Statistics Innovation

ARTICLE

Open Access

https://doi.org/10.48130/stati-0025-0003

*Statistics Innovation* **2025**, 2: e003

# Predictive modeling and inference using deep transfer learning in genetic data analysis

Shan Zhang[1#], Yuan Zhou[2#], Kejin Dong[2], Jinling Liu[3], Pei Geng[4*] and Qing Lu[2]

[1] *Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA*
[2] *Department of Biostatistics, University of Florida, Gainesville, FL, USA*
[3] *Department of Epidemiology, University of Florida, Gainesville, FL, USA*
[4] *Department of Mathematics and Statistics, University of New Hampshire, Druham, NH, USA*
[#] Authors contributed equally: Shan Zhang, Yuan Zhou
[*] Corresponding author, E-mail: pei.geng@unh.edu

## Abstract

Transfer learning has been widely applied in text and image classification, demonstrating its effectiveness in numerous applications. In this paper, we propose a transfer learning procedure for both prediction and association testing between genotypes and phenotypes in a smaller primary data set with an available larger source data set. Specifically, we training a deep neural network model in the source data, transfer a part of the trained weight parameters to the model in the primary data, and complete the training process in the primary data with the remaining free parameters. Furthermore, we develop a permutation-based association test using the trained transfer learning model to identify significant genes in the primary data set. We apply the proposed procedure to two case studies for the investigation of nicotine dependence. These two case studies show that transfer learning can not only improve prediction accuracy but also the power of detecting candidate genes compared to those results without transfer learning.

**Citation:** Zhang S, Zhou Y, Dong K, Liu J, Geng P, et al. 2025. Predictive modeling and inference using deep transfer learning in genetic data analysis. *Statistics Innovation* 2: e003 https://doi.org/10.48130/stati-0025-0003

## Introduction

In recent years, deep learning methods have gained attraction due to the ability to learn complex patterns and representations from large amounts of data, enabling breakthroughs in challenging tasks in image recognition, natural language processing, and speech processing. However, there are two main constraints of deep learning methods: dependency on extensive labeled data and training costs[1]. Often in genetics, the focused problems only have limited labeled phenotypes with high dimensional genetic data. In these cases, transfer learning has the potential of higher prediction accuracy and test power by the shared parameters from a well-trained deep learning model in a massive source problem. For example, with the vast amounts of genetic data collected from biobank projects, an interesting scientific question is whether these resources can be used to enhance genetic analysis in small-scale studies. A common assumption made by most existing approaches is that two studies should be similar (e.g., the same population). However, this assumption could fail in reality. The study design and study population may differ between the two studies (e.g., Caucasian vs African American).

While transfer learning attempts to improve the performance of target learners on target domains by transferring the model parameters contained in different but related source domains, it does not require data from two studies drawn from the same feature space or the same distribution. It learns possibly useful features from source studies and applies learned features based on focused problems. Therefore, it holds great promise in using the enriched resources from large-scale studies for uncovering novel genetic variants in small-scale studies[2].

In recent years, transfer learning has been investigated in biological and medical fields. For example, a data-driven procedure for transfer learning[3], called Trans-Lasso, was proposed in high-dimensional linear regression and applied to understand gene regulation using Genotype-Tissue Expression data. The transfer learning problem under high-dimensional generalized linear models[4] aimed to improve the fit of target data by borrowing information from useful source data. Although several survey articles[5–7] have reviewed recent developments on transfer learning in machine learning methods, including network-based deep transfer learning which reuses the partial of network pre-trained in the source data, the application of deep transfer learning in genetic data analysis is relatively sparse. The transfer learning in convolutional neural networks[8] was developed to predict the progression-free interval of lung cancer with gene expression data. A pipeline of transfer learning for genotype-phenotype prediction[9] was studied using deep learning models with a small number of genotypes.

Given that the transfer methods depend on the models or algorithms being used to learn the tasks, we propose to integrate the idea of transfer learning into deep neural networks for prediction and association analyses with high dimensional genetic data. For example, deep neural networks can be trained in the large-scale UK Biobank dataset for nicotine dependence and transfer the parameter weights to facilitate genetic analysis in small-scale studies. By integrating transfer learning into deep neural networks, we are able to transfer model parameters regarding complex genotype-phenotype relationships (e.g., gene-gene interactions) between two studies.

Besides the proposed predictive modeling using transfer learning in deep neural networks, we further develop a permutation-based association test to detect significant genes in the targeted problem based on the proposed transfer learning. The resulting *p*-values can be interpreted as a measure of feature importance, and they help decide the significance of variables and therefore improve model interpretability[10–12]. The permutation-based association test using transfer learning shows higher statistical power compared to that without transfer learning in two case studies. Moreover, the

permuted data is obtained via randomly shuffling the index of subjects while maintaining their intrinsic genetic structures, hence no model re-fitting is needed which greatly reduces the computation cost.

## Method

Transfer learning applies model parameters gained from one research problem to a different but related problem, and has been widely used in text and image recognition. The basic idea of transfer learning is illustrated in Fig. 1.

In this paper, we illustrate transfer learning in the deep neural networks (DNN) model. In this section, we briefly introduce the DNN model and then integrate the idea transfer learning into DNN (TL-DNN) between two datasets. To distinguish the different data types in the following sections, we use lowercase letters, bold lowercase letters, and uppercase letters to denote scalar, vector, and matrix, respectively.

### Deep neural networks

Suppose our research interest is to find a predictive function $f$ that models the continuous response variable $y$ and predictor variable $x = (x_1, ..., x_q)$, where $q$ is the dimension of input. The true model is written as:

$$y = f(\boldsymbol{x}) + \epsilon$$

where, $\epsilon$ is the noise in this model. We first describe the fully connected DNN algorithm for the model training:

$$\begin{aligned}
\boldsymbol{h}_1 &= \sigma(W_1\boldsymbol{x} + \boldsymbol{b}_1) \\
\boldsymbol{h}_d &= \sigma(W_d\boldsymbol{h}_{d-1} + \boldsymbol{b}_d), \quad d = 2, \ldots, l-1 \\
\hat{y} &= \hat{f}(\boldsymbol{x}) = \boldsymbol{w}_l\boldsymbol{h}_{l-1}
\end{aligned}$$

where, $\boldsymbol{h}_d$ is the hidden layer learned from the primary study and has a dimension of $m_d$, $\sigma$ is a nonlinear activation function such as the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$, $l$ is the number of layers of the DNN model, $W_d$ and $\boldsymbol{b}_d$ are weights and biases, respectively. We denote all the parameters by $\boldsymbol{S} = \{\boldsymbol{W}_d, \boldsymbol{b}_d, \boldsymbol{w}_l \mid d = 1, ..., l-1\}$.

