RESEARCH

Joint learning sample similarity and correlation representation for cancer survival prediction

Yaru Hao[1*], Xiao-Yuan Jing[1,2,3*] and Qixing Sun[1]

*Correspondence:
hyr2018@whu.edu.cn;
jingxy_2000@126.com

[1] School of Computer Science,
Wuhan University, Wuhan, China
[2] Guangdong Provincial Key
Laboratory of Petrochemical
Equipment Fault Diagnosis
and School of Computer,
Guangdong University
of Petrochemical Technology,
Maoming, China
[3] State Key Laboratory for Novel
Software Technology, Nanjing
University, Nanjing, China

Abstract

Background: As a highly aggressive disease, cancer has been becoming the leading death cause around the world. Accurate prediction of the survival expectancy for cancer patients is significant, which can help clinicians make appropriate therapeutic schemes. With the high-throughput sequencing technology becoming more and more cost-effective, integrating multi-type genome-wide data has been a promising method in cancer survival prediction. Based on these genomic data, some data-integration methods for cancer survival prediction have been proposed. However, existing methods fail to simultaneously utilize feature information and structure information of multi-type genome-wide data.

Results: We propose a Multi-type Data Joint Learning (MDJL) approach based on multi-type genome-wide data, which comprehensively exploits feature information and structure information. Specifically, MDJL exploits correlation representations between any two data types by cross-correlation calculation for learning discriminant features. Moreover, based on the learned multiple correlation representations, MDJL constructs sample similarity matrices for capturing global and local structures across different data types. With the learned discriminant representation matrix and fused similarity matrix, MDJL constructs graph convolutional network with Cox loss for survival prediction.

Conclusions: Experimental results demonstrate that our approach substantially outperforms established integrative methods and is effective for cancer survival prediction.

Keywords: Cancer survival prediction, Feature information, Structure information, Correlation representation, Similarity matrix

Introduction

Cancer has been becoming the leading death cause all over the world, which seriously affects human health and living quality [1, 2]. In addition, the mortality rates increase year by year [3–5]. Prognosis prediction can aid physicians significantly in making decisions about care and treatment of cancer patients [6, 7]. Prognosis prediction usually can be described as a censored survival analysis problem, which predicts whether and when

Hao *et al. BMC Bioinformatics*   (2022) 23:553

Page 2 of 20

a death will occur within a given time period [8, 9]. In the past few decade, many survival prediction methods have been proposed, such as standard Cox regression and its extensions [10], tree-based ensemble methods, random survival forests [11], and so on.

Historically, cancer survival prediction works mainly based on histopathological descriptors and low-dimensional clinical data, such as sex, age at diagnosis, cancer grade detail, body fat rate and other clinical features [12–14]. However, clinical practice has found that genomic data tends to contain more molecular biomarkers associated with cancer and thereby can describe the cancer more comprehensively [15, 16]. Meanwhile, with the advance of Human Genome Project, high-throughput sequencing technology becomes cost-effective, which makes it progressively easier to achieve multiple and diverse genome-scale data sets to address clinical and biological questions [17]. In general terms, the above multi-type data describing the same cancer can be regarded as multimodal data. Specifically, multimodal data has two basic characteristics [18–20]. On the one hand, they share the common information both in feature level and structure level. On the other hand, each modality has its own specific information both in feature level and structure level. Compared with single genetic data type, multiple genome-scale data sets can capture more comprehensive information for cancer. Therefore, it is essential and feasible to develop new data-integration algorithms especially for utilizing multi-type high-dimensionality genomic data to capture comprehensive information for cancer.

Motivation

During the past several years, many researchers have been devoted to construct data-integration methods based on binary classification models for cancer survival prediction. In this technology, cancer patients are usually classified to the short or long survival group according to a predefined threshold (e.g., 3 years). For example, Zhang et al. [21] presented a multiple kernel machine learning method combined with min-redundancy max-relevance (mRMR) feature selection algorithm to predict 2-year survival rate of glioblastoma multiforme patients. Zhao et al. [22] studied various prediction methods including ensemble models (Gradient Boosting and Random Forest), support vector machine and artificial neural networks to predict 5-year survival rate of breast cancer by fusing gene expression data, clinical data and pathological images. Unfortunately, this technology reduces the survival analysis to a classification problem, which is counter-practical and far less useful than the estimation of survival times. Another mainstream technology for survival prediction is survival risk regression based methods, such as Cox proportional hazards (Cox-PH) model [23, 24]. Different from binary classification methods, this technology focuses on whether a patient survives at a certain time point rather than when the patient dies, which can handle both uncensored and censored samples. Therefore, patients who survive at a certain time point can be used in modelling patient survivals [25].

Although existing works have promoted the development of data-integration methods in cancer survival prediction, there are two limitations to develop this technology: *(i) simultaneously utilizing structure information and feature information, specifically for small scale dataset; (ii) fully utilizing multi-type data for learning effective discriminant features. Here, structure information points to the information of data distribution within*

data types. Feature information refers to the information contained in the data (such as genes) within a sample. Discriminant features refer to the features learned from original data (such as gene sequences) by utilizing feature learning algorithms, which is useful to separate the samples with different survival time [26]. Existing data-integration methods for cancer survival prediction have yet to address all of these limitations together. In addition, with excellent feature learning ability, the neural network extension of the Cox model has proved its better performance than traditional Cox-PH models in survival prediction, especially for high-throughput sequencing data. Hence, we intend to apply it to our work. In addition, we introduce similarity matrix to exploit structure information, which can access structural information hidden in multi-type data.

Inspired by the above analysis, we intend to design a Multi-type Data Joint Learning (MDJL) approach to obtain a reliable similarity matrix for exploiting structure information and an effective discriminant feature representation for exploiting feature information. In our proposed MDJL, (a) structure information and feature information can be simultaneously utilized; (b) the discriminant feature representations are exploited by learning correlation representations between any two data types, which can ensure the diversity and provide complementary information; (c) the constructed similarity matrices can explore useful structure information even from a small-scale samples.

Contribution

The main contributions of our approach lie in three aspects:

1. Different from existing survival prediction methods, we present a Multi-type genome-wide Data Joint Learning (MDJL) approach for cancer survival prediction, which achieves both a fused similarity matrix and an integrated discriminant feature representation for simultaneously utilizing structure information and feature information.
2. MDJL exploits correlation representations between any two data types by cross-correlation calculation for learning discriminant features. Moreover, based on the learned correlation representations, MDJL constructs sample similarity matrices for capturing global and local structures across different data types. With the learned discriminant representations and similarity matrices, MDJL constructs graph convolutional network with Cox loss for survival prediction.
3. We conduct a number of experiments on four public cancer datasets. Experimental results show that our approach can achieve higher prediction performance than competing methods. Further investigation not only demonstrate the effectiveness of each component for MDJL, i.e., correlation representations extraction component and similarity matrices construction component, but also indicate the robustness.

Organization

The rest of this paper is organized as follows: Sect. Motivation reviews related cancer survival prediction works. The proposed approach and detailed algorithm are introduced in Sect. Contribution. Section Organization talks about the experimental results.

Section Related works conducts further experiments to investigate our approach. Section Binary classification based survival prediction works concludes this paper.

Related works

Binary classification based survival prediction works

In the past few decades, a variety of binary classification based multimodal learning methods for survival prediction have been proposed. In general terms, a modality refers to a kind of data type. These methods mainly focus on learning fused representation from multiple data sources, such as clinical data, histopathological images markers and genomic data [27–31]. With multiple types of data, some data-integration strategies such as joint-based strategy [32, 33] and alignment-based strategy [34–36] have been presented. Joint-based methods utilize multi-type data mainly by concatenating multi-type data into one unified feature matrix. For example, Sun et al. [37] presented a triple model DNN to respectively learn feature representations from gene expression, copy number alteration and clinical data, and then concatenated the learned multiple representations into one unified matrix. To explore the inherent relation between samples and multi-type genomic data, Gao et al. [38] constructed bipartite graphs between patients and gene expression, copy number alteration. Khademi et al. [39] integrated microarray data and clinical data through the probabilistic graph model for prognosis of breast cancer. Methods based on alignment strategy utilize multiple types of data by maximizing the common information across different data types. For example, Wang et al. [40] designed a cluster-boosted multi-task learning approach to exploit the common information across different data types for survival analysis. Although these methods have promoted the development of multimodal cancer survival analysis, they are limited to binary classification problem and are counter-practical.

