438	41

ated with the disease. Hence, the link between the potential genetic representations $\boldsymbol{F}^G$ and disease was weak, and a direct link between them was required for disease-related biomarker detection.

3.2. ADNI analysis

3.2.1. Materials

In this study, the genetic data and longitudinal imaging data of the human brain were obtained from the database of ADNI (http://adni.loni.usc.edu/). ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial MRI, PET and other biological markers are useful in clinical trials of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. ADNI subjects aged 55 to 90 from over 50 sites across the US and Canada participated in the research and more detailed information is available at www.adni-info.org.

T1-weighted MRI images were used in this study. The scanning parameters of 1.5T MRI images can be found in a previous study (Jack et al., 2008). A total of 708 (421 men and 287 women, age 75.61 $\pm$ 6.76 years at baseline visit) subjects with 164 AD, 346 MCI, and 198 NC provided by the ADNI1 dataset were used. For each subject, the MRI scans at baseline visit, 6, 12, 18, 24, and 36 months were used in this study when available. A longitudinal study usually covered a relatively long period in health sciences, and some of the subjects almost always missed their scheduled visits or observation date. Therefore, the sequence of observation times may vary across individuals (van Erp et al., 2013). The details of the subject number at different time points are listed in Table 3.

All MRI data were processed under the following steps: (a) using of a nonparametric nonuniform bias correction for image intensity inhomogeneity correction (Sled et al., 1998); (b) skull stripping (Wang et al., 2014) and warping a labeled template to each skull-stripped image for the removal of the cerebellum (aBEAT in version 1.0, http://www.nitrc.org/projects/abeat); (c) segmenting brain tissues into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) by using the FAST method (Zhang et al., 2001); (d) registering all images into a common template (Kabani, 1998) by using the 4D-HAMMER method proposed in a previous study (Shen and Davatzikos, 2004) (HAMMER in version 1.0, https://www.nitrc.org/projects/hammer/); (e) generating the RAVENS maps (Davatzikos et al., 2001), which can be used to quantify the local volumetric group differences in the whole brain and in each tissue type (WM, GM, CSF, and ventricles), by applying the deformation field; (f) automatic labeling of 93 ROIs on the template (Tzourio-Mazoyer et al., 2002); (f) projecting ROI labels from the template image to each MRI image; and (g) calculating the GM volume of each ROI in the labeled image and normalizing them with the intracranial volume as the ROI-based feature representation for each image. With these processing steps, a 93-dimensional ROI-based feature vector and RAVENS maps can be obtained, for each time point for subsequent analyses.

Human 610-quad Beadchip (Illumina, Inc., San Diego, CA) was used to acquire the SNP data provided by the ADNI dataset. Quality control was initially performed on the SNP data, which included the following steps: (i) gender check, (ii) population stratification, (iii) sibling pair identification, (iv) call rate check for each subject and each SNP marker, (v) marker removal according to minor allele frequency, and (vi) Hardy–Weinberg equilibrium test. Then, SNP screening was conducted by using the following steps: (i) at least 95% retention values, (ii) at least 95% minor allele frequency, and (iii) Hardy–Weinberg equilibrium $p > 10^{-6}$. Finally, the remaining missing genetic data were imputed as the modal value. After these procedures, 708 subjects were retained, and each subject had 501,584 SNPs during the subsequent analyses.

3.2.2. Experimental setup

In the ADNI data analysis, longitudinal ROI-based features were used as phenotypes to evaluate the performance of the proposed method with 708 subjects (164 AD, 346 MCI, and 198 NC). Each subject had 501,584 SNPs and imaging data from various time points, where each time point consisted of 93 ROIs. This study considered two types of hyperparameters; one was related to the network structure (e.g., layer number and node number in each layer), and the other was associated with network learning (e.g., learning rate and batch size). Determining network hyperparameters was challenging because of a considerable amount of experience, guesswork, assumptions, prior knowledge of data, and experiments related to hyperparameter tuning (Zhou et al., 2019b). Moreover, using the nested k-fold cross-validation to determine hyperparameters was not feasible because the number of network parameters required to train a deep neural network for each hyperparameter combination was expensive (Lu et al., 2018; Zhou et al., 2019b). Therefore, a hold-out method was used in the experiments. All subjects were divided into 10 subsets with the same proportion of each class label, of which 8 were applied for training, 1 was utilised for validation, and 1 was used for testing. The training set was used to train the networks; the validation set was utilised to guide the halting of network optimization and prevent overfitting, and the testing set was applied to evaluate the networks. This training process was repeated five times, and the results for validation and testing sets were achieved in each process. The final result was obtained by averaging the five results of the testing set. Our search was limited to a small predefined range to compensate for the high computation cost of turning different hyperparameter combinations. Moreover, the hyperparameters used in the proposed method are summarized in Table 4.

For the GRU-based network in the first step, the best hyperparameter setting was searched via a grid search by using the following hyperparameter combinations: the learning rate ranged from 10^{-1} to 10^{-6} with interval accumulate as multiples of 10^{-1}; the batch size ranged from 4 to 128 with interval accumulate as multiples of 2; the layer number was selected from 1, 2, 3, and 4; the node number of the first GRU layer was selected from 100, 200, 300, 400, 500, 600, and 700; the node number of the second fully connected layer was selected from 50, 100, 200, 300, 400, 500, and 600; the node number of the third fully connected layer was selected from 10, 30, 50, 100, and 200; and the node number of the fourth fully connected layer was selected from 5, 10, 30, 50, and

Table 4
Hyperparameters used in the proposed method, where "UN" represents the unit number in the Nth layer; the superscript j of λ_3^j denotes the unit number in the layer, where the potential genetic representations are mapped into, and "- -" indicates no value in the current parameter.

	GRU-based network	Modified diet network		Fully connected network for AD biomarker detection
		Main network	auxiliary network	
U1	$K_1 = 300$	$K_1 = 300$	$K_5 = 300$	$K_6 = 300$
U2	$K_2 = 400$	$K_2 = 400$	$K_1 = 300$	$K_7 = 10$
U3	$K_3 = 200$	$K_3 = 200$	–	$K_4 = 1$
U4	$K_4 = 1$	$K_4 = 1$	–	–
Learning rate	10^{-4}	10^{-5}		10^{-4}
Batch size	64	64		8
λ_2	–	1		–
λ_3^{400}	–	10		–
λ_3^1	–	20		–

100. The node number of the latter layers was smaller than those of the front layers to maximize potentially integrated information from longitudinal imaging data in the front layers, reduce the node number in the latter layers for overfitting prevention, and narrow the semantic gap between the latter layers and the final sigmoid layer for class-membership probability calculation. Finally, the average value of the classification accuracy can be obtained, and the hyperparameter combination with the highest accuracy was selected.

For the modified diet network in the second step, the best hyperparameter setting was searched by only using some predefined combinations because of the complex network structure and high computation cost in this step compared with those in the two other steps. In comparison with the grid search applied to Step 1, the hyperparameter combination in this step was difficult to optimise. Moreover, the computational burden was further lessened because of the super-high dimension (N_G~10^6) of genetic data. Some tricks were included rather than grid search to optimise our modified diet network. On the one hand, our experiments revealed that the auxiliary network preferred a simple architecture; otherwise, gradient explosion and heavy computation costs may appear during network training. Moreover, the unit number of the last layer of the auxiliary network should be K_1 because the auxiliary network primarily aimed to learn the weights (W_e) of the first layer of the main network. Therefore, the auxiliary network with a number of fully connected layers no more than 3 and a unit number of each layer smaller than K_1 ($K_1 = 300$) was selected. On the other hand, some researchers (Bengio and Lecun, 2007; Bianchini and Scarselli, 2014; Hanin, 2018; Hanin and Rolnick, 2018) pointed out empirically that deep and relatively narrow networks can be readily trained. Hence, the layer number of the main network should be determined first and then the unit number of each layer. Moreover, the layer number was selected from 1 to 3, and the unit number of each layer was selected from one of these elements [100, 200, 300]. For the main network, the following strategy was used to design the spaces where the potential genetic representations should be mapped into: first, the network structure was constructed similar to the GRU-based network; however, the GRU layer was replaced with a fully connected layer with K_1 units to simplify the training process; considering that the prediction layer of the GRU-based network was mostly relative to the class information, the potential genetic representations can be initially mapped into this layer. Second, the outputs were extracted from each layer of the GRU-based network (except for the prediction layer), and a SVM classifier for each output and its combination were subsequently trained to exploit the outputs with the most discriminative class information. Finally, the layer corresponding to the output with the outstanding classification performance was selected as the space where the potential genetic representations should be mapped into. After fixing the structure of the main network, the assessment criteria in this step were designed as the total loss in Eq. (6), and the hyperparameter combination in the modified diet network with the lowest total loss was selected. Moreover, the learning rate and batch size were empirically set to 10^{-5} and 128, respectively, in all the experiments in this step.