In DNN, we first normalize the response variable $y$. Suppose that $\bar{y}$ and $\bar{v}$ are the estimated mean value and standard deviation of $y$, the final model is realized by:

$$\hat{y} = \hat{f}(\boldsymbol{x}) = \bar{v}\boldsymbol{w}_l\boldsymbol{h}_{l-1} + \bar{y}.$$

For simplicity, we her focus on the regression problem and use the $L_2$ loss and mean square error (MSE) to estimate parameters and evaluate the model performance. It can be easily extended to the classification problem, and other loss functions (e.g., cross-entropy loss) and measurements (e.g., misclassification error) can be used.

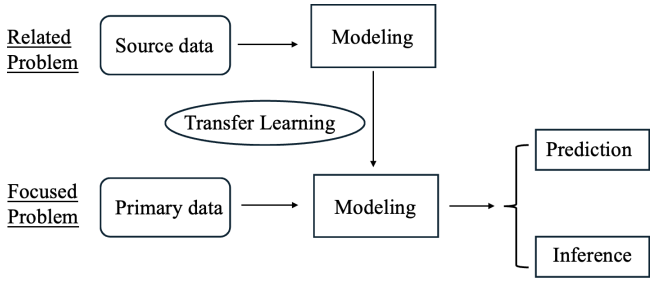


**Fig. 1** An illustration of transfer learning. The goal is to build predictive models and conduct inference in the primary data for the focused problem. The modeling parameters learned from the source data of a related problem are transferred to the modeling process in the primary problem.

In genetic studies, genetic effects are usually smaller than the noise, and therefore heavy penalties are required to avoid overfitting. Various regularization methods, such as penalty regularization, dropout method, and early stopping method, can be used. In this paper, we impose two parameter regularization forms defined below: the square regularization $p_1(\boldsymbol{S};\lambda)$ and the group Lasso regularization[13] $p_2(\boldsymbol{S};\lambda)$, where, $\lambda$ is a hyperparameter controlling the solution space.

$$\begin{aligned}
p_1(\boldsymbol{S};\lambda) &= \lambda\Big(\sum_{d=1}^{l-1} \|W_d\|_2^2 + \|\boldsymbol{w}_l\|_2^2\Big) \\
p_2(\boldsymbol{S};\lambda) &= \lambda\Big(\sum_{d=1}^{l-1} Q(W_d) + Q(\boldsymbol{w}_l)\Big) \\
Q(z) &= \begin{cases} \|z\|, & \text{for} \quad \|z\| \geq \alpha \\ \dfrac{\|z\|^2}{2\alpha} + \dfrac{\alpha}{2}, & \text{for} \quad \|z\| < \alpha \end{cases}, \alpha = 1e^{-3}
\end{aligned} \tag{1}$$

In the following, we apply $p_1(\boldsymbol{S};\lambda)$ for the prediction step while $p_2(\boldsymbol{S};\lambda)$ is used in the association test. Therefore, the solution to the DNN framework is defined as:

$$\hat{\boldsymbol{S}} = arg\min_{\boldsymbol{S}}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - f(\boldsymbol{x}_i))^2 + p(\boldsymbol{S};\lambda)\right)$$

The backpropagation method is typically used to obtain the solution. The selection of hyperparameter $\lambda$ is discussed in the following section.

### Transfer learning in deep neural networks

If we have a source dataset (e.g., UK biobank), in which we model the response variable $y'$ with predictor variable $\boldsymbol{x}' = (x'_1, \ldots, x'_q)$, where $\boldsymbol{x}$ and $\boldsymbol{x}'$ share a similar data structure. Usually, $\boldsymbol{x}$ and $\boldsymbol{x}'$ are the same characteristics of different subjects. We have a similar model:

$$\boldsymbol{h}'_1 = \sigma(W'_1\boldsymbol{x}' + \boldsymbol{b}'_1), \tag{2}$$

$$\boldsymbol{h}'_d = \sigma(W'_d\boldsymbol{h}'_{d-1} + \boldsymbol{b}'_d), \quad d = 2, \ldots, l-1, \tag{3}$$

$$\hat{y}' = \hat{f}'(\boldsymbol{x}') = \bar{v}'\boldsymbol{w}'_l\boldsymbol{h}'_{l-1} + \bar{y}', \tag{4}$$

where, $\bar{y}'$ and $\bar{v}'$ are the estimated mean value and standard deviation of $y'$. The parameters space of this model is denoted by $\boldsymbol{S}'$. The set of their solutions is denoted by $\hat{\boldsymbol{S}}'$.

To prepare for the transfer learning, we divide the solution $\hat{\boldsymbol{S}}'$ into two parts: $\hat{\boldsymbol{S}}'_1 = \{\hat{W}'_d, \hat{\boldsymbol{b}}'_d | d = 1, \ldots, l-1\}$ in Eqns (2) - (3) and $\hat{\boldsymbol{S}}'_2 = \{\hat{\boldsymbol{w}}'_l\}$ in Eqn (4). Given the large scale of the source study for the related problem, it is reasonable to assume that the final hidden layer $\boldsymbol{h}'_{l-1}$ is a good representation of features in $\boldsymbol{x}'$. The relation between $\boldsymbol{x}'$ and $\boldsymbol{h}'_{l-1}$ is realized through $\hat{\boldsymbol{S}}'_1$ while the final hidden layer weight parameter $\hat{\boldsymbol{S}}'_2$ connects the last hidden layer $\boldsymbol{h}'_{l-1}$ with the output. Furthermore, we use $f'_1$ and $f'_2$ to denote the relationships, i.e.:

$$\boldsymbol{h}'_{l-1} = f'_1(\boldsymbol{x}'|\hat{\boldsymbol{S}}'_1), \qquad \hat{y}' = f'_2(\boldsymbol{h}'_{l-1}|\hat{\boldsymbol{S}}'_2).$$

Similarly, the parameter set $\hat{\boldsymbol{S}}$ from the focused problem can also be divided to $\hat{\boldsymbol{S}}_1$ and $\hat{\boldsymbol{S}}_2$. The corresponding relationship is denoted by $f_1$ and $f_2$. The penalty term $p(\boldsymbol{S};\lambda)$ can also be written as $p(\boldsymbol{S}_1, \boldsymbol{S}_2; \lambda)$.