Survival risk regression based survival prediction works

Different from binary classification methods, the survival risk regression methods aim to calculate a risk score for each patient, typically with the Cox-PH model and its extensions [41–43]. For example, to predict an individual survival time, Baek et al. [44] achieved this by integrating hazard network and a distribution function network. Wang et al. [45] proposed a reweighted Lasso-Cox model for cancer survival prediction, which improves the generalization ability of the model by weighing the topologically important genes based on random walk. Considering there are correlations between multi-type genomic data, Bichindaritz et al. [46] presented an adaptive multi-task learning approach for breast cancer survival prediction, which add an auxiliary ordinal loss to the Cox model.

Recently, with the excellent data representation ability and high learning ability, a variety of deep neural networks extension of the Cox-PH model has been proposed [47–50]. For example, instead of learning linear relationship in the Cox-PH model, both Deep-Surv [51] and Cox-nnet [52] introduce neural networks to learn nonlinear feature representation. To fully utilize multi-omics data, Tong et al. [53] designed a concatenation autoencoder to concatenate the learned multiple hidden representations from each data type. In addition, to achieve the consensus representation across multi-omics data, they designed a cross-modality autoencoder to maximize the agreement across modalities.

Hao *et al. BMC Bioinformatics*   (2022) 23:553

Page 5 of 20

Cheerla et al. [54] presented an unsupervised encoder extension of the Cox model to integrate multi-type data into one single feature matrix, which introduces similarity loss to force four data sources align the common information. To eliminate the estimation bias in processing such datasets with a large number of censored samples, Zhang et al. [55] introduced Bayesian Perturbation to approximate the prior knowledge of censored samples to optimize the training process of model. To address the limitation that deep networks tend to fall into over-fitting with small sample size high feature dimension, Qiu et al. [56] present a meta-learning approach based on neural networks for cancer survival prediction. In addition, Kvamme et al. [57] imposed L_1 and L_2 regulation terms on the network parameters to reduce the over-fitting problem. However, these methods mainly exploit feature information but fail to exploit useful structure information.

Similarity matrices construction works

Similarity matrix construction has been widely used in multi-view clustering tasks. Usually, existing methods tend to construct similarity matrix for each data types, based on which they learn a shared similarity matrix of all data types. For example, Zhan et al. [58] learned the consensus similarity graph by minimizing disagreement between different views with a disagreement cost function. To address the limitation that incomplete multi-view clustering fails to exploit hidden information of missing views and handle the information imbalance across different views, Wen et al. [59] designed adaptive weights to balance the importance of different views. Wang et al. [60] designed a multi-view subspace clustering approach, which adopts the Hilbert-Schmidt Independence Criterion to enforce the similarity of similarity matrix have maximum dependence. Chen et al. [61] designed a nonlinear method for multi-view clustering, which jointly learn kernel representation matrix and similarity matrix. Zhang et al. [62] presented an anchor-based approach for multi-view semi-supervised, which constructs the affinity graphs by using an anchor-based strategy and obtains the optimal consensus graph by using feature and label information. Considering that original multi-view data often contain abundant noise and outliers, Xie et al. [63] learned latent feature representation based on the adaptively learned graph. It also introduces Laplacian embedding to maintain the local manifold structure. Zhang et al. [64] constructed a unified similarity matrix for multiple views by utilizing a latent representation explored from the underlying complementary information. Huang et al. [65] integrated similarity learning and local embedding into a unified framework, which constructs a fused similarity matrix and learns a latent low-dimensional representation for capturing the underlying structure. For preserving global structures and obtaining local structures, Wan et al. [66] proposed an embedding method for multi-view clustering, which integrates all views into a combination weight matrix for maintaining global structures and imposes constraint on the learned shared affinity matrix for obtaining the local structure.

Proposed method

In this paper, we propose a Multi-type Data Joint Learning (MDJL) approach for cancer survival prediction based on multi-type genome-wide data. Specifically, instead of exploiting common feature information shared by all data types, we exploit

Hao *et al. BMC Bioinformatics*   (2022) 23:553

Page 6 of 20

Table 1 Summary of the key notations used in the paper

Notations	Explanations
$\mathbf{x}^v \in \mathbb{R}^{d^v \times N}$	Sample set of v-th data type
$x_i^v \in \mathbb{R}^{d^v}$	The i-th sample in v-th data type
f_v	Fully connection neural network used for feature learning
$\mathbf{w}_{f_v}^l \in \mathbb{R}^{m_l \times m_{l-1}}$	The l-th layer weight matrix of neural network f_v
$\mathbf{y}^v \in \mathbb{R}^{d \times N}$	The learned feature representation from $\mathbf{x}^v$ with f_v
$\chi^{v,u} \in \mathbb{R}^{d \times d \times N}$	Interactive map set between data type v and data type u
$\chi_i^{v,u} \in \mathbb{R}^{d \times d}$	Interactive map of i-th sample between data type v and data type u
$\mathbf{y}^{v,u} \in \mathbb{R}^{d \times N}$	The correlation representation of $\mathbf{x}^v$ and $\mathbf{x}^u$
$\mathbf{y} \in \mathbb{R}^{(V(V-1)/2) \times N}$	Fused correlation representation
P^m	Normalized weight matrix
S^m	K nearest similarity matrix
P	Fused similarity matrix
z_i	The output of graph convolutional network
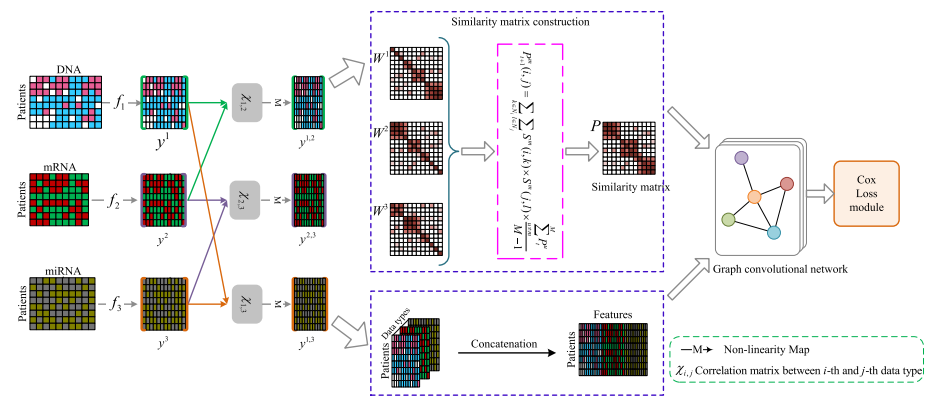
Fig. 1 The architecture of our proposed MDJL

correlation/common feature information between any two data types for exploring diverse and complementary feature information across multiple data types. Secondly, we fully utilize the global and local structure to construct similarity matrices based on the learned multiple correlation representations. Here, global structure refers to the similar structure information across different data types, local structure refers to the neighborhood information within data types. The main architecture of our MDJL approach is illustrated in Fig. 1. MDJL consists of four components: (1) correlation representations extraction component, which is designed for utilizing diverse and complementary feature information across multiple data types by learning correlation representations between any two data types; (2) discriminant representations generation component, which is designed for fusing multiple correlation representations by concatenation; (3) similarity matrices construction component, which is designed for generating sample similarity matrix by fully utilizing both global and local structure across different data types; and (4) graph convolutional network construction

component, which is used for predicting the survival risk for patients. Key notations used in this paper are listed in Table 1.