For the fully connected network in the third step, the best hyperparameter setting was searched via grid search similar to that in the first step from the following hyperparameter combinations: the learning rate ranged from 10^{-1} to 10^{-6} with interval accumulate as multiples of 10^{-1}; the batch size ranged from 4 to 128 with interval accumulate as multiples of 2; the node number of the first fully connected layer was selected from 50, 100, 200, 300, 400, and 500; and the node number of the second fully connected layer was selected from 5, 10, 30, 50, 100, and 200. Few layers and nodes were used in this network because the features might have achieved a high level (more semantic) after the training of the first two networks. Finally, the average classification accuracy can be obtained, and the hyperparameter combination with the highest accuracy was selected.

In all training processes, the maximum number of epochs was 2000, and the maximum patience number of epochs was 100 for early stopping. All of the five hold-out processes were trained with the same network architecture, hyperparameters, and stopping criteria.

3.2.3. Performance of the GRU-based network

A GRU-based network was constructed to integrate longitudinal imaging data. This network was used to bridge longitudinal imaging data and AD-related labels for extracting the potential image representation. Therefore, classification accuracy was utilised to evaluate the performance of the GRU-based network. In this experiment, only AD and NC subjects were applied to perform the AD/NC classification. The average classification accuracy of the testing sets was 0.935 ± 0.049 based on the five hold-out processes, whose accuracy was comparable to those of other previously published AD/NC classification results (Liu et al., 2018b; Lu et al., 2018; Zhu et al., 2019).

We used RNN and long short term memory (LSTM) to construct the network in the first step and evaluate the performance of the proposed GRU-based network. We also compared the classification accuracy, computational time, and parameter number of three different RNN structure-based networks to evaluate their performance. For fair comparison, we designed and selected the hyperparameters with optimal performance. Moreover, the computational time and parameter number were calculated by using the same hyperparameters, such as layer number, node number in each layer, learning rate, and batch size, in three RNN structure-based networks. The comparison results of different RNN structure-based networks are shown in Table 5. At 5% significant level, no difference in classification accuracy was observed among these three

Table 5
Classification accuracy (ACC), computational time (CT), and parameter number (PN) of different RNN structure-based networks. *p*-values of paired *t*-tests comparing the different networks with the GRU-based network are listed in the bottom two rows.

	ACC	CT (s)	PN (K)
RNN	0.935 ± 0.061	46.1	319.0
LSTM	0.930 ± 0.037	101.4	673.6
GRU	0.935 ± 0.049	83.5	555.4
GRU compared to other methods in ACC (p-values)			
RNN	0.996		
LSTM	0.880		

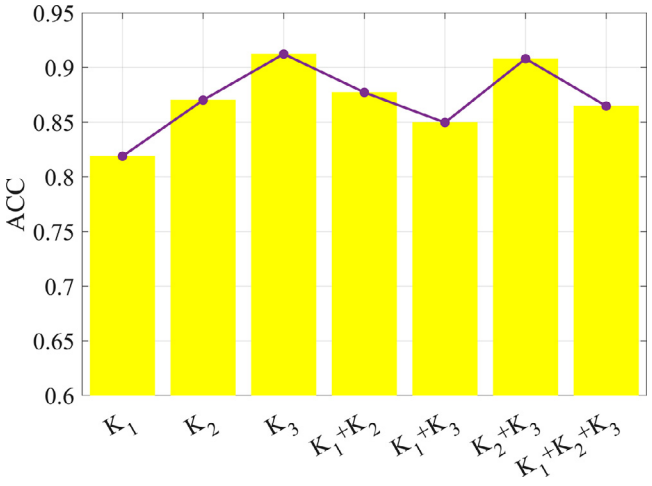

Fig. 4. Classification accuracy (ACC) of outputs of different layers in the GRU-based network obtained by using SVM for the AD/NC classification, where "K_A+ K_B" represents the output concatenation of layers with K_A and K_B units.

networks. Computational time and parameter number of the GRU-based network were lower than those of the LSTM-based network. The lowest computational time and parameter number were obtained by the RNN-based network. However, gradient vanishing often appeared in the RNN, which may result in instability of network results (Cho et al., 2014). Therefore, GRU was applied to balance the computational costs and result stability and provide the effective performance of the whole proposed network.

Moreover, the SVM was trained using outputs and their combinations of each layer of the GRU-based network to determine the suitable layer for mapping the potential genetic representation. As shown in Fig. 4, the output of the third layer with K_3 units achieved the highest classification accuracy; thus, this layer was selected to generate the potential genetic representation. Apart from the third layer, the outputs of the combination of the second and third layers (i.e., K_2+K_3) obtained a superior classification accuracy to those of the other layers. Hence, the second layer of the GRU-based network can also provide rich class information, and $\boldsymbol{F}^G$ was mapped into the spaces of the next and last layers of $\boldsymbol{F}^R$ to extract potential genetic representations (Fig. 2 b).

3.2.4. Effectiveness of using longitudinal imaging data

The proposed method was compared with a network by using the imaging data from a single time point (baseline visit) as phenotype inputs to investigate the effectiveness of using the longitudinal imaging data. In the compared network, the network $\mathcal{M}$ for imaging data was composed of four layers as the proposed GRU-based network, whereas the first hidden layer in $\mathcal{M}$ of the compared network was a fully connected layer with K_1 units. The parameters of the modified diet network and the final classification network were the same as those of the proposed method in the compared network. Moreover, the hyperparameters of the final classification network were determined via grid search for fair comparison. Table 6 shows the classification accuracy of the proposed method with the longitudinal imaging data and the compared network with the imaging data from a single time point. Table 6 also presents that the performance of the proposed method is better than that of the compared network, indicating the effectiveness of using longitudinal imaging data in determining various associations between genetic and imaging data.

Table 6
Classification accuracy (ACC) of the proposed method with longitudinal imaging data (proposed) and the compared network with imaging data from a single time point (compared).

	proposed	compared
ACC	0.709 ± 0.044	0.633 ± 0.070

3.2.5. Effectiveness of the Modified diet network

An ablation experiment was designed as follows to evaluate the effectiveness of the modified diet network. First, the imaging information was ignored by setting $\lambda_4^K = 0$ in Eq. (6), and the loss function of the main network was modified to binary cross entropy (BCE). Therefore, the whole network was returned to the original diet network. Moreover, the weights of BCE and reconstruction losses were set to 20 and 1, respectively, to assess the effectiveness of using imaging data as endophenotypes to link genetic data and diseases. Second, the two auxiliary networks were removed to investigate the effect of using auxiliary networks on the performance of the proposed method. Third, the reconstruction loss was discarded by setting $\lambda_2 = 0$ in Eq. (6) to evaluate the effectiveness of incorporating the reconstruction term of $\boldsymbol{X}$ on the performance of the proposed method. As shown in Fig. 5 (a) and (d), serious overfitting was observed in the absence of imaging information compared with that of the proposed method, indicating that the potential information of the imaging data can guide the proposed network to exploit the latent genetic information. Overfitting was also found in the absence of auxiliary networks (Fig. 5 b). Moreover, the convergent time of the modified diet network without auxiliary networks or without reconstruction term was longer than that of the proposed method (Fig. 5 b–d). These results showed the effectiveness of incorporating the auxiliary networks and reconstruction term on the main network training.

The classification accuracy of these two measurement methods was calculated to compare the advantages of using similarity measurement in Eq. (6) with those in Eq. (5). Using Eq. (5), the structures of the GRU-based network and the two auxiliary networks were the same as those of the proposed method. Moreover, three layers were included in the main network with unit numbers of K_1, K_2, and K_3 in each layer. However, only the outputs of the third layer were mapped into the output space of the corresponding layer in the GRU-based network for the extraction of potential genetic representations. λ_1 was selected from 0.1 to 0.9 with interval accumulate as multiples of 0.2 and fixed at 0.7 because the minimum total loss in Eq. (5) was obtained. Moreover, the hyperparameters of the final classification network were determined via grid search for fair comparison. As shown in Table 7, the performance of the similarity measurement method used in Eq. (6) was better than that used in Eq. (5), indicating that more class-related information can be learned by using Eq. (6) than by utilising Eq. (5).