After training the deep neural network in the related problem, we now apply transfer learning using $\hat{\boldsymbol{S}}'_1$ as a solution to the hidden layer weights in the focused problem. In other words, we adapt the representation of hidden layer weights from the well-trained model in the related problem instead of training it based on limited data in the focused problem. Specifically, similar to the parameter sharing strategy[14], we transfer $\hat{\boldsymbol{S}}'_1$ to the focused problem and only update the weights in the output layers by freezing those transferred hidden layer weights. Finally, we derive the solution of $\boldsymbol{S}_2$ from the

data set of the focused problem. Hence, our final solution in the transfer deep learning procedure is:

$$\begin{aligned}\widetilde{\boldsymbol{S}}_1 \equiv \hat{\boldsymbol{S}}_1 &= \arg\min_{\boldsymbol{S}_1,\boldsymbol{S}_2}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i' - f'(\boldsymbol{x}_i'))^2 + p(\boldsymbol{S}_1,\boldsymbol{S}_2;\lambda)\right)\\ \widetilde{\boldsymbol{S}}_2 &= \arg\min_{\boldsymbol{S}_2}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - f(\boldsymbol{x}_i|\boldsymbol{S}_1 = \widetilde{\boldsymbol{S}}_1))^2 + p(\widetilde{\boldsymbol{S}}_1,\boldsymbol{S}_2;\lambda)\right)\end{aligned} \tag{5}$$

The transfer learning method based on the deep neural networks is illustrated in Fig. 2 in comparison to the direct application of deep neural networks in the focused problem.

We apply the Adam algorithm[15] for optimization in Eqn (5). The backpropagation procedure is implemented as:

$$\begin{aligned}c(\boldsymbol{S}) &= \frac{1}{n}\sum_{i=1}^{n}(y_i - f(\boldsymbol{x}_i))^2 + p(\boldsymbol{S}_1,\boldsymbol{S}_2;\lambda)\\ s &\leftarrow s - r_1\frac{\partial c(\boldsymbol{S})}{\partial s}, \quad s \in \boldsymbol{S}_1\\ s &\leftarrow s - r_2\frac{\partial c(\boldsymbol{S})}{\partial s}, \quad s \in \boldsymbol{S}_2\end{aligned}$$

where, $r_1$ and $r_2$ are the learning rates of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, respectively.

The Adam algorithm is an adaptive optimization method in which the learning rate is determined element-wise. While the default setting of the Adam algorithm works well in most cases, it is not suitable for genetic studies. Due to the heavy penalty in the last layer, the solution of $\boldsymbol{S}$ can be too small compared to the default learning rate of the Adam algorithm. Therefore, we set the learning rate parameter based on $\lambda$ while keeping other parameters as the default value. The iteration process stops when MSE does not decrease for 3,000 epochs. In the next section, we set $l = 3$, and the numbers of hidden units of each layer are 16 and 4.

For modeling performance comparison, we add a baseline model, which is essentially the mean value of the response variable $\bar{y}$. For a predictive model $f$, we propose the following relative efficiency criterion based on the predictive mean square error (MSE) compared to the baseline model:

$$R^2_{pseudo} = 1 - \frac{MSE(f)}{MSE(\bar{y})} = 1 - \frac{\sum_{i=1}^{n}(y_i - f(\boldsymbol{x}_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

The larger the value of $R^2_{pseudo}$, the better the formed model as compared to the baseline model. The values shown in the figures are percentage values, whose maximum possible value is 100.

## Permutation-based association test using transfer learning

Based on the proposed deep transfer learning approach, we further develop a permutation-based association test, called PT-TL-DNN, to identify significant genes for different phenotypes. We adapt the feature selection method[12] and the test procedure is defined as follows. Assume that $\boldsymbol{x}$ is the candidate gene/region to be tested for association with disease $Y$ and we denote $f(\boldsymbol{x}|\widetilde{\boldsymbol{S}}_1,\widetilde{\boldsymbol{S}}_2)$ as the deep transfer learning predictive model based on the train set $D_1$ of the focused problem and $L(Y,f(\boldsymbol{x}))$ as a chosen loss function between the observed $Y$ and the predicted $f(\boldsymbol{x})$ in the test set $D_2$ of the focused problem. To reduce the chance finding due to random splitting, the test statistic is calculated based on $K$-fold cross-validation as follows:

$$\Delta = \frac{1}{K}\sum_{k=1}^{K}\Delta_k = \frac{1}{K}\sum_{k=1}^{K}\sum_{i\in D_{k,2}}\left\{L(y_i, f_k(\boldsymbol{x}_i|\widetilde{\boldsymbol{S}}_1,\widetilde{\boldsymbol{S}}_2)) - E\left[L(y_i, f_k(\boldsymbol{x}_i'|\widetilde{\boldsymbol{S}}_1,\widetilde{\boldsymbol{S}}_2))\right]\right\}$$

where, $\boldsymbol{x}_i'$ represents the permuted data that is obtained via randomly shuffling the index of subjects while maintaining their intrinsic structures (e.g., linkage disequilibrium) and $E[L(y_i, f_k(\boldsymbol{x}_i'|\widetilde{\boldsymbol{S}}_1,\widetilde{\boldsymbol{S}}_2))]$ is the expected loss for the permuted data. If the gene is not significant for the disease, the observed loss would be close to the expected loss by permutation, hence $\Delta$ is close to zero; If the gene is significant for the prediction of the disease, the observed loss would be much smaller than the loss by permutation resulting in $\Delta < 0$. Hence the association test is equivalent to the following hypothesis testing:

$$H_0 : \Delta \geq 0 \qquad vs. \qquad H_a : \Delta < 0$$

Under $H_0$, it was shown that $\Delta \sim N(0,\sigma^2)$ by assuming that the empirical losses come from the same distribution[12] and $\sigma^2$ can be empirically estimated. The $p$-value is computed to conclude the significance of the gene.

During the TL-DNN model training before the permutation step, we apply the smooth group Lasso regularization $p_2$ defined in Eqn (1). Compared to the square regularization, the smooth group Lasso regularization can provide structured sparsity and allow the entire group of features to be removed, which may benefit the permutation testing when we test all variants together as a gene. The regularization parameter is denoted as $\lambda_{SGL}$ to distinguish it from the previous square penalty parameter. The parameter $\lambda_{SGL}$ is selected from the set {0, 0.001, 0.01, 0.05, 0.1, 0.5, 1, 5, 10}. The detailed algorithm of the PT-TL-DNN is described in Table 1.