Correlation representations extraction

Suppose there are N samples and V different data types. Let $\mathbf{x}^v = \left\{ x_i^v \in \mathbb{R}^{d_v} \right\}_{i=1}^N$ be the sample set of the v-th data type, and x_i^v represents the i-th sample of data type v, d_v is the feature dimensionality of $\mathbf{x}^v$, where $v = 1, 2, \ldots, V$. For correlation representation extraction, we firstly define V neural networks $\left\{ f_v \right\}_{v=1}^V$ to conduct feature learning and project $\mathbf{x}^v$ from space $\mathbb{R}^{d_v}$ into space $\mathbb{R}^d$, that is,

$$\mathbf{y}^v = f_v\left(\mathbf{x}^v\right), \tag{1}$$

where $\mathbf{y}^v \in \mathbb{R}^{d \times N}$, and f_v points to a neural network with $L = 3$ layers,

$$\mathbf{h}_{f_v}^l = \sigma\left(\mathbf{w}_{f_v}^l \mathbf{h}_{f_v}^{l-1} + \mathbf{b}_{f_v}^l \right). \tag{2}$$

For the l-th layer $(l = 1, 2, \ldots, L)$, $\mathbf{w}_{f_v}^l \in \mathbb{R}^{m_l \times m_{l-1}}$ denotes the weight matrix $(m_0 = d_v, m_L = d)$, $\mathbf{b}_{f_v}^l \in \mathbb{R}^{m_l}$ is the bias vector, $\mathbf{h}_{f_v}^l \in \mathbb{R}^{m_l}$ denotes the output of the l-th layer $\left(\mathbf{h}_{f_v}^0 = \mathbf{x}^v, \mathbf{h}_{f_v}^L = \mathbf{y}^v \right)$, and σ is the acivation function.

To further explore the correlation representations between any two data types, we borrow correlation computation proposed in [67]. Following work [67], for the i-th sample, the interactive map $\chi_i^{v,u}$ of y_i^v and y_i^u can be defined as,

$$\chi_i^{v,u} = y_i^v \otimes y_i^u, \tag{3}$$

where $v \neq u$, $\otimes$ is outer product, $\chi^{v,u} = \left\{ \chi_i^{v,u} \in \mathbb{R}^{d \times d} \right\}_{i=1}^N$, $\chi^{v,u} = \chi^{u,v}$.

Based on the interactive map set, we further construct a set of neural networks $\psi = \left\{ \psi_{v,u} \right\}_{v,u=\{1,\ldots,V\}, v \neq u}$ to project each $\chi^{v,u}$ from space $\mathbb{R}^{d \times d}$ into an embedded space $\mathbb{R}^d$, which learns deep correlation representations between any two data types. That is,

$$\mathbf{y}^{v,u} = \psi_{v,u}\left(\chi^{v,u}\right) = \sigma\left(\mathbf{w}_{\psi_{v,u}} \text{vec}\left(\chi^{v,u}\right) + \mathbf{b}_{\psi_{v,u}} \right). \tag{4}$$

where $\mathbf{y}^{v,u} \in \mathbb{R}^{d \times N}$ is the correlation representation of $\mathbf{x}^v$ and $\mathbf{x}^u$, $\mathbf{w}_{\psi_{v,u}} \in \mathbb{R}^{d \times d^2}$, $\mathbf{b}_{\psi_{v,u}} \in \mathbb{R}^d$, vec($\cdot$) represents the vectorization of a matrix.

Discriminant representations generation

Based on the above subsections, we have learned multiple correlation representations from multiple data types. The finally fused correlation feature representation from all pairwise data types can be written as,

$$\begin{aligned}
\mathbf{y} = [&\mathbf{y}^{1,2\,\top}, \mathbf{y}^{1,3\,\top}, \ldots, \mathbf{y}^{1,V\,\top}, \\
&\ldots, \\
&\mathbf{y}^{v,v+1\,\top}, \mathbf{y}^{v,v+2\,\top}, \ldots, \mathbf{y}^{v,V\,\top}, \\
&\ldots, \\
&\mathbf{y}^{V-1,V\,\top}]^\top.
\end{aligned} \tag{5}$$

Hao *et al. BMC Bioinformatics*   (2022) 23:553

Page 8 of 20

Similarity learning of global and local structure

As mentioned above, MDJL aims to learn a fused similarity matrix based on multi-type data. The reliability of the similarity matrices constructed from raw data may be polluted severely by noise and outliers. To enhance the ability to resist noise and outliers, we construct similarity matrices based on the learned multiple correlation representations. By correlation information learning, we collect M different correlation feature representations $\left\{ \mathbf{o}^m = \mathbf{y}^{v,u} \in \mathbb{R}^{d \times N} \right\}_{m=1}^{M}$, where $M = V(V-1)/2$. Based on the multiple correlation representations, similarity learning of global and local structure aims to capture a fused similarity matrix, which preserves sufficient local structure information of samples as well as maintains global structure across different data types. First, we construct the similarity matrix $\mathbf{W}^m = [W^m(i,j)]_{N \times N}$ for the m-th correlation representation $\mathbf{o}^m$ by Gaussian kernel. $W^m(i,j)$ represents the similarity between sample x_i^m and x_j^m in the m-th correlation representation. To integrate these similarity matrices constructed from multiple correlation representations, we introduce a normalized weight matrix P^m as follows:

$$
P^m(i,j) = \begin{cases} \frac{W^m(i,j)}{2\sum_{k \neq i}^{N} W^m(i,k)}, j \neq i \\ 1/2, j = i \end{cases}
\tag{6}
$$

where $\sum_{j=1}^{N} P^m(i,j) = 1$.

In order to measure local similarity, we design a sparse kernel based on K nearest neighbors (KNN), that is:

$$
S^m(i,j) = \begin{cases} \frac{W^m(i,j)}{\sum_{k \in N_i^m} W^m(i,k)}, j \in N_i^m \\ 0, otherwise \end{cases}
\tag{7}
$$

where N_i^m is a set of neighbors for y_j^m. This operation sets the similarities of samples that are non-neighboring to zero, which bases on pairwise samples similarity values.

To obtain fused similarity matrix, we iteratively update P^m with its corresponding local similarity matrix S^m and the similarity matrix $\{P^u\}_{u=\{1,...,M\}\backslash m}$ of other data types, so that the updated $P^m|_{m=1}^{M}$ can be more similar to each other, at the same time, local similarity information can also be preserved.

For m-th correlation representation, we iteratively update P^m as follows:

$$
P_{t+1}^m(i,j) = \sum_{k \in N_i} \sum_{l \in N_j} S^m(i,k) \times S^m(j,l) \times \frac{\sum_{u \neq m}^{M} P_t^u}{M-1}
\tag{8}
$$

After T iterations, the learned $P^m|_{m=1}^{M}$ would be enough similar to each other. Then the fused similarity matrix can be defined as the average of $P^m|_{m=1}^{M}$, that is:

$$
P = \frac{\sum_{m=1}^{M} P_T^m}{M}
\tag{9}
$$

Graph convolutional network

According to correlation representations learning, we obtain the fused discriminant representation matrix $\mathbf{y}$. According to similarity matrices construction, we obtain the fused similarity matrix P. Then the $\mathbf{y}$ and P were used as the input of graph convolutional network for model training and prediction. In this paper, we construct the graph convolutional network $G = f(\mathbf{y}, P)$ with three layers for training and prediction, that is,

$$H_g^{l+1} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{P} \tilde{D}^{-\frac{1}{2}} H_g^l W_g^l \right) \tag{10}$$

where $\tilde{P} = P + I_N$ denotes the adjacency matrix of the undirected graph G with added self-connections. I_N represents identity matrix, $\tilde{D}_{(i,i)} = \sum_j \tilde{P}_{(i,j)}$, W_g^l is trainable weight matrix of the l-th layer, H_g^l points to the matrix of activations in the l-th layer ($H_g^0 = \mathbf{y}$), and σ is the activation function.

To describe the effectiveness of quantitative variables on survival time, we introduce Cox loss as loss function [25], that is,

$$L(\beta) = - \sum_{i:C(i)=1} \left[\phi_i - log \left(\sum_{t_j \geqslant t_i} e^{\phi_j} \right) \right] \tag{11}$$

$$\phi_i = z_i \beta \tag{12}$$

where ϕ_i denotes the log hazard ratio for sample i, z_i denotes the learned vector from graph convolutional network, β represents coefficient weight vector between z_i and the output ϕ_i. $C(i)$ is the censorship flag. If sample i is uncensored sample, $C(i) = 1$, otherwise, if sample i is censored sample, $C(i) = 0$. t_i points to the survival time for patient i, where patient i should be uncensored samples. $t_j \geqslant t_i$ points to the survival time of j-th sample is longer than that of i-th sample, where patient j can comes from either uncensored samples or censored samples.