In assessing the effect of different distances on the performance of the proposed method, Pearson correlation (PC) was used to replace mean square error (MSE) and measure the similarity between imaging and genetic representations. The objective function

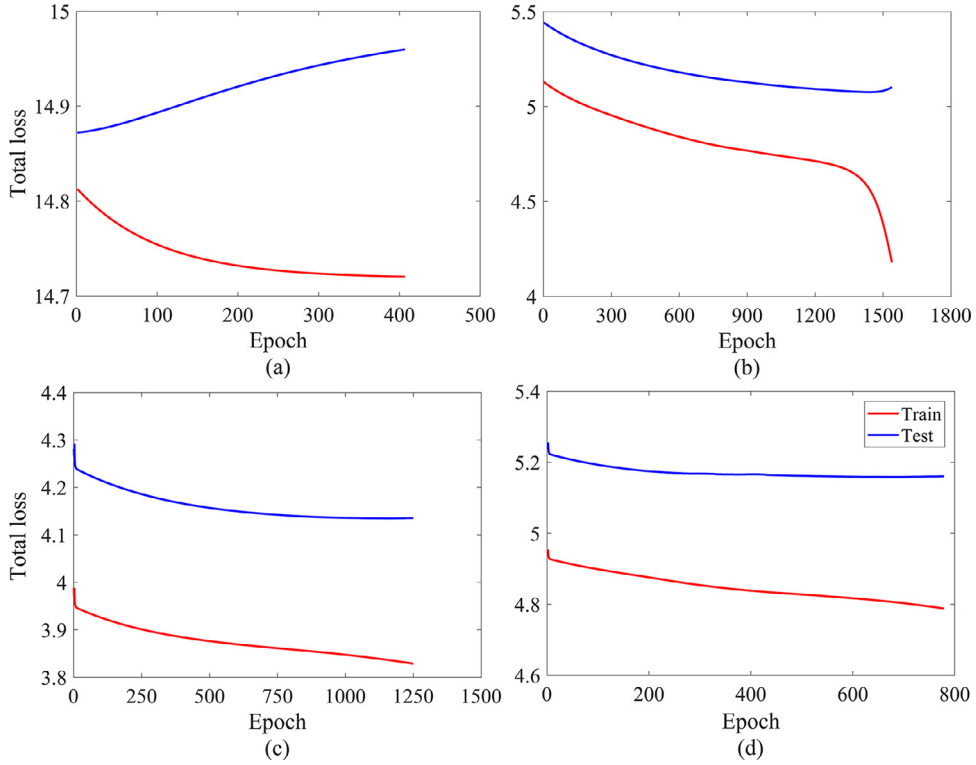

Fig. 5. Loss plots of (a) the original diet network, (b) modified diet network without auxiliary networks, (c) modified diet network without reconstruction loss, and (d) proposed modified diet network.

Table 7
Classification accuracy (ACC) of the proposed method using similarity measurement shown in Eq. (6; proposed) and Eq. (5; compared).

	proposed	compared
ACC	0.709 ± 0.044	0.600 ± 0.035

Table 8
Classification accuracy (ACC) of the proposed method using MSE and PC losses in Eq. (6) and Eq. (7), respectively.

	MSE	PC
ACC	0.709 ± 0.044	0.698 ± 0.052

was designed as follows:

$$\frac{1}{2n}\left(\lambda_2 \sum_{i=1,\ldots,n} \left\|X_i - \hat{X}_i\right\|^2 + \sum_j \lambda_3^j \left(1 - P\left(f_i^R, f_i^G\right)\right)\right) \tag{7}$$

where $P(\boldsymbol{a}, \boldsymbol{b})$ is the PC coefficient between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Considering that the PC coefficient cannot be calculated between two numerical values, the labels were represented as a form of one hot (e.g., AD and NC labels are changed from 1 and 0 to 01 and 10, respectively), and the final layer of the GRU-based network was substituted with a softmax layer. λ_2, λ_3^{400}, and λ_3^2 were set to 1, 10, and 20, respectively. The parameters of the modified diet network with PC loss in Eq. (7) were the same as those with MSE in Eq. (6). Moreover, the hyperparameters of the final classification network were determined via grid search for fair comparison. Table 8 lists the classification accuracy of the proposed method with MSE and PC losses. As shown in Table 8, the classification accuracy of the proposed method with MSE loss was slightly higher than that of the proposed method with PC loss, which indicated that the effect of these two losses was similar to that of the proposed method.

3.2.6. Biomarker detection

We examined the genetic effect of each of the 501,584 SNPs on phenotypes and detected some potential biomarkers associated with AD. The contribution scores $\boldsymbol{c}^{(k)}(g)$ and $\boldsymbol{c}_t^{(k)}(r)$ of each of the five hold-out processes were calculated using the strategy in Section 2.5 by averaging the five contribution scores: $\boldsymbol{c}(g) = \{\boldsymbol{c}^{(k)}(g)\}_{k=1,\ldots,5}$ and $\boldsymbol{c}_t(r) = \{\boldsymbol{c}_t^{(k)}(r)\}_{k=1,\ldots,5}$.

Table 9 lists the top 30 SNPs with the highest scores in $\boldsymbol{c}(g)$, including the corresponding SNPs, chromosome (CHR) IDs, base pair (BP) values, and genes. Consequently, the following genes were detected: ZNF827 (CHR4) was related to the APoE4 noncarriers of AD (Jiang et al., 2015); MACROD2 (CHR 20) and CYP2B6 (CHR 19) were reported to be associated with AD (Cacabelos, 2010; Kohannim et al., 2012); PPM1H (CHR12) was linked to an increased AD risk (Badhwar et al., 2017); DLG2 (CHR 11) was connected with increased AD-related neuroanatomical shape asymmetries (Wachinger et al., 2018); ARID3B (CHR 15) was related to growth retardation and intellectual disability (Donahue et al., 2017); RUNX1 (CHR 21) was associated with neuronal development (Logan et al., 2013); LINGO2 (CHR 9) was related to Parkinson's disease (Su et al., 2012), and SKIV2L (CHR 6) was linked to brain development (Rosenfeld et al., 2020).

Fig. 6 shows the top 10 ROIs in the MRI scans at different time points as sorted by $\boldsymbol{c}_t(r)$. Among these ROIs, the hippocampus and amygdala were related to memory; the middle temporal gyrus was linked to language processes; and the precuneus was associated with a high level of cognitive function, such as episodic memory retrieval, mental imagery strategies, and cue reactivity. The reliability of these findings was confirmed by their consistency with AD prediction and AD imaging genetic studies (Huang et al., 2015; Ning et al., 2018; Zhou et al., 2019a). As shown in Fig. 6, the contribution scores of top 10 ROIs in the MRI scans in 6 months were constantly higher than those in other time points (except for the lateral ventricle at the right hemisphere), which indicated their

Table 9
ADNI ROI volume GWAS: top 30 selected SNPs associated with the 93 ROIs. "- -" indicates that the item does not to correspond to the genes.

	SNP	CHR	BP	Gene		SNP	CHR	BP	Gene
1	rs3184088	15	74,890,354	ARID3B	16	rs10506635	12	71,898,821	LGR5
2	rs1472866	4	146,849,715	ZNF827	17	rs11178812	12	71,864,192	LGR5
3	rs2059409	2	218,720,042	TNS1	18	rs3958102	12	83,481,871	TMTC2
4	rs10940415	5	53,985,003	LOC102467080	19	rs1517820	12	83,488,072	TMTC2
5	rs200762	20	15,609,236	MACROD2	20	rs771972	12	63,275,017	PPM1H
6	rs2834675	21	36,292,500	RUNX1	21	rs34555781	6	31,925,697	SKIV2L
7	rs2072164	7	143,657,427	OR2F1	22	rs548927	16	3,257,868	–
8	rs17128435	8	13,147,657	DLC1	23	rs479067	1	27,375,997	LOC101928391
9	rs10503446	8	13,147,702	DLC1	24	rs9327478	5	127,850,129	FBN2
10	rs7852959	9	28,471,251	LINGO2	25	rs12807906	11	84,064,102	DLG2
11	rs2291287	19	41,522,451	CYP2B6	26	rs7115655	11	132,024,460	NTM
12	rs1190053	6	105,745188	PREP	27	rs16880442	5	52,185,695	ITGA1
13	rs11008540	10	31,894,818	–	28	rs1962230	4	54,021,591	SCFD2
14	rs728581	14	52,258,905	LOC101927598	29	rs7128766	11	20,908,913	NELL1
15	rs12504009	4	125,553,855	–	30	rs11694726	2	1,460,220	TPO

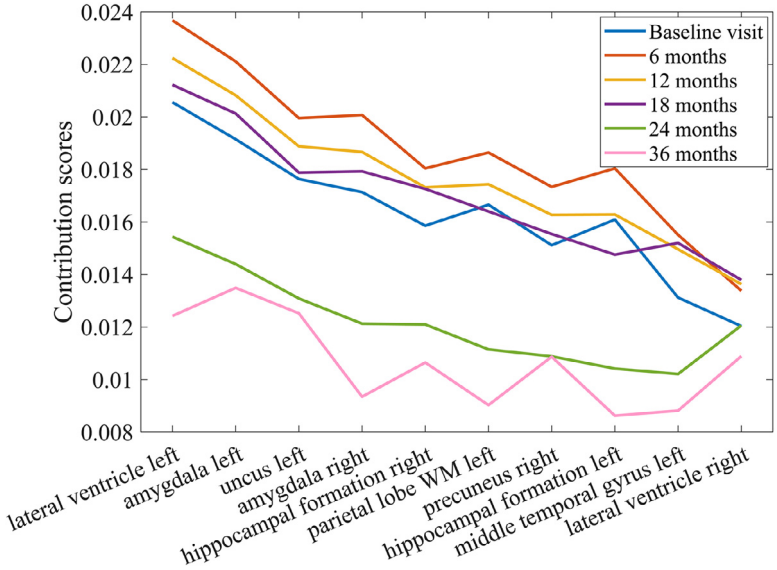

Fig. 6. Contribution scores of top 10 ROIs in the MRI scans at different time points.

crucial changes. Their contribution scores in 12 months were also higher than those at baseline visit. Therefore, important information on AD progression may be lost when imaging data were used at a single time point (e.g., baseline visit). Moreover, AD-related variations as discovered by longitudinal imaging data application can be determined by imaging genetics when the associations between imaging and genetic data were analysed. Hence, longitudinal imaging data may be used in determining the effects of genetic data on disease progression trajectory and enhancing the detection power of potential biomarkers.