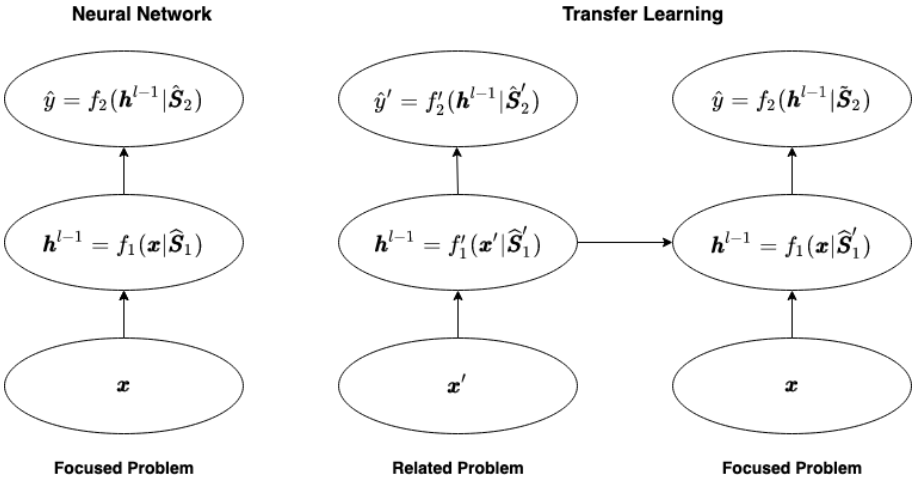


**Fig. 2** Transfer learning in deep neural networks. The left panel shows the application of neural networks in the focused problem while the right panel displays the proposed transfer learning method based on training the neural networks in the related problem. Here $\hat{\boldsymbol{S}}_1'$ are the transferred weight parameters in hidden layers. The output layer weights $\tilde{\boldsymbol{S}}_2$ are updated during the training while $\hat{\boldsymbol{S}}_1'$ are frozen.

**Table 1.** Algorithm for the permutation-based association test using transfer learning.

| Permutation-based test using transfer learning with K-fold cross-validation |
|---|
| **Input:** Genetic variants of a gene **x**, Phenotype **y**, a set of candidate smooth Group Lasso regularization parameters $\boldsymbol{\lambda}_{\mathbf{SGL}}$. |
| **Output:** Empirical p-value of the gene |
| **Step 1:** Construct a TL-DNN model $f(x)$ with 2 hidden layers. |
| **Step 2: For** $k \leftarrow 1, ..., K$ **do** |
| 1: Split ($\boldsymbol{x}_{train}$, $\boldsymbol{y}_{train}$), ($\boldsymbol{x}_{test}$, $\boldsymbol{y}_{test}$). |
| **For** each $\lambda_i$ in $\lambda$ **do** |
| a: Input ($\boldsymbol{x}_{train}$, $\boldsymbol{y}_{train}$) and train f($x_{train}$, $y_{train}$;$\lambda_i$) with smooth Group Lasso regularization parameter $\lambda_i$, output $\hat{f}$. |
| b: Evaluate Mean Square Error on ($\boldsymbol{x}_{train}$, $\boldsymbol{y}_{train}$) , $MSE(y_{test}, \hat{f}(x_{test}; \lambda_i))$. |
| **end.** |
| 2: Choose $\lambda_{\text{opt}}$ with the lowest *MSE*, output $\hat{y}_{test} = \hat{f}(x_{test}; \lambda_{opt})$ and calculate $MSE(y_{test}, \hat{f}(x_{test}; \lambda_{opt}))$. |
| 3: Permute $x_{test}$ by row, denoted as $x'_{test}$, calculate $\hat{y}'_{test} = \hat{f}(x'_{test}; \lambda_{opt})$, $MSE(y_{test}, \hat{y}'_{test})$ and $l = MSE(y_{test}, \hat{y}_{test}) - MSE(y_{test}, \hat{y}'_{test})$. |
| 4: Repeat 3 for B times, obtain $l_1, ..., l_B$. Calculate $\Delta_k = \frac{1}{B}\sum_{b=1}^{B} l_b$ and $\hat{\sigma}_k^2 = var(l_1, ..., l_B)$. |
| end. |
| **Step 3:** Calculate statistic $\Delta = \frac{1}{K}\sum \Delta_k$ with limiting distribution $N(0, \sigma^2 = \frac{1}{K}\sum_1^K \sigma_k^2)$, calculate and output p value. |
| **end**. |

For comparison, we also perform the permutation-based association test using the direct application of deep neural networks (PT-DNN) in the focused problem. In other words, we replace the transfer learning predictive model $f_k(\boldsymbol{x}_i|\widetilde{\boldsymbol{S}}_1, \widetilde{\boldsymbol{S}}_2)$ by $f_k(\boldsymbol{x}_i|\widetilde{\boldsymbol{S}}_1, \widetilde{\boldsymbol{S}}_2)$, the deep neural network model solely trained in the focused problem.

The advantage of the proposed test lies in the predictability of the transfer learning method using the candidate gene. On the one hand, when the transfer learning exhibits higher prediction accuracy, it can further reduce the value of the first loss function in $\Delta$, resulting in a smaller p-value. On the other hand, when the transfer learning shows prediction robustness with small prediction standard error, it yields small $\sigma^2$ hence a smaller p value.

## Real data

In this section, we apply the proposed transfer learning method and the association test to investigate the relationships between nicotine addiction and candidate genes based on two relevant projects. Specifically, we conducted two transfer learning case studies: a cross-project case study and a cross-ethnicity case study. We show that TL with large-scale data sets helps improve predictive accuracy and gene selection.

### Data description

Cigarette smoking is one of the leading causes of preventable disease, contributing to 5 million deaths worldwide each year[16]. During the last decade, a great deal of progress has been made in identifying genetic variants associated with smoking. Among those findings, the nAChRs subunit genes (e.g., *CHRNA5*) have been identified and confirmed in several large-scale studies[17]. In this application, we apply the proposed transfer learning to study the complex relationships between the nAChRs subunit genes and nicotine dependence.

The datasets to be analyzed are the large-scale UK Biobank (UKB) dataset and the relatively small-scale dataset from the Study of Addiction: Genetics and Environment (SAGE). UKB is a population-based prospective cohort of nearly 500,000 individuals recruited in the United Kingdom who were aged 40–69 years. UKB contains a wealth of data on detailed clinical information, genome-wide genotype data, and whole-exome sequencing data[18]. SAGE is one of the most comprehensive studies conducted to date, aimed at discovering genetic contributions to substance use disorders. It included about 4,000 participants to study the genetic association with multiple phenotypes, including alcohol, nicotine, and other drug dependence. For the focus of our gene-based analysis, we used cigarettes per day (CPD), available in both SAGE and UKB, as the phenotype and considered genes from two clusters, *CHRNA5-CHRNA3-CHRNB4*[19] and *CHRNB3-CHRNA6*[20] that are associated with nicotine dependence. Before the analysis, we re-assessed the quality of the data (e.g., checking for successful genotype calls, missing rates, deviation from the Hardy-Weinberg equilibrium, and unexpected relationships).