Optimization

Feedforward and calculate the loss

For each of the V data types, the sample set $\mathbf{x}^v$ are fed forward to the MDJL as in Eq. 1, and the output of the MDJL is denoted as $\{z_i\}_{i=1}^N$. The loss of the whole network is calculated as in Eq. 11, denoted as $L(\beta) = - \sum_{i:C(i)=1} \left[\phi_i - log \left(\sum_{t_j \geqslant t_i} e^{\phi_j} \right) \right]$.

Update neural networks

$\left\{ \{f_v\}_{v=1}^V, \{\psi_{v,u}\}_{v,u=\{1,\dots,V\},v\neq u}, G \right\}$.   The   network   parameters   of $\left\{ \{f_v\}_{v=1}^V, \{\psi_{v,u}\}_{v,u=\{1,\dots,V\},v\neq u}, G \right\}$ can be jonintly optimized by minimizing Eq. 11. We perform batch gradient descent with the whole dataset in each iteration for network training.

Table 2 A brief statistics of all datasets

Cancer type	GBM	KRCCC	LSCC	BIC
Instances	215	122	106	105
Uncensored	199	79	66	83
Description of data types	DNA (12042-D)	DNA (17899-D)	DNA (12042-D)	DNA (17814-D)
	mRNA (534-D)	mRNA (329-D)	mRNA (352-D)	mRNA (354-D)
	miRNA (1305-D)	miRNA (24960-D)	miRNA (23074-D)	miRNA (23094-D)
Average survival time (months)	19.6	42.8	24.3	34.8

Algorithm 1 Algorithm for MDJL

Input: sample set $\left\{ \mathbf{x}^v \in \mathbb{R}^{d_v \times N} \right\}_{v=1}^{V}$, sample survival time set, sample survival status set.

Initialize: hyperparameters K, T.

Update until convergence:

    Forward propagation:

      1. Perform f_v with Eq.1 and then obtain $\mathbf{y}^v$.

      2. Compute interactive map $\chi^{v,u}$ with Eq.3.

      3. Obtain correlation representations $\mathbf{y}^{v,u}$ with Eq.4.

      4. Obtain fused correlation representations with Eq.5.

      5. Construct normalized weight matrix $P^m|_{m=1}^{M}$ with Eq.6.

      6. Construct sparse kernel matrix $S^m|_{m=1}^{M}$ with Eq.7.

      7. Iteratively update $P^m|_{m=1}^{M}$ with Eq.8.

      8. Obtain fused similarity matrix P with Eq.9.

      9. Construct gaph convolutional network G with Eq.10.

    Back propagation:

      Update network parameters of $\{f_v\}_{v=1}^{V}$, $\left\{ \psi^{v,u} \right\}_{v,u=\{1,2,...,V\}, v \neq u}$
      and G by minimizing Eq.11.

Output: The predicted hazard ratios of testing samples.

Algorithm 1 describes the process of cancer survival prediction by using MDJL.

Experiments

Datasets

Four cancer datasets[1] including glioblastoma multiforme (GBM), kidney renal clear cell carcinoma (KRCCC), lung squamous cell carcinoma (LSCC) and breast invasive carcinoma (BIC) are used to evaluate our MDJL approach. For each dataset, we collect three types of genomic data, including DNA methylation, mRNA expression and miRNA expression data. The datasets used in this paper are obtained from http://compbio.cs.toronto.edu/SNF/, which are provided and preprocessed by work [68]. It downloads these data from The Cancer Genome Atlas (TCGA) website and performs three steps of preprocessing: sample selection, missing-data imputation and normalization. Detailed preprocessing process is described as follows: (i) if one patient sample has more than 20% missing data in any data type, then this sample will be removed; (ii) if a certain gene has more than 20% missing values, then this gene will be filtered,
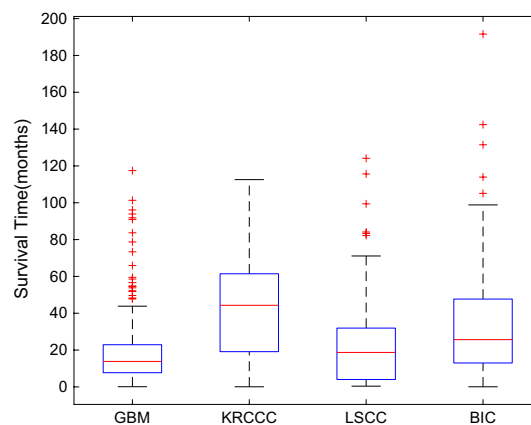
---

[1] http://compbio.cs.toronto.edu/SNF/

Fig. 2 Survival time distribution of patients for four cancers

otherwise, the k-nearest interpolation is used for complementing this gene; (iii) the z-score transformation is used for normalizing the data samples. Table 2 summaries the detailed information of datasets used in experiments. Figure 2 describes the survival time distribution for each cancer, which is represented by box plot.

Experimental settings

Compared methods

To evaluate the performance of our MDJL approach, we compare it with several state-of-the-art cancer survival prediction methods:

- MKL + Cox loss (MKL-Cox). MKL is a multiple kernel learning based binary classification method for cancer survival prediction, which fuses multi-type data using joint strategy [21]. For a fair comparison, we extend MKL with Cox loss.
- MDNNMD + Cox loss (MDNNMD-Cox). MDNNMD is a multimodal deep neural network based binary classification method for cancer survival prediction, which fuses multi-type data using joint strategy [37]. For a fair comparison, we extend MDNNMD with Cox loss.
- DLMR. DLMR is a multimodal deep neural network extension of the Cox model for cancer survival prediction, which fuses multi-type data using alignment strategy [54].
- CrossAE. CrossAE is a cross-modality autoencoder based survival prediction method for utilizing the consensus representations across multi-type data [53].
- VAECox. VAECox is a deep transfer learning architecture for cancer survival prediction based on alignment strategy [25].
- DeepSurv. DeepSurv is a deep learning generalization of the Cox proportional hazards model, which predicts survival risks based on single-type data [51]. For comparison, we use the unified feature matrix concatenated from DNA, mRAN and miRAN as the input for DeepSurv.
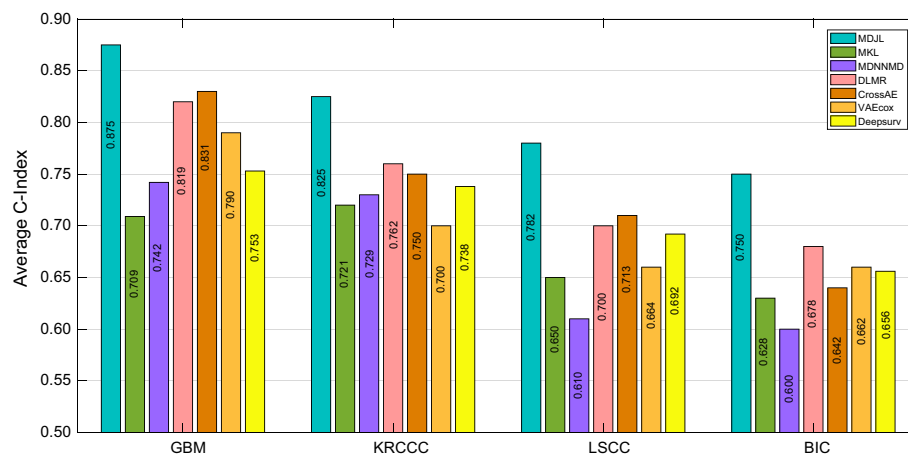
Hao *et al. BMC Bioinformatics*   (2022) 23:553

Page 12 of 20

Fig. 3 The performance of our approach and compared methods on all the datasets

The implementations of MDNNMD-Cox, DLMR, CrossAE, VAECox and DeepSurv are downloaded from the websites provided by their authors. With there are no public codes for MKL-Cox, we implement MKL-Cox by ourselves.