AD/NC classification was performed on the top SNPs (or ROIs) by using a SVM classifier to investigate the associations between top SNPs (or ROIs) and diagnosis. Only ROIs obtained from baseline visit were used for AD/NC classification. Fig. 7 (a) and (b) show the classification results with different number of top ROIs and SNPs, respectively. For ROI (Fig. 7 a), the highest classification accuracy was 0.897 by using the top 25 ROIs from baseline visit, which was higher than that by using all 93 ROIs from baseline visit. For SNP (Fig. 7 b), the classification accuracy of using different number of top SNPs was higher than that of using the total super-high-dimensional SNPs (~10^6). Moreover, the highest classification accuracy of SNP was 0.851 by using the top 500 SNPs. The classification accuracy outperformed several relevant methods (Huang et al., 2019b; Ning et al., 2018; Zhou et al., 2019b), which involved a set of preselected SNPs as predictors to classify subjects into AD and NC. These results indicated that the proposed method could detect some AD-related SNP/ROI biomarkers.

3.2.7. Detection of AD-related pairs of ROI and SNP

The associations between imaging and genetic data were exploited by measuring the similarity between imaging and genetic representations ($\boldsymbol{F}^R$ and $\boldsymbol{F}^G$), as described in Step 2 of the proposed method. For a given unit p in one layer of a multi-layer fully connected network, the derivative of p of a subject i with AD was obtained with regard to an input ($\boldsymbol{x}(g)$ or $\boldsymbol{y}_t(r)$) based on the chain rule for derivatives. For instance, the derivatives of units in the second layer (FC2 in Fig. 2) with regard to the inputs can be obtained for imaging and genetic data, denoted as $h_2(g) = \frac{\partial \boldsymbol{H}_2^G}{\partial \boldsymbol{x}(g)}$ and $h_2(r) = \frac{\partial \boldsymbol{H}_2^R}{\partial \boldsymbol{y}_t(r)}$, respectively, where $\boldsymbol{H}_2^G$and $\boldsymbol{H}_2^R$ represent the outputs of FC2 for the GRU-based network and modified diet network, respectively: $h_2(g) \in \mathbb{R}^{1\times K_2}$, and $h_2(r) \in \mathbb{R}^{1\times K_2}$. Therefore, the association score between the gth SNP and the rth ROI in the second layer can be achieved by $\boldsymbol{c}_2(g,r) = E[A(h_2(g) \times h_2(r)^T)]$. The derivatives of units in the fourth layer (prediction layer in Fig. 2) with regard to the inputs can also be calculated, and the association score between the gth SNP and rth ROI in the fourth layer can be achieved by $\boldsymbol{c}_4(g,r) = E[A(h_4(g) \times h_4(r)^T)]$. Considering that the

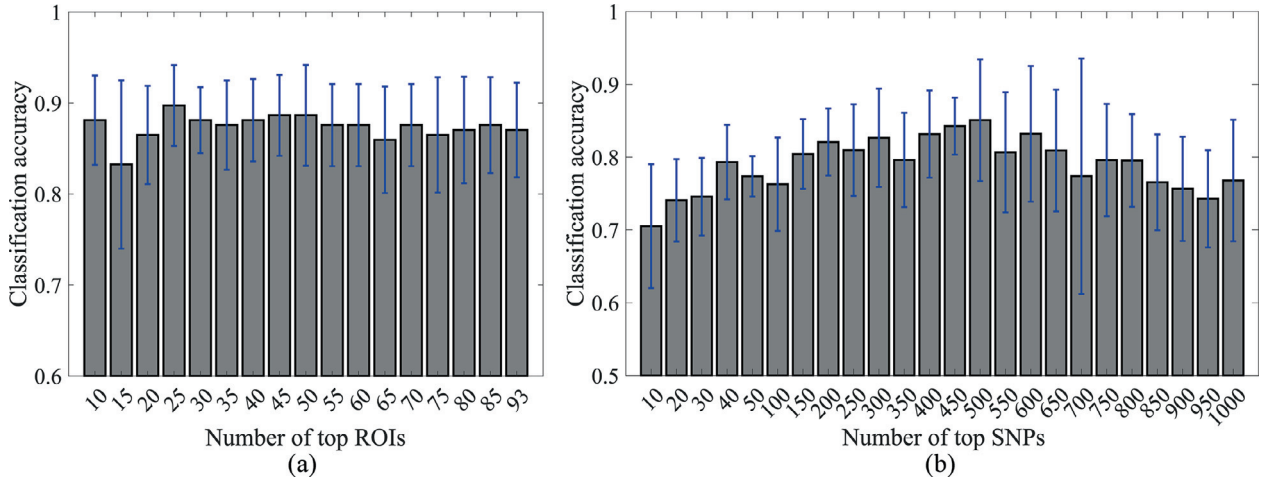

Fig. 7. Classification accuracy of (a) top ROIs and (b) top SNPs.

Table 10
Top 16 SNP-ROI pairs with the largest association scores, where amyg.R/L and hiopp.R/L are abbreviations of amygdala right/left and hippocampal formation right/left, respectively.

		SNP(Gene)				Mean ± std
		rs3184088 (ARID3B)	rs2059409 (TNS1)	rs2072164 (OR2F1)	rs7852959 (LINGO2)	
ROI	amyg.R	0.999	0.997	0.988	0.973	0.492 ± 0.240
	amyg.L	0.918	0.917	0.908	0.896	0.449 ± 0.223
	hiopp.L	0.889	0.886	0.880	0.862	0.438 ± 0.213
	hiopp.R	0.812	0.810	0.805	0.790	0.403 ± 0.198
Mean ± std		0.417 ± 0.166	0.415 ± 0.166	0.412 ± 0.164	0.406 ± 0.161	

distance was calculated between the imaging and genetic outputs of the second and fourth layers (Eq. (6)), respectively, the association scores between imaging and genetic data can be obtained by $\boldsymbol{c}(g,r) = (\boldsymbol{c}_2(g,r) + \boldsymbol{c}_4(g,r))/2$.

We first selected the top 10 SNPs and ROIs with the highest contribution scores calculated by using the strategy proposed in Section 2.5 to illustrate the complex associations between SNPs and ROIs. Then, we calculated the association scores by using the above-mentioned strategy to obtain the associations between the top 10 SNPs and top 10 ROIs. Table 10 shows the top 16 SNP-ROI pairs with the largest association scores. As shown in Table 10, the association scores were similar among different SNPs for a ROI, which may result from the weak and average effects of the top SNPs. However, the association scores of the top SNPs were remarkably larger than the mean association scores of total SNPs (~10^6). The same results were found in the association scores of the top ROIs compared with the mean association scores of total ROIs. Therefore, the important ROI-SNP pairs can be detected by using the proposed method. Among these ROI-SNP pairs, the hippocampus and amygdala were related to memory; ARID3B (CHR 15) was related to growth retardation and intellectual disability (Donahue et al., 2017); LINGO2 (CHR 9) was related to Parkinson's disease, and their associations need further investigation and verification (Su et al., 2012).