Two studies of transfer learning are investigated in this section. One is a cross-project study and the other is a cross-ethnicity study. In the first study, we apply the transfer learning from the white British population to the black and the white Irish populations. In the second study, we transfer the model trained from UKB to SAGE to demonstrate the performance of the transfer learning approach.

To illustrate the implementation and evaluation of the transfer learning procedure, we split the dataset into the train set (80%) and the test set (20%). We train the original NN model using the source data, transfer the trained model to the train set, and compare their performance on the test set. To avoid chance findings due to the data splitting, we repeat the random splitting 100 times, train the model on the train set of each split, and then average the evaluation metrics on the test sets to assess model performance more reliably. A validation process is used to determine the value of $\lambda$. In the validation process, we split our train set into the subtrain and validation sets with a ratio of 4:1. We evaluate a range of possible values of $\lambda$, in the subtrain set, and then select the optimal $\lambda$ based on the validation set. After a careful quality assessment, the hyperparameter we consider for lambda is $\{10^{-1}, 10^{-0.5}, 1, 10^{0.5}, 10^{1}\}$. The learning rates $r_1$ and $r_2$ are both set as $10^{-3}$. The initial values of parameters are generated by a normal distribution whose mean and standard deviation are 0 and $10^{-3}$, respectively.

### Cross-project transfer learning

In this section, we transfer the model parameters from the large-scale UKB data to the relatively small-scale SAGE data. For this analysis, we focus on the Caucasian population in both datasets. The sample sizes of the population are 288,039 in UKB and 2,517 in SAGE.

For model prediction performance, we compute the relative efficiency criterion $R^2_{pseudo}$ defined in Eqn (6) in both the train and test sets. The performance of the proposed TL-DNN and the direct application of DNN without transfer learning is shown in Fig. 3.

As we can see from Fig. 3, the TL-DNN model outperforms the DNN with higher prediction efficiency in the test set for all five candidate genes. The TL-DNN also shows its robustness to overfitting while the DNN suffers from overfitting for most genes.

The permutation-based association test is used to evaluate the association of the five candidate genes for nicotine dependence. Table 2 summarizes the results of the PT-DNN applied in the UKB Caucasian data, indicating all five candidate genes are associated with nicotine dependence at the significance level of 0.05.

When testing the genetic associations in the smaller SAGE data set, as shown in Table 3, PT-DNN without the transfer learning step identified four candidate genes but failed to identify *CHRNB4*

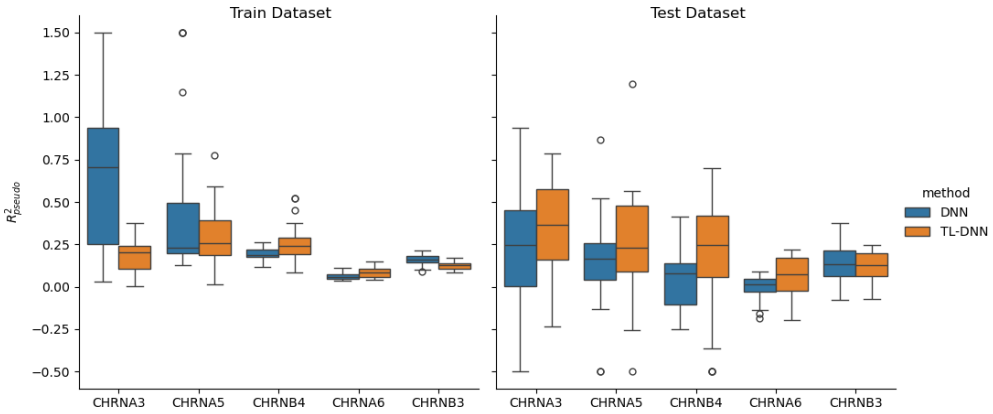


**Fig. 3** Prediction comparison regarding relative efficiency in the SAGE data set between the transfer learning (TL-DNN) from UK Biobank and the direct application of DNN without transfer learning.

($p$-value = 0.636); with the transfer learning step, PT-TL-DNN identified that all five genes are significantly associated with nicotine dependence with *CHRNB4* having a $p$-value of 0.0256. Moreover, PT-TL-DNN provided smaller p-values for all genes except *CHRNB3* compared to PT-DNN. It is also found that PT-TL-DNN yielded smaller estimated standard deviations $\hat{\sigma}$ for most genes compared to PT-DNN.

**Table 2.** PT-DNN results from the association of the five candidate genes in the UKB Caucasian sample.

| Gene | $\Delta$ | $\hat{\sigma}$ | $p$-value |
|---|---|---|---|
| *CHRNA3* | $-1.33e^{-3}$ | $1.10e^{-4}$ | 0 |
| *CHRNA5* | $-1.13e^{-3}$ | $1.01e^{-4}$ | 0 |
| *CHRNA6* | $-8.24e^{-5}$ | $3.19e^{-5}$ | $4.88e^{-3}$ |
| *CHRNB3* | $-1.20e^{-4}$ | $4.13e^{-5}$ | $2.19e^{-3}$ |
| *CHRNB4* | $-1.4e^{-3}$ | $1.11e^{-4}$ | 0 |

**Table 3.** Comparison between the permutation-based test without transfer learning (PT-DNN) and with transfer learning (PT-TL-DNN) in the SAGE data set.

| Gene | PT-DNN | | | PT-TL-DNN | | |
|---|---|---|---|---|---|---|
| | $\Delta$ | $\hat{\sigma}$ | $p$-value | $\Delta$ | $\hat{\sigma}$ | $p$-value |
| *CHRNA3* | −0.0112 | $3.81e^{-3}$ | $1.66e^{-3}$ | $-8.28e^{-3}$ | $2.78e^{-3}$ | $1.48e^{-3}$ |
| *CHRNA5* | $-8.64e^{-3}$ | $3.63e^{-3}$ | $8.58e^{-3}$ | $-7.79e^{-3}$ | $3.26e^{-3}$ | $8.41e^{-3}$ |
| *CHRNA6* | $-9.16e^{-3}$ | $3.18e^{-3}$ | $1.97e^{-3}$ | $-6.54e^{-3}$ | $2.26e^{-3}$ | $1.91e^{-3}$ |
| *CHRNB3* | −0.0139 | $3.20e^{-3}$ | $7.35e^{-6}$ | $-7.75e^{-3}$ | $2.53e^{-3}$ | $1.09e^{-3}$ |
| *CHRNB4* | $4.85e^{-8}$ | $1.39e^{-7}$ | 0.636 | $-5.15e^{-3}$ | $2.64e^{-3}$ | 0.0256 |

## Cross-ethnicity transfer learning

The vast amount of genetic data collected from the Caucasian population provides us with a great resource for genetic research in other populations, especially minority populations. In the UKB dataset, there are 271,240 white British, 7,349 white Irish, and 6,219 black. In this case study, our target populations are the white Irish population and the black population, while the white British population is used as the source population. We transfer the model parameters trained from the white British population to either the white Irish or black population.