Implementation details

All these methods are evaluated on GBM, KRCCC, LSCC and BIC datasets. For each cancer dataset, we randomly select 70% data for training and utilize the rest of 30% for testing. The details of network architecture for MDJL are as follows: For feature learning, we design the networks $\{f_v\}_{v=1}^{V}$ with second and third layer of size 512 and 128. For prediction, we construct a three-layer graph convolutional network with hidden layer containing 32 nodes. For the network architecture, we adopt Adam optimizer and set the learning rate as 0.0001. In addition, we set hyper-parameters $K=20$, and $T=30$ in similarity matrix fusion algorithm. In this paper, the concordance index (C-index) is adopted to evaluate the performance of the competing survival prediction models, which mainly measures the proportion of all sample pairs for which the predictions and actual results are consistent. In order to guarantee fairness and robustness of research methods, for each dataset, we conduct 20 trials for each compared method, and the average performance of 20 trials is reported. For each trial, we would re-split the training and testing sets with 70% data for training and 30% data for testing, and re-fit the models. The corresponding Python code for carrying out our method is available at https://github.com/githyr/MDJL_Survival.

Experimental results

The predictive results of all competing methods are reported in Fig. 3, from which we can observe that our MDJL approach outperforms other competing methods on four cancer types in terms of average concordance index (C-index). In general, compared with the second best method, our approach improves the average prediction performance by 4.40%, 6.30%, 6.90% and 7.2% on the GBM, KRCCC, LSCC and BIC datasets, respectively. The reasons are two-fold: Firstly, our approach exploits correlation information
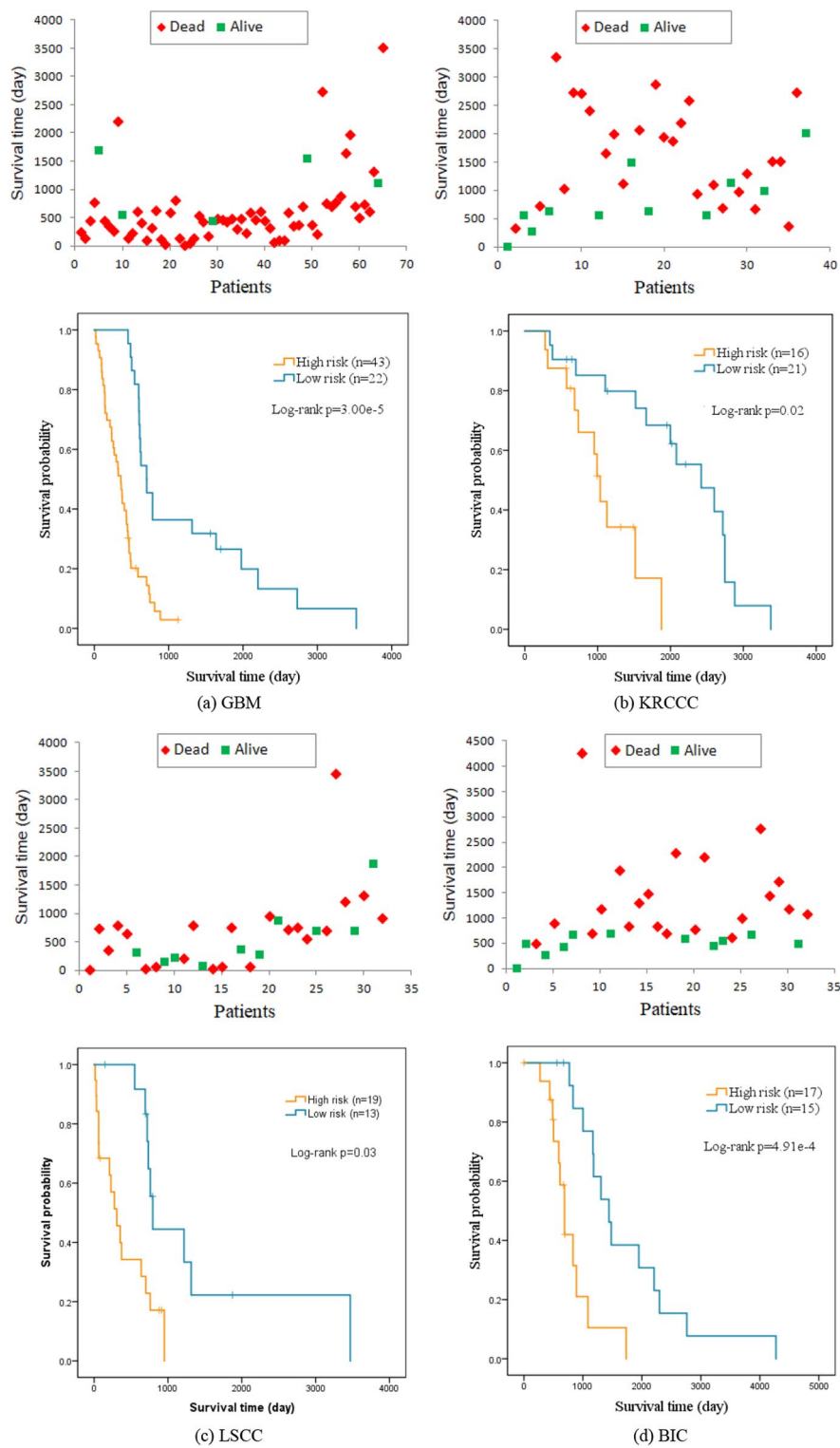
Fig. 4 Survival analysis of test sets for four cancers. The top plots in **a**–**d** present the survival time of test sets samples with the red dots pointing to dead patients and the green dots pointing to censored patients. The bottom plots in **a**–**d** show the KM survival curves
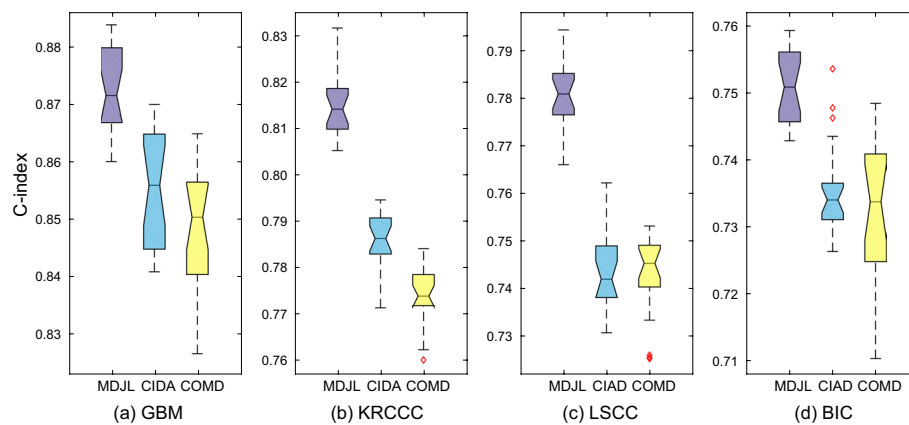
Fig. 5 The box plot for the performance of different discriminant feature learning strategies

between any two data types, which can learn more useful information as well as reduce noise more thoroughly than joint based and alignment based methods. In addition, we further explore structural information, which can help learn effective feature representations with small sample size.

We further investigate our MDJL approach with survival analysis which can be regarded as a statistical method considering both results and survival time. The patient samples for each cancer type would be divided into high-risk and low-risk groups based on their predicted hazard ratios. For example, a patient sample would be assigned to high-risk group if his hazard ratio is higher than the median hazard ratios of all patient samples, otherwise, he would be included in low-risk group. We illustrate the Kaplan-Meier (KM) curves in Fig. 4, which can reflect the survival condition of a group. The survival curve is a broken line, with each step corresponding to a time point of death and each mark pointing to a sample censoring, and P values are computed according to the curves. From the figure, we can observe that the survival probability of each group gradually drops with the increase of survival time, and the P-values for GBM, KRCCC, LSCC and BIC are 3.00×10^{-5}, 0.02, 0.03 and 4.91×10^{-4}, respectively, which are all smaller than 0.05. From the KM curves and the P-values, we can conclude that our approach can achieve a convinced result for predicting the high-risk or low-risk of one patient sample.