3.2.8. Effect of using RAVENS maps as inputs

Whole-brain images (RAVENS maps) were used as inputs of the proposed method to investigate the effect of different inputs on the performance of the proposed method, and a three-step training strategy was used to extract the potentially integrated information of the longitudinal imaging data for image representation ($\boldsymbol{F}^R$). For the first step, a 3D AD/NC classification framework was constructed for the imaging data, and its structure is listed in Table 11. As shown in Table 11, the 3D classification framework, namely, ResNet3D, was extended from 2D ResNet50 (He et al., 2016) (i.e., all convolution and pooling layers were changed from 2D to 3D), and a fully connected layer with 128 units (FC128) was added behind the average pooling layer. With the well-trained ResNet3D, the feature extracted from the output of FC128 was used as image measurement ($\boldsymbol{Y}_{it} \in \mathbb{R}^{128\times 1}$). In this step, all brain images of AD and NC subjects with different time points in the training set were used to train the network. For the second step, the parameters in the ResNet3D were fixed, and $\boldsymbol{Y}_{it}$ extracted from ResNet3D was applied as input to train the proposed GRU-based network. For the third step, the ResNet3D without the sigmoid layer was concatenated with the GRU-based network, and an end-to-end network fine-tuning was performed to optimise the whole network for extraction of image representations. For the whole network (including the ResNet3D and GRU-based network), the average classification accuracy of the whole-brain images in the testing sets was 0.930 ± 0.041 based on the five hold-out processes, whose accuracy was similar to that of using ROIs as inputs, and was comparable to those of other previously published AD/NC classification results (Liu et al., 2018b; Lu et al., 2018; Zhu et al., 2019).

With the well-trained network for the whole-brain images, the image representations can be extracted, and the following two steps, including association analysis between imaging and genetic data and biomarker detection, can be performed. Finally, the AD/NC classification accuracy of the genetic representation ex-

Table 11
Structures of the ResNet3D for extraction of image measurements. The symbol $[A] \times D$ denotes that block A is repeated for D times. "$p^w \times q$ conv" represents the sequence Conv-BN layer with convolutional kernel size of p^w and q features. "FC(t)" corresponds to a fully connected layer with t units. "–" denotes no structure in the current stage.

	Output size	ResNet3D
Input	$97^3 \times 1$	–
Convolution	$91^3 \times 64$	$7^3 \times 64$ conv, ReLU, stride 1
Pooling1	$46^3 \times 64$	$3^3 \times 64$ Max Pooling, stride 2
ResBlock1	$23^3 \times 256$	$[(1^3 \times 64$ conv), $(3^3 \times 64$ conv), $(1^3 \times 256$ conv), ReLU$] \times 3$
ResBlock2	$12^3 \times 512$	$[(1^3 \times 128$ conv), $(3^3 \times 128$ conv), $(1^3 \times 512$ conv), ReLU$] \times 4$
ResBlock3	$6^3 \times 1024$	$[(1^3 \times 256$ conv), $(3^3 \times 256$ conv), $(1^3 \times 1024$ conv), ReLU$] \times 6$
ResBlock4	$3^3 \times 2048$	$[(1^3 \times 512$ conv), $(3^3 \times 512$ conv), $(1^3 \times 2048$ conv), ReLU$] \times 3$
Pooling2	$1^3 \times 2048$	$3^3 \times 2048$ Average Pooling, stride 1
FC	128	FC(128)

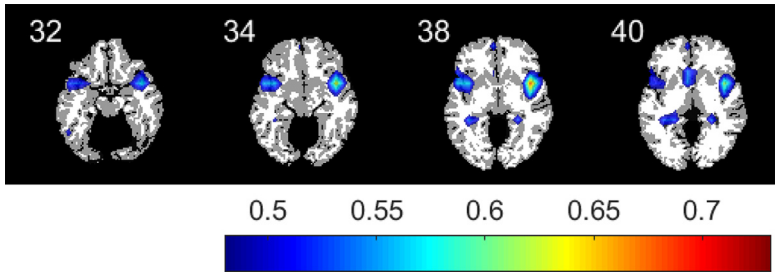

Fig. 8. Selected slices of the contribution scores of voxels calculated in the proposed network for extraction of image representation.

Table 12
ADNI whole-brain image GWAS: top 30 selected SNPs associated with the whole-brain images.

	SNP	CHR	BP	Gene		SNP	CHR	BP	Gene
1	rs6986134	8	106,672,991	ZFPM2	16	rs925717	12	79,379,050	SYT1
2	rs4970843	1	109,887,191	SORT1	17	rs12597935	16	73,147,484	ZFHX3
3	rs4837062	9	120,031,365	ASTN2	18	rs7179325	15	101,066,053	CERS3
4	rs4943302	13	35,986,558	NBEA	19	rs3738773	1	110,048,527	AMIGO1
5	rs2262425	2	238,984,129	SCLY	20	rs7337025	13	87,481,267	LOC105370300
6	rs2264132	2	238,994,676	SCLY	21	rs17246639	11	122,279,782	MIR100HG
7	rs821480	2	238,972,827	SCLY	22	rs2076954	15	33,987,341	RYR3
8	rs745611	18	45,909,040	ZBTB7C	23	rs3817043	10	1,256,632	ADARB2
9	rs1847325	15	101,067,542	CERS3	24	rs1442330	8	106,673,800	ZFPM2
10	rs9395285	6	47,554,177	CD2AP	25	rs12919255	16	83,535,543	CDH13
11	rs6507842	18	45,887,992	ZBTB7C	26	rs10105946	8	106,664,873	ZFPM2
12	rs972731	15	100,997,743	CERS3	27	rs1489310	16	83511242	CDH13
13	rs8097442	18	45,885,036	ZBTB7C	28	rs2306907	16	83,520,078	CDH13
14	rs11247214	15	101,055,938	CERS3	29	rs6030129	20	40,946,744	PTPRT
15	rs989891	21	197,08347	TMPRSS15	30	rs4149580	12	6,446,990	TNFRSF1A

tracted by the modified diet network was 0.703. The contribution score of each voxel was calculated by using Grad-CAM++ to illustrate the contribution of each voxel in the whole-brain images at all time points to the network outputs (Fig. 8) (Chattopadhay et al., 2018), and is shown in Fig. 8. Moreover, Table 12 lists the top 30 SNPs with the highest contribution scores. As shown in Fig. 8, several voxels in major ROIs including the amygdala, hippocampus, superior temporal gyrus, middle temporal gyrus, putamen, and thalamus, have previously demonstrated their associations with AD (Huang et al., 2021; Wen et al., 2020). Among the top 30 SNPs listed in Table 12, the following genes were detected: SORT1 (CHR 1) was related to the risk of developing AD (Andersson et al., 2016); CD2AP (CHR 6) was associated with high risk of AD (Tao et al., 2019); RYR3 (CHR 15), CDH13 (CHR 16), and TNFRSF1A (CHR12) were related to AD (Bruno et al., 2012; Liu et al., 2018a; Shang et al., 2015); ZFHX3 (CHR 16) was related to dementia (Rollo et al., 2015); ASTN2 (CHR 9) was connected to the development of the cortical regions of the brain (Wilson et al., 2010); ZBTB7C (CHR 18) was related to mental retardation (Gilling et al., 2011); TMPRSS15 (CHR 21) was implicated in neuronal function (Morey et al., 2017); AMIGO1 (CHR 1) was associated with brain development (Zhao et al., 2011); ADARB2 (CHR 10) was linked to the changes of temporal lobe volume extracted from AD patients (Kohannim et al., 2012); PTPRT had a positive effect on dendritic arborization and synapse formation in cultured hippocampal neurons (Thomas et al., 2013); ZFPM2 (CHR 8) was associated with schizophrenia (Greenbaum et al., 2012), and NBEA (CHR 13) was related to autism (Castermans et al., 2010).

4. Discussion

In this study, a novel dGDN-GWAS method is introduced to detect potential AD biomarkers. A GRU-based network is applied without imputing incomplete longitudinal imaging data to integrate longitudinal data with variable lengths and extract an image representation. Considering the high parameters produced by the genetic network that takes super-high-dimensional genetic data (~10^6) as input, a modified diet network that can considerably reduce the number of parameters in the genetic network is proposed to perform the GWAS between the genetic data and image representation. Consequently, the genetic representation can be extracted for the following analyses. Finally, a link between the ge-

netic representation and AD is created for potential AD biomarker detection. The proposed method is also evaluated on a set of simulated data and 708 subjects obtained from a real AD dataset. Based on the simulated experiments, the results show that the proposed method can accurately detect relevant biomarkers and achieve high classification accuracy. Based on the real AD data analysis, the results show that the proposed method has high accuracy in AD/NC classification, and this method can detect some AD-related biomarkers. Therefore, the proposed method is an important tool for the prediction, diagnosis, and monitoring of AD.