Similar to above, the prediction performance of the proposed TL-DNN and the direct application of DNN without transfer learning is shown in Fig. 4 for the white Irish population, and Fig. 5 for the black population.

Both Figs 4 and 5 show that transfer learning achieves higher relative efficiency for *CHRNA3, CHRNA5,* and *CHRNB4*, while it shows similar or slightly worse performance for *CHRNA6* and *CHRNB3*. These results may be related to the genetic heterogeneity of *CHRNA6* and *CHRNB3*, while the heterogeneity is not significant in the other three genes, which is consistent with our previous finding[21].

For the association analysis, Table 4 summarizes the result of the PT-DNN in the UKB white British sample, indicating all five candidate genes are significantly associated with nicotine dependence.

However, when applying PT-DNN to the UKB white Irish and black samples, whose sample sizes are limited, it yielded less significant results. Nevertheless, the proposed PT-TL-DNN had improved power to test all the candidate genes by using the model parameters learned from the white British sample. Tables 5 and 6 present the

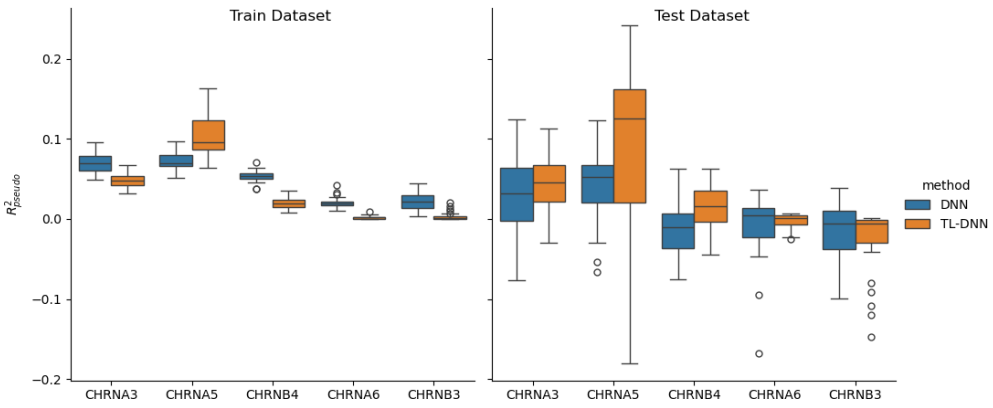


**Fig. 4** Prediction comparison regarding relative efficiency in the SAGE data set between the transfer learning (TL-DNN) from UK Biobank and the direct application of DNN without transfer learning.

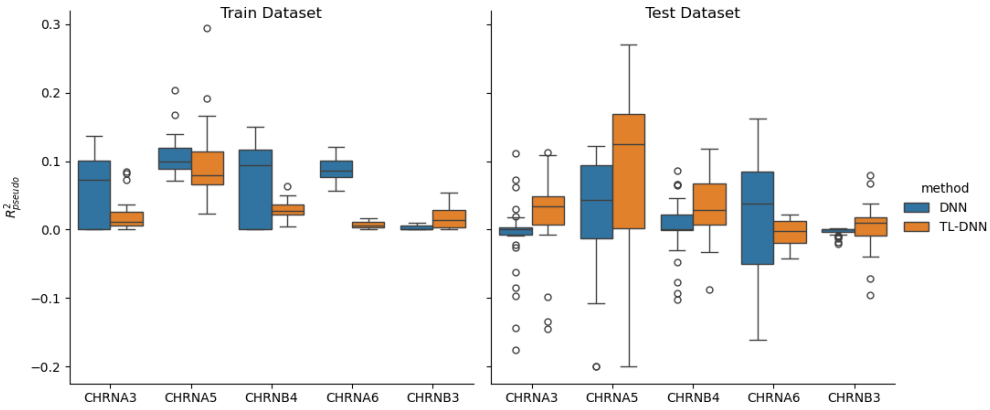


**Fig. 5** Prediction comparison regarding relative efficiency in the UKB black population between the transfer learning (TL-DNN) from the white British population and the direct application of DNN without transfer learning.

**Table 4.** PT-DNN results from the association analysis of five candidate genes in the UKB white British sample.

| Gene | $\Delta$ | $\hat{\sigma}$ | *p*-value |
|---|---|---|---|
| *CHRNA3* | $-6.6e^{-4}$ | $9.38e^{-5}$ | $1.02e^{-12}$ |
| *CHRNA5* | $-6.20e^{-4}$ | $1.07e^{-4}$ | $3.83e^{-9}$ |
| *CHRNA6* | $-6.09e^{-5}$ | $2.39e^{-5}$ | $5.474e^{-3}$ |
| *CHRNB3* | $-1.00e^{-4}$ | $3.88e^{-5}$ | $4.63e^{-3}$ |
| *CHRNB4* | $-1.02e^{-3}$ | $1.27e^{-4}$ | $5.00e^{-16}$ |

testing results from PT-DNN and PT-TL-DNN based on the UKB white Irish and black, respectively. As seen in Table 5, PT-TL-DNN identified all five genes are significantly associated with nicotine dependence in the white Irish sample while PT-DNN failed to detect the association of three genes (*CHRNA5*, *CHRNA6,* and *CHRNB4*). Similarly, for the association test in the UKB black sample, PT-DNN only identified one gene (*CHRNA3*) but our proposed PT-TL-DNN successfully identified four genes, (*CHRNA3*, *CHRNA5*, *CHRNA6,* and *CHRNB4*), associated with nicotine dependence.

## Discussion

Our capacity to detect novel genes in small-scale studies or minority populations (e.g., African Americans) is often limited by the small sample size. The vast amount of genetic data collected from biobank projects and the Caucasian population provides us with an additional resource for genetic research in small-scale studies or minority populations. In this paper, we have proposed a transfer learning procedure in deep neural networks for genetic risk prediction and genetic association analyses in small-scale studies or different populations. Through a cross-project study and a cross-ethnicity study, we demonstrate the advantages of transfer learning in terms of prediction accuracy and testing power.

To further address the potential genetic difference among ethnic groups, we investigated the individual SNP *p*-values and effect sizes of all five genes in the white British, white Irish, and black populations, respectively, using software PLINK[22] and R package CMPlot[23]. From Fig. 6, we observe that the three populations show different patterns in genes *CHRNA6* and *CHRNB3*. Moreover, in the black population, there are more significant SNPs in these two genes compared to the other three genes. These findings indicate that the DNN without transfer learning performs more effectively compared to the transfer learning in the black samples. On the other hand, for the three genes with more significant SNPs in the white British population, transfer learning attains better performance than DNN alone. These observations may partially explain the difference in model prediction performances in the cross-ethnicity case study.