Further investigation

Effectiveness of correlation representation extraction

In this section, we verify the effectiveness of correlation representation extraction. In this paper, we integrate multiple data types for exploiting discriminant features by exploiting correlation information between any two data types, instead of exploiting common information shared by all data types or directly concatenating original multiple data types. In this paper, we call the version of exploiting common information shared by all data types for learning discriminant feature representations as CIAD, and the version of directly concatenating original multiple data types for learning discriminant feature representations as COMD. For CIAD, we exploit shared feature matrix by constructing feature learning networks for each data type and imposing Euclidean distance constraint
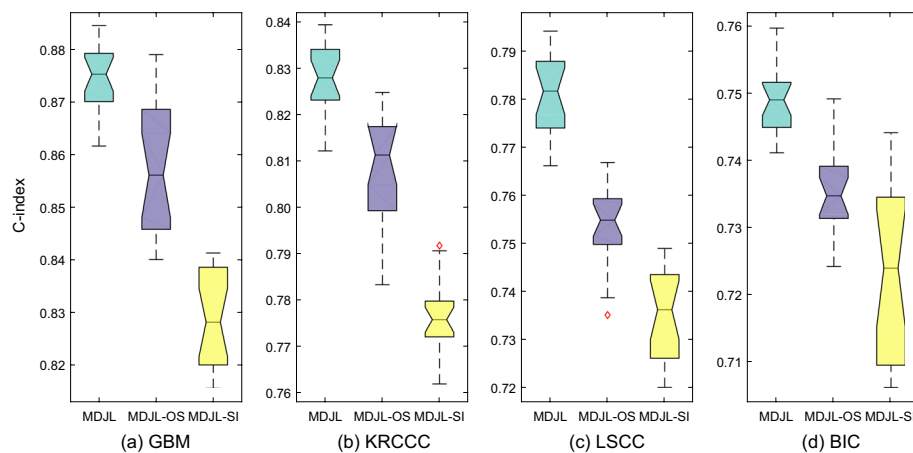
Fig. 6 The box plot for the effectiveness of learning structure information (C-index)

between the learned feature representations of any two data types, and construct similarity matrices based on original multiple data types. For COMD, we concatenate original multiple data types into a unified feature matrix, and construct similarity matrices based on original multiple data types.

We perform MDJL, CIAD and COMD on each cancer dataset respectively for 20 trials and record the C-index score for each performance. For each trial, we would re-split the training and testing sets with 70% data for training and 30% data for testing, and re-fit the models. Figure 5 illustrates the C-index for 20 times with box plot. From the figure, we can observe that our approach outperforms the other two versions on four cancer types. As a summary, learning discriminant feature representations by exploiting correlation information between any two data types can achieve better performance than exploiting common information shared by all data types or directly concatenating original multiple data types.

Effectiveness of learning structure information

In this section, we verify the effectiveness of learning structure information based on correlation representations. We respectively perform the model with learning structure information based on correlation representations, the model with learning structure information based on original data, and the model without learning structure information. We call the version that utilizes original multi-type data to construct similarity matrices as MDJL-OS, and call the version of MDJL without learning structure information as MDJL-SI. For MDJL-OS, we utilize original multi-type data to construct similarity matrices and exploit discriminant feature representations by learning correlation information between any two data types. For MDJL-SI, we exploit discriminant feature representations by learning correlation information between any two data types and replace the graph convolutional network with a three-layer fully connected network.

We perform MDJL, MDJL-OS and MDJL-SI on each cancer dataset respectively for 20 trials and record the C-index score for each performance. For each trial, we would re-split the training and testing sets with 70% data for training and 30% data for testing, and re-fit the models. Figure 6 reports the C-index scores for 20 times with box
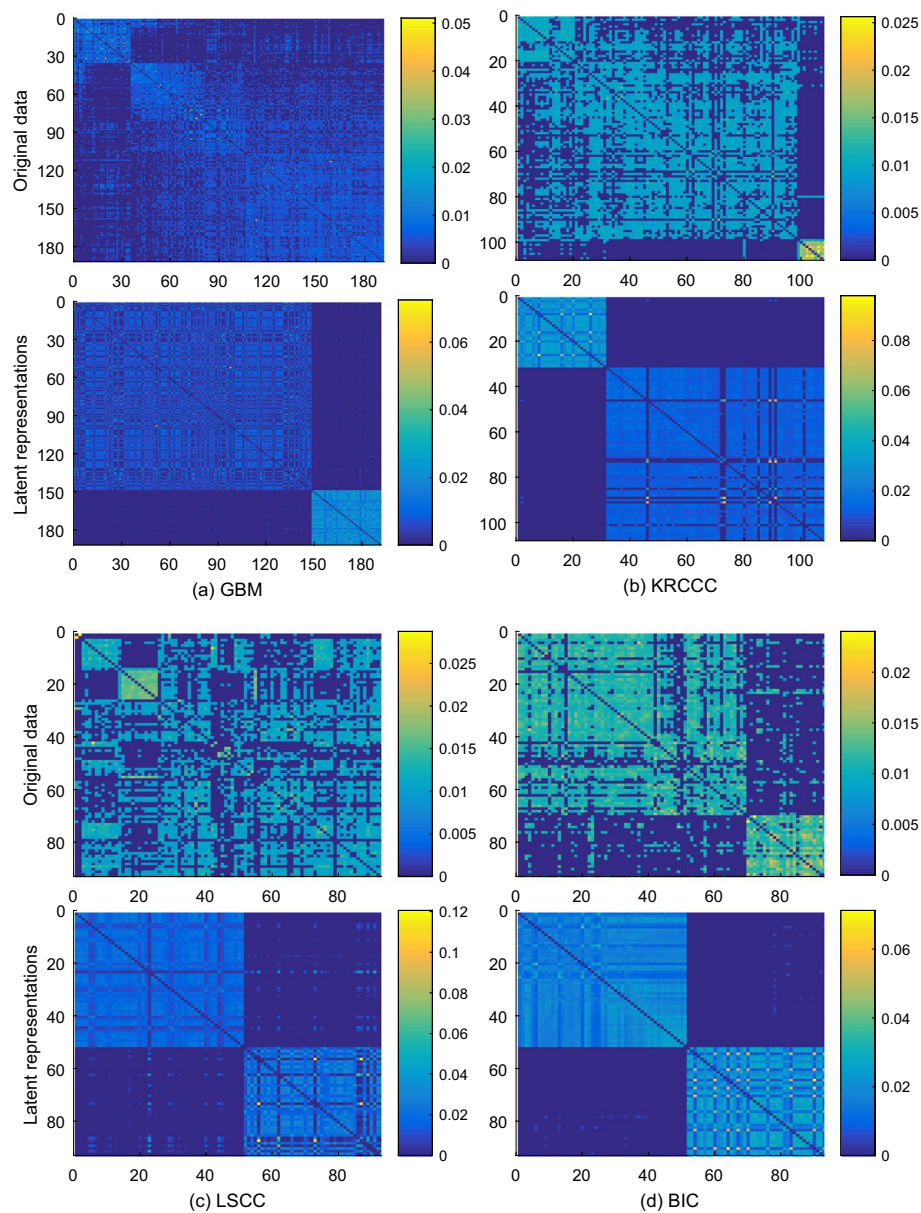
Fig. 7 Experimental visualizations of the fused similarity matrices on four cancer datasets. The top plots in **a**–**d** present the fused similarity matrices learned from original data. The bottom plots in **a**–**d** illustrate the fused similarity matrices learned from correlation information