The proposed method primarily consists of three steps, where the first two steps are used to conduct GWAS. In the first step, the potentially integrated information of the longitudinal imaging data for image representation $\boldsymbol{F}^R$ is extracted by using a GRU-based network. $\boldsymbol{F}^R$ extraction can be simply formulated as $\boldsymbol{F}^R = \varphi(\boldsymbol{Y})$, where $\varphi(\boldsymbol{.})$ is a nonlinear function. In the second step, the modified diet network is used to extract the potential genetic information $\boldsymbol{F}^G$ and examine the association between the imaging and genetic data. Similar to $\boldsymbol{F}^R$ extraction, $\boldsymbol{F}^G$ extraction can be formulated as $\boldsymbol{F}^G = \phi(\boldsymbol{X})$, where $\phi(\boldsymbol{.})$ is considered as a nonlinear function. Therefore, the association between the imaging and genetic data is nonlinear, and it can be represented as $\varphi(\boldsymbol{Y}) = \phi(\boldsymbol{X})$. From this association formulation between the imaging and genetic data, we can observe that the correlations among imaging data, the correlations among genetic data, and the association between imaging and genetic data are considered by using the proposed method. Therefore, the proposed method performs GWAS between longitudinal imaging and super-high-dimensional genetic data in a multivariate way.

Longitudinal imaging data are important for discovering the effect of genetic data on the trajectory of disease progression and enhancing the detection power of potential biomarkers. Table 6 shows that the classification accuracy of the proposed method with longitudinal imaging data is higher than that of a network with imaging data from a single time point (paired t-test $p = 0.031$). As shown in Table 6, the AD/NC classification accuracy with genetic data as inputs is lower than that with imaging data as inputs (0.709 vs. 0.935), which may result from the super-high-dimensional features (approximately 10^6) and weak genetic effects of the genetic data. In some previously published methods (Ning et al., 2018; Zhang et al., 2014), which also involve SNP as a predictor in the AD/NC classification, their AD/NC classification accuracy is below 0.7 with a set of AD-related SNPs (SNP number is less than 6000). However, under the guidance of the GRU-based network, the AD/NC classification accuracy of the super-high-dimensional genetic data can be improved using the proposed method. More AD-related SNP biomarkers may be detected by using the proposed method compared with other studies with a small set of SNPs. Moreover, AD/NC classification is performed on the top SNPs (or ROIs) by using a SVM classifier to investigate the associations between top SNPs (or ROIs) and diagnosis (Fig. 7). As shown in Fig. 7, the classification accuracy of top SNPs (or ROIs) is higher than that of all SNPs (or all ROIs), and it outperforms some relevant methods (Huang et al., 2019b; Ning et al., 2018; Zhou et al., 2019b), which indicates that the proposed method can detect some AD-related SNP/ROI biomarkers. As shown in Fig. 6, the contribution scores of top 10 ROIs in the MRI scans in 6 and 12 months are higher than those in the baseline visit. Therefore, if imaging data in the baseline visit are only used in most previous studies, then important information on AD progression may be lost. Several studies have also incorporated longitudinal data in analysing the associations between imaging and genetic data and demonstrated an improved performance (Du et al., 2019; Hao et al., 2017; Marchetti-Bowick et al., 2016b). However, the number of time points of longitudinal data should be the same in most of these studies, but this requirement is difficult to achieve in a longitudinal study with a relatively long period (such as AD in this study).

In this study, a modified diet network is incorporated into the proposed network to reduce the network parameters for extracting potential genetic information and effectively conducting GWAS. In the modified diet network, the per-class histograms of the genetic data in the training set are used as feature embeddings, which may incorporate class information to guide the discovery of potential information in the genetic data. In addition to an association loss between the imaging and genetic data in traditional GWAS, a reconstruction loss of genetic data is added to the objective function of the modified diet network. As shown in Fig. 5, incorporating reconstruction term can improve the convergence of the training of the network possibly because some important information of the genetic data is reserved using the reconstruction term. Thus, the potential information of the genetic data can be extracted for the following analyses. In comparison with the proposed method, the original diet network suffers from serious overfitting, indicating that extracting potential AD-related information from high-dimensional SNP data is difficult without imaging data. Moreover, a comparative experiment is performed by using Eqs. (6) and (5) to evaluate different similarity measurements on the performance of the AD/NC classification. As shown in Table 7, the classification accuracy obtained by using Eq. (6) is superior to that calculated by using Eq. (5), indicating that the distance between imaging and genetic representations may be shortened via the similarity measurement in Eq. (6) for the extraction of potential genetic information. PC is applied to replace MSE in Eq. (6) to measure the similarity between imaging and genetic representations and evaluate the performance of using different distances on the proposed method. As shown in Table 8, the classification accuracy of the proposed method with MSE or PC loss is similar (pair t-test $p = 0.74$), suggesting that the effects of using different distance on the proposed method may be weaker than other factors, such as using longitudinal imaging data and using similarity measurement (Eq. (6)).

Apart from using longitudinal ROIs as inputs of the proposed method, whole-brain images (RAVENS maps) are also applied to assess the effect of different inputs on the performance of the proposed method. In the biomarker detection experiments, amygdala, hippocampus, and middle temporal gyrus are detected by using ROIs or whole-brain images. Moreover, some different brain regions are detected by using ROIs (such as the lateral ventricle, uncus, and precuneus) and whole-brain images (such as the superior temporal gyrus, putamen, and thalamus). For genetic biomarker detection, several AD-related genes, including ZNF827 (CHR4), MACROD2 (CHR 20), CYP2B6 (CHR 19), PPM1H (CHR12), and DLG2 (CHR 11), are found by using ROIs as inputs, whereas some other AD-related genes, including SORT1 (CHR 1), CD2AP (CHR 6), RYR3 (CHR 15), CDH13 (CHR 16), and TNFRSF1A (CHR12), are identified by using whole-brain images as inputs, suggesting that different image measurements may provide different information to guide the association analysis between imaging and genetic data and lead to different results of biomarker detection. Moreover, several SNPs with potential risks are discovered by using ROIs and whole-brain images as inputs, but their effect on AD has not been revealed in literature. Therefore, these genes related to AD progression should be further investigated. Such findings may improve the discovery of new AD-related genetic biomarkers and the early prediction and treatment of this disease. Although different top 30 SNPs are detected, respectively, a similar phenomenon is found by using ROIs and whole-brain images as inputs (Tables 9 and 12). Among the top 30 SNPs listed in Table 9, rs17128435 and rs10503446 belong to the DLC1 gene, rs10506635 and rs11178812 belong to the LGR5 gene, and rs3958102 and rs1517820 belong to the TMTC2 gene. Moreover, among the top 30 SNPs listed in Table 12, rs6986134, rs1442330, and rs10105946 belong to the

ZFPM2 gene; rs2262425, rs2264132, and rs821480 belong to the SCLY gene; rs745611, rs6507842, and rs8097442 belong to the ZBTB7C gene; rs972731, rs11247214, and rs7179325 belong to the CERS3 gene, and rs12919255, rs1489310, and rs2306907 belong to the CDH13 gene. This result may be attributed to the correlations among adjacent SNPs, which are considered in the proposed method. Thus, adjacent SNPs, which are associated with AD, can be detected using the proposed method.

Several issues should be addressed in our future research. First, the proposed deep-learning network is separated into three parts and trained step by step, thereby achieving a suboptimal outcome. Therefore, an end-to-end framework and an alternative optimal method should be incorporated into the proposed method. Second, remarkable results may be achieved when different image phenotypes are combined with our research because various characteristics of image phenotypes can be obtained from different neuroimaging modalities (e.g., functional MRI, PET, and diffusion tensor imaging).

Declaration of Competing Interest

None.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 81601562, 81974275, and U1501256), the Science and Technology Planning Project of Guangzhou (No. 201904010417), the Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515012011), and the Science and Technology Planning Project of Guangdong Province (No. 2015B010131011). Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The investigators within the ADNI contributed to the design and implementation of ADNI but did not participate in the analysis or writing of this paper. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how to apply/ADNI Acknowledgement List.pdf.

References

Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad-CAM++: generalised gradient-based visual explanations for deep convolutional networks. In: Proceeding of the IEEE Winter Conference on Applications of Computer Vision, pp. 839–847.

Andersson, C.H., Hansson, O., Minthon, L., Andreasen, N., Blennow, K., Zetterberg, H., Skoog, I., Wallin, A., Nilsson, S., Kettunen, P., 2016. A genetic variant of the sortilin 1 gene is associated with reduced risk of Alzheimer's disease. J. Alzheimers Dis. 53, 1353–1363.

Alzheimer's Association, 2018. 2018 Alzheimer's disease facts and figures. Alzheimers Dement. 14, 367–425.

Badhwar, A., Brown, R., Stanimirovic, D.B., Haqqani, A.S., Hamel, E., 2017. Proteomic differences in brain vessels of Alzheimer's disease mice: normalization by PPAR agonist pioglitazone. J. Cerebr. Blood Flow Met. 37, 1120–1136.

Y. Bengio and Y. LeCun. Scaling learning algorithms towards AI. In L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, editors, Large Scale Kernel Machines. MIT Press, 2007.