Transfer learning is expected to become one of the key drivers of machine learning success. It does not require data from two studies drawn from the same feature space and the same distribution[5]. In most cases, as long as part of the model parameters are shared between two studies, transfer learning can help improve performance. In our studies, we transferred the model parameters among different studies and populations. In most scenarios, we found that transfer learning improved the models' accuracy and enhanced the test's power. Nevertheless, when a large discrepancy between the source data and the primary data exists, we expect poor performance of the transfer learning. Under such a circumstance, both datasets should be carefully examined, and additional procedures (e.g., testing the heterogeneity of data resources) need to be implemented in the transfer learning approach[14]. Regarding the theoretical basis of transfer learning, we refer to previous studies[24–26].

Besides the application of transfer learning in small-scale studies or minority populations, it can be used for other purposes, such as transfer learning from different species. While human studies play a significant role in genetic research, human research can be

**Table 5.** Comparison between the permutation-based test without transfer learning (PT-DNN) and with transfer learning (PT-TL-DNN) in the UKB white Irish sample.

| Gene | PT-DNN | | | PT-TL-DNN | | |
|---|---|---|---|---|---|---|
| | $\Delta$ | $\hat{\sigma}$ | *p*-value | $\Delta$ | $\hat{\sigma}$ | *p*-value |
| *CHRNA3* | $-1.15e^{-3}$ | $6.48e^{-4}$ | 0.0378 | $-6.95e^{-4}$ | $3.67e^{-4}$ | 0.0291 |
| *CHRNA5* | $-8.40e^{-4}$ | $5.22e^{-4}$ | 0.0529 | $-1.81e^{-3}$ | $7.9e^{-4}$ | 0.0110 |
| *CHRNA6* | $-4.20e^{-4}$ | $5.00e^{-4}$ | 0.201 | $-1.07e^{-3}$ | $3.97e^{-4}$ | $3.58e^{-3}$ |
| *CHRNB3* | $-1.58e^{-3}$ | $7.14e^{-4}$ | 0.0132 | $-2.59e^{-3}$ | $9.68e^{-4}$ | $3.75e^{-3}$ |
| *CHRNB4* | $8.60e^{-4}$ | $6.09e^{-4}$ | 0.0789 | $-1.02e^{-3}$ | $4.68e^{-4}$ | 0.0145 |

**Table 6.** Comparison between the permutation-based test without transfer learning (PT-DNN) and with transfer learning (PT-TL-DNN) in the UKB black sample.

| Gene | PT-DNN | | | PT-TL-DNN | | |
|---|---|---|---|---|---|---|
| | $\Delta$ | $\hat{\sigma}$ | *p*-value | $\Delta$ | $\hat{\sigma}$ | *p*-value |
| *CHRNA3* | $-3.40e^{-3}$ | $1.73e^{-3}$ | 0.0247 | $-2.9e^{-3}$ | $9.2e^{-4}$ | $9.1e^{-4}$ |
| *CHRNA5* | $-2.50e^{-4}$ | $4.94e^{-4}$ | 0.305 | $-5.00e^{-3}$ | $1.80e^{-3}$ | $2.69e^{-3}$ |
| *CHRNA6* | $-1.09e^{-5}$ | $2.34e^{-5}$ | 0.679 | $-3.60e^{-3}$ | $1.12e^{-3}$ | $5.90e^{-4}$ |
| *CHRNB3* | $1.88e^{-11}$ | $4.35e^{-11}$ | 0.667 | $-1.20e^{-3}$ | $1.28e^{-3}$ | 0.183 |
| *CHRNB4* | $-1.62e^{-3}$ | $1.07e^{-3}$ | 0.0651 | $-2.80e^{-3}$ | $1.53e^{-3}$ | 0.0336 |