plot, from which we can see that: (1) the performance for MDJL is better than that for MDJL-OS and MDJL-SI; (2) the performance for MDJL-OS is better than that for MDJL-SI. These results in this figure confirm that: (1) compared with only utilizing feature information, joint learning structure information and feature information can achieve better performance; (2) compared with constructing similarity matrices with original data, constructing similarity matrices with the learned correlation features can achieve better performance.
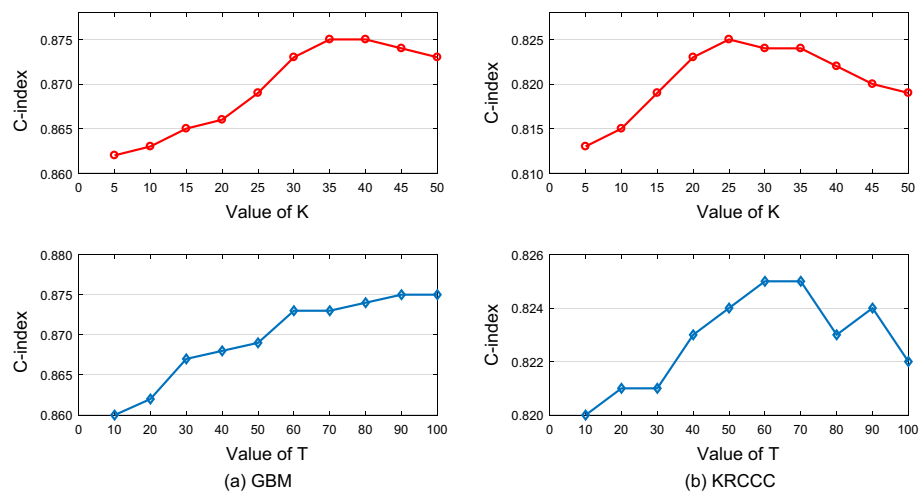
Fig. 8 The performance of our approach with different values of K or T on GBM and KRCCC

Table 3 Comparison of training time (Seconds)

Datasets	Methods						
	MDJL	MKL	MDNNMD	DLMR	CrossAE	VAEcox	DeepSurv
GBM	81.57	59.74	72.86	63.85	86.23	52.76	41.52
KRCCC	118.31	90.44	103.81	96.28	130.84	79.56	64.23
LSCC	71.26	54.17	62.92	57.36	80.17	48.57	36.76
BIC	116.45	84.23	95.34	90.04	130.81	73.58	59.64

To further investigate the effective of the fused similarity matrices respectively learned from multiple correlation representations, we exhibit the fused similarity matrices of the training sets on four cancer datasets in Fig. 7. From the figure, we can observe that the outline of the similarity matrices learned from multiple correlation representations are obvious than these learned from original multiple data types on all four cancer datasets. The reason is that the original data is unfavorable to the estimation of similarity matrices.

Parameter analysis

In this section, we investigate the sensitivity for hyper-parameters K and T with fixing any one hyper-parameter and changing the value of another hyper-parameter. When K is evaluated, we set T as 50. When T is evaluated, we set K as 20. We repeat each execution 20 times and record the average C-index. For each trial, we would re-split the training and testing sets with 70% data for training and 30% data for testing, and re-fit the models. Figure 8 shows the C-index of our MDJL approach versus different values of K and T on GBM and KRCCC. From the figure, we can observe that the C-index of MDJL on GBM and KRCCC datasets have a small fluctuation range (< 0.2). In general, the proposed approach is insensitive to hyper-parameters K ranging from 5 to 50 and T ranging from 10 to 100.

Hao *et al. BMC Bioinformatics*   (2022) 23:553

Page 18 of 20

Computing time

In this section, we use the model training time iterating over all the datasets 200 times to measure the computing time of MDJL and other baselines. Computing time of all compared methods is collected from a computer with an Intel i7 quadcore 3.6GHz CPU, a NVIDIA GTX1080Ti GPU, and 16GB memory. As seen from Table 3, the computing time of MDJL is acceptable.

Conclusion

In this paper, we propose a novel multi-type data joint learning approach, and apply it to the cancer survival prediction task. MDJL integrates correlation representation learning, similarity learning and graph convolutional network construction into a unified framework. Correlation feature representations between any two data types are effectively and fully exploited to learn discriminant feature representations. Global and local structure information among samples is fully exploited to learn the relationships among samples.

Extensive experiments on four public cancer datasets demonstrate that our approach can achieve better performance than other competing cancer survival prediction methods. In addition, experiments also demonstrate the effectiveness of the designed modules of our approach.

Availability of data and materials
The datasets generated and analysed during the current study are available with http://compbio.cs.toronto.edu/SNF/.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, Wender RC. Cancer screening in the united states, 2019: a review of current American cancer society guidelines and current issues in cancer screening. CA Cancer J Clin. 2019;69(3):184–210.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69–90.
3. Balacescu O, Balacescu L, Virtic O, Visan S, Gherman C, Drigla F, Pop L, Bolba-Morar G, Lisencu C, Fetica B, et al. Blood genome-wide transcriptional profiles of her2 negative breast cancers patients. Mediators Inflamm. 2016;2016(2):1–12.
4. Liao Z, Li D, Wang X, Li L, Zou Q. Cancer diagnosis through isomir expression with machine learning method. Curr Bioinform. 2018;13(1):57–63.

5.   Yu L, Huang J, Ma Z, Zhang J, Zou Y, Gao L. Inferring drug-disease associations based on known protein complexes. BMC Med Genomics. 2015;8(S2):1–13.

6.   Yu L, Ma X, Zhang L, Zhang J, Gao L. Prediction of new drug indications based on clinical data and network modularity. Sci Rep. 2016;6(32530):1–12.

7.   Sun Z, Dong W, Shi J, He K, Huang Z. Attention-based deep recurrent model for survival prediction. ACM Trans Comput Heal. 2021;2(4):35–13518.

8.   Kim DW, Lee S, Kwon S, Nam W, Cha I-H, Kim HJ. Deep learning-based survival prediction of oral cancer patients. Sci Rep. 2019;9(6994):1–10.

9.   Doppalapudi S, Qiu RG, Badr Y. Lung cancer survival period prediction and understanding: Deep learning approaches. Int J Med Inform. 2021;148: 104371.

10. Zhao L. Deep neural networks for predicting restricted mean survival times. Bioinformatics. 2021;36(24):5672–7.

11. Delgado R, Núñez-González JD, Yébenes JC, Lavado Á. Survival in the intensive care unit: a prognosis model based on Bayesian classifiers. Artif Intell Med. 2021;115: 102054.

12. Louis DN, Perry A, Reifenberger G, Von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol. 2016;131(6):803–20.

13. Ding D, Lang T, Zou D, Tan J, Chen J, Zhou L, Wang D, Li R, Li Y, Liu J, Ma C, Zhou Q. Machine learning-based prediction of survival prognosis in cervical cancer. BMC Bioinform. 2021;22(1):331.

14. Ksiazek W, Gandor M, Plawiak P. Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma. Comput Biol Med. 2021;134: 104431.

15. Wang J, Chen Y. Network-adjusted Kendall's tau measure for feature screening with application to high-dimensional survival genomic data. Bioinformatics. 2021;37(15):2150–6.

16. Bichindaritz I, Liu G, Bartlett CL. Survival analysis of breast cancer utilizing integrated features with ordinal cox model and auxiliary loss. In: Perner P, editor. ICDM. Ibai Publishing; 2020. p. 105–27.

17. Yu L, Zhao J, Gao L. Drug repositioning based on triangularly balanced structure for tissue-specific diseases in incomplete interactome. Artif Intell Med. 2017;77:53–63.

18. Jia X, Jing X, Zhu X, Chen S, Du B, Cai Z, He Z, Yue D. Semi-supervised multi-view deep discriminant representation learning. IEEE Trans Pattern Anal Mach Intell. 2021;43(7):2496–509.

19. Li Y, Yang M, Zhang Z. A survey of multi-view representation learning. IEEE Trans Knowl Data Eng. 2019;31(10):1863–83.

20. Wan Y, Sun S, Zeng C. Adaptive similarity embedding for unsupervised multi-view feature selection. IEEE Trans Knowl Data Eng. 2021;33(10):3338–50.

21. Zhang Y, Li A, Peng C, Wang M. Improve glioblastoma multiforme prognosis prediction by using feature selection and multiple kernel learning. IEEE/ACM Trans Comput Biol Bioinf. 2016;13(5):825–35.