Bi, X., Yang, L., Li, T., Wang, B., Zhu, H., Zhang, H., 2017. Genome-wide mediation analysis of psychiatric and cognitive traits through imaging phenotypes. Hum. Brain Mapp. 38, 4088–4097.

Bianchini, M., Scarselli, F., 2014. On the complexity of neural network classifiers: a comparison between shallow and deep architectures. IEEE Trans. Neural Netw. Learn. Syst. 25, 1553–1565.

Bruno, A.M., Huang, J.Y., Bennett, D.A., Marr, R.A., Hastings, M.L., Stutzmann, G.E., 2012. Altered ryanodine receptor expression in mild cognitive impairment and Alzheimer's disease. Neurobiol. Aging 33.

Cacabelos, R., 2010. Pharmacogenomic protocols in CNS disorders and dementia. Neurodegener. Dis. 7, 167–169.

Castermans, D., Volders, K., Crepel, A., Backx, L., De Vos, R., Freson, K., Meulemans, S., Vermeesch, J.R., Schrander-Stumpel, C.T.R.M., De Rijk, P., Del-Favero, J., Van Geet, C., Van De Ven, W.J.M., Steyaert, J.G., Devriendt, K., Creemers, J.W.M., 2010. SCAMP5, NBEA and AMISYN: three candidate genes for autism involved in secretion of large dense-core vesicles. Hum. Mol. Genet. 19, 1368–1378.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A Simple Framework for Contrastive Learning of Visual Representations. ICML 1597–1607.

Cho, K., Merrienboer, B.V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empitical Methods in Natural Language Processing, EMNLP, 1724-1734.

Davatzikos, C., Genc, A., Xu, D., Resnick, S.M., 2001. Voxel-based morphometry using the Ravens maps: methods and validation using simulated longitudinal atrophy. Neuroimage 14, 1361–1369.

Donahue, J., Krähenbühl, P., Darrell, T., 2017. Adversarial feature learning. Proceeding of the International Conference on Learning Representations.

Du, L., Liu, K.F., Zhu, L., Yao, X.H., Risacher, S.L., Guo, L., Saykin, A.J., Shen, L.Alzheimer's Disease Neuroimaging Initiative, 2019. Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: a longitudinal study of the ADNI cohort. Bioinformatics 35, I474–I483.

Dudbridge, F., 2016. Polygenic epidemiology. Genet. Epidemiol. 40, 268–272.

Ghazi, M.M., Nielsen, M., Pai, A., Cardoso, M.J., Modat, M., Ourselin, S., Sorensen, L.Alzheimer's Disease Neuroimaging Initiative, 2019. Training recurrent neural networks robust to incomplete data: application to Alzheimer's disease progression modeling. Med. Image Anal. 53, 39–46.

Gilling, M., Lind-Thomsen, A., Mang, Y., Bak, M., Moller, M., Ullmann, R., Kristoffersson, U., Kalscheuer, V.M., Henriksen, K.F., Bugge, M., Tumer, Z., Tommerup, N., 2011. Biparental inheritance of chromosomal abnormalities in male twins with non-syndromic mental retardation. Eur. J. Med. Genet. 54, E383–E388.

Greenbaum, L., Simth, R.C., Lorberboym, M., Alkelai, A., Zozulinsky, P., Lifshytz, T., Kohn, Y., Djaldetti, R., Lerer, B., 2012. Association of the ZFPM2 gene with antipsychotic-induced parkinsonism in schizophrenia patients. Psychopharmacology 220, 519–528 (Berl.).

Guo, Y.J., Wu, C.X., Guo, M.Z., Zou, Q., Liu, X.Y., Keinan, A., 2019. Combining sparse group lasso and linear mixed model improves power to detect genetic variants underlying quantitative traits. Front. Genet. 10.

Hanin, B., 2018. Which neural net architectures give rise to exploding and vanishing gradients? In: Proceeding of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018). Curran Associates, Inc., Montréal, Canada, pp. 582–591.

Hanin, B., Rolnick, D., 2018. How to start training: the effect of initialization and architecture. In: Proceeding of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018). Curran Associates, Inc., Montréal, Canada, pp. 571–581.

Hao, X.K., Li, C.X., Yan, J.W., Yao, X.H., Risacher, S.L., Saykin, A.J., Shen, L., Zhang, D.Q.Alzheimer's Disease Neuroimaging Initiative, 2017. Identification of associations between genotypes and longitudinal phenotypes via temporally-constrained group sparse canonical correlation analysis. Bioinformatics 33, I341–I349.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770–778.

Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. Comput. Sci. 14, 38–39.

Hu, W., Meng, X., Bai, Y., Zhang, A., Qu, G., Cai, B., Zhang, G., Wilson, T.W., Stephen, J.M., Calhoun, V.D., Wang, Y., 2021. Interpretable multimodal fusion networks reveal mechanisms of brain cognition. IEEE Trans. Med. Imaging 40, 1474–1483.

Hu, W.X., Cai, B., Zhang, A.Y., Calhoun, V.D., Wang, Y.P., 2019. Deep collaborative learning with application to the study of multimodal brain development. IEEE Trans. Biomed. Eng. 66, 3346–3359.

Huang, C., Thompson, P., Wang, Y., Yu, Y., Zhang, J., Kong, D., Colen, R.R., Knickmeyer, R.C., Zhu, H.Alzheimer's Disease Neuroimaging, I, 2017. FGWAS: functional genome wide association analysis. Neuroimage 159, 107–121.

Huang, M., Chen, X., Yu, Y., Lai, H., Feng, Q., 2021. Imaging genetics study based on a temporal group sparse regression and additive model for biomarker detection of Alzheimer's disease. IEEE Trans. Med. Imaging 40, 1461–1473.

Huang, M., Deng, C., Yu, Y., Lian, T., Yang, W., Feng, Q., 2019a. Spatial correlations exploitation based on nonlocal voxel-wise GWAS for biomarker detection of AD. NeuroImage Clin. 21.

Huang, M., Nichols, T., Huang, C., Yu, Y., Lu, Z., Knickmeyer, R.C., Feng, Q., Zhu, H.Alzheimer's Disease Neuroimaging, I., 2015. FVGWAS: fast voxelwise genome wide association analysis of large-scale imaging genetic data. Neuroimage 118, 613–627.

Huang, M., Yu, Y., Yang, W., Feng, Q., 2019b. Incorporating spatial-anatomical similarity into the VGWAS framework for AD biomarker detection. Bioinformatics 35, 5271–5280.

Huang, M., Yu, Y., Yang, W., Feng, Q., 2019c. Incorporating spatial–anatomical similarity into the VGWAS framework for AD biomarker detection. Bioinformatics.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., Study, A., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Jiang, S., Yang, W.L., Qiu, Y., Chen, H.Z., Adni, 2015. Identification of novel quantitative traits-associated susceptibility Loci for APOE epsilon 4 non-carriers of Alzheimer's Disease. Curr. Alzheimer Res. 12, 218–227.