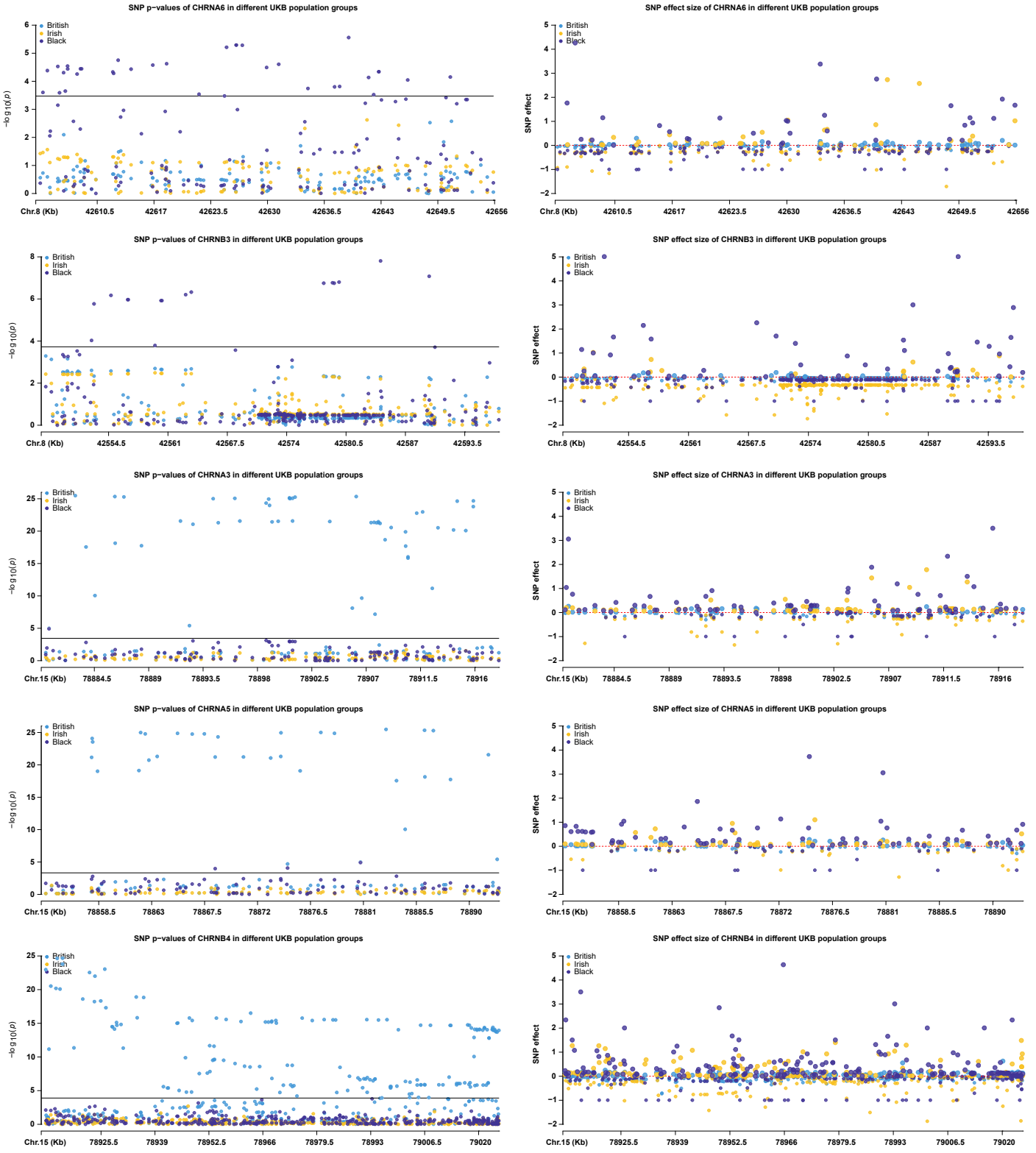


**Fig. 6** Illustration of genetic heterogeneity among the ethnic UKB population groups regarding individual SNP *p*-values and affect sizes of the five genes in populations white British, white Irish, and black.

restricted due to study design and high cost. In contrast, animal research is more flexible and can adopt different designs. Given the experimental results generated from the well-controlled conditions, it would be valuable to further transfer these results to humans.

Limitations of the study include the limited sample sizes and the generality of the results in the two case studies. The proposed transfer learning techniques can be further explored and applied to a wider range of applications, such as different types of omics data, other complex traits, and diverse populations. Just like most applications of transfer learning, how to measure the transferability across domains and avoid negative transfer remains an important issue. Finally, despite the advances in the theoretical basis of transfer

learning, theoretical studies for the proposed deep transfer learning can be further conducted in the future to better understand the properties of the proposed deep transfer learning.

## Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Lu Q, Geng P; analysis: Zhang S, Zhou Y; interpretation of results: Lu Q, Geng P, Dong K, Liu J; manuscript preparation: Geng P, Zhang S, Zhou Y. All authors reviewed the results and approved the final version of the manuscript.

## Data availability

All data generated or analyzed during this study are included in this published article, and Python codes are available on GitHub (https://github.com/DianaYuanZhou/PT-TL-DNN).

## Acknowledgments

The authors wish to thank the two reviewers for their comments which greatly improved the manuscript.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Dates

Received 1 April 2025; Revised 30 May 2025; Accepted 10 June 2025; Published online 28 July 2025

## References

1. Iman M, Arabnia HR, Rasheed K. 2023. A review of deep transfer learning and recent advancements. *Technologies* 11(2):40
2. Torrey L, Shavlik J. 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, eds. Soria E, Martín-Guerrero JD, Martinez M, Magdalena R, Serrano AJ. USA: IGI Global Scientific Publishing. pp. 242–64. doi: 10.4018/978-1-60566-766-9.ch011
3. Li S, Cai TT, Li H. 2022. Transfer learning for high-dimensional linear regression: prediction, estimation and minimax optimality. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84(1):149–73
4. Tian Y, Feng Y. 2023. Transfer learning under high-dimensional generalized linear models. *Journal of the American Statistical Association* 118(544):2684–97
5. Pan SJ, Yang Q. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10):1345–59
6. Ribani R, Marengoni M. 2019. A survey of transfer learning for convolutional neural networks. *2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T), Rio de Janeiro, Brazil, 2019*. USA: IEEE. pp.47–57. doi: 10.1109/SIBGRAPI-T.2019.00010
7. Tan C, Sun F, Kong T, Zhang W, Yang C, et al. 2018. A survey on deep transfer learning. *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks Proceedings*. Cham: Springer. pp. 270–79. doi: 10.1007/978-3-030-01424-7_27
8. López-García G, Jerez JM, Franco L, Veredas FJ. 2020. Transfer learning with convolutional neural networks for cancer survival prediction using gene-expression data. *PLoS One* 15(3):e0230536
9. Muneeb M, Feng S, Henschel A. 2022. Transfer learning for genotype–phenotype prediction using deep learning models. *BMC Bioinformatics* 23(1):511
10. Altmann A, Toloşi L, Sander O, Lengauer T. 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* 26(10):1340–47
11. Mi X, Zou B, Zou F, Hu J. 2021. Permutation-based identification of important biomarkers for complex diseases via machine learning models. *Nature communications* 12(1):3008
12. Liu L, Meng Q, Weng C, Lu Q, Wang T, et al. 2022. Explainable deep transfer learning model for disease risk prediction using high-dimensional genomic data. *PLoS Computational Biology* 18(7):e1010328
13. Wang J, Zhang H, Wang J, Pu Y, Pal NR. 2021. Feature selection using a neural network with group lasso regularization and controlled redundancy. *IEEE Transactions on Neural Networks and Learning Systems* 32(3):1110–23
14. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, et al. 2020. A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109(1):43–76
15. Kingma DP. 2014. Adam: a method for stochastic optimization. *arXiv* Preprint
16. Mathers CD, Loncar D. 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine* 3(11):e442
17. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, et al. 2010. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature Genetics* 42(5):436–40
18. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562:203–9
19. Weiss RB, Baker TB, Cannon DS, von Niederhausern A, Dunn DM, et al. 2008. A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genetics* 4(7):e1000125
20. Zeiger JS, Haberstick BC, Schlaepfer I, Collins AC, Corley RP, et al. 2008. The neuronal nicotinic receptor subunit genes (CHRNA6 and CHRNB3) are associated with subjective responses to tobacco. *Human Molecular Genetics* 17(5):724–34
21. Zhang X, Lan T, Wang T, Xue W, Tong X, et al. 2019. Considering genetic heterogeneity in the association analysis finds genes associated with nicotine dependence. *Frontiers in Genetics* 10:448
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81(3):559–75
23. Yin L, Zhang H, Tang Z, Xu J, Yin D, et al. 2021. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, proteomics & bioinformatics* 19(4):619–28
24. Cody T, Beling PA. 2023. A systems theory of transfer learning. *IEEE Systems Journal* 17(1):26–37
25. Tripuraneni N, Jordan M, Jin C. 2020. On the theory of transfer learning: the importance of task diversity. *Advances in Neural Information Processing Systems: Annual Conference on Neural Information Processing Systems (NeurIPS 2020)*. virtual. pp. 7852–62
26. Yang L, Hanneke S, Carbonell J. 2013. A theory of transfer learning with applications to active learning. *Machine Learning* 90:161–89