22. Zhao M, Tang Y, Kim H, Hasegawa K. Machine learning with k-means dimensional reduction for predicting survival outcomes in patients with breast cancer. Cancer Inform. 2018;17:1–7.

23. Yousefi S, Amrollahi F, Amgad M, Dong C, Lewis JE, Song C, Gutman DA, Halani SH, Vega J, Brat DJ. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Sci Rep. 2017;7:1–11.

24. Mobadersany P, Wang J, Zhang M, Xu M, Zhang Z. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci. 2018;115:2970–9.

25. Kim S, Kim K, Choe J, Lee I, Kang J. Improved survival analysis by learning shared genomic information from pan-cancer data. Bioinformation. 2020;36(1):389–98.

26. Jing X, Liu Q, Wu F, Xu B, Zhu Y, Chen S. Web page classification based on uncorrelated semi-supervised intra-view and inter-view manifold discriminant feature extraction. In: IJCAI. 2015:2255–2261.

27. Chen W, Lv H, Nie F, Lin H. i6ma-pred: identifying dna n6-methyladenine sites in the rice genome. Bioinformatics. 2019;35(16):2796–800.

28. Chen W, Yang H, Feng P, Ding H, Lin H. idna4mc: identifying dna n4-methylcytosine sites based on nucleotide chemical properties. Bioinformatics. 2017;33(22):3518–23.

29. Gevaert O, Smet FD, Timmerman D, Moreau Y, Moor BD. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics. 2006;22(14):184–90.

30. Das J, Gayvert KM, Bunea F, Wegkamp MH, Yu H. Encapp: elastic-net-based prognosis prediction and biomarker discovery for human cancers. BMC Genomics. 2015;16:263.

31. Xiao Y, Wu J, Lin Z, Zhao X. A deep learning-based multi-model ensemble method for cancer prediction. Comput Methods Progr Biomed. 2018;153:1–9.

32. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res. 2018;24(6):1248–59.

33. Mishra S, Kaddi CD, Wang MD. Pan-cancer analysis for studying cancer stage using protein and gene expression data. In: Engineering in Medicine and Biology Society (EMBC). 2016:2440–2443.

34. Nguyen C, Wang Y, Nguyen HN. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. J Biomed Sci Eng. 2013;6(5):551–60.

35. Li Y, Wang L, Wang J, Ye J, Reddy CK. Transfer learning for survival analysis via efficient l2, 1-norm regularized cox regression. In: International Conference on Data Mining, 2016:231–240.

36. Ching T, Zhu X, Garmire LX. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. PLoS Comput Biol. 2018;14(4):1–18.

37. Sun D, Wang M, Li A. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. IEEE/ACM Trans Comput Biol Bioinf. 2018;16(3):841–50.

38. Gao J, Lyu T, Xiong F, Wang J, Ke W, Li Z. Mgnn: a multimodal graph neural network for predicting the survival of cancer patients. In: ACM SIGIR Conference on Research and Development in Information Retrieval, 2020:1697–1700.

39. Khademi M, Nedialkov NS. Probabilistic graphical models and deep belief networks for prognosis of breast cancer. In: International Conference on Machine Learning and Applications (ICMLA), 2015:727–732.

40. Wang L, Chignell MH, Jiang H, Charoenkitkarn N. Cluster-boosted multi-task learning framework for survival analysis. In: International Conference on Bioinformatics and Bioengineering. 2020:255–262.
41. Dang X, Huang S, Qian X. Penalized cox's proportional hazards model for high-dimensional survival data with grouped predictors. Stat Comput. 2021;31(6):77.
42. Li R, Tanigawa Y, Justesen JM, Taylor J, Hastie T, Tibshirani R, Rivas MA. Survival analysis on rare events using group-regularized multi-response cox regression. Bioinform. 2021;37(23):4437–43.
43. Zhang W, Zhang Y. Integrated survival analysis of mrna and microrna signature of patients with breast cancer based on cox model. J Comput Biol. 2020;27(9):1486–94.
44. Baek E, Yang HJ, Kim S, Lee G, Oh I, Kang S, Min J. Survival time prediction by integrating cox proportional hazards network and distribution function network. BMC Bioinform. 2021;22(1):192.
45. Wang W, Liu W. Integration of gene interaction information into a reweighted lasso-cox model for accurate survival prediction. Bioinformatics. 2021;36(22–23):5405–14.
46. Bichindaritz I, Liu G, Bartlett CL. Integrative survival analysis of breast cancer with gene expression and DNA methylation data. Bioinformatics. 2021;37(17):2601–8.
47. Li X, Krivtsov V, Arora K. Attention-based deep survival model for time series data. Reliab Eng Syst Saf. 2022;217:108033.
48. Hathaway QA, Yanamala N, Budoff MJ, Sengupta PP, Zeb I. Deep neural survival networks for cardiovascular risk prediction: the multi-ethnic study of atherosclerosis (MESA). Comput Biol Med. 2021;139: 104983.
49. Hassanzadeh HR, Wang MD. An integrated deep network for cancer survival prediction using omics data. Frontiers Big Data. 2021;4: 568352.
50. Arya N, Saha S. Multi-modal advanced deep learning architectures for breast cancer survival prediction. Knowl Based Syst. 2021;221: 106965.
51. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18(1):1–12.
52. Ching T, Zhu X, Garmire LX. Cox-nnet: an artificial neural network method for prognosis prediction of high-through-put omics data. PLoS Comput Biol. 2018;14(4): e1006076.
53. Tong L, Mitchel J, Chatlin K, Wang MD. Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis. BMC Med Inform Decis Mak. 2020;20(1):225.
54. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics. 2019;35(14):446–54.
55. Zhang Z, Chai H, Wang Y, Pan Z, Yang Y. Cancer survival prognosis with deep Bayesian perturbation cox network. Comput Biol Med. 2022;141:105012.
56. Qiu YL, Zheng H, Devos A, Selby H, Gevaert O. A meta-learning approach for genomic survival analysis. Nat Commun. 2020;11(6350):1–11.
57. Kvamme H, Borgan Ø, Scheel I. Time-to-event prediction with neural networks and cox regression. J. Mach. Learn. Res. 2019;20(129).
58. Zhan K, Nie F, Wang J, Yang Y. Multiview consensus graph clustering. IEEE Trans Image Process. 2019;28(3):1261–70.
59. Wen J, Yan K, Zhang Z, Xu Y, Wang J, Fei L, Zhang B. Adaptive graph completion based incomplete multi-view clustering. IEEE Trans Multimed. 2021;23:2493–504.
60. Wang X, Lei Z, Guo X, Zhang C, Shi H, Li SZ. Multi-view subspace clustering with intactness-aware similarity. Pattern Recognit. 2019;88:50–63.
61. Chen Y, Xiao X, Zhou Y. Jointly learning kernel representation tensor and affinity matrix for multi-view clustering. IEEE Trans Multimed. 2020;22(8):1985–97.
62. Zhang B, Qiang Q, Wang F, Nie F. Fast multi-view semi-supervised learning with learned graph. IEEE Trans Knowl Data Eng. 2022;34(1):286–99.
63. Xie D, Gao Q, Wang Q, Zhang X, Gao X. Adaptive latent similarity learning for multi-view clustering. Neural Netw. 2020;121:409–18.
64. Zhang C, Fu H, Hu Q, Cao X, Xie Y, Tao D, Xu D. Generalized latent multi-view subspace clustering. IEEE Trans Pattern Anal Mach Intell. 2020;42(1):86–99.
65. Huang A, Chen W, Zhao T, Chen CW. Joint learning of latent similarity and local embedding for multi-view clustering. IEEE Trans Image Process. 2021;30:6772–84.
66. Wan Y, Sun S, Zeng C. Adaptive similarity embedding for unsupervised multi-view feature selection. IEEE Trans Knowl Data Eng. 2021;33(10):3338–50.
67. Xu J, Li W, Liu X, Zhang D, Liu J, Han J. Deep embedded complementary and interactive information for multi-view classification. In: AAAI. 2020;6494–6501.
68. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014;11(3):333–7.

Publisher's Note