Kabani, N.J., 1998. 3D anatomical atlas of the human brain. Neuroimage 7, P–0717.
Kim, M., Won, J.H., Youn, J., Park, H., 2020. Joint-connectivity-based sparse canonical correlation analysis of imaging genetics for detecting biomarkers of Parkinson's disease. IEEE Trans. Med. Imaging 39, 23–34.
Kohannim, O., Hibar, D.P., Stein, J.L., Jahanshad, N., Hua, X., Rajagopalan, P., Toga, A.W., Jack, C.R., Weiner, M.W., de Zubicaray, G.I., McMahon, K.L., Hansell, N.K., Martin, N.G., Wright, M.J., Thompson, P.M., Initia, A.D.N., 2012. Discovery and replication of gene influences on brain structure using LASSO regression. Front. Neurosci. 6 -Switz.
Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thornton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.T., Choi, S.H., Reitz, C., Pasquier, F., Hollingworth, P., Ramirez, A., Hanon, O., Fitzpatrick, A.L., Buxbaum, J.D., Campion, D., Crane, P.K., Baldwin, C., Becker, T., Gudnason, V., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Moron, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fievet, N., Huentelman, M.J., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuinness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossu, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Naranjo, M.C.D., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannfelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F.G., Valladares, O., Younkin, S.G., Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson, P.V., Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley, T.H., Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F.A.G., Passmore, P., Montine, T.J., Bettens, K., Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie, K., Lunetta, K.L., Kauwe, J.S.K., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltunen, M., Martin, E.R., Schmidt, R., Rujescu, D., Wang, L.S., Dartigues, J.F., Mayeux, R., Tzourio, C., Hofman, A., Nothen, M.M., Graff, C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer, L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D., Amouyel, P., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1458.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lee, G., Kang, B., Nho, K., Sohn, K.A., Kim, D., 2019. MildInt: deep learning-based multimodal longitudinal data integration framework. Front. Genet. 10.
Li, G., Han, D., Wang, C., Hu, W., Calhoun, V.D., Wang, Y., 2020. Application of deep canonically correlated sparse autoencoder for the classification of schizophrenia. Comput. Methods Prog. Biomed. 183, 105073. doi:10.1016/j.cmpb.2019.105073.
Li, P., Guo, M.Z., Wang, C.Y., Liu, X.Y., Zou, Q., 2015. An overview of SNP interactions in genome-wide association studies. Brief. Funct. Genomics 14, 143–155.
Lin, W.M., Tong, T., Gao, Q.Q., Guo, D., Du, X.F., Yang, Y.G., Guo, G., Xiao, M., Du, M., Qu, X.B., Neuroimaging, A.D., 2018. Convolutional neural networks-based MRI Image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Front. Neurosci. 12 -Switz.
Liu, F.F., Zhang, Z., Chen, W., Gu, H.Y., Yan, Q.J., 2018a. Regulatory mechanism of microRNA-377 on CDH13 expression in the cell model of Alzheimer's disease. Eur. Rev. Med. Pharmacol. 22, 2801–2808.
Liu, M., Zhang, D., Shen, D.Alzheimer's Disease Neuroimaging, I., 2015. View-centralised multi-atlas classification for Alzheimer's disease diagnosis. Hum. Brain Mapp. 36, 1847–1865.
Liu, M.X., Zhang, J., Adeli, E., Shen, D.G., 2018b. Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. 43, 157–168.
Logan, T.T., Villapol, S., Symes, A.J., 2013. TGF-beta superfamily gene expression and induction of the Runx1 transcription factor in adult neurogenic regions after brain injury. PLoS One 8.
Lu, D.H., Popuri, K., Ding, G.W., Balachandar, R., Beg, M.F., Initia, A.D.N., 2018. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. Med. Image Anal. 46, 26–34.
Marchetti-Bowick, M., Yin, J., Howrylak, J.A., Xing, E.R., 2016a. A time-varying group sparse additive model for genome-wide association studies of dynamic complex traits. Bioinformatics 32, 2903–2910.
Marchetti-Bowick, M., Yin, J.M., Howrylak, J.A., Xing, E.R., 2016b. A time-varying group sparse additive model for genome-wide association studies of dynamic complex traits. Bioinformatics 32, 2903–2910.
Morey, R.A., Davis, S.L., Garrett, M.E., Haswell, C.C., Marx, C.E., Beckham, J.C., McCarthy, G., Hauser, M.A., Ashley-Koch, A.E.Mid-Atlantic MIRECC, 2017. Genome-wide association study of subcortical brain volume in PTSD cases and trauma-exposed controls. Trans. Psychiatry 7.
Ning, K.D., Chen, B., Sun, F.Z., Hobel, Z., Zhao, L., Matloff, W., Toga, A.W., Neuroimaging, A.s.D., 2018. Classifying Alzheimer's disease with brain imaging and genetic data using a neural network framework. Neurobiol. Aging 68, 151–158.
Rollo, J., Knight, S., May, H.T., Anderson, J.L., Muhlestein, J.B., Bunch, T.J., Carlquist, J., 2015. Incidence of dementia in relation to genetic variants at PITX2, ZFHX3, and ApoE epsilon 4 in atrial fibrillation patients. Pace 38, 171–177.
Romero, A., Carrier, P.L., Erraqabi, A., Sylvain, T., Auvolat, A., Dejoie, E., Legault, M.A., Dube, M.P., Hussin, J.G., Bengio, Y., 2017. Diet networks: thin parameters for fat genomics. Proceeding of the International Conference on Learning Representations.
Rosenfeld, C.S., Hekman, J.P., Johnson, J.L., Lyu, Z., Ortega, M.T., Joshi, T., Mao, J., Vladimirova, A.V., Gulevich, R.G., Kharlamova, A.V., Acland, G.M., Hecht, E.E., Wang, X., Clark, A.G., Trut, L.N., Behura, S.K., Kukekova, A.V., 2020. Hypothalamic transcriptome of tame and aggressive silver foxes (Vulpes vulpes) identifies gene expression differences shared across brain regions. Genes Brain Behav. 19, e12614.
Shang, Z.W., Lv, H.C., Zhang, M.M., Duan, L., Wang, S.T., Li, J., Liu, G.Y., Zhang, R.J., Jiang, Y.S., 2015. Genome-wide haplotype association study identify TNFRSF1A, CASP7, LRP1B, CDH1 and TG genes associated with Alzheimer's disease in caribbean hispanic individuals. Oncotarget 6, 42504–42514.
Shen, D., Davatzikos, C., 2004. Measuring temporal morphological changes robustly in brain MR images via 4-dimensional template warping. Neuroimage 21, 1508–1517.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87.
Su, F.C., Chen, C.M., Chen, Y.C., Wu, Y.R., 2012. LINGO-2 polymorphism and the risk of Parkinson's disease in Taiwan. Parkinsonism Relat. Disord. 18, 609–611.
Tao, Q.Q., Chen, Y.C., Wu, Z.Y., 2019. The role of CD2AP in the pathogenesis of Alzheimer's disease. Aging Dis. 10, 901–907.
Thomas, J., Garg, M.L., Smith, D.W., 2013. Altered expression of histone and synaptic plasticity associated genes in the hippocampus of streptozotocin-induced diabetic mice. Metab. Brain Dis. 28, 613–618.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.
van Erp, T.G.M., Hibar, D.P., Rasmussen, J., Andreassen, O.A., Haukvik, U.K., Agartz, I., Potkin, S.G., Hulshoff-Pol, H., Ophoff, R., van Haren, N.E.M., Gruber, O., Kramer, B., Erlich, S., Hass, J., Wang, L., Alpert, K., Thompson, P.M., Turner, J.A.ENIGMA Schizophrenia Working Group, 2013. A large-scale meta–analysis of subcortical brain volume abnormalities in schizophrenia via the enigma consortium. Biol. Psychiatry 73, 299s–300s.
Wachinger, C., Nho, K., Saykin, A.J., Reuter, M., Rieckmann, A., Neuroimaging, A.D., 2018. A longitudinal imaging genetics study of neuroanatomical asymmetry in Alzheimer's disease. Biol. Psychiatry 84, 522–530.
Wang, Y., Nie, J., Yap, P.T., Li, G., Shi, F., Geng, X., Guo, L., Shen, D., 2014. Knowledge-guided robust MRI brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates. PLoS One 9, e77810.
Wen, C., Ba, H., Pan, W., Huang, M., 2020. Co-sparse reduced-rank regression for association analysis between imaging phenotypes and genetic variants. Bioinformatics.
Wilson, P.M., Fryer, R.H., Fang, Y., Hatten, M.E., 2010. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during Glial-Guided Neuronal Migration. J. Neurosci. 30, 8529–8540.
Xu, Z., Wu, C., Pan, W., Alzheimer's Dis Neuroimaging, I., 2017. Imaging-wide association study: integrating imaging endophenotypes in GWAS. Neuroimage 159, 159–169.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.
Zhang, Z.M., Huang, H., Shen, D.G., Initia, A.D.N., 2014. Integrative analysis of multi-dimensional imaging genomics data for Alzheimer's disease prediction. Front. Aging Neurosci. 6.
Zhao, X., Kuja-Panula, J., Rouhiainen, A., Chen, Y.C., Panula, P., Rauvala, H., 2011. High mobility group box-1 (HMGB1; amphoterin) is required for zebrafish brain development. J. Biol. Chem. 286, 23200–23213.
Zhou, T., Thung, K.H., Liu, M.X., Shen, D., 2019a. Brain-wide genome-wide association study for Alzheimer's disease via joint projection learning and sparse regression model. Ieee Trans. BioMed. Eng. 66, 165–175.
Zhou, T., Thung, K.H., Zhu, X.F., Shen, D.G., 2019b. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum. Brain Mapp. 40, 1001–1016.
Zhu, Q., Yuan, N., Huang, J., Hao, X., Zhang, D., 2019. Multi-modal AD classification via self-paced latent correlation analysis. Neurocomputing 355, 143–154.
Zhu, X., Suk, H.I., Huang, H., Shen, D., 2017. Low-rank graph-regularised structured sparse regression for identifying genetic biomarkers. IEEE Trans. Big Data 3, 405–414.
Zhu, X.F., Zhang, W.H., Fan, Y., Initi, A.s.D.N., 2018. A robust reduced rank graph regression method for neuroimaging genetic analysis. Neuroinformatics 16, 351–361.
Zhuo, S., Qiao, Y., Lelieveldt, B.P.F., Staring, M.Alzheimer's Disease Neuroimaging, I., 2018. Integrating spatial-anatomical regularization and structure sparsity into SVM: improving interpretation of Alzheimer's disease classification. Neuroimage 178, 445–460.